Cognitive Technologies

Wolfgang Wahlster
Hans-Joachim Grallert
Stefan Wess
Hermann Friedrich
Thomas Widenka *Editors*

Towards the Internet of Services: The THESEUS Research Program

Springer

Cognitive Technologies

For further volumes:
http://www.springer.com/series/5216

Wolfgang Wahlster • Hans-Joachim Grallert •
Stefan Wess • Hermann Friedrich • Thomas Widenka
Editors

Towards the Internet of Services
The THESEUS Research Program

Springer

Editors
Wolfgang Wahlster
Deutsches Forschungszentrum für
Künstliche Intelligenz (DFKI) GmbH
Saarbrücken, Germany

Hans-Joachim Grallert
Fraunhofer Heinrich-Hertz-Institut
Berlin, Germany

Stefan Wess
Empolis Information Management GmbH
Kaiserslautern, Germany

Hermann Friedrich
Corporate Technology
Siemens AG
Munich, Germany

Thomas Widenka
Strategy Advisory
SAP Deutschland AG & Co. KG
Walldorf, Germany

Managing Editors
Dov M. Gabbay
Augustus De Morgan Professor of Logic
Department of Computer Science
King's College London
London, UK

Jörg Siekmann
Forschungsbereich Deduktions- und
Multiagentensysteme
DFKI
Saarbrücken, Germany

ISSN 1611-2482 Cognitive Technologies
ISSN 2197-6635 (electronic)
ISBN 978-3-319-06754-4
ISBN 978-3-319-06755-1 (eBook)
DOI 10.1007/978-3-319-06755-1
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014947346

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

Information and communication technologies (ICT) provide tremendous opportunities for business and society. This is linked to their ability to help create sustainable innovation, growth and employment. The new Internet protocol IPv6 is just one such technology that offers significant growth potential. It can be used to generate a virtually unlimited number of Internet addresses and to open up new fields of application right through to the linking of different objects. The result is a massive range of new business opportunities for Internet-based services, e.g., the linking, monitoring and control of industrial production processes (e.g., Industrie 4.0), as well as applications in the health and energy sectors (e.g., the Smart Grid). In order to simplify access to information, to link existing data to new knowledge and to explore the potential of the Internet, 60 research partners from science and business joined within the ICT research program THESEUS to successfully develop new technologies for the Internet of Services.

The year 2012 saw the conclusion of the THESEUS research program, which ran for a period of 5 years. The results that were generated by the ICT research program are extremely positive: THESEUS produced a multitude of new technologies for the Internet of Services, implemented 20 standardization activities, initiated 20 development partnerships, and successfully secured 39 follow-up projects. Over 50 patents and other protected results were registered within the framework of the research program. This was in addition to the development of over 130 prototypes and the publishing of more than 800 scientific papers. In total, more than 1,600 individual results were generated, of which around 1,100 can be accessed by the public in the THESEUS results catalogue, which is also available online.

THESEUS is one of the central areas within the government's ICT strategy "Germany Digital 2015", as well as in its "2020 High-Tech Strategy". The program received total funding of 200 million € – half of which was provided by the German Federal Ministry for Economic Affairs and Energy, the rest by program partners from the fields of industry and research. The key areas of research were defined as the development and testing of basic technologies, standards and promising applications for new Internet-based products and services. Through the work carried out in the THESEUS research program, Germany is taking on a pioneering and internationally admired role in the use of semantic technologies and the creation of new standards for the Internet of Services. This work has also served as the basis for the "Future Internet" program set up by the European Commission.

The research efforts focus on basic semantic technologies which collect, analyze, classify and link the meaning of content. Although standard search technologies are already able to provide good access to content on websites and in textual documents, automatic processing of visual, audio and audiovisual data content is much less successful. The knowledge contained in these types of data often remains hidden. Technologies from the THESEUS research program help to make this hidden knowledge usable. All companies and institutions that have to manage and structure large quantities of data can ultimately benefit from these technologies. Automatic indexing, for example, is useful for storing both doctor's X-ray images as well as TV producer's archive material. Bookkeepers need text recognition to scan their invoices, and librarians need this technology to read their index cards. The quick and automatic identification of well-known persons, places or events can assist an editor as he prepares his current news program, as well as a historian who needs to evaluate and link book contributions, newspaper articles and websites. The ability to find synonyms and similarities not only helps chemists with patent analysis, but also aids engineers who are looking to compare different bids.

THESEUS partners used this basic technology and went on to deliver tools and platforms for the development of new services and business models on the Internet. Individual software components and services were integrated with one another, and companies were able to link individual components using complex yet flexible solutions. A multitude of market players have been able to use online development platforms to develop and provide very simple web-enabled services. Conventional industrial companies have thereby become providers of web-based services. In this way, THESEUS has helped to turn the Internet into an ever greater basis for building ICT applications, infrastructure and services. Already, it is, for instance, possible to quickly find suitable services and to link these even more closely together. We also see increased use of cloud computing platforms which will not only offer software components, but also capacity and storage. Web-based services could even be used to tap potential in the area of traditional engineering and plant construction. In the publishing and media industries, web-based services can furthermore be seen as key factors for market growth. Art and cultural treasures can be preserved through digitization and can be processed to make them accessible in new environments. At hospitals, cloud computing technology can be used to help doctors compile information from a variety of different sources in a more efficient manner. In the future, the Internet of Services will provide both the general public and the economy with opportunities to develop new ideas.

In order to simplify the access of SME to technologies from the THESEUS research program, the German Federal Ministry for Economic Affairs and Energy launched the competition "THESEUS Mittelstand 2009" 3 years ago. The companies selected were given early access to the various technologies developed as part of THESEUS and were able to test and use them to create new products and processes. The work undertaken provided support for the development of services and business models and strengthened the competitiveness of small and medium-sized enterprises. The copying and replication of these achievements have served to strengthen growth and employment in Germany. The fact that five successful spin-offs were set up within the THESEUS research program shows how quickly it was possible to build bridges between research and commercial use.

Furthermore, the competition of ideas, "THESEUS Talents", offered young scientists, students and independent developers fresh opportunities to use their creativity and programming skills to contribute to THESEUS developments. Specific questions linked to the range of topics covered by the THESEUS research program have encouraged these groups to actively shape the research into semantic technologies. As it brought upcoming young talent into contact with scientific excellence and commercial potential, the project proved very successful in linking up pioneering research with broad-based training.

The THESEUS Innovation Center for the Internet of Services in Berlin hosts demonstrations and presents prototypes. Here, interested parties can test for themselves the new services and tools that have been developed as part of the research program since 2007. In the future, the Innovation Center is to become a Center of Excellence for the dissemination of THESEUS tools and THESEUS knowledge.

In addition, the THESEUS mobile tablet PC application provides information about the research program as a whole and allows users to view all of the results generated so far using the integrated catalogue of results of accompanying research. The intelligent research tool is in fact itself a product of THESEUS, using semantic technologies developed within the research program.

Only recently, the German Federal Government adopted ten forward-looking projects to implement the 2020 High-Tech Strategy. These forward-looking projects are to bring together the German Federal Government's research and innovation activities in especially promising fields of technology. The Internet is increasingly emerging as a hub for the supply and demand of services. German firms can go on to draw even greater benefit from this than they have done so far. The German Federal Ministry for Economic Affairs and Energy is therefore involved in the project "Internet-based Business Services", which is largely based on the results and findings from THESEUS. The research program will bring commerce and science together to work on the provision of new forms of high-value online services. The central focus will be on cloud computing.

The research program "THESEUS – New Technologies for the Internet of Services" provided significant impetus for further fields of research, e.g. in areas related to cloud computing and big data. This book provides detailed information on the THESEUS research program and the results generated, makes new connections, and looks ahead to the future. It is thus a source of information for both scientists and commercial users on the developments made within the THESEUS research program.

Berlin, Germany

Dr. Andreas Goerdeler
Deputy Director General Information Society, Media
German Federal Ministry for Economic Affairs and Energy

Preface

A book such as this one could obviously not be put together without the help and cooperation of many people.

I am particularly indebted to the authors who graciously made their contributions available in a timely fashion.

I would like to thank Dr. Anselm Blocher for his excellent editorial assistance and the production of the final camera-ready copy. Special praise goes to Mona El Hadidy and Renato Orsini for their assistance in formatting and copy-editing the book. Special thanks go to Ronan Nugent from Springer for his continuous publishing support.

The THESEUS research program was made possible by funding from the German Federal Ministry for Economic Affairs and Energy (BMWi). I would like to thank Dr. Andreas Goerdeler, Deputy Director General Information Society, Media at BMWi, for his constant and tireless support of the THESEUS research program.

Saarbrücken, Germany

Wolfgang Wahlster
CEO of the German Research Center
for Artificial Intelligence (DFKI)

Contents

Part I
Principal Challenges

Semantic Technologies for Mass Customization

Wolfgang Wahlster

Abstract We discuss the key role of semantic technologies for the mass customization of smart products, smart data, and smart services. It is shown that new semantic representation languages for the description of services and product memories like USDL and OMM provide the glue for integrating the Internet of Things, the Internet of Data, and the Internet of Services. Semantic service matchmaking in cyber-physical production systems is presented as a key enabler of the disruptive change in the production logic for Industrie 4.0. Finally, we discuss the platform stack for mass customization and show how a customized smart product can serve as a platform for personalized smart services that are based on smart data provided by the connected product.

1 Introduction

Semantic technologies are enabling mass customization for the delivery of goods and services that meet individual customers' needs and tastes with near mass production efficiency and reliability (Tseng and Jiao 2007). Mass customization creates a competitive advantage in the industrial economy, the service economy, and the emerging data economy. Mass customization leads to smart products, smart services, and smart data (see Fig. 1). The more they are customized, the smarter they get, since they are also adaptable to specific tasks, locations, situations, and contexts.

The Internet of Things, the Internet of Services (Heuser and Wahlster 2011), and the Internet of Data provide the technical infrastructure for the mass customization of products, services, and data. But only semantic technologies can provide

W. Wahlster (✉)
DFKI GmbH, German Research Center for Artificial Intelligence, Saarbrücken, Germany
e-mail: Wolfgang.Wahlster@dfki.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_1,

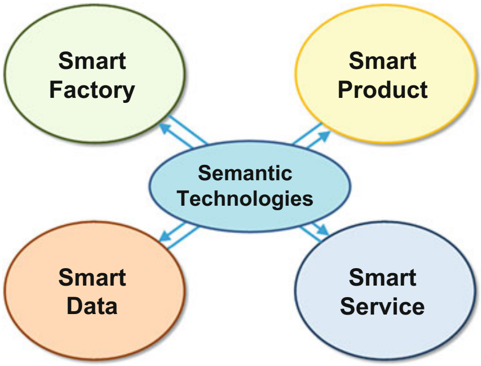


Fig. 1 Customization based on semantic technologies

the flexibility, adaptability, agility, and interoperability in manufacturing, service creation and delivery, as well as data analytics that are needed for the efficient customization of products, services, and big data sources. Semantic technologies allow us to describe, revise, and adapt the characteristics, functions, processes, and usage patterns of customization targets on the basis of a machine-understandable content representation that enables automated processing and information sharing between human and software agents (Fensel et al. 2003).

Smart Factories use cyber-physical production systems, semantic machine-to-machine (M2M) communication, and semantic product memories (Wahlster 2013a) to create smart products. These smart products are the basis for smart services that use them as a physical platform. Since smart products and smart factories include many sensors, the streams of sensor data collected by them can be fused and transformed into smart data that in turn can increase the efficiency of smart factories, smart products, and smart services.

Therefore, the results of the THESEUS research program described in this book, including methods, toolsets, and standards for semantic technologies together with insights gained into their application potential through the various use cases, form a solid basis for Industrie 4.0 and the fourth industrial revolution, the hybrid service economy, and the transformation of big data into useful smart data – as three major new R&D funding programs in Germany.

2 The Role of Semantic Product Memories for Mass Customization

Active semantic product memories will play a key role in the upcoming fourth industrial revolution based on cyber-physical production systems, which allow mass customization at an affordable prize. Industrial mass customization means low-volume high-mix production. In the extreme case, this leads to a batch size of 1, but quite often custom products are manufactured in larger quantities for an individual customer. In addition, the volatility of markets and the ever-decreasing product life spans require multi-adaptive and multi-product factories with very short reconfiguration times.

Low-cost and compact digital storage, sensors and radio modules make it possible to embed a digital memory into a product for recording all relevant events throughout the entire lifecycle of the artifact. By capturing and interpreting ambient conditions and user actions, such computationally enhanced products have a data shadow and are able to perceive and control their environment, to analyze their observations, and to communicate with other smart objects and human users about their lifelog data.

Cyber-physical systems and the Internet of Things lead to a disruptive change in the production architecture: the workpiece, the emerging product or the product packing container navigate through a highly instrumented smart factory and try to find the production services that they need in order to meet their individual product specifications stored on the product memory. In contrast to the classical centralized production planning and manufacturing execution systems, this leads to a decentralized production logic, where the emerging product with its object memory is not only a central information container, but also an observer, a negotiator, and an agent in the production process. A semantic service architecture based on a production ontology and ubiquitous microweb servers realizes intelligent matchmaking processes between emerging products and production tools (see also Loskyll 2013; Loskyll et al. 2012).

Let us illustrate the principle of mass customization based on semantic product memories using a simple example from modern food production. The German company mymuesli[1] produces custom-mixed cereals. The customer can select from a variety of cereal base mixes and refine them in small steps down to 15 g, choosing between 80 ingredients like various nuts and fruits. Thus, more than 560 billion variants are offered by mymuesli via a web-based graphical configurator on the company's portal (see Fig. 2). The consumer's individual product specification is stored on a physical product memory tag. This enables the transfer from the Internet of Services, which is used for the B2C ordering and delivery processes, to the Internet of Things, which is used for manufacturing the individualized food product. Thus the virtual and the real world are integrated in the business model of mass customization at this point.

The machine-understandable product memory describing the particular mixture ordered by the customer is attached to the paper tube box that is used for shipping the muesli to the end consumer. The product memory is read by each of the 80 filling machines, as the tube box is moving along the row of machines. For example, it will pick up 35 g of cherries, skip the mangos, and then pick up 20 g of coconut chips from the next machine and so on, until the right amounts of all selected ingredients

[1] http://www.mymuesli.com

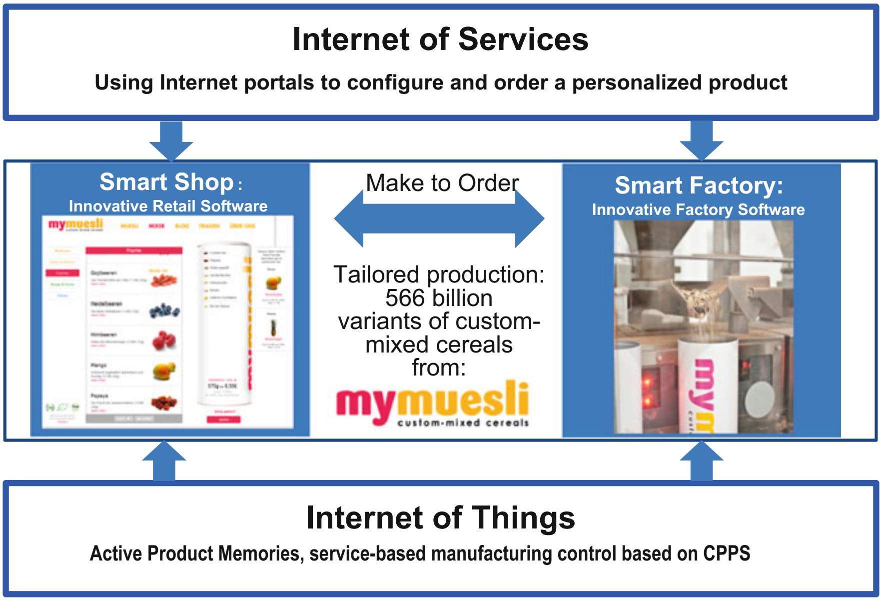


Fig. 2 Mass customization in food production

stored in the product memory are filled into the paper tube box. This means that the semantic product memory guides the flow across conveyor belts and filling stations in the smart factory (Wahlster 2013b).

By social media mining and information extraction from customer blogs big data can be collected about the customers' preferences for ingredients and their combination. Big data analytics transforms this customer data into smart data that can help to optimize the supply chain, so that always the right amount of ingredients is available in the warehouse for replenishing the filling machines. In addition, special marketing actions for selling ready-made mixtures can be launched, when a high demand can be predicted based on the smart data harvested for particular combinations. In addition, the production process could be optimized by such smart data analytics in real-time by the intelligent reordering of the sequence of filling machines, so that the average path length for the tube boxes is minimized.

The decentralized production control is also very useful, when a new filling station with a new ingredient is added due to market demands. The service-oriented architecture for smart factories in the Industrie 4.0 paradigm allows a very simple and fast solution. The new machine simply announces its service in the new row of filling stations, and as soon as the first tube box requesting the new ingredient is passing along, the machine is used in the reconfigurated factory. This illustrates the versatility and adaptability of this new factory paradigm, where new machines can be integrated in a plug-and-produce fashion minimizing manufacturing downtime.

3 Semantic Service Descriptions for Decentralized Production Control

As we have shown in the previous section, mass customization is achieved in smart factories organized like markets with many booths, where vendors advertise their goods and services. The emerging product can be viewed as a customer who is attracted to particular booths, since they offer goods and services that are on the customer's shopping list. The semantic representation languages used for the service description (see Fig. 3) allow for service discovery and selection, even in cases where no exact match is possible, but a subsumption relation can be established between the vendor's offer and the customer's needs, or a semantic equivalence relation between both descriptions can be established by automated reasoning based on the model-theoretic semantics of the markup-language. This means that the emerging product is no longer passive, but its active semantic memory and M2M communication bring it to life so that it can instruct the production machine as to how it should look. This is in sharp contrast to the traditional manufacturing paradigm where the product remained passive and all machines were centrally controlled by a manufacturing execution system (MES) generating a large programming overhead for every process change. The disruptive innovation in Industrie 4.0 is the inversion of the production logic based on semantic technologies: no longer does a MES dictate what happens with a workpiece, but the workpiece tells the machines what it expects from them.

As shown in Fig. 3, OWL-S is used as a semantic service description language in DFKI's smart factory installations. OWL-S is an ontology based on W3C's Ontology Web Language (OWL) that enables users and software agents to automatically discover, invoke, compose, and monitor Web resources offering services. Since OWL representations can be mapped automatically onto triples of the Resource Description Framework (RDF), there is a natural link to the Universal Service Description Language (USDL, see Oberle 2014) developed in the THESEUS research program as a building block of the Internet of Services. USDL focuses on the specification of various business aspects of services, like pricing, licensing, and service-level agreements and can also be mapped onto RDF, so that OWL-S and USDL can be easily connected to bridge service descriptions from the Internet of Things and the Internet of Services. Since RDF is also being viewed for big data analytics, it can be used as a kind of universal low-level assembly language of semantic technologies for smart factories, smart products, smart data, and smart services (see Fig. 1).

USDL 3.0 was described in the final report of the W3C incubator group in 2011. The Internet of Services requires a way of describing services as economic or social transactions within a broader context, including the price schemes, or the terms and conditions when consuming the service and paying for it. Also third-party intermediaries, such as brokers, or cloud providers are interested in the business and legal details as well as the functional details of a service in order to augment services and to monetize them.

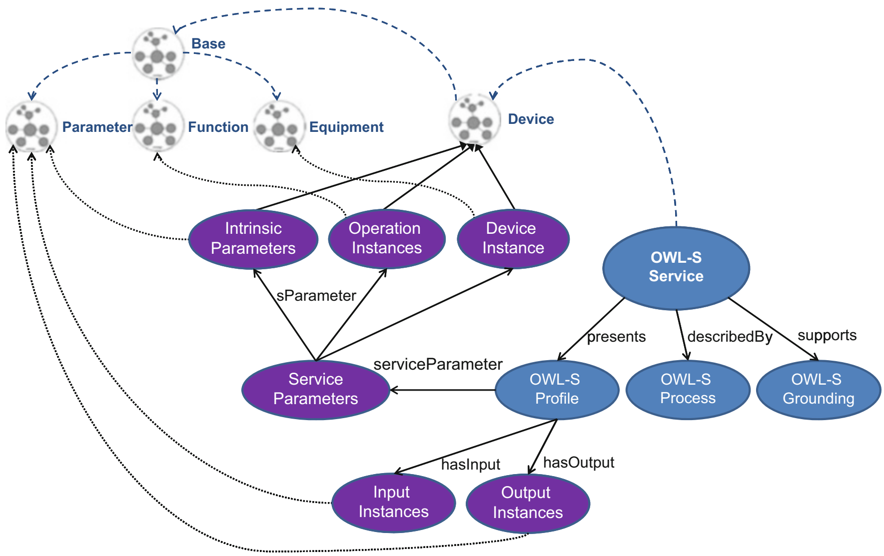


Fig. 3 Semantic service descriptions for the smart factory (Loskyll 2013)

USDL creates a kind of "commercial envelope" around a service, so that technical and functional service descriptions in OWL-S may be lifted to business services by adding USDL descriptions. As many services have a hybrid character with a digital and/or physical and/or a manual footprint, USDL can facilitate the combination and mash-up of such services. USDL has been expanded to a European scale in the context of the Future Internet PPP on the basis of the so-called Linked USDL, which is a remodeled version of USDL that builds upon the Linked Data principles. USDL also allows us to attach service offerings to sensor networks in smart factories, which can be used for internal accounting or selling sensor network services.

As mentioned above, active semantic object memories play a key role for connecting physical artifacts with smart data analytics and smart services. With the Object Memory Model (OMM) we have designed a generic framework in a W3C incubator group for implementing active semantic product memories (Kröner et al. 2013). OMM partitions a semantic product object memory into several blocks. Each block contains a specific information fragment and provides a set of metadata for search tasks in the memory. This list of blocks is supplemented with an optional table of contents and a header section. This header specifies the version of the OMM, a primary unique identifier for this object memory and an optional link to additional external block sources. Each object memory block contains information about a specific aspect or phase of the product's life. For the identification of relevant blocks by users and applications, a set of block metadata indicates the block's topic, which is stored in the block payload. If the block payload is not embedded directly into the

XML structure, then a link can be provided to indicate a relation to an outsourced block payload at any location in the cloud.

The block payload is the information container for the product memory entries and ideally is encoded in Semantic Web languages like RDF or OWL, so that a machine-understandable ontology and standardized epistemological primitives can be used for automatic processing. Semantic technologies embedded into OMM (Haupert 2013) guarantee interoperability of the product memory during the complete lifecycle of smart products and enable ubiquitous access by smart data analytics, smart services, and end users to the smart product's lifelog.

Thus, OWL-S, OMM, USDL and RDF provide the basic toolset to establish semantic technologies as the glue for integrating the Internet of Things, the Internet of Data, and the Internet of Services for Mass Customization.

4 Semantic Service Matchmaking in Cyber-Physical Production Systems

Of course, the semantic matchmaking techniques for cyber-physical production systems illustrated in Sect. 2 can be scaled to much more complex customized products that are produced on demand, like kitchens and cars. For example, the German kitchen manufacturer Nobilia automatically produces 2,800 customized kitchens per day allowing for 14 million variants. Every premium car of BMW is theoretically available in 10^{31} variants, and since there are close to 100 networked embedded systems in such a car, with frequent software updates, it is obvious that already today each car leaving the factory is a unique artifact.

Since the service descriptions for machines in a particular industrial domain are linked to ontologies (see Fig. 3) for the various production functions offered by the automation components (like adhere, weld, grasp), for the field devices that can be used to realize these functions (like sensor, drive, gripper), and for the parameters that are relevant for a particular process (like angle, velocity, force), the semantic descriptions can be easily reused, adapted and extended for new devices or new production systems in other domains.

As an example of cyber-physical production architectures with active semantic product memories, we have developed a production line for customized smart keyfinders in DFKI's smart factory within the framework of the Industrie 4.0 project RES-COM (Wahlster 2013b). The abstract production process specification is stored on the memory chip inside the backcover plastic frame (see Fig. 4). Using Bluetooth technology, the keyfinder will alert its owner any time his keys are more than 30 ft. away from his smartphone. The smart keyfinder is itself a simple example of a cyber-physical system, is produced in a smart factory, and provides various smart services. Please note that the massive amount of active product memories in a smart factory generates an enormous big data stream that can be harvested and analyzed for resource-efficient and high-quality production.

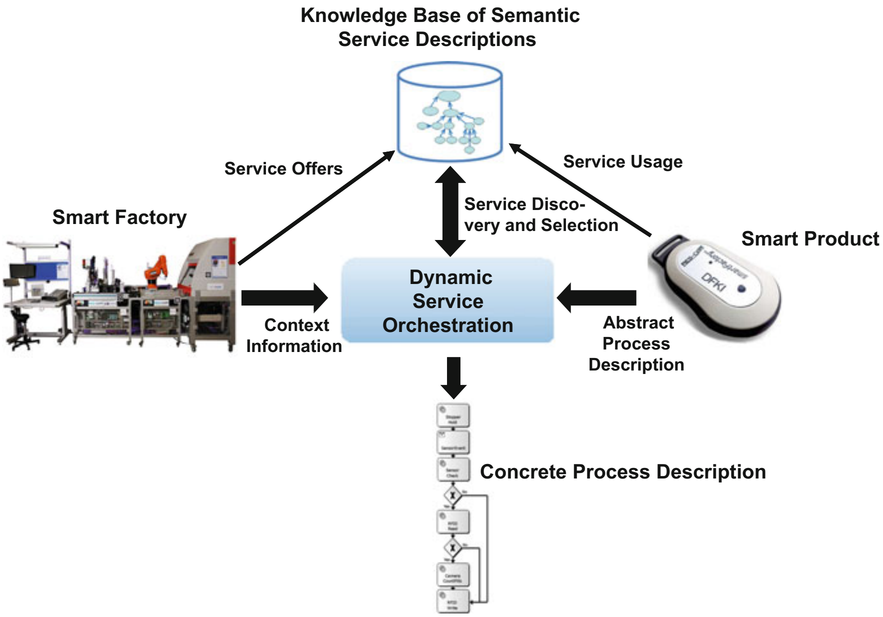


Fig. 4 Orchestration of Web services in the smart factory (Loskyll 2013)

5 From Connected Smart Products to Hybrid Smart Services

A car can be manufactured as a smart product in a cyber-physical production environment based on the specification stored on its semantic product memory that is attached as a black box to the car's chassis, before the assembly process is started. The semantic product memory not only drives the production flow, but also records all production steps, the resources, parts, and materials used. Since the semantic product memory is connected to the embedded systems and the various in-car bus systems, or in the near future to the IP-based in-car network, it can record all sensor and actuator signals of the car.

The logging data stored in the semantic product memory of the car can be used for various advanced services. Since cars are connected also with other vehicles and the road-side infrastructure via mobile Internet in Car2X networks, unimaginable volumes of anonymized traffic data can be created and stored in cloud-based mobility management systems. Since the cost of memory has declined by a factor of 500 million during the last 50 years and in-memory and massively parallel cloud computing is available, we can now analyze and respond to such big data streams in real time. This software-defined platform layer (see Fig. 5) is the basis for new service platforms that open up innovative business models based on mass customization.

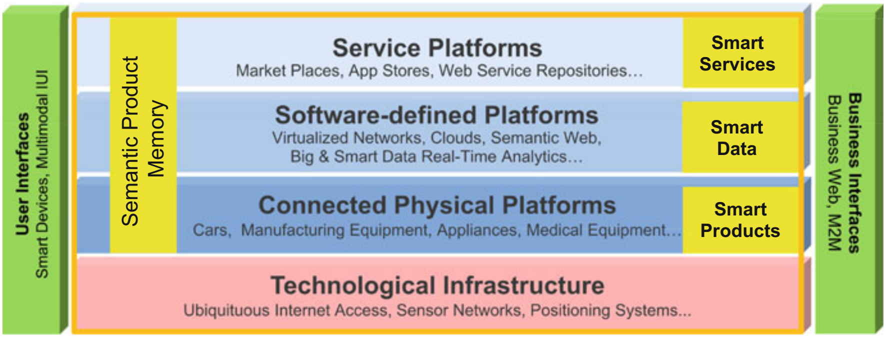


Fig. 5 Platform stack for customized smart products, data and services

Thus, the car as a connected physical platform can be used for various smart services. For example, the pay-how-you-drive functionality provided by auto insurance companies offers the driver lower fees or a bonus when they modify their driving behavior. The semantic product memory of the car provides real-time logging of the acceleration and braking behavior, the number of instances exceeding the speed limit, the number of night-time drives and other risk factors that are aggregated into a score that is delivered in a privacy-preserving trusted way to the insurance company, where the individual incentive is computed. This example shows clearly how a customized smart product can serve as a platform for personalized smart services that are based on smart data. It also illustrates the key contribution of semantic product memories to the mass customization of products as well as services.

The maturity of mobility with Internet connectivity being ubiquitous on a global scale and the consumerization of end-user devices including sensors and positioning systems that has driven the bring-your-own-device movement in workplaces characterizes the technological infrastructure for the path from smart products to smart services (see the first layer in Fig. 5).

End-user devices like smartphones, tablets, and wearables that are used in North America and Asia as connected physical platforms for services distributed via App stores are no longer produced in Germany. However, connected cars, connected factories, connected homes, and connected healthcare systems are physical platforms that are produced by Germany's export economy with the premium quality branding "Made in Germany" (see the second layer in Fig. 5).

The development of large-scale software-defined platforms, which support the mass adoption of trusted cloud computing as well as the storage and manipulation of big data streams enabling semantic analysis in real time, has accelerated the migration of value from hardware to software through virtualization, as hardware differentiation vanishes, submerged by low-margin commodity components manufactured at consumer scale (see the third layer in Fig. 5).

Market places for smart data and App stores or web service repositories for smart services constitute the fourth layer of service platforms on the platform stack (see Fig. 5).

Of course, the ultimate success of end-user applications depends to a large degree on the quality of the user interfaces. Intelligent multimodal interfaces based on semantic technologies were another important contribution of the THESEUS research program. SemVox[2] is one of the very successful spin-off companies of DFKI, that used results from THESEUS to integrate multimodal and spoken dialog systems into connected cars, homes, healthcare systems, and factories.

6 Conclusion

We have shown that in the era of Industrie 4.0, smart factories use cyber-physical production systems, semantic machine-to-machine communication, and semantic product memories to create smart products by mass customization. These smart products are the basis for smart services that use them as a physical platform. Since smart products and smart factories include many sensors, the streams of sensor data collected by them can be fused and transformed into smart data that in turn can improve the efficiency of smart factories, the functionality of smart products and the attractiveness of smart services. Therefore, the results the THESEUS research program described in this book, including methods, toolsets, and standards for semantic technologies, form a solid basis for Industrie 4.0 and the fourth industrial revolution, the hybrid service economy, and the transformation of big data into useful smart data for the emerging data economy.

Acknowledgements This work was partially funded by the THESEUS research program from the German Federal Ministry for Economic Affairs and Energy (BMWi) under contract number 01MQ07016, as well as by the German Federal Ministry of Education and Research (BMBF) in the projects SemProM under contract number 01IA08002, RES-COM under contract number 01IA11001, and SmartF-IT under contract number 01IS13015.

References

D. Fensel, J. Hendler, H. Lieberman, W. Wahlster, Introduction to the semantic web, in *Spinning the Semantic Web. Bringing the World Wide Web to Its Full Potential*, ed. by D. Fensel, J. Hendler, H. Lieberman, W. Wahlster (MIT, Cambridge, 2003), pp. 1–25

J. Haupert, DOMeMan: a framework for representation, management, and utilization of digital object memories, in *Proceedings of the 9th International Conference on Intelligent Environments*, Athens (IEEE Computer Society, 2013), pp. 84–91

L. Heuser, W. Wahlster (eds.), *Internet der Dienste* (Springer, Berlin/Heidelberg/New York, 2011)

[2] http://www.semvox.de

A. Kröner, J. Haupert, C. Hauck, M. Deru, S. Bergweiler, Fostering access to data collections in the internet of things, in *Proceedings of UBICOMM 2013, the Seventh International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, Porto, Sept 2013, ed. by W. Narzt, A. Gordon-Ross (IARIA, 2013), pp. 65–68

M. Loskyll, *Entwicklung einer Methodik zur dynamischen kontextbasierten Orchestrierung semantischer Feldgerätefunktionalitäten*. Doctoral dissertation, TU Kaiserslautern, May 2013

M. Loskyll, I. Dahmann, J. Schlick, M. Schwarz, Context-based orchestration for control of resource-efficient manufacturing processes. Future Internet **4**(3), 737–761, 8 (2012)

D. Oberle, A unified description language for the internet of services, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

M.M. Tseng, J. Jiao, Mass customization, in *Handbook of Industrial Engineering: Technology and Operations Management*, 3rd edn., ed. by G. Salvendy (Wiley, New York, 2007)

W. Wahlster, Digitize your factory for smart products, in *The Digital Enterprise*, ed. by K.H. Streibich (Software AG, Darmstadt, 2013a), pp. 144–149

W. Wahlster, The semantic product memory: an interactive black box for smart objects, in *SemProM: Foundations of Semantic Product Memories for the Internet of Things*, ed. by W. Wahlster (Springer, Berlin/Heidelberg/New York, 2013b), pp. 3–21

Challenges of the Internet of Services

Stephan Fischer

Abstract The past decade was marked by a profound change in the business world. Globalization was already taking off in the 1990s, but it reached a new quality when people and companies got used to web-based collaboration across the globe. Started as a well-defined technical infrastructure, the Internet has become the world's encompassing communication infrastructure. Year by year, the network grew, more and more nodes were added, humans, machines, and businesses were linked, and finally the Internet boomed. It emerged into the world's business backbone.

1 Introduction

In 2005, Thomas L. Friedman coined the term of a "flat world" we are living in Friedman (2005). Since the 1990s, the globe has started to become flatter, caused by changes in the political landscape, business innovations, and web-based tools for collaboration. As these tools became mainstream, competition and collaboration became more equal on a global scale and hierarchies started to erode. A playing field was created that allowed for new forms of collaboration in real time, regardless of geography and distance. One of the well known results was outsourcing, which became a hot topic at this time. When Friedman revisited the topic in 2012, the narrative of a flat world seemed to have reached an ending point (Friedman 2012). Today's CEOs rarely talk about "outsourcing" anymore, says Friedman. The business world has become so integrated that products are engineered, built, and marketed through truly global supply chains using the most efficient resources wherever they are available in the world. More and more of today's products are "made in the world", as Friedman puts it.

S. Fischer (✉)
SAP AG, SAP Research, Walldorf, Germany
e-mail: S.Fischer@sap.com

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_2,

The growing importance of services will lead to a similar development in the service sector. In Germany, as in most industrial nations, the service sector is growing fast and will be a main source of further industrialization and globalization. In Germany, around 73 % of the economic value creation was generated from the service sector in 2009 (going up from 62 % in 1991). In the same year, around 73 % of all employees worked for the service sector (going up from 59 % in 1991).[1] This increase is accompanied by changes in the quality of services and their provisioning, and IT plays a critical role in making the change happen. More and more companies offer products combined with services such as monitoring or maintenance services. Some companies focus mainly on offering a service, e.g. renting out machine tools. It is a logical idea to offer these services through the Internet. This includes web services and other automated services as well as traditional business services.

We only see a very first glimpse of how the future of the web-based service economy will look like, but evidence matures that an Internet of Services is emerging. Enterprises will use the Internet to build and provide huge numbers of new kinds of services that go beyond booking flights or purchasing books. Services that are available on the Web separately will be combined and linked with one another resulting in aggregated value added services.[2] Using web technologies, services will become more widely and easily available. Enterprises will open their business processes to others to form value networks which will be, in the end, a necessity for success in most markets. The Internet of Services is expected to ensure profitability and further growth of the service sector. It is, so to say, *the* instrument of globalization of the upcoming decade.

However, the Internet of Services has not become reality yet – at least not in the full scale researchers have envisioned it. What challenges do we face regarding the rise of the web-based service economy? This paper will review this question from different angles: Research, daily life, IT trends as well as concrete projects that translate the Internet of Services vision into prototypes and field tests.

2 Landscapes of a Web-Based Service Economy

In 2005, Thomas L. Friedman felt that the Internet became the central globalization engine and foundation of the "flat world". He then looked back and identified ten "flatteners". These are innovations that helped to remove barriers for global collaboration. With regard to the emerging Internet of Services we might not know the exact flatteners of service-oriented challenges; however, we can use this idea to discuss areas where the new flatteners might come from.

[1] http://www.bmwi.de/BMWi/Navigation/Wirtschaft/dienstleistungswirtschaft,did=239886.html

[2] This is the Internet of Services vision that lies behind the THESEUS research program and the TEXO use case that developed an integrated platform for providing, managing and combining Internet-based services; see http://www.theseus-program.de/en/about.php

When taking a bird's-eye view, the world of web-based services is not flat. It is rather structured along larger players that seem to dominate the business, complemented by larger ecosystems, loose networks, or smaller gated communities. It is a long known imperative to maintain openness and fairness in the Internet, politically as well as economically.[3] One challenge of the service sector is to ensure low entrance barriers with access to platforms, reliable and secure infrastructures, and standards. But there are several more areas that need to be looked at when discussing the Internet of Services.

We will review five landscapes that belong to the emerging world of web-based services or are tightly interconnected with it:

1. The scientific world that discovers the Internet and web-based services as appealing objects of research.
2. Our daily life that is penetrated with new technologies that will spread into the business world.
3. Today's IT trends that already contain traces of the upcoming world of web-based services.
4. The applied industrial research that builds on these trends and prepares the Internet of Services with prototypes, standards and platforms.
5. Innovation projects that translate research results into tangible software pilots for testing the market and evaluating first steps of a roadmap to the Internet of Services.

3 Research Focus

We started with the notion that today's Internet emerged from a well-designed infrastructure into something new and much more complex, namely a business platform and mission-critical infrastructure.[4] This development is mirrored by the research questions about the Internet. In the 1990s, most Internet research focused on technical artifacts such as new protocols and components.[5]

Around 2000, the notion changed. In his famous book "Linked", the physicist and network scientist Albert-László Barabási noted that a growing number of researchers asked a then unexpected question: What is it that we have created with the Internet (Barabási 2003)? Obviously the Internet went beyond the original scope it was designed for. This question caught the interest of researchers and is still a source of a growing body of knowledge about the Internet. New disciplines like Web Science or Network Science developed, with contributions from computer

[3]Most recently stated by Google's co-founder Sergej Brin who blamed Facebook for inhibiting innovation due to its closed networks (Katz 2012).

[4]Some authors use the term Societal Operation Platform (Couturier et al. 2011).

[5]This tradition is still followed by larger research programs working on advancements of the Future Internet.

science, mathematics, physics, economics and many more fields. Today, there are several lively research communities that chose the Internet as their scientific object.

The Service Science can be counted among these research disciplines although it originated from ideas that are not directly connected to the Internet. In 2006, Henry Chesbrough and Jim Spohrer introduced this discipline by publishing the famous research manifesto for service science (Chesbrough and Spohrer 2006). They saw the increasing importance of the service sector for growth and prosperity and proposed a new, interdisciplinary research on services in order to better understand services and advance service innovation. Using Information and Communication Technology (ICT) and the Internet is only one aspect of the service economy, but an important one. Chesbrough and Spohrer identified ICT as the critical enabler of today's service innovations.[6] Service Science shares two important aspects with other disciplines that conduct research on the Internet. First, all these disciplines tackle grand challenges both from a societal and a technological perspective, best labeled with the term "complexity". "The past decade has seen a growing public fascination with the complex 'connectedness' of modern society", state David Easley and Jon Kleinberg in their rich book *Networks, Crowds, and Markets: Reasoning about a Highly Connected World* (Easley and Kleinberg 2010). Albert-László Barabási identified the "data explosion" as a driving force for further research and mentions data coming from social media, sensors, business activities, or cell biology (Barabási 2003). Complexity, in his view, "takes a driving seat in science, engineering, and business" (Barabási 2003).

Due to the complexity of the Internet's networks, communities and business models, the Internet is seen as an adequate scientific object also by natural scientists. It is, so to speak, today's universe waiting to be explored. At the same time, research seems more and more headed towards creating business impact. This is a second aspect which all Internet sciences share: "Our children no longer want to become physicists and astronauts. They want to invent the next Facebook instead", says Barabási (2003). This mindset of tackling hard problems while heading towards business creativity is a great asset for understanding and realizing the Internet of Services.

4 Role of Everyday Experiences

The fascination of the Internet is not bound to the researchers or entrepreneurs. The Internet has become part of our daily life for a long time. We are using it to share news, pictures, and videos, communicate and socialize, and entertain ourselves. This influences the way we interact with technologies, which leads to changes in enterprise computing, supporting the rise of an Internet of Services.

[6] The THESEUS research program in Germany, the Smart Services CRC in Australia and parts of the Future Internet research on European level are all based on the notion of a Service Science.

The various changes the Internet has undergone or initiated have often been described as a pendulum swinging back and forth between the consumer market and business-to-business (B2B) models. This surely influences how services will be handled online. Initially, the pendulum of the Internet started to swing in the early 1990s and was primarily targeted at end users by granting them easy access to now structured content. Following this innovation, the pendulum swung back the other way by focusing on business-to-business segments, for instance e-commerce, and facilitating as well as digitizing easy shopping experiences. Then, the quick emergence of social networking sites illustrated the pendulum's next swing, again towards customers: Facebook and Twitter are just two examples of this dramatic change that took place (Hagemann Snabe and Couturier 2011).

As pointed out by IT experts and developers, the spotlight of the Internet is shifting again. One of the major trends for the Future Internet includes the comprehensive alignment of people and processes. Accordingly, existing social networks will smoothly merge with business networks, opening up new possibilities for individuals and enterprises. A fundamental change will transform social networks to multilayered structures that extend their reach beyond the ability to connect with friends and can be used as real business platforms; for business users, the friendship relationship will be extended to cover also project-based work, skills and competencies, hierarchies, as well as customer and partner contacts (Hagemann Snabe and Couturier 2011).

In the future, enterprises will open their business processes to others to form value networks when barriers for developing, providing and managing services are removed (Couturier et al. 2011). This might lead to "flash companies" that are built from freelancers and small service providers to fulfill a project goal (Petrie 2010).

Driven by the ubiquity of mobile phones, smart phones and their integration into our daily lives, the development processes for business applications will be speeded up drastically, from 2 years to 2 months, and will primarily target end users. This "economy of apps" will support the third trend that both Internet and the real world will become one entity. The Internet will be characterized by collaboration and interconnections of services and things. Seen from the end user perspective, the Internet of Services will be at the horizon when the Internet will be an omnipresent fact of life that permeates virtually every human activity.[7]

5 IT and Business Trends

Between the scientific endeavors and the daily life of consumers lies a middle ground handled by engineers and business executives. The radiuses of action for both groups are bound by the technologies at hand and the markets of today. However, both technology and markets are evolving fast. SAP has identified four

[7] See Couturier et al. (2011) for an account of the even broader term Future Internet.

major IT trends that drive new technologies and business innovations: (1) Cloud Computing, (2) Business Networks, (3) Big Data Analytics, (4) Mobility (Sikka 2012). They are all connected to the Internet of Services. In the following, we will elucidate how each of the four areas plays a vital role in materializing the vision of the Internet of Services.[8]

Cloud Computing Cloud Computing and Virtualization have become established concepts used for operating data centers and building new kinds of business software. The main advantage is that cloud offerings can be consumed almost instantly by enterprises that are highly elastic. Enterprises can start small and scale globally quickly if needed. This marks a paradigm shift in the way applications and IT services are developed, sold, deployed, maintained, and consumed. At the same time, the Cloud Computing concept can serve as a blueprint for providing any other technical service – be it global connectivity to smart phones and machines, integration services, brokering between service providers, and many more.

Business Networks Business Networks are becoming an additional focus of enterprise systems. All enterprises operate within a complex network of relationships. These networks include short-term and long-term relationships, some of them enabled through automated processes, others built on human collaboration. Today's enterprise software captures only a small part of these relationships. Future business network technology, based on a common foundation of core services (for example, sign-up or identity management), will enable businesses to discover their existing business relationships easily, and to self-organize into multiple short-term and long-term collaborations. Many-to-many connections between enterprises and multiparty collaborations will lead to new applications and value-added services provided by a network of companies. Future platforms that are open and global will support broad ecosystems of partners to drive great innovations. The challenge of the Internet of Services is to find ways to establish trusted networks of service providers and support them to easily find, combine, and operate services which are linked to a common platform.

Big Data Big Data Analytics evolved into a promising topic for enterprise computing. More and more data is being produced, and more of it is relevant for enterprises. New sources of data emerged complementing data coming from within a company or from networks of a few businesses. Organizations are seeing the value in all kinds of new data sources, from logs and machine-generated data to unstructured data such as e-mail and tweets. Across all industries, in-memory computing is simultaneously enabling new use cases, including forward-looking analytics and simulations that provide businesses with the right data and analytics to react in real time when the unexpected happens.

[8]The following is adapted from Sikka (2012).

Mobility Mobile Computing was mentioned earlier. Its global adoption outpaced the adoption of any other technology in history. Business users will handle more and more tasks with mobile devices. This will lead to the availability of business applications in multiple forms on multiple devices, which is a challenge from a developer's point of view. At the same time, new kinds of applications and services will emerge, consisting of small and lightweight applications that are developed fast and directly address the needs of the end users. In combination with machine-to-machine communication, sensors, and smart items, completely new services will be offered, for example, location-based services in supermarkets, for logistics and asset management. Both enterprises and software vendors struggle with the fact that enterprise IT systems, mostly introduced for longer periods of time, are exposed to disruptive innovative technologies which have to be introduced to stay competitive in the market. SAP recognized this challenge and addressed it with a concept called *timeless software*, ensuring that the technology platforms in use are open for future extensions without disruption to enterprise applications (Friesen et al. 2012). Increasing the flexibility of traditional enterprise systems has long been discussed with regard to service-oriented architectures and the virtualization of IT infrastructures. The above mentioned trends add four dimensions that outline the service-oriented challenge: More and more IT-related services will be integrated into the paradigm of service-orientation. This is reflected in the term "everything as a service" ("XaaS"). At the same time, enterprise systems must learn to manage also with non-technical services and relationships between companies. The third dimension is the mobilization of enterprise computing, showing the need to reduce complexity of applications by bundling existing services for specific tasks rather than provide larger desktop-like user interfaces. These developments are accompanied by the need to enable enterprises to act safely in these new, flexible cloud-based environments and far-reaching collaboration in business eco-systems.

6 Results from Research Projects: The TEXO Case

We saw that today's IT and business trends already contain traces of the upcoming world of web-based services. How can these trends be translated into artifacts that support Internet-based forms of value creation? The TEXO use case as part of the THESEUS research program provided answers to this question by studying the technical challenges of the Internet of Services to prepare prototypes, standards and platforms, and to evaluate potential business models.[9] Funded by the German Federal Ministry for Economic Affairs and Energy, the THESEUS research program was conducted by a large consortium of enterprises and research organizations.

[9]The following is adapted from Terzidis et al. (2011); for a detailed account of TEXO results see Kuhlmann et al. (2014).

The starting point of the project was the promise of web-based marketplaces that was up in the air at the beginning of the project (and still is). The main goal was to specify the opportunity of making services tradable goods by exploring the infrastructure for service marketplaces in the Internet. Guiding principles were automation, standardization and specialization as they are the three main levers that lead to an increase of productivity in industrial production. The question was how these principles can bring the service's economy to the next level by combining characteristics of the Internet, the SOA paradigm, semantic technologies, and the new ways of value co-creation known from Web 2.0.

An example of export services can illustrate concrete challenges of the Internet of Services. If you look at the business processes around export, you quickly realize that many services are involved. For the goods transfer, you need services like land and water logistics; you need insurances; you also need customs as public services and potentially certificates like the German TÜV. For the financial transfer, there are services provided by your bank, but also instruments like export credit guarantees. And of course, there are small and medium-sized enterprises (SME) at hand that support you in handling an export. The central question to be addressed is how far an offering of such services over the Web can go.

Service Descriptions One challenge is to find an appropriate understanding of the term "services". The technical meaning refers to "web services", "network services", and so on. An additional business meaning refers to "professional services", the "services industry", etc. In the context of the Internet of Services the term "services" has to be used with both connotations. In the following, we will examine both connotations of the term "service".

In order to form the Internet of Services, services have to be described in a way that the business dimension and the technology dimension come together. Enterprises have to describe the business aspect of a service while at the same time services have to be described in a way that computers can automatically understand and link services. This requires a language that can be processed automatically. The TEXO use case team came up with the *Unified Service Description Language (USDL)* which provides a uniform conception of services across all walks of life, including the mediation – the way in which you find and order services – delivery, and service-level agreements (Barros and Oberle 2012). TEXO brought USDL into a W3C incubator group to support standardization and adoption. In addition, the project provided a tool chain (editor, repository and marketplace) and made these tools available as open source solutions.

The standardization of describing services is essential for the aggregation of services on Internet-based marketplaces. This can best be illustrated with the export process example. The entire export process may be considered as a "service bundle". This is already known from package tours that offer the airport transfer, the flight, the hotel, possibly a guided tour and so on. Similarly, there may be "all-inclusive" offers for an export to, say, Brazil, made up of many different components as mentioned above. The example illustrates how service ecosystems will complement traditional enterprise software in the future. Only if services can be combined easily with several different other services, can the Internet of Services enable the full economic potential, especially for SMEs.

Market Roles The service-oriented architectures use the following three roles: service provider, service broker, and service consumer. Research has shown that the Internet of Services will need additional roles as bundling and channeling of added-value services, among others, require independent players. The TEXO use case assumes four additional roles: service host, service gateway, service aggregator, and channel makers (ensuring the utilization of services by end users).

The project team developed a prototypical tool chain that supports exactly these roles with tools for service engineering, search, matchmaking (including concepts for handling legal issues), bundling of services, brokerage, implementation and usage of service bundles including monitoring, e.g., SLA compliance.

Experimentation and Consulting Five projects with SMEs were conducted to test the concepts and gain more insights from the practice. The SMEs were provided with information and training about the concepts, methodologies and technologies developed within the TEXO use case. In order to facilitate their application and scenarios, TEXO provided a testing infrastructure called TEXO *Lab*. The SMEs used the TEXO Lab's infrastructure to test the relevant technology components in a controlled environment.

The SME projects provided valuable insights into service-oriented challenges that complement technological challenges. First, businesses must envision their own market role in the Internet of Services (Service provider or broker? Gateway or channel maker?) Second, the enterprises need to have clarity about which services they would like to offer or trade via the Internet. This demands a good understanding of the company's own processes. Third, the reach of the business activities of the enterprise must be sorted out. The current network of customers and partners need to be reviewed, new relationships explored. An example derived from the export scenario can illustrate this point: A service provider that is specialized in export service into a specific country could decide to offer export services also for additional countries. He probably will need to partner with other service providers extending his network to new geographies. The goal would be to bundle services from his own company with services from new partners with specific expertise in a region not covered by the own expertise. This needs a clear understanding of the service composition, and the possibilities for replacing parts of the service and a new value-add service.

Through the open and plural marketplace concept, TEXO empowers the genesis of new marketplaces and thus opens up new markets for small and medium-sized service providers. By providing a comprehensive service description language, an overview of evolving market roles, a tool environment covering the whole service lifecycle and an environment to test ideas for new service and experiment with business models, the TEXO use case created a blueprint of the Internet of Services.

7 Translation of Vision into Products and Services: The Business Web Initiative

Building on concepts and results of TEXO and other projects, SAP Research shaped a future product vision called Business Web. Although the general theme of supporting business networks is still present, the Business Web narrows down the vision of the Internet of Services to application scenarios which allows for testing aspects of the Internet of Services in real-world settings, with a focus on machine-to-machine communication leveraging Telco services. Scenarios include precision retailing (a mobile shopping assistant for end consumers), smart asset tracking (making ice cream cabinets smart), product maintenance, and small business webs for very small companies in emerging economies.

Cloud-Based Business Environment It is the goal of the Business Web to provide a cloud-based business environment enabling access to the necessary infrastructure, applications, content and connectivity to deliver end-to-end business applications and services optimized for mobility. The key motivation for SAP Research's Business Web is today's shift in the business software market as mentioned before when talking about today's IT trends. Enabled by Cloud Computing and enterprise mobility, a new kind of business application and service has emerged. Business applications follow the line of mobile apps and are becoming smaller, developed faster, and consumed mobile. Lightweight applications that address the end users will be a given for business software providers (B2B2C). New technologies provided by partners make it possible to integrate machines more easily and cost efficiently into business processes (B2M, M2M). In-memory technologies help in dealing with big data and offer real-time analytics.

An example can best illustrate the kind of applications and services the Business Web team works on. The example concerns traffic control and vehicle management at a major European port as well as supply chain visibility across different parties. It is the main objective of this project to optimize the loading and unloading of docks by avoiding traffic jams and reducing waiting time for truck drivers. The trucks are equipped with a telematics unit so that their location can be tracked during the transport. Information about the availability of appropriate loading docks and expected loading/unloading slots in the terminals are imported into the Business Web. The drivers can access this information on their mobile devices. They can optimize routes interactively and see when a dock is free for clearing the containers. The service is seamlessly integrated into the existing enterprise systems and brings the necessary information directly to the truck drivers' mobile phone. Due to the high number of users and the high volume of machine data, the service is dealing with big data combined from several sources and being analyzed in real time. This scenario is developed together with a major European port and is steered by Deutsche Telekom, SAP and selected partners.

One-Stop Shopping Model Although focusing on pilot applications like the smart logistics scenario, the Business Web initiative is deeply rooted in projects such as

TEXO and brings some central concepts of the Internet of Services into the new settings of the pilots. In general, the Business Web could eventually become a "go to enterprise cloud" for developers, partners, and service providers to rapidly build and deliver solutions, and especially get mobile connectivity everywhere as part of applications and services. There are two main characteristics the Business Web aims to bring forward: one-stop shop model and trusted network. One goal is to integrate services from different players including cloud providers, CSPs, application providers and content providers to form a standardized platform for mobile business and orchestrate end-to-end business processes completing the M2M value chain. Hence, the Business Web will be a platform where both web services and business services are bundled to hide complexity from the enterprise customers and end users. CSPs can use the Business Web to develop, deploy and operate M2M applications and services. Solution offerings might include M2M apps by verticals, cloud infrastructures, Mobile Network as a Service (MNaaS), real-time M2M intelligence, high performance billing via the cloud, and secure connectivity. This is complemented by SAP's services and a wide spectrum of partnerships with developers, system integrators, Communication Service Providers, cloud services partners, hardware partners, content providers, and so on. In addition, the Business Web is envisioned as a trusted network of device, machine and service providers, creating a business network and a trusted platform for the development of M2M applications for a large developer ecosystem. Within the Business Web work, SAP Research will use the aforementioned scenarios to further develop the vision of the Internet of Services vision, for example by developing pilots and testing them in selected markets. This will ultimately result in a Business Web platform including a one-stop-shop model and a trusted network of service providers. Step by step, the focus will be widened from the current focus on M2M applications and services to larger parts of the Internet of Services vision.

8 Conclusion

We started with the vision of the Internet of Services and introduced the idea of "flatteners". This idea served as a starting point and as a guiding metaphor for discussing service-oriented challenges. Thomas L. Friedman identified ten groups of flatteners of the globalizing world in his book *The World Is Flat* (Friedman 2005). They include a diverse group of innovations and historical milestones, such as the going-live of Google and other search engines, the Voice over Internet Protocol, Instant Messaging, and file sharing, blogs, and Wikipedia. The historical milestones include the end of the Cold War and China's entrance in the World Trade Organization, which both enlarged the economic playing field and allowed for greater competition. Our short walk through five landscapes of the Internet of Services pointed us to challenges of bringing the web-based service economy to life. At the same time, structures became visible that provide us with a glimpse into the future. Reducing barriers can stir creativity, innovation and growth. Barriers include technological challenges such as security, reliability, extendibility, and flexibility of infrastructures, both of enterprises and the Internet. Cloud Computing established a new paradigm of providing IT services online. Following the notions Infrastructure-as-a-Service, Platforms-as-a-Service and Software-as-a-Service, more and more ICT services will be offered online such as Mobile-Networks-as-a-Service envisioned by the Business Web initiative carried out by SAP. Additional barriers might include divergent national laws and business cultures, and lack of experience of small and medium companies to find the best market role in the web-based service economy, and to identify and shape market services on a global scale.

We do not know how the flatteners of the service economy will look. What we can be sure about is that innovations will most likely come from the following fields of action condensed from the discussion above: Research can play a decisive role by addressing the complexity that is inherent in the Internet and that made the Internet an appealing object of scientists. As current IT and business trends contain already traces of the Internet of Services, research can help in carving out future infrastructures, standards and best practices, and in both establishing platforms for experiments and developing pilot applications and services to test the markets. Facebook might be one flattener, as well as Cloud platforms, semantic technologies, the first marketplace using standardized service descriptions, and the one-stop-shop for M2M applications and services as envisioned by the Business Web initiative. There might be also political changes that result in new markets, greater competition or mind shifts towards services-orientation.

References

A.L. Barabási, *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. A Plume Book (Plume and B&T, New York, 2003)

A. Barros, D. Oberle (eds.), *Handbook of Service Description: USDL and Its Methods* (Springer, Berlin/Heidelberg/New York, 2012)

H. Chesbrough, J. Spohrer, A research manifesto for services science. Commun. ACM Serv. Sci. **49**(7), 35–40 (2006)

H. Couturier, B. Neidecker-Lutz, V.A. Schmidt, D. Woods, *Understanding the Future Internet*. International Research Forum 2010 (Evolved Technologist Press, New York, 2011)

D.A. Easley, J.M. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World* (Cambridge University Press, Cambridge, UK, 2010)

T.L. Friedman, *The World Is Flat: A Brief History of the Twenty-first Century* (Farrar, Straus and Giroux, New York, 2005)

T.L. Friedman, Made in the World. The New York Times, Jan 2012, http://www.nytimes.com/2012/01/29/opinion/sunday/friedman-made-in-the-world.html?_r=0

A. Friesen, W. Theilmann, M. Heller, J. Lemcke, C. Momm, On some challenges in business systems management and engineering for the networked enterprise of the future, in *Business System Management and Engineering: From Open Issues to Applications*, ed. by C.A. Ardagna, E. Damiani, L.A. Maciaszek, M. Missikoff, M. Parkin. Volume 7350 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2012), pp. 1–15

J. Hagemann Snabe, H. Couturier, Future internet – the business web, in *Internet der Dienste*, ed. by L. Heuser, W. Wahlster (Springer, Berlin/Heidelberg/New York, 2011), pp. 49–54

I. Katz, Web freedom faces greatest threat ever, warns Google's Sergey Brin. Exclusive: threats range from governments trying to control citizens to the rise of facebook and apple-style 'walled gardens'. The Guardian, 15th Apr 2012, http://www.theguardian.com/technology/2012/apr/15/web-freedom-threat-google-brin

F. Kuhlmann, J. Hannemann, M. Traub, C. Böhme, S. Zillner, A. Cavallaro, S. Seifert, B. Decker, R. Traphöner, S. Kayser, U. Lindemann, S. Prasse, G. Marczinski, R. Grützner, A. Fasse, D. Oberle, The THESEUS use cases, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

C.J. Petrie, Plenty of room outside the firm. IEEE Internet Comput. **14**(1), 92–96 (2010), http://dblp.uni-trier.de/db/journals/internet/internet14.html#Petrie10

V. Sikka, SAP technology vision: leveraging technology innovation to reshape your business. Sap thought leadership paper: technology, SAP AG (2012), http://www.sdn.sap.com/irj/scn/go/portal/prtroot/docs/library/uuid/00054fef-b324-2f10-7b88-9eba7efcb556?QuickLink=index&overridelayout=true

O. Terzidis, A. Fasse, B. Flügge, M. Heller, K. Kadner, D. Oberle, T. Sandfuchs, Texo: Wie Theseus das Internet der Dienste gestaltet – Perspektiven der Verwertung, in *Internet der Dienste*, ed. by L. Heuser, W. Wahlster (Springer, Berlin/Heidelberg/New York, 2011), pp. 141–160

Further Reading

U. Kubach, R. Ackermann, M. Ameling, V. Lotz, D. Oberle, V.A. Schmidt, Business Web: Cloud-basierte Flexibilisierung und Mobilisierung von Geschäftsprozessen, in *Smart Mobile Apps: Mit Business-Apps ins Zeitalter mobiler Geschäftsprozesse*, chapter 25 (Springer, Berlin/Heidelberg/New York, 2012), pp. 385–399

Industry in the Network of Tomorrow

Norbert Luetke-Entrup

Abstract Computers analyze information and understand the content. New services provide us with the information that we actually need. The Internet of Services will change many things and represents a major opportunity for Germany, as an economic center, to establish itself as a pioneer in this important development.

1 Introduction

The Internet Age, information society, flood of information – these are just some of the striking catch phrases that capture current developments in concise terms. One thing is true: The Internet itself and the volume of data that can continue to provide us with information about almost anything are growing exponentially. New and innovative applications are being created that are interesting for private individuals and industry alike. The Internet of Services is the logical progression of recent developments, and it can be broken down into four clear trends.

First Trend: From a Publication Medium to an Interactive Platform In its early days, the Internet was used almost exclusively to provide information. Those who wanted to publish information on the Internet used software tools to create a web page. All other Internet users could only read the pages. Communication took place via e-mail. In the years leading up to the turn of the millennium, the first more extensive Internet applications established themselves; the best-known of those is probably e-shopping. Since then, Internet store customers have been able to use their PCs to find out which goods are in stock. They can easily select the items they want, and they no longer have to go through the tedious process of entering long order numbers on paper forms and mailing them off. In shopping portals, development has

N. Luetke-Entrup (✉)
Siemens AG, Munich, Germany
e-mail: hermann.friedrich@siemens.com

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_3,

not remained static: Many new possibilities have been added. Web 2.0, a contrived world based on software versioning, characterizes the new web. The Internet has become a complex network of interactive applications – ranging from platforms for sharing and editing videos and photos to online communities in which users can share product reviews. Users are no longer simply readers but are instead active participants who constantly generate new content for the Internet.

Second Trend: Shared Knowledge Is Knowledge Doubled Thanks to its diverse content, the Internet is continuing to develop into a global knowledge platform. Many individuals already take advantage of the general knowledge of Internet users: Via wikis and blogs, people can share their knowledge and network with others, using social software. Wikipedia, an online encyclopedia, is already far more extensive than any lexicon and also contains information about many specialized areas. Thanks to feedback from the many readers, the quality is no worse than in printed media. By conducting online research, users can often quickly find answers to even non-trivial questions. If shoppers want to buy a digital camera, they might like to first get an overview of the quality of different alternatives. The Internet is a veritable treasure trove of different opinions and reviews, which combine to form an overall picture that is often more useful than an editorial test report in a magazine. The Internet is becoming a platform that can be used to generate new knowledge and distribute existing knowledge.

Third Trend: Mobile Access Mobile access to the Internet is becoming increasingly frequent. Every year, the sales figures for mobile devices set new records. Apps – small programs for smart phones and tablet computers – provide access to numerous services while on the go. Today, it is already nearly impossible to keep track of the number of apps that are available – either free ones or those with a prize tag. Apps range from highly generic information services to weather forecasts and pollen counts to special applications that tell entertaining stories about noteworthy sights during a city tour. For their part, intelligent mobile devices store data on the Internet, which can then be made available to other applications. Examples include location information in the form of GPS data, or photos that have been taken with the camera of a mobile device.

Fourth Trend: New Services Control the Flood of Data Commissioned by EMC, the "IDC Digital Universe Study: Extracting Value from Chaos" from the year 2011 estimated that in 2012 we would already have had access to more than 2 ZB of data worldwide (zetta corresponds to 10^{21}). That is an inconceivably large amount of data that will continue to grow. By the year 2020, the study predicts that there will be 50 times more data than in 2012. Mobile devices are not the only factors responsible for this impressive growth. An increasing number of small systems – often just individual, network-capable sensors – have their own IP addresses, communicate with each other in the "Internet of Things," and hence contribute to the flood of data. Innovative services with intelligent algorithms will be able to separate "information nuggets" from the data trash. New knowledge can arise if information is correlated in a suitable manner. It is very difficult for people to read machine data, and it is impossible for them to quickly analyze large volumes of data. The help of automated procedures and services that can analyze, link, and interpret the data in a suitable manner will be essential in order to control the flood of information.

2 Internet of Services: New Opportunities for Industry

The THESEUS research program focuses on the Internet trends described above. Intelligent procedures analyze data and make the results available via the Internet of Services. Industry, in particular, will benefit from this if new business opportunities can be tapped based on new professional Internet services.

Services Today The Internet already offers simple services that concisely summarize information and relieve users of tiresome searches. The 'assistants' for opening hours – which can display up-to-date information at a moment's notice showing which doctors, pharmacies, restaurants, swimming pools, etc. are open – are one example of this. Integrated route planners show the shortest route to a given destination. These services are particularly practical for anyone who needs cough syrup or pain pills, for instance, on a Saturday evening. Hikers benefit from the fact that GPS data for a particular route can be precisely documented and shared over the Internet. Automatic links to maps display the exact route covered, calculate distances, and show the overall elevation changes.

Tomorrow's Services In addition to these simple services for consumers, new complex services are developing that have potential for the industry. Services in the Internet can link the data of sensors and terminals. Building on this, new services can be offered. The number of services used by industry will rapidly increase in the years ahead. Services from different providers can be sold and linked on domain-specific service marketplaces. The previously noted trends are already showing the way. Through interactive applications in the Internet, companies can offer services to more people and directly, easily, and rapidly interact with their customers. Market research and product planning will no longer be carried out in an isolated fashion; instead, these activities will be implemented collaboratively in customer communities. Today, Internet services link broader and more general knowledge. In the future, we will see the increasing development of special subject-specific expert services. Applications from the open innovation field (for example Jam, idea competitions, etc.) will be used to tie in external experts in order to address company-specific questions. Through mobile access to the Internet, services can be offered on-site, and local information (for example about the facility) can be included. This makes it possible, for example, for service technicians inspecting equipment to be supported by context-specific knowledge.

Today, health platforms already exist in which patients share information regarding symptoms and therapies. In the future, these platforms can develop into powerful links between doctors, patients, and healthcare providers if the personal health history of patients can be linked to current vital signs and then analyzed. City services represent another possibility: Google already offers a service that analyzes the position reports from cellular phones to identify and assess traffic jams. Similarly, this same technology can be used to evaluate vehicle and train data to obtain a complete overview of a city's traffic situation. Corresponding services can be made available to traffic planners or passengers. In addition to the numerous services in the Internet, industrial companies can take advantage of the potential benefits offered by corresponding applications on the Intranet. From product development to product maintenance, business processes will always be accompanied by the search for relevant information or for specialized experts.

Technologies and Challenges Many Internet services are based on the same fundamental idea. They connect information that already exists but that is stored at different locations on the Internet. The technical challenge lies in developing procedures to intelligently link the data and extract the essential content from the large volumes of data. Today, we generally find information in the Internet by using search engines that search for specific words in files and web pages. The meaning of the content is not analyzed. For search engines, files are only full texts in which the search word in any context is counted as a hit. For example, if a search engine searches for the word "chamber", it might turn up architectural references, references to classical music, or even references to chambers of the heart.

Semantic Models In the Internet of Services, computer programs will independently analyze and interpret data and information. With semantic modeling, a key technology exists that can be used to handle the machine processing of facts and their correlations. The number of facts and correlations increases the complexity of a semantic model. The quality of the model influences how well a machine will be able to "understand" information at a later point. The new services will only meet our requirements if they are based upon high-quality, mature, and error-free semantic models.

Artificial Intelligence Creating semantic models manually is very time-consuming and prone to errors. As a result, industry is calling for automated procedures that can create and maintain models. Approaches from Artificial Intelligence – for example machine learning, data mining, and neurolinguistic programming – play a key role in this endeavor. These approaches can be used, for example, to identify contexts and patterns that are mapped as relations in a semantic model. Only if large parts of semantic models can be created and adapted automatically will their use generate a positive benefit. Artificial Intelligence approaches can support new Internet services even beyond semantic modeling – in text comprehension, in the interpretation of data, or in the recommendation of actions.

Information Security and Data Protection The Internet of Services makes it easier for users to find specific information. However, some information should not be accessible to everyone. Private individuals want to protect their privacy on the Internet. Companies need to safeguard their investments in research and development.

As a result, new services will only be established after requirements have been met regarding security and the protection of intellectual property and privacy. At the same time, it can make strategic sense for industry to share knowledge with others. At first glance, it may seem counterproductive to offer up development results to the general public. However, often this kind of business decision results not in business opportunities being given away, but in fact created. The expertise of an external expert can provide new impetus. A good system that is free of charge may establish itself more easily as a new standard than a cost-based system, thereby becoming the foundation for promising new business ideas. In the future, a balance must be struck between the diverging requirements for open cooperation with parties outside a company and for the protection of confidential data.

New Business Opportunities for Industry The development trends and results from THESEUS are a task for industry. With the Internet of Services, a new era of information supply may be dawning that offers the industry promising new possibilities and business fields. German industry has a tremendous opportunity to establish itself as a pioneer and expand upon its existing strengths:

- Historically, German industry has been shaped by engineering and by mechanical engineering in particular. Even today, production is still being carried out at many locations. In modern factories, sensors monitor production and computer-supported procedures control the processes and generate a permanent flow of data. This data can form the foundation for new services. For example, it may be possible to predict bottlenecks in the production plant, avoid downtimes by redirecting material flows, and increase efficiency.
- Germany has a first-class healthcare system. Modern medicine, especially in imaging procedures, works with large volumes of data. New services support the work of doctors and facilitate diagnosis and therapy. This makes it possible to improve the quality of healthcare while at the same time reduce costs.
- Our energy networks are changing. Smart grids are replacing the existing network infrastructure. What is needed are intelligent services that will ensure energy efficiency in the future and safeguard a stable energy supply.
- Germany is a country of experts; as a result, excellent knowledge is available in many areas. By improving cooperation among scientists, for example, it will be possible to better leverage their extensive specialized knowledge in different application areas. With corresponding new services for cooperation, the ideal conditions will be in place for Germany to take a leading role as a knowledge society.

3 Future Scenarios

The Internet of Services will create new “Pictures of the Future” in industry’s application domains. Examples from widely different business fields, such as energy supply, medicine, and urban development, show the wide range of possibilities.

Energy Supply Our energy networks are changing. The Internet of Services can help us develop new and intelligent power networks, also called smart grids. In the past, it was large power plants that were primarily responsible for generating electricity. The amount of power was adapted to demand, which fluctuated depending on the time of day, to ensure that the line voltage remained constant and to maintain a stable energy supply. Today – and this will become even more pronounced in the future – smaller power producers using renewable energy sources (such as wind turbines and solar systems, for example) exist alongside large power plants. Depending on the weather, they supply varying levels of energy – levels that can only be predicted to a certain extent. Electric vehicles will become more common in the years ahead and will place a strain on energy networks. After all, when batteries are being charged, an electric vehicle consumes as much power as an entire household. Under these types of conditions, classical energy networks cannot guarantee a stable voltage, and a blackout could occur.

Smart grids safeguard the energy supply by providing intelligent storage and distribution procedures. The Internet of Services will help to control smart grids. Variable pricing, for example, will enable electricity providers to relieve the load on energy networks at peak times. Household consumers and electric vehicles in particular could avoid peak times whenever possible and take advantage of lower rates when general power consumption drops or when a particularly large supply is available through solar and wind sources. Automatic linking to weather data provides for reliable weather forecasts – at least short-term – making it easier to plan the performance of renewable energy sources. In addition, the batteries in electric vehicles can be used to stabilize the line voltage during peak times by feeding power back into the grid when the rates are high. This creates a complex system of supply and demand. The drivers of these vehicles do not have to worry about any of this. An intelligent service ensures that, based on the usage patterns of each vehicle, the batteries will be charged as optimally as possible, regardless of whether the car is parked in its own garage or in a public parking space. At the same time, the power is paid for automatically, securely, and electronically.

Another value-added service can automatically provide a detailed evaluation of a user's power consumption in order to identify possible energy savings. By comparing results with a reference group, an incentive could additionally be offered that would encourage users to be better than the average. The operators of energy networks can also take advantage of the possibilities offered by the Internet of Services. New solar energy systems, for example, can automatically register with energy operators, and the accounting of remuneration for electricity fed into the grid can be handled over the Internet.

Medicine: Precise Diagnosis and Optimal Therapy Unclear symptoms, protracted chronic illnesses, or rare illnesses for which there is no established therapy – illnesses are as varied as people themselves. Much of the diagnostic data from different studies can result in medical diagnostics becoming a large puzzle. Many individual factors must be pieced together to form an overall picture that enables the cause of the illness to be recognized. This is essential for determining the optimal therapy. In recent years, the number of pieces of the puzzle has been growing. Imaging procedures in particular, such as magnetic resonance imaging and CT scans, produce huge volumes of data. The procedures themselves have become very powerful and can depict even small pathological features measuring just a few millimeters. Patients benefit from this only if these anomalies can also be discovered in the large volume of data. For the human eye, it is difficult to view numerous images together with other data. Doctors, already under significant time and cost pressure, will find it increasingly difficult to conduct their own analyses and error-free evaluations of the growing volume of data and images. An automatic image data analysis system can support physicians. If a computer can view an image – for instance of a liver – as context-related content rather than just a mass of pixels, a semantic model can be used to carry out many automatic and precise analyses.

For example, a computer might evaluate the size of the organ and its tissue structure. A comparison with reference data helps to find and mark typical changes. Inconspicuous images will then no longer need to be examined intensively. Thanks to the machine-based preparatory work, the doctors work with data that is already prepared and much more concise. In addition, the Internet of Services can support the diagnosis of illnesses and the therapy of patients by providing context-related medical knowledge. New services can, for example, associate a patient's medical file with the latest relevant research results as well as with progress reports and therapy reports of similar cases. This is additional, valuable information for physicians who must determine a treatment plan while at the same time working in the hectic environment of a hospital. This kind of support can provide better care, particularly to patients with unclear symptoms or rare illnesses. Data protection and information security are particularly sensitive issues when it comes to medical data. If there is even the slightest perception that health records could fall into the wrong hands, the corresponding services will not be adopted. For this reason, the highest standards for information security and data protection must be adopted in this application area in particular.

The City of the Future More than half of the global population lives in cities, and large cities in particular are developing rapidly. Millions of people live in close quarters – residing, working, and spending their free time there. They expect a safe environment, a stable supply of water and energy, and an efficient transportation system. They need businesses, doctors, and schools, and they would like interesting cultural events. Modern cities represent both a challenge and an opportunity for industry. Attractive living spaces need to be created that also meet the principles of sustainability, for example low CO_2 emissions.

The Internet of Services can help contribute to the smooth co-existence of an entire city. The urban living spaces of the future need comprehensive procedures that can coordinate all the components, such as buildings, transportation systems, technical infrastructure, etc., in order to achieve an optimum balance. In the future, cameras, sensors, and other information sources distributed throughout the city will be able to record the existing situation. Using Artificial Intelligence methods, this data can be automatically analyzed and evaluated, and prompt action can be taken if necessary. There are many possible applications for these services – reducing traffic congestion, for example. Every day, traffic jams waste significant amounts of time and energy and place added stress on people. New services can provide drivers with more precise and up-to-date information regarding traffic congestion than was previously possible and can recommend alternative routes. Using this kind of accurate information, city governments could often even prevent traffic jams and, if needed, temporarily open break-down lanes on highways or reroute traffic.

The Internet of Services can coordinate urban processes better than ever before, thereby saving significant amounts of energy – an important factor for the administration of a city. Cities today consume 75 % of the energy and are responsible for 80 % of CO_2 emissions. Only if we succeed in conserving energy in cities will we be able to significantly reduce energy consumption worldwide and slow down global warming.

4 Conclusion

In conjunction with the Internet of Services, information and communication technology (ICT) and digital media offer tremendous potential for added value and are important drivers for more innovation, growth, and employment in Germany. In addition, the Internet of Services will make an important contribution to a European information society. It will change a great deal and represents a tremendous opportunity for Germany, as an economic center, to establish itself as a pioneer in this important development.

New Dimensions in Semantic Knowledge Management

Rudi Studer, Catherina Burghart, Nenad Stojanovic, Tran Thanh, and Valentin Zacharias

Abstract With the Internet becoming part of everyday life, technologies enriching available data and software services with metadata gain importance. These semantic technologies for modeling knowledge bases and managing them have already started to be applied in the World Wide Web and in enterprises for some years. But, up to now, semantic technologies are mainly associated with semantic search and automatic extraction of lightweight semantics (metadata) from Wikipedia and other text sources available in the World Wide Web. New challenges are the design of analysis tools and the modeling of further characteristics of data for automatic information processing and interpretation: Aspects like relations between people, entities and instances as well as all different types of contexts, also comprising big data, come into focus. In this article the authors illustrate these new dimensions with respect to context addressed and show some prospective applications.

1 Introduction

Almost ubiquitous access to the Internet proliferates the usage of smart devices and Internet services in business and everyday life. Apps, small software services, are generated to the thousands per day. With ongoing availability of Internet access, data and services, these apps as well as more complicated Internet services have a demand for tools generating metadata and managing relations between data. These tools are called semantic technologies; they incorporate methods like

R. Studer (✉) • C. Burghart • N. Stojanovic • V. Zacharias
FZI, Forschungszentrum Informatik, Karlsruhe, Germany
e-mail: studer@fzi.de; burghart@fzi.de; nstojano@fzi.de; zach@fzi.de

T. Thanh
KIT, Karlsruhe Institute of Technology, Karlsruhe, Germany
e-mail: ducthanh.tran@kit.edu

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_4,

automatic design of metadata, automatic reasoning, mapping of knowledge bases, complex event processing, semantic search or tracking and repair of changes within knowledge bases.

Modern systems for intelligent knowledge management also tend to take into account the actual context, especially when interpreting complex situations or when being tailored towards personalized usage. Modern intelligent systems allow a correlation between high-level, low-level, internal and external data. An example shows the desired alignment of different contexts in a personalized mobile application:

Imagine stopping at an unknown airport in order to catch a connecting flight at a different terminal. Your navigation system in your smart phone helps you to get to the new terminal on time while guiding you by a newspaper store stocking your favorite journal, passing you by a rack with your perfume brand at the duty-free shop on the way and showing you the way to tidy restrooms close to your gate.

In this scenario, there are several contexts to be taken into account: On the one hand, there is the personal element or profile, which can be enriched by learning algorithms automatically detecting some habits (i.e. walking the same route at specific times or buying specific products). Further contexts are time- and place-dependent: using micro navigation and the map of the airport your location is detected. Your goal can be derived from your ticket and a flight information system. A route is planned taking into account your preferences combined with experiences collected by crowdsourcing, such as evaluating the cleanliness of the restrooms.

The types of context are manifold and can be categorized as follows (Strube et al. 1996): Besides linguistic contexts (which are not further considered), situational contexts are of great importance in order to interpret situations by using semantic tools; they are contexts referring to topics (here, domain knowledge is relevant), interactions, institutions, systems (abstract, virtual or real), and cultures. Social semantics and crowdsourcing are considered as being part of a cultural context.

Context can still be further or differently classified (Mehra 2012): a distinction between small context, i.e. events or time and location, and large context, as available in Linked Open Data, can be made. Other categories, vertical to this distinction, are time and location, contracts and commitments, web of things, web of people, web of knowledge, and emotions and outcome. These contexts can also have smaller entities such as particular human situations. As can be easily seen, these categories can be aligned with the definitions of contexts in Strube et al. (1996).

The challenge now is to formally represent these different types of context in order to enable usage of knowledge. With respect to domain knowledge and institutional knowledge that has already begun, also sparsely with respect to time and place as can i.e. be seen in Yndurain et al. (2012), Henson et al. (2012) and Boyaci et al. (2012). Challenges with respect to semantic search, proactiveness and social semantics and their intended usage of contexts are described in the following sections. Future intelligent systems aim at taking into account the personal context of a user as well as linking large areas of knowledge in the web of knowledge for a more adequate search as described in Sect. 2. Intelligent proactive event processing wants to correlate patterns from various contexts to interpret situations and to predict possible outcomes (Sect. 3). Additionally, future social semantics have the goal of combining personal and interaction contexts with the web of knowledge as well as with place and time (Sect. 4).

2 Semantic Data Management

Modern search engines are efficient in retrieving documents (web pages in the case of web search) from large collections based on *matching the text* of the user's query with the textual content of the document or the words used to label the links to the document (anchor text). While they have been effective in addressing common needs, most categories of queries are in the "long tail" of web query logs (88 % of the unique queries that appear only once in the year (Baeza-Yates et al. 2007)) are still unsolved. Particularly difficult are *ambiguous queries*, which arise naturally when the user is not able to name precisely the item that is sought. Also, *complex queries*, which go beyond a reference to a single named entity, and involve several entities and relationships between them, pose challenges to current search solutions. While these problematic types of queries are not new, the renewed interest in trying to address them can be explained by the increasing availability of semantic resources in terms of document annotations (also called metadata), large datasets and schemas. Many *semantic search systems* (Finin et al. 2011) have emerged specifically to exploit these resources. Thesauri and particularly schemas capture semantics that can be used to understand the query and data and to *resolve ambiguities*. Instead of returning textual data (web pages, documents), precise facts capturing entity-related information and their relationships can be directly retrieved from semantic annotations and data to *provide direct answers to complex queries*. Here we discuss the broad concept of semantic search and present several types of semantic search systems.

Overview While all semantic search approaches involve some kind of explicit semantics, the retrieval contexts and the specific semantic models used to deal with the meaning behind the query intent and data vary. In particular, we identified the following aspects, based on which we will characterize and distinguish existing solutions:

- The type of targeted *information needs*,
- The representation of information resources (*data*),
- The representation of the information need (*query*) or the underlying method for querying the data (*querying paradigms*),
- The *semantic models* used to understand and to represent the meaning behind query and data,
- The framework for *matching* queries against data, which also involves *interpreting* the data and query intent as well as *ranking* the results.

Figure 1 illustrates the specific differences in queries, data, semantics, results (corresponding to information needs) and solutions for the subproblems that

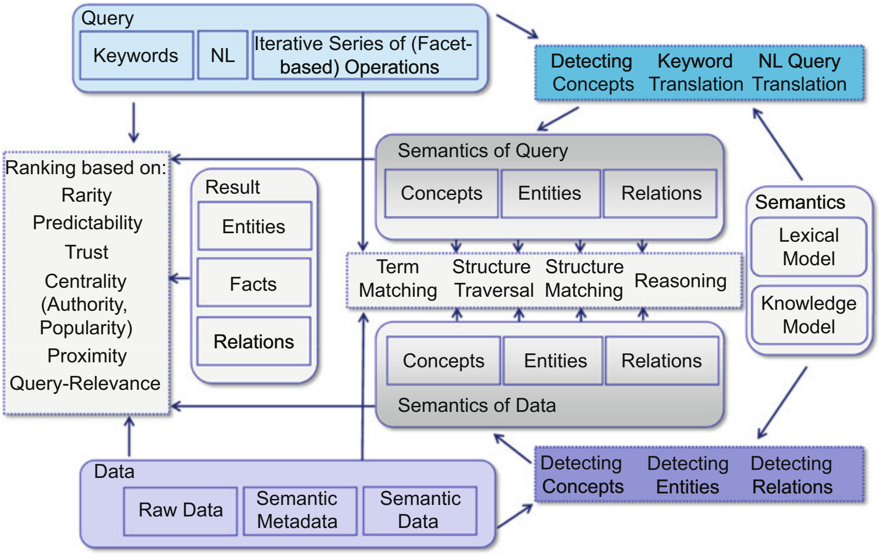


Fig. 1 Overview of semantic search approaches

distinguish existing semantic search approaches. We can see that central to semantic search is the use of semantics (depicted as gray boxes in Fig. 1). In particular, semantic resources represented by lexical models (thesauri, taxonomies, dictionary of words), knowledge models (ontologies, concept networks, schemas), as well as semantic data (RDF) and metadata (RDF data describing the content of documents, called RDFa) are used for the subproblems of understanding raw data (textual data, image objects) and queries. The resulting semantics of data and queries are then employed for solving the subproblems of matching the queries against the data and ranking the results. We will now discuss the main types of semantic search approaches.

Concept-Based Document Retrieval This is the classic type of semantic search systems, which has been studied already in the early years of IR research. Researchers in this community have recognized the need for a richer representation of the information needs and data that goes beyond the bag-of-words model. Thus, lightweight *lexical models* have been employed to interpret the semantics of query and documents as concepts, and used within the standard bag-of-words IR paradigm (e.g. vector space model and language model).

However, there are *no clear evidences* suggesting conceptual IR can outperform standard IR techniques. Even when concepts are selected by hand, it helps only when the query is an incomplete description of the need (Voorhees 1994).

Annotation-Based Document Retrieval These approaches exploit the advancement in IE technologies to obtain a richer representation of queries and documents, namely as *entities* and *relations* represented as *annotations* (query and document annotations, respectively); this query interpretation has been applied both to keywords and NL questions.

It has been shown that high precision is possible for queries for which corresponding annotations exist (Chu-Carroll et al. 2006). However, high recall is difficult because it requires the information extraction program to recognize arbitrary types of entities and relations. Along this line, there is a study showing that the quality of IEs significantly impacts semantic search performance (Chu-Carroll and Prager 2007).

Entity Search In contrast to document retrieval, this approach searches for *entities* representing real-world objects (anything but documents). It includes search over (1) documents (e.g. for processing entity-related facts and listing queries of the TREC Question Answering track (Chu-Carroll et al. 2006), or expert search queries of the TREC Enterprise track (Balog et al. 2009)); (2) Wikipedia data, which, as discussed, can be seen as a kind of semantic data (e.g. for answering entity-related queries of the INEX Entity Ranking track (Kaptein et al. 2010; Pehcevski et al. 2008)); or (3) pure semantic data crawled from the Web (e.g. for answering entity queries of the SemSearch Entity track (Blanco et al. 2011)).

Some *type(1)* systems can also be categorized as annotation-based document retrieval systems because they detect entities and relations in data and queries (Cheng et al. 2007; Chu-Carroll et al. 2006; Nie et al. 2007). Thus, analogous to these systems, precision can be very high when quality extraction outputs are available here.

The use of semantics in *type(2)* systems is limited to viewing Wikipedia pages as entities (Kaptein et al. 2010), exploiting categories associated with them (Balog et al. 2011) as well as links between them (Pehcevski et al. 2008). As a source of semantics, experiments have shown that category information helps to interpret the entity query and to enrich the Wikipedia documents and, as a result, to outperform standard text-based retrieval (Balog et al. 2011; Kaptein et al. 2010).

While many *type(3)* systems (Cheng and Qu 2009; Tran et al. 2009a; Tummarello et al. 2010) have been built to search for entities over semantic data from the Web, solutions for ranking and the evaluation of their results have attracted interest only recently. Recently, the SemSearch challenge attracted a large number of participants, which submitted results produced by different ranking schemes based on traditional IR concepts. To date, the best performed method Blanco et al. (2011) uses an adaptation of BM25F.

Relational Keyword Search This category comprises all approaches which process *keyword queries over semantic data* (or structured data capturing entities and relations) (Ladwig and Tran 2011; Tran et al. 2009b). While the results here include entities, the focus is to find possibly complex subgraphs encompassing several entities and relations between them (i.e. to support relational search). Regarding result quality, a benchmark study Coffman and Weaver (2010) has shown that the proposed ranking strategies (e.g. Liu et al. 2006; Tran et al. 2009b) based on the adoption of proximity, TF-IDF and PageRank do not perform well when considering a broad range of queries. Recently, it has been shown that high quality results can be achieved through the use of edge-specific language models that are constructed for results and queries (Bicer et al. 2011). Also, the learning to rank strategy, which combines TF-IDF-based ranking, proximity and other factors, yields high performance (Coffman and Weaver 2011).

Relational NL Search As opposed to NL search over documents, which is mainly studied by IR researchers, this topic attracts a separate community that deals with interfaces over databases and knowledge bases. The underlying problem is about *searching over semantic data*. We refer to it as relational NL search because relations in the data have to be processed to answer complex queries, possibly asking for relations between several entities.

Traditionally, this type of query interface is used for expert systems targeting one particular domain. The questions of portability, and the problem of building domain-independent search systems, have attracted attention in the past few years (Cimiano et al. 2008; Li et al. 2007; Wang et al. 2007). While this problem of open-domain NL search, e.g. NL search over large amounts of heterogeneous web data, seems to be largely unsolved, relatively good results have been reported for single domain tasks. However, while there is TREC-based evaluation for NL search over documents (TREC QA track), there exists no such thing close to a standardized evaluation in this community.

3 Proactiveness

Due to an enormous increase of information sources relevant for making decisions and ever shorter time for acting, the efficiency of information management has moved the focus from finding relevant information in to anticipating potentially relevant situations that will enable proactive delivery of information (Engel and Etzion 2011). A common example is the prediction of the delays in the traffic and working on the mitigation before they happen. Obviously, searching for an existing traffic jam on the planned route is an easily manageable task, but the anticipation that because of the heavy rains at around 5 p.m. in a part of the route close to a big city there might be a traffic jam in about 30 min is a much more challenging task. Even more demanding is the anticipation based on the unusual trends in the social media streams, such as the increase of the number of tweets related to the traffic in an area, that might indicate some emerging problems on the road. There are two main challenges required for realizing such a scenario:

1. Real-time detection of interesting situations in very huge and dynamically changing data streams, and
2. Anticipation of events through the proactive delivery of information to relevant consumers or contexts.

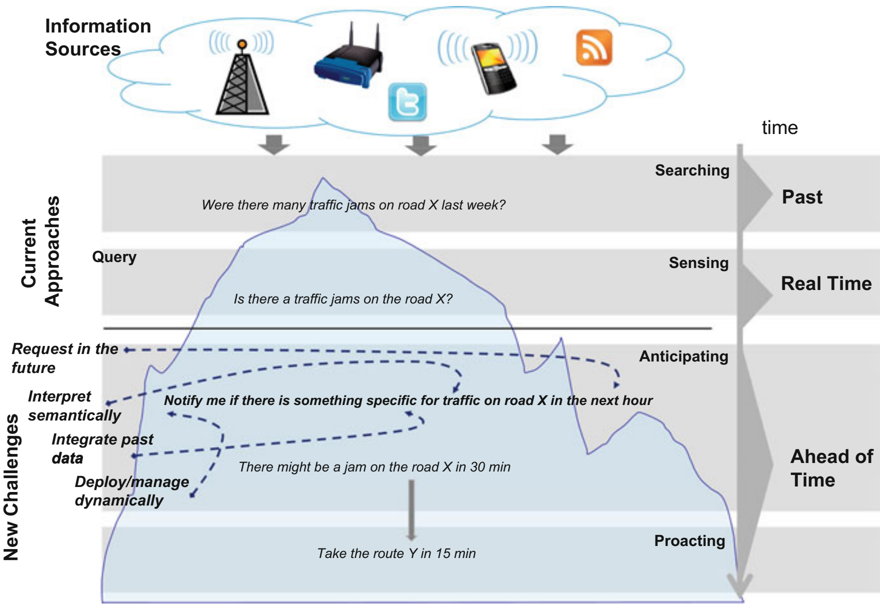


Fig. 2 Challenges for realizing proactiveness

More concretely, the problems are coming from the fact that the relevance of a piece of information can be determined only through the real-time combination with other information, which implies the need for data-driven processing. In such a processing scenario, a relevant situation is emerging from the data and will be discovered immediately after being completed (differently from the detection in the query time – for query-driven processing) (Silberstein et al. 2007). Another issue is, that we do not expect that a user can know and subscribe to all possible interesting situations, but rather has to be notified about them once they seem to be relevant for her/him. Otherwise, the user will suffer from an enormous information overload, which is much more difficult to overcome than in the traditional search due to the paradigm shift since "the user is not searching for relevant information, but information is searching for the relevant user". This leads to the need for a very flexible and dynamic mechanism for the selection of situations to be detected.

We argue that semantic processing is of crucial importance for resolving these challenges for proactive computing due to its ability to interpret data in a broader (semantic) context and support the detection on a higher abstraction level. Figure 2 illustrates the process of proactive processing of real-time data sources and summarizes the tasks required for realizing an anticipatory behavior that will be described in the rest of this section.

Queries for Searching in the Future Streams In contrast to a database/web search, where queries are formulated on an already existing dataset, queries for real-time processing are created at first and then applied on a continuous stream of upcoming events. In web search, web data is usually crawled and then indexed for enabling efficient search on the data corpus. The efficiency of search can be highly improved by applying powerful but time-consuming optimizations on storage and indexing. From a user perspective, web search is done to fulfil an information need immediately. For that reason, there is less time to improve queries and most effort is put into index optimization (see Sect. 2). The problems appearing in stream search are completely different. Here, queries are formulated at first and then matched against streaming data. While manipulations on incoming data are time-consuming, the largest potential for improvement lies in semantic processing of queries at design time in order to get better stream search results. Basically, the domain knowledge is used for extending the original queries with domain knowledge. However, in order to cope with the information overload, these extensions must be put in the context of the importance for the original query. For that reason, relevancy-based semantic requests at this level are extended with frequent terms in these concepts based on domain knowledge. This can be done by calculating the weight of relationships between entities (that have been gathered during runtime) and only considering entities with weights exceeding a certain threshold in order to refine patterns (Fang and Lei 2011). With respect to the dynamic flow of information in real-time social media streams, where messages might become important within a very short period in time (e.g., as often seen in disaster reporting), this requires a highly scalable computation of the current importance of a message.

Predictions: Searching in the Future Using Past Data In order to understand what can be a trend in the future an efficient introspection of the past is needed. There are several applications domains where the possibility to act based on predictions (i.e. ahead of time) is getting ever more important, like Smart Grid and Smart Cities (Etzion 2010; Stojanovic et al. 2011a). This phase model concentrates on the mining of historical knowledge gathered from previously matched patterns. Based on the partial fulfilment degree of current patterns and the probability of complete fulfilment in history, unusual messages can be detected. Unusuality from our point of view is based on two features, unexpectedness and importance (Sen et al. 2010). The traditional way of querying RDF data is a blocking get operation. However, real-time applications need an asynchronous query mode to be more responsive on arrival of RDF data. Publish/subscribe (pub/sub) is a messaging pattern where publishers and subscribers communicate in a loosely coupled fashion. Subscribers can express their interests in certain kinds of data by registering a subscription (continuous query) and be notified asynchronously of any information (called an event) generated by the publishers that matches those interests. Notifications are made possible thanks to a matching algorithm that puts in relation publications and subscriptions. One of the challenges that faces this huge amount of data is the ability to combine the storage (associated with traditional, synchronous, query mode) and asynchronous delivery of RDF data (pub/sub protocol), whereas the deployment in a cloud environment should ensure the elasticity of the solution.

Dynamic Management of Future Queries Another challenge for the information overload is the possibility to mange the deployment of the queries in a way that ensures that only relevant requests will be active at each point of time (Stojanovic et al. 2011b). This leads to the dynamic subscriptions, based on the real-time contextualization of the queries that will invoke queries only if the right context has been achieved. Similarly, the queries will be active only until the right context exists. Moreover, the validity of the queries should be verified in a long run, and in the case of deviation from the expected behavior, corresponding changes should be proposed. In a similar way, based on the data flow, new relevant queries can be generated on the fly and introduced in the system, reflecting the ultra dynamics of the real-time streams. The challenge for the semantics lies in the real-time mining of interesting queries and detection of the situations when they should be introduced in the system. Finally, managing proactiveness is an issue of enabling real-time acting in order to realize potential business opportunities or mitigate possible conflicts, anticipated dynamically from the data flow.

4 Scaling Social Semantics

In a nutshell semantic technologies strive to store and process meaning separately from the program itself and from other data. By doing so they enable the creation of systems that:

- Can cope better with changes that affect the meaning of the data – because the program itself is not (so much) affected by such a change,
- Can better work with heterogeneous data – again because the core program is not affected by this heterogeneity because it does not directly encode the meaning of the data,
- Can use powerful algorithms such as semantic search that make use of the meaning of the data – because the meaning of the data is available for processing and not hidden in a program's code.

These properties are only important for certain application areas – those areas where data changes frequently, where heterogeneous data abounds and where complex processing of data is needed.

For example semantic technologies may not be a good match for a system that manages the well understood and slowly changing inventory of a medium-sized company that produces and sells different kinds of paper – the flexibility offered by semantics is just not needed and all kinds of more complex processing can be done just as well directly in the software. If the company starts to add a new kind of paper having really different properties it will have to change so many things anyway that an adaption of the inventory software is comparably small and no hindrance to the speed with which the company can act and adapt. Contrast this with a system that manages information about actors in the health care sector in Europe and that is maintained by geographically distributed researchers in this area. Here change is permanent, the managed data is heterogeneous (there is not just one class of actor but actors can be all kinds of different things) and powerful semantic tools are needed to make this complex data accessible.

These are settings which are also high in social complexity; settings where multiple stakeholders constantly introduce new concepts, where a single schema cannot be enforced and in any case would be outdated by the time it is and where complex processing is needed to enable people with different levels of expertise to work with the data.

Going back to the examples of the paper factory one can see that this setting is low in social complexity with one company as one stakeholder, a single schema representing the inventory that is slowly changing. In contrast the hypothetical system managing the data about Europe's health care actors is used in a setting high in social complexity – researchers from many locations contribute to it, nobody knows enough about all health care actors to even device a complete schema (which would be outdated in mere days anyway) and there might be conflicting views about what the actors are and what their role is. Also, for this system to be useful to the general public there is probably the need for the system to mediate between the language of the health care experts and that of the broader public.

Hence: for successful semantic systems to emerge in places where semantic technologies can offer the most benefit it is important that semantic technology research focuses not just on scaling to large datasets, but also on scaling to application areas high in social complexity.

Building semantic systems that scale to high social complexity needs research into:

- Systems that are able to deal with *multiple points of view*, that can handle even conflicting knowledge and use it to produce results reflecting both the most likely answer as well as the diversity of opinion (e.g. Flöck et al. 2011).
- Systems that support a user group in coming to an agreement and *contributing in a structured way*. Systems that for example use crowdsourcing tools and methodologies to involve a large group of users in the creation of knowledge (e.g. Di Iorio et al. 2010).
- Systems that seriously *consider the relationships between users*, for example by including privacy and data ownership as first-order citizens (e.g. Duma et al. 2007).
- *Support diverse groups of users* with different world views in first understanding the available data, then formulating queries and finally understanding the result. With large heterogeneous and fast changing datasets and heterogeneous users it can no longer be expected that users have knowledge about the kinds of data that are available or have the knowledge to understand this data. To address this, a new generation of systems must support exploration of these datasets and make use of available semantics to enable users at different levels of expertise to understand the data (e.g. Schenk et al. 2009).

One example of such work is the ongoing Render project (Hasan et al. 2011; Thalhammer et al. 2011). In this project, methods and techniques are researched that allow us to leverage the diversity of the web as a crucial source of innovation and creativity. More concretely, new search and aggregation techniques are developed that show not just the majority opinion or the opinion closest to the user but the whole range of sentiments. Such techniques are particularly important on the Internet, where pervasive recommender systems mean more and more that users only see the results confirming their views, thereby leading to users living in a kind of "Filter Bubble" isolated from conflicting views – something that may lead them to become ever more extreme in their views (Pariser 2011).

Such research – as well as that into the other challenges mentioned above – is a prerequisite for employing semantic technologies in the places where they are most needed and can bring the most benefit: applications areas that are high in social complexity.

5 Conclusion

Research in semantic technologies strives to use context-awareness in all its facets within future applications. New dimensions are gained as soon as complex situations respecting large and small contexts on the one hand and dynamic, volatile contexts with changing references on the other hand can be adequately interpreted or inferred. Future challenges also are the well balanced usage of semantic technologies like semantic search, proactive complex event processing or social semantics whenever applicable and suitable to achieve better results or performance.

Personalization and individualization serve as a motor to promote research in modelling and extracting all forms of contexts. Future applications will show the potential still to be uncovered by context-aware semantic technologies.

References

R. Baeza-Yates, A. Gionis, F. Junqueira, V. Murdock, V. Plachouras, F. Silvestri, The impact of caching on search engines, in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'07)*, Amsterdam (ACM, New York, 2007), pp. 183–190. http://portal.acm.org/citation.cfm?id=1277741.1277775&coll=ACM&dl=ACM&CFID=36690833&CFTOKEN=36487312

K. Balog, L. Azzopardi, M. de Rijke, A language modeling framework for expert finding. Inf. Process. Manage. **45**(1), 1–19 (2009). http://dblp.uni-trier.de/db/journals/ipm/ipm45.html#BalogAR09

K. Balog, M. Bron, M. de Rijke, Query modeling for entity search based on terms, categories, examples. ACM Trans. Inf. Syst. (TOIS) **29**(4) (2011). http://dblp.uni-trier.de/db/journals/tois/tois29.html#BalogBR11

V. Bicer, T. Tran, R. Nedkov, Ranking support for keyword search on structured data using relevance models, in *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM'11)*, Glasgow, ed. by C. Macdonald, I. Ounis, I. Ruthven

(ACM, New York, 2011), pp. 1669–1678. http://dblp.uni-trier.de/db/conf/cikm/cikm2011.html#BicerTN11

R. Blanco, P. Mika, S. Vigna, Effective and efficient entity search in RDF data, in *The Semantic Web – ISWC 2011*, Bonn, ed. by L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N.F. Noy, E. Blomqvist. Volume 7031 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011), pp. 83–97. http://dblp.uni-trier.de/db/conf/semweb/iswc2011-1.html#BlancoMV11

O. Boyaci, V. Beltran Martinez, H. Schulzrinne, Bridging communications and the physical world. IEEE Internet Comput. **16**(2), 35–43 (2012). http://dblp.uni-trier.de/db/journals/internet/internet16.html#BoyaciBS12

G. Cheng, Y. Qu, Searching linked objects with falcons: approach, implementation and evaluation. Int. J. Semant. Web Inf. Syst. **5**(3), 49–70 (2009). http://dblp.uni-trier.de/db/journals/ijswis/ijswis5.html#ChengQ09

T. Cheng, X. Yan, K.C.C. Chang, EntityRank: searching entities directly and holistically, in *Proceedings of the 33rd International Conference on Very Large Data Bases (VLDB'07)*, Vienna, ed. by C. Koch, J. Gehrke, M.N. Garofalakis, D. Srivastava, K. Aberer, A. Deshpande, D. Florescu, C.Y. Chan, V. Ganti, C.C. Kanne, W. Klas, E.J. Neuhold (VLDB Endowment, 2007), pp. 387–398. http://dblp.uni-trier.de/db/conf/vldb/vldb2007.html#ChengYC07

J. Chu-Carroll, J.M. Prager, An experimental study of the impact of information extraction accuracy on semantic search performance, in *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management (CIKM'07)*, Lisbon, ed. by M.J. Silva, A.H.F. Laender, R.A. Baeza-Yates, D.L. McGuinness, B. Olstad, Y.H. Olsen, A.O. Falcão (ACM, New York, 2007), pp. 505–514. http://dblp.uni-trier.de/db/conf/cikm/cikm2007.html#Chu-CarrollP07

J. Chu-Carroll, J.M. Prager, K. Czuba, D.A. Ferrucci, P.A. Duboué, Semantic search via XML fragments: a high-precision approach to IR, in *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'06)*, Seattle, ed. by E.N. Efthimiadis, S.T. Dumais, D. Hawking, K. Järvelin (ACM, New York, 2006), pp. 445–452. http://dblp.uni-trier.de/db/conf/sigir/sigir2006.html#Chu-CarrollPCFD06

P. Cimiano, P. Haase, J. Heizmann, M. Mantel, R. Studer, Towards portable natural language interfaces to knowledge bases – the case of the ORAKEL system. Data Knowl. Eng. **65**(2), 325–354 (2008). http://dblp.uni-trier.de/db/journals/dke/dke65.html#CimianoHHMS08

J. Coffman, A.C. Weaver, A framework for evaluating database keyword search strategies, in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM'10)*, Toronto, ed. by J. Huang, N. Koudas, G.J.F. Jones, X. Wu, K. Collins-Thompson, A. An (ACM, New York, 2010), pp. 729–738. http://dblp.uni-trier.de/db/conf/cikm/cikm2010.html#CoffmanW10

J. Coffman, A.C. Weaver, Learning to rank results in relational keyword search, in *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM'11)*, Glasgow, ed. by C. Macdonald, I. Ounis, I. Ruthven (ACM, New York, 2011), pp. 1689–1698. http://dblp.uni-trier.de/db/conf/cikm/cikm2011.html#CoffmanW11

A. Di Iorio, A. Musetti, S. Peroni, F. Vitali, Crowdsourcing semantic content: a model and two applications, in *3rd Conference on Human System Interactions (HSI)*, Rzeszów, 2010, pp. 563–570

C. Duma, A. Herzog, N. Shahmehri, Privacy in the semantic web: what policy languages have to offer, in *Eighth IEEE International Workshop on Policies for Distributed Systems and Networks*, Bologna (IEEE Computer Society, 2007), pp. 109–118. http://dblp.uni-trier.de/db/conf/policy/policy2007.html#DumaHS07

Y. Engel, O. Etzion, Towards proactive event-driven computing, in *Proceedings of the 5th ACM International Conference on Distributed Event-Based System (DEBS'11)*, New York, ed. by D.M. Eyers, O. Etzion, A. Gal, S.B. Zdonik, P. Vincent (ACM, New York, 2011), pp. 125–136. http://dblp.uni-trier.de/db/conf/debs/debs2011.html#EngelE11

O. Etzion, Event processing: past, present and future, in *Proceedings of the VLDB Endowment*, Singapore, vol. 3, 2010, pp. 1651–1652

J. Fang, G. Lei, Calculation of weight of entities in ontologies by using usage information, in *3rd International Workshop on Intelligent Systems and Applications (ISA)*, Wuhan, 2011, pp. 1–4

T. Finin, R. Mizoguchi, S. Staab (eds.), *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 9(4) (Elsevier, Amsterdam, 2011)

F. Flöck, D. Vrandečić, E. Simperl, Towards a diversity-minded Wikipedia, in *Proceedings of the ACM 3rd International Conference on Web Science*, Koblenz, ed. by E. Munson, D.W. Hall, D. de Roure, S. Staab, 2011

R. Hasan, K. Siorpaes, R. Krummenacher, F. Flöck, Towards a knowledge diversity model, in *First International Workshop on Knowledge Diversity on the Web (DiversiWeb'11), 20th World Wide Web Conference (WWW'11)*, Hyderabad, ed. by E. Simperl, D. Madalli, D. Vrandecic, E. Alfonseca, 2011

C.A. Henson, A.P. Sheth, K. Thirunarayan, Semantic perception: converting sensory observations to abstractions. IEEE Internet Comput. **16**(2), 26–34 (2012). http://dblp.uni-trier.de/db/journals/internet/internet16.html#HensonST12

R. Kaptein, P. Serdyukov, A.P. de Vries, J. Kamps, Entity ranking using Wikipedia as a pivot, in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM'10)*, Toronto, ed. by J. Huang, N. Koudas, G.J.F. Jones, X. Wu, K. Collins-Thompson, A. An (ACM, New York, 2010), pp. 69–78. http://dblp.uni-trier.de/db/conf/cikm/cikm2010.html#KapteinSVK10

G. Ladwig, T. Tran, Index structures and top-k join algorithms for native keyword search databases, in *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM'11)*, Glasgow, C. Macdonald, I. Ounis, I. Ruthven (ACM, New York, 2011), pp. 1505–1514. http://dblp.uni-trier.de/db/conf/cikm/cikm2011.html#LadwigT11

Y. Li, I. Chaudhuri, H. Yang, S. Singh, H.V. Jagadish, Enabling domain-awareness for a generic natural language interface, in *Proceedings of the 22nd National Conference on Artificial Intelligence (AAAI'07)*, Vancouver (AAAI, 2007), pp. 833–838. http://dblp.uni-trier.de/db/conf/aaai/aaai2007.html#LiCYSJ07

F. Liu, C.T. Yu, W. Meng, A. Chowdhury, Effective keyword search in relational databases, in *Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data (SIGMOD'06)*, Chicago, ed. by S. Chaudhuri, V. Hristidis, N. Polyzotis (ACM, New York, 2006), pp. 563–574. http://dblp.uni-trier.de/db/conf/sigmod/sigmod2006.html#LiuYMC06

P. Mehra, Context-aware computing: beyond search and location-based services. IEEE Internet Comput. **16**(2), 12–16 (2012). http://dblp.uni-trier.de/db/journals/internet/internet16.html#Mehra12

Z. Nie, Y. Ma, S. Shi, J.R. Wen, W.Y. Ma, Web object retrieval, in *Proceedings of the 16th International Conference on World Wide Web (WWW'07)*, Banff (ACM, New York, 2007), pp. 81–90

E. Pariser, *The Filter Bubble: What the Internet Is Hiding from You* (The Penguin Press HC, New York, 2011)

J. Pehcevski, A.M. Vercoustre, J.A. Thom, exploiting locality of Wikipedia links in entity ranking, in *Advances in Information Retrieval*, ed. by C. Macdonald, I. Ounis, V. Plachouras, I. Ruthven, R.W. White. Volume 4956 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2008), pp. 258–269. http://dblp.uni-trier.de/db/conf/ecir/ecir2008.html#PehcevskiVT08

S. Schenk, C. Saathoff, S. Staab, A. Scherp, SemaPlorer – interactive semantic exploration of data and media based on a federated cloud infrastructure. J. Web Semant. **7**(4), 298–304 (2009)

S. Sen, N. Stojanovic, B.F. Shemrani, EchoPAT: a system for real-time complex event pattern monitoring, in J. Bacon, P.R. Pietzuch, J. Sventek, U. Çetintemel, *Proceedings of the Fourth ACM International Conference on Distributed Event-Based Systems*, Cambridge (ACM, New York, 2010) pp. 107–108. http://dblp.uni-trier.de/db/conf/debs/debs2010.html#SenSS10

A. Silberstein, G. Filpus, K. Munagala, J. Yang, Data-driven processing in sensor networks, in *Proceedings of the 2007 Conference on Innovative Data Systems Research*, Asilomar, 2007, pp. 10–21. http://dblp.uni-trier.de/db/conf/cidr/cidr2007.html#SilbersteinFMY07

N. Stojanovic, D. Milenovic, Y. Xu, L. Stojanovic, D. Anicic, R. Studer, An intelligent event-driven approach for efficient energy consumption in commercial buildings: smart office use case, in *Proceedings of the 5th ACM International Conference on Distributed Event-Based System (DEBS'11)*, New York, ed. by D.M. Eyers, O. Etzion, A. Gal, S.B. Zdonik, P. Vincent (ACM, New York, 2011a), pp. 303–312. http://dblp.uni-trier.de/db/conf/debs/debs2011.html#StojanovicMXSAS11

N. Stojanovic, L. Stojanovic, D. Anicic, J. Ma, S. Sen, R. Stühmer, Semantic complex event reasoning – beyond complex event processing, in *Foundations for the Web of Information and Services*, ed. by D. Fensel (Springer, Berlin/Heidelberg/New York, 2011b), pp. 253–279. http://dblp.uni-trier.de/db/conf/birthday/studer2011.html#StojanovicSAMSS11

G. Strube, B. Becker, C. Freksa, *Wörterbuch der Kognitionswissenschaft* (Klett-Cotta, Stuttgart, 1996)

A. Thalhammer, I. Toma, R. Hasan, E. Simperl, D. Vrandečić, How to represent knowledge diversity. Poster at 10th international semantic web conference (ISWC11), Bonn, Oct 2011

T. Tran, H. Wang, P. Haase, Hermes: data web search on a pay-as-you-go integration infrastructure. J. Web Semant. **7**(3), 189–203 (2009a). http://dblp.uni-trier.de/db/journals/ws/ws7.html#TranWH09

T. Tran, H. Wang, S. Rudolph, P. Cimiano, Top-k exploration of query candidates for efficient keyword search on graph-shaped (RDF) data, in *Proceedings of the 2009 IEEE International Conference on Data Engineering (ICDE'09)*, Shanghai, ed. by Y.E. Loannidis, D.L. Lee, R.T. Ng (IEEE Computer Society, Washington, DC, 2009b), pp. 405–416. http://dblp.uni-trier.de/db/conf/icde/icde2009.html#TranWRC09

G. Tummarello, R. Cyganiak, M. Catasta, S. Danielczyk, R. Delbru, S. Decker, Sig.ma: live views on the web of data. J. Web Semant. **8**(4), 355–364 (2010). http://dblp.uni-trier.de/db/journals/ws/ws8.html#TummarelloCCDDD10

E.M. Voorhees, Query expansion using lexical-semantic relations, in *Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval (SIGIR'94)*, Dublin (Springer, Berlin/Heidelberg/New York, 1994), pp. 61–69. http://dblp.uni-trier.de/db/conf/sigir/sigir94.html#Voorhees94

C. Wang, M. Xiong, Q. Zhou, Y. Yu, PANTO: a portable natural language interface to ontologies, in *The Semantic Web: Research and Applications*, ed. by E. Franconi, M. Kifer, W. May. Volume 4519 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2007), pp. 473–487. http://dblp.uni-trier.de/db/conf/esws/eswc2007.html#WangXZY07

E. Yndurain, D. Bernhardt, C. Campo, Augmenting mobile search engines to leverage context awareness. IEEE Internet Comput. **16**(2), 17–25 (2012). http://dblp.uni-trier.de/db/journals/internet/internet16.html#YndurainBC12

THESEUS: A Successful First Step

Henning Kagermann

Through the work completed as part of the THESEUS research program, Germany is taking on a pioneering role in the use of semantic technologies and the creation of new standards for the Internet of Services. This work has gone on to gain international recognition.

THESEUS has produced a host of results and technical innovations. So far, over 50 patents and other protected results have been registered as part of the research program, and a multitude of standardization activities have been launched. This was in addition to the development of over 100 prototypes, the founding of five companies, and the publishing of more than 800 scientific papers. The research program has shown the success that can be achieved through effective cooperation between business, science and government, and that this can serve to keep Germany competitive in the future.

Indeed THESEUS can boast very successful results, but it is, in fact, just a successful first step on a longer term roadmap: THESEUS is also an important milestone on the pathway to the future of a web- and knowledge-based services society. Germany's success as an exporting country is based on its strong industrial core. Thanks to a high level of competitiveness, it has been able to successfully overcome the economic and financial crisis of the recent past. However, the solid German economy also faces great future challenges. Prosperity and social security in Europe can only be safeguarded if the strong industrial heartlands are successfully complemented with services that are close to the needs of business, and with new Internet-based business models which enable the development of new markets and new stimuli for growth and added value.

THESEUS is thus securing the future of the German model of success since its results form a platform for the next paradigm shift, which can be described as "Industrie 4.0". This means that our economy is standing on the threshold to

H. Kagermann (✉)
acatech, Munich, Germany
e-mail: kagermann@acatech.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_5,

the Fourth Industrial Revolution. Thanks to the power of the Internet, the real and the virtual worlds are growing closer together. The Internet of Things will be used in the factory. How quickly this will happen, and how much is at stake, can be seen in an example from another sector – smartphones, where telecommunications technologies merge together with information technologies. The result was that, in the space of just a few years, a large part of the value added from Europe flowed across into Silicon Valley, because it was there that Apple and Google were able to develop solutions which met customers' needs.

"Industrie 4.0" describes the future form of industrial production, which is characterized by a high level of product customization made possible by high-flexibility, resource and energy-efficient production. This involves the extensive integration of customers and business partners in dynamic, real-time-optimized value chains and the coupling of production and quality services to produce hybrid products. "Industrie 4.0" can thus be used to address many challenges of our time in a new network economy of greater automation, higher flexibility, lower use of resources, and age-appropriate jobs close to urban areas. German industry now has the opportunity to use the results of THESEUS as a basis to actively shape the Fourth Industrial Revolution.

Although THESEUS represents just an intermediate step on the roadmap to the Fourth Industrial Revolution, it is worth casting a look back over where we have come from. Have the respective goals, assumptions and expectations held by government and industrial companies/SMEs gone on to be met?

At the first IT Summit in 2006, the German Federal Government laid the foundations for a marked improvement in Germany's international competitive position in the field of ICT. The need for this initiative was quite clear: in the relevant international rankings, Germany was lagging far behind its international competitors. However, information and communication technologies are drivers of innovation and growth for the whole of the German economy, and are thus decisive for the international competitiveness of German industry. Today, it is plain to see that THESEUS has been able to play a part in this.

The specific goals of the research program were developed 2 years before its actual inception. At this time, the research program was launched in partnership with France under the title "Quaero". France later decided to focus upon the development of search engines – a path which it went on to pursue alone. However, German researchers agreed that it was not further search engines that were needed, but the development of a "response engine". The strategic goal formulated at this time was to develop products, business models and markets that would enable companies and consumers everywhere to access services, content and knowledge at any time. Six years on, we can now see that THESEUS has enabled these goals to be achieved.

The pertinent questions that dictate discussions on the development of the industry today were already evident at that time. In terms of priority-setting, a large proportion of companies across the world favor the development of business model innovations above that of product innovations. Then as now, hardware and software developers merged to form typical hybrid service providers. Since the start of the research program, the focus has been on tailored solutions and not on individual products. Six years ago, it was also already apparent that the development and winning potential of Service Oriented Architecture (SOA) threatened to overthrow the conventional business model of software localization. The theory at that time was that the new convention of using something without actually owning it was something that would prevail, and would revolutionize typical usage, simply because it enabled such high efficiency gains. We all know that we are well on the way towards achieving this. Today, we are able to see once again that this original question and the other key questions set by the THESEUS research program at the very beginning were still the right ones. And many of these questions could be answered.

The aim of THESEUS was to create a new Internet-based knowledge infrastructure which allowed the simplified and targeted use of knowledge available on the Internet. The classic web has drastically improved worldwide access to digitally stored information. But availability of information is not everything. Information does not become knowledge until there are technologies that are able to offer high-precision responses. However, still today, content is machine-*readable* but cannot be *understood* by machines.

In contrast, the Semantic Web is based on the substantive description of digital documents with standardized vocabularies, which use semantics that can be understood by machines. The transformation towards the use of semantics takes us from a syntactic web with floods of information, via the reduction of textual documents to meaningless combinations of letters, to high-precision responses in a web based upon relationships of meaning.

Knowledge that is processed in this way can be scanned more quickly and efficiently, and also offers a means of assimilating information that is adapted to human needs and user habits – namely the creation of associations between new knowledge and existing knowledge.

The development of new semantic technologies is therefore of central importance for the Internet of Services. In general semiotics, the word semantics denotes the theory of meaning. Semantic technologies bridge the gap between specialist language within informatics and the languages of their users, allowing translation between different systems of terms, without the loss of meaning. This means that in the future, a specialist can formulate his knowledge in the relevant specialist language; he does not then need a computer scientist to translate everything once more. This will serve to considerably accelerate innovations.

The research program has therefore developed basic technologies for the Internet of Services which, in the future, will enable machines to understand content and models that they will be able to mechanically learn, and ultimately use to communicate with one another and to make autonomous decisions.

In addition to semantic technologies, the Internet of Services also requires the possibility of creating digital services (Software as a Service/SaaS) straightforwardly in order to establish this Internet-based services sector. Once THESEUS has been used to assign an adequate semantic description to these services, they can then be combined in a way which allows them to be reused, and can also be extended in a flexible manner. The service user pays a usage-based fee for usage and operational costs, instead of paying expensive licence fees for software packages. He or she also receives a high level of transparency and flexibility and can focus on his or her core business while preserving liquidity.

For many consumers and companies, the search for appropriate services on the Internet has been a laborious and drawn-out process. Innovative and complex business models often require the combination of diverse "part-services" from different providers. The ability to combine these on a single platform would greatly simplify the trade in services.

While the online trade in content such as music and videos is already routine, web-based services have not yet been produced, offered, linked, and used to the same degree. iTunes, Internet video rental services, and App stores already offer consumers comparable advantages that are enjoyed on a daily basis, but such benefits have been wholly unavailable to small and medium-sized companies. THESEUS now makes it possible to establish an infrastructure on the Internet which enables services online to be much more easily found, combined, used, and paid for. Like apps, these digital services will bundle existing information on the Internet and will use it for a specific, clear, and closely programmed task. This will be an important step on the pathway towards the Internet of Services. The key factor in this process is that these technologies, business apps, and services must be accessible on an open platform in a cloud so that all developers and market participants can quickly and simply build web-based applications, services and new business models (Platform as a Service/PaaS). This will enable products from the real economy to be linked with high-quality services, thus paving the way for global networks and new forms of collaboration between various international partners. This makes the interoperability, transparency, and complete openness of these electronic platforms and the scalability of capacity and storage crucially important for the Internet of Services. In addition, data security, and the availability of a comprehensive infrastructure that provides the necessary transmission speeds, are both vital. Private and public investors in Germany and Europe should therefore invest in this infrastructure so that a secure and high-performance cloud can be created that also conforms to our regulations and strict standards.

In view of the importance of cloud computing for Germany as a center for technology and business, the German Federal Ministry for Economic Affairs and Energy has launched the cloud computing action program in order to speed up its expansion. A key objective within this program is to use the traditional strengths of German IT providers for the development of cloud computing. These notably include reliability, quality, security and conformity with data protection laws, and closeness to research institutes. The trademark "Cloud Computing – Made in Germany" is to build upon this basis.

The expansion of the cloud infrastructure is a considerable milestone for both Europe and Germany on the pathway towards the Internet of Services. In the meantime, the task is to prepare the commercialization of the THESEUS research results, and to prove that such a program is both practicable and profitable.

The restructuring of Germany's energy supply could play a key role in this task. Smart meters (smart grids) require the technologies developed within the THESEUS research program. The restructuring of Germany's energy supply means that in the future, the new distributed energy supply involving numerous consumers and producers must be intelligently managed. This challenge could become one of the first major tests of the Internet of Services.

THESEUS opens up particularly big opportunities for small and medium-sized companies to offer their services via the Internet. It will enable them to present these services worldwide and win new customers. Small and medium-sized companies can also easily extend their own range of services by combining them with those of other providers. These innovative service packages will certainly enable SMEs to compete against large corporations and to tap into new markets.

THESEUS has enabled Germany to acquire valuable knowledge. The research program had the right vision and strategic direction. We have successfully attained important technological results that address a number of economic and social issues. This short-term advantage must be retained. To ensure that we do this, it would make good sense to establish a center of excellence that rolls out the new technologies and acts as a central contact point for all smaller companies. The new knowledge should also be anchored in university education and professional training, and should be internationally protected through the use of standards.

THESEUS has not only created the basis for the Internet of Services, but has also opened wide the door to other future developments. The merging of the physical world with cyberspace is continuing. Embedded systems of electronics and software play a key role as drivers of innovation for export and high-growth markets for German industry. They significantly raise the functionality and thus the utility value – as well as increase the added value – of vehicles, aeroplanes, medical devices, production plants, and household appliances.

Today, around 98 % of microprocessors are embedded and are connected to the outside world via sensors and actuators. They are increasingly being networked with one another and connected online. This results in Cyber Physical Systems (CPSs), which are part of the globally networked world that we will see in the future, in which products, devices, and objects with embedded hardware and software will interact across application boundaries. With the help of sensors, these systems process data from the physical world and make it available for Internet-based services that can directly act upon processes in the physical world.

Cyber Physical Systems challenge both the classic limits delineating sector and specialization, as well as established business models. But there are sizable challenges associated with CPSs, such as operational safety, IT security and the protection of privacy, which are among the key issues in this context.

Anyone who seeks to gain international technological leadership and to close in on the German lead must demonstrate unparalleled achievement in system integration. The strategic and economic benefits for Germany are enormous since our strengths lie in integrated IT solutions. Cyber Physical Systems are making enormous efficiency gains, not least in production. Thanks to CPS production, the factory of the future will be able to integrate supply chains and individual customer requests in real time.

CPSs are also becoming drivers of innovation in the fields of energy and climate, mobility and logistics, and health and healthcare. Future Cyber Physical Systems will help us to achieve better security, efficiency, comfort and health in ways that have, thus far, been almost unimaginable. In this way, they will play a role in solving the central challenges facing our society, such as the ageing population, scarcity of resources, mobility, and the restructuring of Germany's energy supply, to name but a few fields of application.

Cyber Physical Systems will connect the physical world with the virtual world to create an Internet of Things, Data, and Services. They will also increasingly do the same for humans, who will be connected to this virtual world through the use of ever more diverse human-machine interfaces extending far beyond today's smartphone. Furthermore, complex Cyber Physical Systems also represent a strong competitive advantage and key factor within the labor market – even for Germany as an established location for high technology. This is seen both in cross-domain sales and employment figures, as well as in the increasing demand for German expertise in this area coming from foreign investors.

Medium-sized and large companies are using these technologies to develop and gain an ever stronger position in new markets worldwide. The THESEUS research program has opened up the way, and will continue to play an important role in the future of a web- and knowledge-based services society. Without THESEUS, this fascinating future of the Fourth Industrial Revolution would be unthinkable. THESEUS has thus made a major contribution to the Internet of Services and to the establishment of the web-based services society, which secures the level of growth necessary to uphold and further expand Germany's economic growth and increase its competitiveness.

Part II
Core Technologies

Core Technologies for the Internet of Services

Tilman Becker, Catherina Burghart, Kawa Nazemi, Patrick Ndjiki-Nya, Thomas Riegel, Ralf Schäfer, Thomas Sporer, Volker Tresp, and Jens Wissmann

Abstract Information and knowledge are growing permanently and represent valuable resources for many enterprises. The efficient access to knowledge of an enterprise like expertise, contact persons, project and milestone plans, etc. may simplify business processes and lead to time and cost savings. Semantic technologies offer numerous possibilities to enrich data with background information about their meaning. Such semantic relations do not only lead to more efficient search in larger information repositories but they also assist the user in diverse processes like editing, annotation and processing of information. In addition they offer new means of access and transfer of knowledge. Each information unit is linked to other units in the same domain, which allows faster search and offers a way of information access that is close to the habits of humans, i.e. the creation of new

T. Becker (✉)
DFKI GmbH, German Research Center of Artificial Intelligence, Saarbrücken, Germany
e-mail: tilman.becker@dfki.de

C. Burghart • J. Wissmann
FZI, Forschungszentrum Informatik, Karlsruhe, Germany
e-mail: burghart@fzi.de; wissmann@fzi.de

K. Nazemi
Fraunhofer Institute for Computer Graphics Research (IGD), Darmstadt, Germany
e-mail: kawa.nazemi@igd.fraunhofer.de

P. Ndjiki-Nya • R. Schäfer
Fraunhofer Institute for Telecommunications (HHI), Berlin, Germany
e-mail: patrick.ndjiki-nya@hhi.fraunhofer.de; ralf.schaefer@hhi.fraunhofer.de

T. Riegel • V. Tresp
Siemens AG, Munich, Germany
e-mail: thomas.riegel@siemens.com; Volker.Tresp@siemens.com

T. Sporer
Fraunhofer Institute for Digital Media Technology (IDMT), Ilmenau, Germany
e-mail: spo@idmt.fraunhofer.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_6,

knowledge and its association to already existing knowledge. This approach of the whole THESEUS research program was also the basis of the CORE TECHNOLOGY CLUSTER of THESEUS.

1 Introduction

The CORE TECHNOLOGY CLUSTER (CTC) had been established in parallel to the different use cases in order to research and develop innovative technologies, which are common and required by several of these use cases in order to avoid parallel and separate developments. The CTC was subdivided into *Work Packages*, five of which (WP2–WP6) were dedicated to specific technologies and one (WP8) to the evaluation of these technologies according to international benchmarks such as ImageClef (Müller and Tsikrika 2010) or TRECVID (Smeaton et al. 2006):

- WP1 – "Program Management"
- WP2 – "Video, Metadata, Platforms – Processing of Multimedia Data"
- WP3 – "Ontology Management"
- WP4 – "Situation aware Dialog Processing"
- WP5 – "Innovative User Interfaces and Visualization"
- WP6 – "Statistical Machine Learning"
- WP7 – (canceled in the planning phase)
- WP8 – "Evaluation"

Accordingly, this article is subdivided into six sections, each of them dealing with a technology mentioned above. The first section deals with new methods for image and video processing, and for semantic search in multimedia archives. Ontologies are formal knowledge models that conceptually represent the knowledge within a given subject area and make it possible to process that knowledge automatically at the level of meaning. Different tools for ontology design, for the mapping of different, heterogeneous knowledge structures, for the tracking of modifications (evolution) and for reasoning have been developed and will be explained in detail. The next section describes a dialog shell, which is an architecture and development platform for new means of interaction. A platform for multimodal and situation aware dialog systems has been created, whose architecture and some applications are described. A further core technology is semantic visualization, which is a graphical presentation of information to enhance decision making through visual processing of complex interdependencies. Last but not least machine learning methods have been developed, which are intelligent data analysis processes that facilitate automatic recognition of data relationships and interconnections so that they can be modeled and structured. These methods are applied to texts, images and audio and video data, and they help identify relationships between different types of data. The last section is dedicated to the evolution of the above mentioned technologies. Experts were assessing the quality of the basic technologies developed within the framework of THESEUS. The developed technologies have been tested to determine their reliability, functionality and suitability, in an effort to ensure that the research meets international quality standards. The results of this evaluation were also taken into account in the research and development process, helping to further optimize end results.

The CTC started its work in February 2008 and the last activities ended in March 2012. The following companies and organizations were involved in the CTC (lead marked in bold):

- DFKI GmbH, German Research Center of Artificial Intelligence, Saarbrücken and Kaiserslautern (WP3, **WP4**),
- Fraunhofer Institute for Computer Architecture and Software Technology (FIRST), Berlin (WP6, WP8),
- Fraunhofer Institute for Telecommunications (HHI), Berlin (**CTC**, **WP1**, **WP2**),
- Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS), Birlinghoven (WP6),
- Fraunhofer Institute for Computer Graphics Research (IGD), Darmstadt (WP2, **WP5**),
- Fraunhofer Institute for Digital Media Technology (IDMT), Ilmenau (**WP8**),
- Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB), Karlsruhe (WP3),
- FZI, Forschungszentrum Informatik, Karlsruhe (**WP3**, WP4),
- Ludwig Maximilian University, Munich (WP6),
- Siemens AG, Munich (**WP6**),
- Machine Learning Group, TU Berlin (WP6).

2 WP2: Video, Metadata, Platforms: Processing of Multimedia Data

The amount of digital images and movies is ever increasing. To effectively manage this flood of content in archives, new methods for image and video processing and for semantic search in multimedia archives are required. The Work Package 2 has therefore developed innovative solutions for the processing of multimedia content within the THESEUS CORE TECHNOLOGY CLUSTER.

An overview of the WP2 tasks is given in Fig. 1. It is agreed that manual annotation of digital images and movies would yield tremendous costs. Therefore, only a very limited amount of metadata is available for a vast proportion of audio-visual media.

To effectively manage this flood of content in archives, new methods for image and video processing and for semantic search in multimedia archives were required. With respect to related key challenges, the work in WP2 was structured around four major axes:

- Image and Video Recognition
- Perceptual Hashing

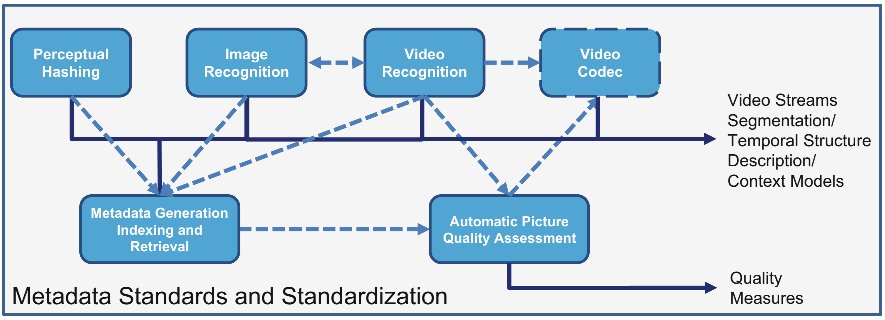


Fig. 1 Summary of the WP2 tasks

- Automatic Picture Quality Assessment
- Metadata Generation, Indexing and Retrieval.

Some peripheral activity was also conducted on video coding in order to generate transmittable video streams.

The *Image and Video Recognition* techniques aim at automatically extracting (vision-based) *metadata*, the fuel of any semantic search engine. Another kind of meta-information was explored in the *Perceptual Hashing* task, where the identification of several versions (e.g. image/video compression or changes in brightness) of the same content was the relevant use case. Automatic methods for the *Quality Assessment* of images/videos for their qualitative comparison were also addressed.

Searching videos and images in a semantic manner requires the consideration, utilization, and integration of the entire spectrum of the available information, from image analysis results to relevant context and background knowledge. To achieve this, a *modular framework for solving semantic queries* in videos and images has been developed.

Recognize and Classify Manual annotation of digital images and videos is extremely expensive. Often, only a very limited amount of metadata is available for audio visual media. Within WP2, one important goal was to develop image and video recognition techniques that are able to automatically extract important metadata. One of the fields focused on was semantic classification of images. The computer uses machine learning algorithms to learn which images belong to or contain given concepts, such as “outdoors”, “vehicle”, or “city.” Afterwards, the trained system can be used to detect these concepts in new, previously unseen images. Besides these general purpose concept detectors, very specific model-based concept detectors were developed, e.g. for the detection of faces within images and videos. The proposed face detector is very robust with respect to lighting conditions. Besides the position of faces, it can also classify their pose, i.e. the direction of the face.

Different approaches can be used to transfer algorithms designed for images into the video domain. One possibility is to properly extract a set of (key-) frames and use algorithms on these extracted frames (images). Within WP2, a temporal video segmentation algorithm was developed that is able to detect the temporal hierarchical structure of videos. It can extract scenes, shots, and sub-shots and is able to also extract key-frames from these temporal units. These key-frames are then utilized for further analysis or for the construction of a visual table-of-contents that assists the user in browsing and navigating through the video.

Enabling computer programs to automatically detect the content of a scene, the various settings and the genre of a film sequence was one further goal. In addition, mechanisms were investigated and implemented, which combine low-level features of the content (e.g. color, contour and motion) with information about the context and thereby derive complex relationships.

Effective Feature Storage and Indexing For search engines, two main development goals have to be reached: The results have to be of high quality and the amount of time needed to do the search and return results has to be low. The second goal can be addressed by the development of compression and indexing methods.

Compression methods are used to reduce the dimensionality of features, i.e. the amount of metadata per item to be compared during a search is reduced. This has to be done, however, without sacrificing too much in terms of search quality.

Indexing methods are the second way of limiting the amount of time needed for a search. A linear search for example has a time duration or hardware complexity that grows linearly with the number of items in the database. Some archives have grown to a point where this is infeasible. Therefore, algorithms are needed that perform faster than linear variants. In WP2, a tree-based indexing method was developed that has a logarithmic complexity during search, which means doubling the archive size does not double the time needed for a search. Metadata nodes are chosen that are representative of other metadata nodes. In the search, nodes can be skipped if their representative nodes are not similar enough to what is searched for. This way, only a small set of the metadata in the database actually has to be compared for a search. Most of it can be skipped which considerably reduces search complexity (Ciaccia et al. 1997; Skopal et al. 2003).

Framework for Semantic Search in Media Data Searching videos and images in a semantic manner requires the consideration, utilization, and integration of the entire spectrum of the available information, from image analysis results to relevant context and background knowledge. To achieve this, a modular framework for solving semantic queries in videos and images has been developed. It allows for exploiting the (inter-) relations between objects/events across multiple images, the situational context, and the application context, e.g. for identifying persons in video sequences by considering contextual information or detecting events in a video sequence without direct observation. This evidence-based detection/recognition is carried out using a novel calculus approach ("Subjective Logic"), which allows us to aggregate different evidence for and against a hypothesis, hence deriving a degree of belief for it. This is achieved by taking into account all the available evidence, including the inherent uncertainties (accounting for the vagueness of image/video analysis results).

For a unified access to distributed and heterogeneous (multimedia) data resources, a middleware component has been developed. This “Query Broker” acts as a mediator between a client and multiple heterogeneous retrieval engines or database systems, encapsulating all retrieval functionalities in one component and exposing a standardized query format. The retrieval process is further supported by modules for intelligent query segmentation and distribution, as well as federated result set aggregation.

The Query Broker was designed to support the integration of heterogeneous knowledge bases into the global query evaluation process, enabling semantically enriched retrieval. In addition to features available in meta-search engines, the Query Broker also supports specific multimedia retrieval paradigms (e.g., query by example), as well as cross system multimedia retrieval (i.e. both cross metadata and cross query language). In this context, WP2 additionally promoted the international standardization of MPQF (at ISO/MPEG) on search support for semantic concepts and on JPSearch (at ISO/JPEG) through successful contributions.

Quality Assessment Quality assessment plays an important role in various image and video processing applications. Although human observers can effectively and easily judge the quality of images and videos, subjective measurements cannot be integrated into real-time image/video processing systems because of high runtimes and processing costs. Therefore, a lot of research has focused on objective metrics with the intention of addressing the drawbacks of the subjective metrics. Nevertheless, it is difficult for objective metrics to assess image quality in a manner similar to human perception.

Three measurement classes, i.e. full-reference (FR), reduced-reference (RR) and no-reference (NR), are proposed in the area of objective image quality assessment. The FR metrics predict the quality of an image based on differences to a reference image. Mean Square Error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Structure Similarity (SSIM) are widely used FR metrics. On the contrary to RR metrics, which use indirect information as a reference, NR metrics predict the quality of images by extracting and modeling prior knowledge on distortions (Wang and Bovik 2005).

The quality of images and videos is evaluated and assessed by humans in a highly complex process within a split second. Mimicking this natural process using technology is one of the most difficult tasks in image and video processing. WP2 had the objective to develop automatic methods for the quality assessment of images, thereby facilitating, inter alia, the qualitative comparison of different images. The automatic metrics incorporate perceptual quality measures by modeling human visual system characteristics.

Copy Detection In order to identify several versions of the same content in archives quickly, perceptual hashing algorithms can be applied. These algorithms extract characteristic features from multimedia content and combine them to form a small "digital fingerprint", the so-called perceptual hash. This hash value is robust against several modifications of the content, such as image/video compression or changes in brightness. This property is crucial, as it enables the detection of content after it has undergone processing steps which are usual in image or video processing. Various application scenarios exist for this method. In addition to the identification of content in archives, it is also possible to detect copyright infringements on the Internet, to monitor the broadcast of commercials, or to link the content with additional information (such as content owner or copyrights) stored in a database. Furthermore, overlapping sections of video snippets can be detected, which allows for a reconstitution of the full video.

The robustness against distortions distinguishes these methods from cryptographic hash values, like SHA-1 (National Institute of Standards and Technology (NIST) 2002). The objective of those methods is to detect if the content has been modified. Therefore, they are bit sensitive, meaning that the hash value changes completely when a single bit of the input data changes. In contrast, perceptual hashing algorithms should be robust against small modifications. Content originating from the same source should get the same or a similar hash value. This enables identifying copies of the same content based on the distance between the hash values. The advantage compared to watermarking techniques, which have similar application areas, is that the content does not have to be labeled before it can be detected. Therefore, it is possible to detect already distributed content. WP2 has developed robust perceptual hashing methods which enable the identification of images and videos.

3 WP3: Ontology Management

Research and design of semantic technologies based on ontologies is one of the most prominent research topics within the THESEUS research program. These technologies enable computers to "understand" the meaning of an object or symbols with the help of ontologies, which are knowledge bases of formal models conceptually describing domain knowledge. Automatic processing of ontologies allows semantic interpretations – a typical characteristic of a human being. Using methodologies for structural processing and logical deductions permits ontology-based information systems to achieve intelligent behavior: i.e. within THESEUS anatomical relations between organic structures are modeled in order to enhance search in medical information systems. Another use case supports knowledge management within production and distribution of goods by describing functions and characteristics of industrial automation.

Ontology management actually denotes the handling of knowledge bases. Knowledge itself is formally represented and modelled by using either first order logic or an ontology language: within THESEUS we have agreed to support ontologies modelled by one of the W3C standards OWL or RDF(S).

The goal of WP3 is to provide an *ontology management* (OM) toolbox to be used by use cases and other CTC work packages for the handling of ontologies and semantic metadata. Management of ontologies comprises four different tasks within the THESEUS research program: ontology design, mapping of heterogeneous knowledge structures to one another, tracking of changes to ontologies (evolution), automatic deductions (reasoning) and semantic search. Within these four tasks several components were developed which can be used by other partners of THESEUS to manage ontologies; some of the tools are available in different platforms like NEON, webGenesis, information workbench or SMILA. Several of these tools can also be found as open-source plug-ins on sourceforge.net.

We partition the field of OM into several subjects concerned with the maintenance and the processing of ontologies:

- *Persistence* In addition to programmatic in-memory manipulation, ontologies need to be stored in databases for their persistence and efficient access. In particular for large ontologies a database persistence layer is required to ensure their scalable handling.
- *Change Management* The changes made to ontologies over time need to be maintained explicitly in order to allow for features like versioning and intelligent update management. This also comprises ontology evolution techniques that take into account the semantics of changes, such as quality management and semantic consolidation.
- *Alignment* For bridging the heterogeneity between several ontology models, techniques for ontology alignment need to be provided. They comprise ontology mapping, for finding corresponding entities across ontologies, as well as ontology merging, for integrating ontologies based on such correspondences.
- *Reasoning* The process of reasoning checks an ontology for its consistency or derives new knowledge that logically follows from what was explicitly stated in the ontology. *Reasoning brokerage* allows for the dynamic combined use of different features and advantages of several reasoners, while *approximate reasoning* addresses scalability issues by trading correctness for speed.
- *Querying* Mostly related to reasoning, the process of querying retrieves relevant entities from ontologies. Typically, derived knowledge is taken into account for the generation of answers.
- *Entity Disambiguation* Apart from direct manipulation of ontologies, the process of ontology engineering needs to be supported in various ways. One such way of support is to use a tool that disambiguates the meaning of entities that occur in ontologies (or related text documents) based on their context.

In the following, the main components[1] designed and implemented within the THESEUS research program are sketched:

[1]The OM components are named after gods or heroes of Greek mythology in accordance with naming in the THESEUS research program, and with some meaning related to the component's functionality.

- *MNEMOSYNE* augments the OWL API by a persistence layer for the (native) storage of OWL ontologies in a relational database. This aims towards the scalable handling of large ontology models, for which the current in-memory implementation in the OWL API is insufficient.
- *ARETE* performs an automated recognition of references between natural language text elements and entities in an ontology.
- *ACHILLES* provides an adapter from the OWL API to RDF stores. It enables the programmatic handling of RDFS ontologies from within the OWL API, and thus bridges between the ontology languages OWL and RDF(S).
- *HARMONIA* addresses the problem of semi-automated ontology mapping for identifying correspondences between entities in different ontologies that cover similar or overlapping domains of interest.
- *CHRONOS* aims at providing a change management system for OWL ontologies that covers aspects of versioning and quality management. In its current state it offers tools for taxonomy clean-up and redundancy consolidation.
- *HERAKLES* provides a reasoning broker system that allows the use of various OWL reasoners in a combined way. It supports scenarios where several reasoners run in parallel on remote machines and where appropriate reasoners are selected at runtime.
- *DELPHI* encompasses systems for approximate reasoning with OWL ontologies that trade a certain degree of correctness of reasoning results for a speed-up in runtime performance. This component has been integrated into the reasoning broker HERAKLES.
- *ATLAS* is a reasoning framework for OWL 2 Full ontologies and RDFS. It is based on the working principle of theorem provers and thus offers a reasoning methodology different from known OWL-DL reasoners.
- *PYTHIA* provides the possibility of accessing OWL ontologies through structured queries stated in the query language SPARQL.
- *KOIOS* is a semantic search engine that enables keyword search on graph-shaped (RDF) Data. For a given keyword query, KOIOS computes k relevant SPARQL queries, answers to which are retrieved from an inverted keyword index and a special summary graph of the origin data. Additionally, KOIOS combines search within ontologies, databases and texts. Search results are returned according to relevance.
- *HEPHAISTOS* supports a rule-based transformation of RDF(S) data in the context of generating explanations for reasoning results.

Figure 2 details the subjects of ontology management, arranged on the two layers of maintenance activities and processing activities for ontologies, respectively.

In the field of ontology design three components have been developed: two for persistent storage of large ontologies, another for disambiguation of entities in texts using associated ontologies. Up to now, it is rather tiresome and problematic to load large scale ontologies into an ontology editor and manipulate them. MNEMOSYNE solves this issue by using an object relational representation based on the open-source framework Hibernate and translating accesses to the ontology directly into

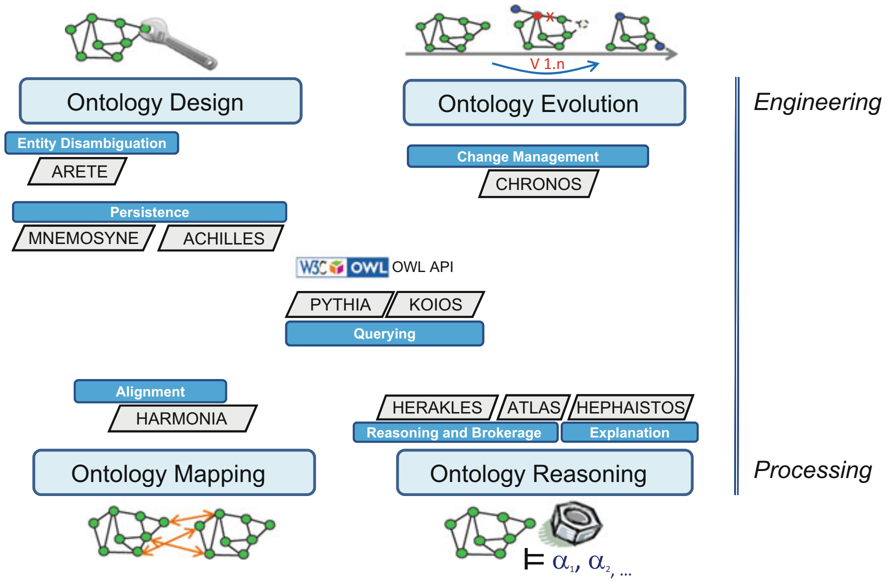


Fig. 2 Subjects of ontology management and their relation to WP3 components and Semantic Web technology

database queries via the OWL-API. The result of the query is transformed into axiom objects and entity objects. Therefore, only presently used objects of the ontology have to be kept in the memory when manipulating the ontologies. Graph-based queries in SPARQL can be made using the component PYTHIA, which directly interacts with MNEMOSYNE. The third component, called ARETE, is used for disambiguation of occurring ambiguities within texts, provided found text snippets refer to entities of an ontology at the same time, i.e. an ontology for geolocations can be used to distinguish between names referring to totally different locations, towns or cities. Disambiguation is achieved by using the spreading activation technique, which operates on partial graphs of the associated ontology with the help of statistical methods (Kleb and Abecker 2010).

Ontology mapping plays an important role in order to align different knowledge bases describing the same application domain but representing heterogeneous foci. Within this task a new component was designed and implemented using two different technologies: on the one hand a particle swarm algorithm is used to obtain alignments between two different ontologies (Bock 2010; Bock and Hettenhausen 2012); on the other hand correspondences between knowledge bases can be found with the help of evolutionary techniques. Both methodologies are incorporated within the component HARMONIA. Ontologies usually are built by different designers or ontology experts and thus have a subset of common representations or overlaps. These correspondences have to be automatically identified using HARMONIA. The results of this alignment process can then be used to form new, bigger ontologies by adding the alignments to the two analyzed ontologies and obtaining a new, composite knowledge base. In this way the search space for queries can be enlarged to an enormous extent. This can be also illustrated by an example within the SUI (Search in Environmental Information Systems) research program using the components developed within THESEUS. Here, three different ontologies (GEMIT, catalogue of object types and ontology of life situations) are combined in order to search a much larger set of data when looking for building space within a given region. Without mapping these ontologies a person wishing to build a house in a specific region would only receive a subset or no results when entering the request in the environmental portal of the SUI server. For 3 years HARMONIA has also participated in the annual OAEI (ontology alignment) contest in the directory track and has there proven its capability to hold its own with the best mapping software programs. Additionally, the designed mapping component can exploit parallel computing infrastructures due to the techniques applied. Therefore, large ontologies can be aligned by exploiting parallelization and thus allowing good scalability. The algorithms can also be used in a cloud and have already been tested there (Bock et al. 2010).

There are many scenarios within which ontologies are developed on a collaborative basis, but if many different persons change, add or delete entities and data within a knowledge base, inconsistencies, mistakes and important differences can be the result. This issue is addressed by the third task of ontology management: ontology evolution. Here, a newly developed component called CHRONOS keeps track of different versions of an ontology, computes differences between different versions, performs repairs, executes unit tests, and corrects staging (Grimm and Wissmann 2011). CHRONOS is responsible for the grooming of ontologies and thus for quality assurance and quality control. With the help of a specially designed graphical interface users can design individual grooming processes for their ontologies. Additionally, the pursuit of development within a specific branch of an ontology is supported as well as the division of an ontology into different modules and the integration of new knowledge domains.

The fourth task within the ontology management work package focuses on ontology reasoning and semantic search. A specially designed reasoning broker called HERAKLES enables a processing of an ontology by several reasoning tools in parallel as well as the selection of the most suitable reasoning tool, if it is connected to the reasoning broker. Here, HERAKLES performs the communication between broker kernel and external reasoners via OWLLink. In this way, external reasoners can be addressed without explicit integration into the reasoning broker (Bock et al. 2009). Depending on the structure of the ontology some reasoners are more applicable than others. If there is a complex structure of several concepts and relations (called T-Box) another reasoner is chosen than if there were large scale data and a rather simple structure of the ontology (this is called A-Box). By analyzing the structure HERAKLES selects the appropriate reasoning tool and reasoning strategy. Two new reasoning techniques were also added to the broker system: approximate reasoning and anytime reasoning. In cases of very large knowledge bases it can be useful to start the processing of reasoning results before the actual reasoning process has ended. In this case all results are continuously published while the reasoning tool is still working. Approximate reasoning produces either results which are all correct but do not cover the complete result set or all correct results plus some additional incorrect results. As this reasoning tool runs faster than others a slower reasoner running in parallel can be used to produce missing results or verify the output.

Finally, semantic search is the last semantic tool category designed and implemented within the ontology management work package. The component KOIOS performs keyword-based search on so-called linked data sets; these are data related to each other via a net of RDF triples. The keyword query is expanded by additional relations or details which the user can select before it is translated into a SPARQL query. Additionally, KOIOS searches for further existing links in the RDF graph, thus providing data not explicitly specified in the expanded query but relevant due to existing relations between data. The relations are computed based on feature vectors and weighted by probabilities (Bicer et al. 2011). For example, a movie can be classified as a German movie, although it is not labeled as such, solely by comparing producers, actors, directors and original language with those of other German movies.

All the above described components were developed within the THESEUS research program; some of them have been implemented in to several integration platforms, some of them are used in use cases; others have even been integrated by industry partners. Designing and implementing these components is an essential building block in getting enterprises to use semantic tools for knowledge management as future projects will surely show.

4 WP4: Situation Aware Dialog Processing

The Internet of Services poses new challenges, including the development of new interaction paradigms for the man-machine interface. As part of the THESEUS CORE TECHNOLOGY CLUSTER, we have developed a dialog shell, investigating new interaction styles and providing software tools for the rapid and efficient development of multimodal and situation-aware dialog systems.

Given semantically annotated content and access to semantic services, it becomes possible to extend existing interaction approaches that are menu-like and system-driven to a more natural interaction that is aware of the current situation and context. The dialog shell developed in WP4 enables flexible dialogs that integrate the user into complex business processes; see, e.g., the demonstrator for the Uses Case TEXO as described in Porta et al. (2014). Another interaction style, explorative browsing, enabled by semantically annotated content and services in conjunction with the dialog shell is demonstrated by the CALISTO system, described in Löckelt et al. (2014).

We followed an integrated approach, providing a comprehensive toolkit, the Dialog Shell, for the THESEUS research program and beyond (Sonntag et al. 2010). It addresses two basic research questions: how can semantic services in the Internet

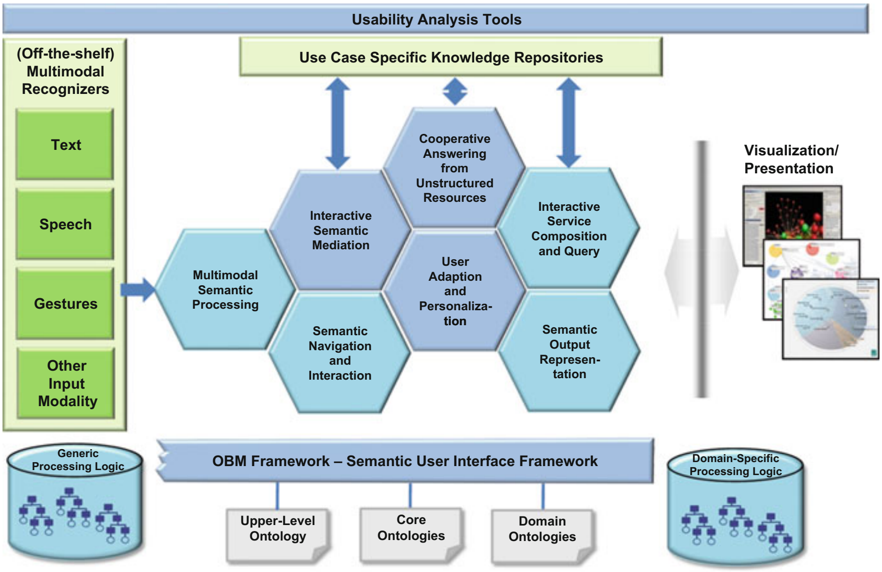


Fig. 3 Architecture of WP4

be found, addressed, and combined and what new forms of interaction are possible with them.

Ontology-Based Dialog Platform The core result of this work is the ontology-based dialog platform (ODP), which is used in a number of use cases of THESEUS and has been successfully marketed by the spin-off company SemVox since 2008.

The starting point for the development of ODP was the standard architecture for multimodal dialog systems as described in Wahlster (2006) and Becker (2010). It comprises a number of components that are organized along a core processing pipeline. To use this architecture for the Internet of Services, it has been extended and adapted as shown in Fig. 3. The main steps are sketched here and elaborated in Porta et al. (2014).

User input can come from various modalities in paralle, e.g., spoken language and pointing gestures as in "Show me more videos about this!". After separate mono-modal interpretations of the input, the fusion module computes the combined intention of the user, taking context and a user model into account. Once the intention has been recognized, dialog and interaction management computes the system reaction. Typically, this is a call to a service, e.g. to provide requested information. Which service is applicable and if the request can only be fulfilled by the combination of multiple services is decided in the modules for semantic mediation and service composition, which call the corresponding services. The results are processed and presented to the user by the presentation planning module, which also allocates the information to the available media channels. It could be the spoken output "17 videos were found" combined with a list of the videos on the screen.

The ontology-based dialog platform (ODP) provides modules for all of these processing steps. These modules are connected through a common, semantic data representation formalism, the so-called extended typed feature structures (eTFSs) (Pfleger 2007). This eliminates the frequent transformation steps between modules that are typical for earlier dialog systems. Not only does this speed up processing, it also supports error-free internal communication. To work with this common semantic representation formalism, the modules are provided with a common API that implements some powerful computations on eTFSs (Alexandersson et al. 2006). Additionally, all modules have access to PATE, a rule-based abstract machine based on cognitive models (Pfleger 2007) that allows writing processing rules at a high level of abstraction and yet features efficient processing. Through common tools in all modules, the dialog shell provides – for the first time – a developer with an integrated development environment (IDE) for the various, diverse tasks in a dialog system.

This specialized IDE is an important part of the ODP platform and it is based on freely available standards and systems. All development tools are integrated into the Eclipse platform and support the development of code and knowledge sources, e.g., the rules for speech interpretation or the PATE system. Wizards and specialized editors support the efficient setup of new ODP projects and maintenance of knowledge bases and processing rules for the modules. The workbench also provides specialized debugging and testing tools.

For the connection between front end user interface and back end semantic services, WP4 has developed the so-called "Joint Service Engine" (JSE) that provides service grounding and supports the semantic annotation process for services with a declarative data type mapping mechanism. It integrates a range of different representation formalisms, e.g., RDF, XML(Atom), or JSON.

Multimodal Interaction As an example for multimodal interaction beyond the boundaries of devices, we have developed an information kiosk for the THESEUS Innovation Center in Berlin that allows search for and interaction with media content (images, videos and texts) on the exhibits of the center and on the early history of computing and computer pioneer Konrad Zuse. For example, the starting point of a search can be an image that is stored on a mobile device and sent to the information terminal with a "frisbee gesture", i.e., a simulated throwing of the images with the mobile device. On the console of the information kiosk, the multitouch screen can then be used to activate various interaction windows (spotlets); see Fig. 4. The central spotlet is used for semantic search, triggered by simple drag and drop of semantic concepts or simply media objects. Search results are shown according to media type, and the related semantic tags can be inspected and used for continuing the search and browsing activity. Besides watching the media objects, e.g. with the video player, other Internet-based services are connected. All objects that have geospatial metadata can be shown on a map and the Internet message service Twitter is connected by a spotlet. The functionalities can also be addressed through spoken

Fig. 4 Touchscreen of the information kiosk CALISTO on "The progress from the mechanical calculator to the Internet of Services"

interaction, e.g. by saying "Please show the THESEUS Innovation Center on the map."

The components in WP4 provide functionality for spoken and typed input, for gestures such as pointing, selection, and drag and drop and the symbolic throwing gesture (the so-called "frisbee gesture" where a media object is selected on a mobile device and "thrown"). Modality-specific interpretation components map the input into a common semantic representation and a fusion component derives an integrated interpretation of the user input. System output includes written and spoken text, various graphical presentation forms, including outputs from WP5 and a desktop-like environment, as well as haptic feedback, e.g. the vibration in a mobile device. Rendering components exist for a multitude of operating systems and devices. Except for the latter, all components are device independent and the dialog shell has been employed on desktop PCs, Android, and iPhone devices as well as tablets and large multi-touch kiosks.

This new mode of interaction, being a variant of browsing in semantic datasets, in combination with the Internet of Services allows for a faster navigation through the exhibits in the THESEUS Innovation Center and guarantees a high relevance of the search results through the semantic annotation.

The integrated dialog shell is an innovative basis for new user interfaces for the numerous applications of the Internet of Services. It is already successfully marketed by SemVox GmbH, a spin-off company of DFKI GmbH, demonstrating the market readiness of the technology. The results of this work have been published in books, journals and at conferences (a total of 36 publications) and led to two Ph.D. and four Master's theses, eight lectures and seminars, four contributions to standards and a patent. Based on the dialog shell, three industrial projects, six EU projects and three national projects have been started and contribute to DFKI's participation in the EIT ICT Lab. Current work further extends the approach of the dialog shell: a common semantic representation, strong modularization, efficient semantic processing throughout the system, adaptation to the situation and personalization and a strongly coupled development of mediation and composition of semantic services.

5 WP5: Innovative User Interfaces and Visualizations

Work Package 5 investigated the graphical representations of semantic and massive data in two main areas of research: Semantics Visualization and Visual Analytics. Semantics Visualization focuses on human-centered graphical representations of complex semantic structures or formal notations of ontologies. Research on this topic incorporates the investigation of data manipulation and enrichment of graphical algorithms for representing the structure of the semantic conceptualization. Furthermore the question of how to present the graphical information, e.g. by visual attributes or data filtering and recommendation, were topics of research.

The second main topic of WP5 was the research and interaction with huge data amounts. Visual Analytics in WP5 investigated different methods for analyzing, structuring and visualizing to support decision-making and analyzing tasks.

CTC work packages aim at providing core technologies for an instantiation in the THESEUS use cases. In WP5 the human factor played a key role during the process of research and development. One of the main challenges was to develop systems that provide generic characteristics to be adaptable to various usage scenarios and to consider the variety of human perception, preferences and knowledge. Thus the THESEUS use cases did not only investigate different domains of knowledge, e.g. medical computing, Internet of Services or social and semantic web, their scenarios involved heterogeneous users. The users within just one use case differed significantly, e.g. starting from service engineers to service providers and ending in service consumers. Facing the heterogeneity of users in visualization, systems by adaptable and adaptive visualizations was the main challenge of this work package.

Semantics Visualization WP5 has investigated techniques and concepts to visualize semantic data for heterogeneous users, usage scenarios and semantic characteristics. The challenge of the research was the conceptualization and development of a technology that is user-oriented on the one hand and features the characteristics of a core technology on the other hand. Therefore the adaptable and adaptive SemaVis technology was developed, which covers both the features of a user, use case and context adaptable core technology and a user-centered and adaptive Semantics Visualization.

SemaVis integrates a method for orchestrating visualizations in a multiple visualization user interface as a visualization cockpit (Nazemi et al. 2010a). A set of about 15 different visualization techniques can be positioned in a user interface for an enhanced juxtaposed aspect-oriented visualization. Each of the visualizations focuses on certain characteristics of semantic data. The SemaVis cockpit metaphor is more than the combination of multiple visualizations, as known from Brushing & Linking. It provides the separation of the semantic information into several view-modes for supporting heterogeneous tasks, e.g. analytical comparison, knowledge exploration or information acquisition.

SemaVis integrates the visualizations and the general user interface using a layer-based approach for a fine-granular separation of visual information. The entire information visualization process is subdivided into different layers of abstraction. Visualizations can be described according to what is displayed, where with it is displayed, and how it is displayed. Accordingly, each visualization in SemaVis implements the layers semantics, layout and presentation (Nazemi et al. 2011).

Semantics defines which data is visualized. It contains information about the data, its structure and its implicit and explicit properties. Layout defines the graphical layout algorithm to be used. And presentation defines more precisely how the data will be presented. It is the visual layer of the visualizations and parametrizes the visual look. The input of the presentation is the geometry as calculated by layout.

The layer-based adaptability of SemaVis provides an adequate approach for generating heterogeneous user interfaces by manipulating the entire spectrum of the visualization characteristics. This ability is used to further provide an automatic adaptation to several impact factors, e.g. user interaction or data structure. The goal of the automatic adaptation of the visualization is to support the user in his navigation process, during the exploration of information. Therefore a new approach was conceptualized, where the impact factors are captured and enriched with a semantic structure provided by the data. A user interaction, for instance, is determined as a three-dimensional conceptualization of the interaction taxonomy. Thereby the semantic taxonomy of data is used and enriched with a semantic hierarchy (Nazemi et al. 2010b). Further the characteristics of the data, especially the semantic structure and the explicit and implicit properties, are considered in a similar way. Therefore a data semantics is generated that provides information about the current visible data.

Visual Analytics An ideal search and retrieval engine structures and organizes information in the way needed by a user in his current situation and task. It shows the right information at the right time – in the right way. While representing, refining and using such a structure was the main goal of the aforementioned achievements, the goal of Visual Analytics in THESEUS was the creation of this structure. For unstructured or weakly structured data – most raw texts fall in this category – this is a difficult feat. The ideal structure does not only depend on properties of text, like vocabulary, topic and context; in addition it should reflect how the user thinks about the data.

Purely automated methods for data analysis and machine learning reflect how the developer thinks about the data, but there is never a guarantee for a match between the user's and the developer's perspective. Visual Analysis combines techniques for visual interactive and automated analysis. The goal of this combination is to compensate for the weakness of one technology with the strength of the other. Analysis becomes flexible by introducing the user into the process via visualization: visual representations and interaction.

Visual Analytics research in THESEUS explored how one of the most valuable assets could be introduced to the analysis – the knowledge and abilities of the human user. For example, visualization enables the user to detect patterns and structures in noisy and ambiguous data. It enables the user to evaluate and refine analytical results and to dynamically control the process depending on the users' knowledge. These leverage points have been investigated and structured within the Visual Analytics Framework. In addition, WP5 investigated strategies for how the leverage points were made accessible even to non-experts to analysis, allowing the users to interact with the data on their own terms.

Visualization Use Cases WP5 technologies were implemented in several usage scenarios and applications within and beyond the THESEUS research program. For instance SemaVis was integrated into different applications of the Internet of Services. The architectural design of the technology allows the instantiation of the visualizations as services. WP5 forced the idea to use Visualizations as a Service (VaaS) for providing rich and interactive graphical representations of highly complex and semantics data. The modular architecture allows a differentiated implementation for various data and for different users. Existing applications of SemaVis provide exploratory searching, learning and analysis tasks, e.g. in joint application with the THESEUS use cases ORDO and PROCESSUS (Stab et al. 2012). Further the exploration of services as a service browser provides the opportunity to examine virtual services for different roles, e.g. service consumer or service engineer. An open-access version of SemaVis opens the technology for everyone and provides a visual interactive interface to the most common semantic data repositories and search environments. In future applications SemaVis will be enhanced to do more than visualize semantically annotated information and especially to provide support for visual analysis and decision making tasks. Therefore existing open social media repositories will be used to visualize sentiments and opinions about topics of interest.

The visual interfaces of both researched areas, Semantics Visualization and Visual Analytics, will be more and more applied to the policy modeling life cycle. The visualization of the lacks in existing policies and the visualization of the impacts of new policies will be a main area of interest in the early future. Further related areas, e.g. the visualization of law and law consequences, will be very important to the related areas.

Semantics Visualization and Visual Analytics provide interaction abilities with graphical representations of information. With the changes in mobile computing and alternative interaction techniques, the graphical interaction will have a sustainable role in our way of interacting with information.

6 WP6: Statistical Machine Learning

Currently the Internet is dominated by unstructured content such as texts, images and videos, all of which are representations of information that are quite suitable to a human user but not to a machine. In the evolving Internet of Services, machines need to make decisions, which requires that machines have some form of understanding of unstructured content. A classical example, relevant even before the golden age of the Internet, but still of high importance nowadays, is postal automation, where a machine needs to understand the handwritten postal code and address on the envelope of a letter in order to properly sort the letter. Content extraction evolved into a major task on the Internet, where a text, image or video needs to be annotated with the right labels such that it can be found by a person with the corresponding information need. Statistical machine learning has become instrumental to solve these tasks and in WP6 a number of novel approaches have been developed.

Textual information is one of them. A first step to understanding a text is to extract the major named entities that are mentioned, such as persons, locations, companies, diseases and genes. Based on sophisticated statistical methods a number of approaches were developed for this task that are even able to deal with multilingual documents. More useful but also more challenging is the extraction of textual statements in the form of relation extraction and semantic role labeling. Thus, whereas named entity recognition identifies Kohl as a person's name and China as a country (and not as a ceramic), relation extraction would be able to extract from the text the statement (Kohl, visits, China) as the content of a news story. Finally, statistical topic modeling describes a document by extracting its main themes, and in the work package existing approaches were refined to combine the strengths of topic modeling and text annotation; see also Bundschus et al. (2008, 2009).

The developed methods were applied to news texts in the CONTENTUS Use Case, for the extractions of the relationships between genes and diseases out of medical abstracts in the MEDICO use case, and for the detection of the relationship between companies (GE, competitor of Siemens) out of news texts for the ORDO use case. The approaches have also been extended to detect opinions or sentiments in texts. Thus a company's product problems can quickly be detected by analyzing blogs and news texts. Finally, one can get insight into a document by extracting the main topics covered. We consider two applications developed with the THESEUS technologies. The QUOTE application regularly analyzes fresh news texts by relation extraction technology and detects persons and quotes by those persons, which is much more relevant than reported speech. A second application "Eat & Drink" scans the comments in a large repository of restaurant reviews and extracts important phrases highlighting some aspect of a restaurant. Hence the often very large number of reviews is condensed to a few highlighting citations. Together with statistical information on the reviews this allows the users to see the strengths and weaknesses of a restaurant at a glance and alleviates the decision for the user. Both applications are provided as apps for Android and the iPhone.

Based on the topic models and the other developed technologies on text analysis, a demonstrator for the handling of claims was developed and presented at CeBIT. The demonstrator was also using results of WP6 developed for the analysis of document structures and for the recognition of handwritten text. For the recognition of handwritten text, a new learning approach based on recurrent hidden Markov models has been developed and has already been successfully integrated in various commercial applications. For the analysis of document structure, a self-adaptive solution has been developed. Consider that existing approaches for automatic logical structure analysis in documents have in common that they were developed for a specific task and document type. Adapting such a method to a different task requires modifying the existing set of rules or grammars, which is a laborious manual task. The document structure analysis developed within WP6 uses machine learning techniques instead of manually created rules or grammars. The module learns a structure model that takes layout, formatting and content features into account. The learned structure model is then applied in order to analyze the structure of a given document. For more details, see Schambach (2009) and Stoffel et al. (2010).

Machine learning is also very effective for the extraction of information from images. Image annotation is the basis for content-oriented retrieval of images, and highly competitive approaches for generic images were developed in WP6. As a special case we considered medical tomographic images, the focus of the use case MEDICO. We developed solutions for semantic localization, the task of automatically finding body regions which contain certain anatomical concepts like organs, bone structures or body landmarks. This information is very useful when working with CT Scans and for connecting them to other medical information. Thus, semantic localization is a powerful tool to connect the world of medical 3D images to the world of ontologies and semantics. Another important functionality is to align two CT scans for differential diagnosis and to optimize retrieval times from the picture archiving and communication system to the workstation of a physician. Compared to other solutions to this problem our new approach suffices with very limited information, i.e. a single 2D slice from the CT scan to precisely predict positions in the human body. Whereas some of our solutions are specialized for detecting a certain type of anatomical concept, for example the location of the vertebrae, we also developed a very generic approach mapping the positions of single slices to a general model of the human body. The model consists of a standardized coordinate system of the human body and statistical information about the location of multiple anatomical concepts within this coordinate system. Employing a further method being developed within the work package, it is now possible to align a CT volume to the coordinate system and to predict the most likely location of the anatomical concepts within the CT volume. A developed use case permits a physician to precisely localize a desired tomographic slide in the human body without loading the whole set of all tomographic slices, thus reducing loading times and network load substantially. This solution became part of the MEDICO use case and found great interest in the medical community in general. For more details, see Grafa et al. (2009) and Graf et al. (2011).

Another special case constitutes the annotation and ranking of images for generic visual concepts. Such annotations may be employed for the search of images containing certain visual concepts based on the image content rather than keywords linked to an image. Solutions developed within the work package have achieved top ranks in international benchmark competitions with undisclosed ground truth such as the Pascal VOC Classification Challenge and yielded first-ranked submissions in the ImageCLEF2011 Photo Annotation Challenge. The methodological challenge consisted in the ability to deal with a large set of differing visual concepts, ranging from localized objects to overall emotional impressions; the large variability of image appearance present within general visual concepts beyond simple objects like Partylife, BeachHoliday, Mountains, Indoor, Euphoric, and AestheticImpression; and the disagreement of human annotators in labeling images for such concepts. Statistical methods are robust against annotator disagreement, labeling errors, varying image qualities and scales of visible cues, differing lighting conditions, occlusion and clutter in images. They are able to extract relevant structures based on implicit and statistical definitions, labeling example images as belonging to the same concept, even when there is no way to define deterministic rules for what a concept should be. This makes statistical learning methods a valuable complement to explicit and deterministic knowledge modeling. The results of this work are described in detail in Binder et al. (2014).

Finally, WP6 considered video information. The main focus here was the extraction of textual information in videos. The developed solutions are applied to the detection of logos in videos, e.g., for detecting placed advertisement in sport reports and for the detection of licence plates and vehicle types in traffic surveillance systems. The logo recognition technology developed within this work package works on previously learned logos and allows fast recognition and high recognition performance, enabling various real-time recognition applications. For example, in postal automation the online detection of postal stickers and value symbols like stamps, service symbols, return address logos or similar logos is already successfully in operation.

In addition to unstructured data, machine learning can be applied to structured data as well. Consider the Linked Open Data (LOD) initiative, where data from various domains is made available in the format of the Semantic Web. In addition, information sources cross-reference one another such that entities become unambiguous: e.g., if two data sources talk about Paris, it is clear if they both mean the capital of France and not a person popular with the media. Structured data show regularities that can be exploited via machine learning. In WP6 we have developed various approaches for machine learning in semantic domains in the form of the SUNS framework (Huang et al. 2010; Tresp et al. 2009). One application in MEDICO permits the prediction of diseases that might be affected by a given gene mutation. Other applications are in the recommendation of web services and in the recommendation of procedures in medical domains, as decision support component for physicians. Details of our approach can be found in Tresp et al. (2014).

7 WP8: Evaluation

There are two key paradigms researchers can follow: research just to learn, and research to solve practical problems. The first one in general is called basic research, the latter applied research. The result of applied research is usually something like a prototype which might be rather close to a final product. Most of the work in THESEUS was focused on applied research, but the work flow from technology to prototype was split into two separate functions: The basic technologies necessary for at least two application areas have been developed in the CORE TECHNOLOGY CLUSTER; the integration and prototype development was concentrated in the use cases. That way synergy effects between application areas could be exploited.

Doing applied research is always an important topic for evaluation: Only if there are clear definitions of "good" and "better" is it possible to improve a technology. The evaluation tasks are different in the CTC and in the use cases: In a use case the complete system can be accessed concerning performance and usability. The weighting of different dimensions of quality often depends on subjective aspects and is application specific. In the CTC usually only technology components are available for testing. Each dimension is evaluated separately. An example can illustrate this: In the final application a question typed in by the user should be answered by the system. For the end user it is essential that he receives a useful answer in reasonable time. If the answer needs a lot of time or if there are too many results which do not make sense to the user he will not accept such a system. From a CTC point of view this system will consist of several parts: Algorithms for the analysis of the query, algorithms for the pre-processing of the database used in the query, algorithms to search in that database and tools to present the result of the search. For each of these components there are different performance parameters like computational time, storage requirements and correctness. Not all of the components are necessary for each query, but might influence the total performance. In the CTC it was essential to have the detailed performance parameters to monitor and control the development process. Note that the user interface, which is a key component in appearance for the end user, most of the time is application specific and is therefore part of the use case.

Due to the fact that each use case had different requirements on the overall performance, its evaluation was done in each use case separately. The evaluation of technology components developed in the CTC was concentrated in WP8. To fulfill the final requirements of the use case all evaluations had to consider the intended use. Special care had to be taken to use test data similar to the data in the use case. Due to the fact that the CTC developed a variety of different algorithms also a large number of test methods had to be implemented and performed. To exploit synergy effects expert groups performing the evaluations were organized in tasks which did not match to the structure of the CTC 1:1. To monitor progress and give hints about weak points in the algorithms all evaluations have been repeated annually during the existence of each development task. The following list gives some examples of evaluations performed. Note, that due to the limitations of this article results cannot be detailed.

Databases In many evaluations audiovisual content was necessary in a raw impaired form. The database service task was responsible for the collection, creation and processing of data to be used in the evaluations. An important aspect for testing the performance of systems is that the databases used for the evaluation should not be used for the optimization of the algorithms. Therefore several new databases have been generated and kept secret from the partners. To enable the international comparison of results the task also organized challenges in international benchmarks under the umbrella theme of ImageCLEF (see also Liebetrau et al. 2014). Processing of data included the adaptation to different resolution for images and movies but also the controlled impairment like adding noise, coding, smearing and rotation of pictures.

Text Analysis Older data collections in libraries and medical archives are only available on file cards. They often are only available on paper and are even handwritten. Usually they have fixed layouts which simplify the automatic processing. Different font types can be used to recognize the function of recognized test. Within the CTC, algorithms for text recognition and text type recognition have been developed. To evaluate these algorithms, databases from medical records and libraries have been used. To simulate the effects of old paper and bad scanning, different impairments like adding of noise and rotating of picture have been used.

Media Data Analysis For the management of media databases it is essential to recognize duplicated pictures and videos. An important aspect is that copies which would be recognized as "identical scenes" by humans should be detected, while copies humans would annotate as "different takes by same actor" should be classified as different. The algorithms evaluated therefore created perceptual hashes for each item. For the evaluation modified copies of images and videos were used. Such modifications included rotation, changes of aspect ratio and blurring.

Picture Analysis In THESEUS, understanding the content of pictures is necessary in many use cases. Due to different demands a large number of CTC tasks performed work on different algorithms. The algorithms for face detection had to find one or several faces in pictures with different light and background. Evaluation criteria were that all faces present on the picture were found and marked and that no other structures were wrongly marked to be a face. The task of picture segmentation is to mark objects in pictures. Evaluation criteria were both the detection ratio and the precision for finding the form of the marked objects. In the context of analysis of films, algorithms for the detection of scenes are important. Detection might be slightly before or after the "switch" between scenes. The evaluation was counting the correct and false positive detections within a short time window around the correct scene borders. To find pictures with certain features, like "summer", "vehicles", "mountain" or "family&friends" in large databases of images, automatic algorithms for image annotation have been developed. Manually annotated databases have been used for the evaluation. To enable international comparison, an evaluation task in the ImageCLEF workshop was organized. For further information about the test data see also Liebetrau et al. (2014). Algorithms for video genre classification (film, cartoon, news, commercial, music) were evaluated, based on video clips either containing only one genre ("pure") or an artificial combination of scenes of different genre ("mixed"). Each genre was represented by about 5 h of content. For "pure" data the analysis of whole clips had higher precision than the analysis of single video frames. The performance for the "mixed" content was worse than for the "pure" content. However in a real world usage scenario video genre classification would be combined with scene detection: As a result the algorithm would always work on "pure" data. Special algorithms in picture analysis are necessary for the localization of Computer Tomography slices. Such algorithms are useful in several medical applications (see Kuhlmann et al. 2014). In the evaluation 34 datasets from 24 patients and different regions of the body (pars cervicalis, thorax) were used. Depending on the region, different algorithms provided the best results.

Audio Quality Originally it was planned that the selection of algorithms for speech recognition and text2speech will be necessary for innovative dialog shells. In an early phase of THESEUS it was decided that such algorithms were application specific and are no longer on the list of research to be performed in the CTC.

Video Quality Algorithms for the efficient representation of video content at different resolutions are important aspects for many applications in THESEUS. The final instance for the assessment of video quality is the visual test with test subjects. Such tests are expensive and time consuming. Therefore WP2 conducted work towards the automatic measurement of the perceived video quality. Automatic measurement normally is based on the comparison of an unimpaired reference with the output of a technical system. Such algorithms can work based on either "full reference" or "reduced reference". Both configurations have the disadvantage that such measurement algorithms are not adequate in applications where the reference is not accessible or does even not exist at all (example: copy of an old film). The algorithm developed in WP2 therefore estimates the quality parameters "blur" and "blocky" by analyzing the output data only.

WP8 evaluated coding schemes and measurement schemes developed in WP2: One of the databases used to evaluate the performance of the video codec was used also to evaluate the performance of the video measurement algorithm. The two non-reference quality parameters "blur" and "blocky" in general showed a similar rank order correlation as the state of the state-of the-art full-reference algorithms PSNR and SSIM.

Iterative System Design For fulfilling the wide range of tasks, the CTC had to deal with ontologies, infrastructural measures, various visualization modes, and the annotation of various file types (images, documents), including machine learning aspects. WP8 thus aimed at identifying, preparing, performing, and analyzing appropriate evaluation means, methods, tools, and procedures for all the aforementioned tasks (components), frequently reporting the results back to the developers and into their iterative development cycles. A total of more than 25 CTC tasks contributed to the list of components to evaluate.

The evaluation of the WP3 tasks was based on the *ontology management components* called DELPHI and HERAKLES (Ontology Reasoning), HARMONIA (Ontology Mapping) and MNEMOSYNE (Ontology Design and Ontology Evolution). The efficient manipulation and usage of large OWL ontologies is a major problem, because large ontologies often exceed the available main memory. This problem is caused by the fact that ontologies have to be parsed and loaded in the memory before they can be used. Performance and scalability are thus two major criteria for describing the quality of a persistence system. The evaluation measured the respective response times for several query tasks. Both the performance and the scalability of the systems have been tested against real-world ontologies of varying sizes and complexity.

The WP4 focused on a flexible *dialog shell* for the implementation of applications with multimodal user interaction. It consisted of several different integrated functional modules that can be connected in adaptable and flexible combinations to each other. The developed test framework also fulfilled basic requirements that pointed toward the possibility of measuring the time performance of the functional modules, forming the dialog shell, as well as debugging and tracking the event flow. The evaluation aimed at successfully running both single component tests as well as an end-to-end test, particularly for components relevant for the use case MEDICO.

The objective of WP5 has been the development of a reliable and *useful user interface* that is appropriate to the user. The user interface and the underlying *visualization* technologies developed aimed at being generic in order to be used in several Use Cases. The evaluation based on various cycles and a set of tasks to be fulfilled by human testers made sure that the modular approach was able to satisfy both functional and technical requirements of the different use cases. Apart from technical tests, a large usability evaluation completed the analysis of the WP5 components.

The evaluation of WP6 contained components concerning algorithms that extracted a latent semantic representation from *textual corpora* as well as *self-learning algorithms* for large-scale textual archives. The tests addressed mainly the correctness of the results, and made use of available tools and corpora as well as self-created tools and manually annotated corpora. The strengths and weaknesses of the components have been spotted and, where appropriate, the quality has been compared to state-of-the-art approaches. Further aspects under consideration have always been the scalability of the solutions and the required amount of annotated data for fulfilling the requirements from the use cases.

Privacy and Security Privacy and security aspects have to be considered in system designs from the beginning. In the THESEUS context this is especially important for the use case where core technologies are integrated into systems. Within the CTC only a few technologies have privacy aspects, and these tasks have been analyzed based on technical and legal requirements. In a final phase of THESEUS a workshop concerning privacy aspects for the whole THESEUS research program was organized in cooperation with the ULD Schleswig-Holstein. An important result of this workshop was that in subsequent projects privacy has been considered as a separate key action across the whole program.

Field Tests The CTC presented its huge variety of developed technology components in the context of specific demonstrators at project meetings, trade shows, and various domain-related fairs. For figuring out both the functional and the non-functional capabilities of those components, a series of tests was necessary to perform aiming at much more than just pure functionality. The wide range of tests performed in that context was especially focused on, but not limited to, usability aspects. From that viewpoint, the standardized evaluation procedures along with selected methods and characteristics addressed preferably the entire chain of components (e.g. from input to analysis, to search, to processing, to visualization) integrated into a demonstrator application. Therefore non-functional aspects like usability, always based on standardized evaluation methodologies, characteristics, and sub-characteristics, have been evaluated.

8 Results

Altogether 45 different basic technologies have been developed within the THESEUS CORE TECHNOLOGY CLUSTER. A number of outstanding scientific results have been achieved, which is documented by a number of top rankings in international challenges. Furthermore the number of publications is remarkable. It includes 35 book chapters, 26 journal papers, 235 contributions to conference proceedings, 82 additional publications, 26 patents and 29 contributions to standardization bodies. In addition 20 doctoral theses have been completed and CTC results are used for teaching, including 25 diploma, bachelor and master theses and 3 lectures.

Besides the research and development described above, the CTC dedicated some effort towards the development of joint demonstrators. The reason for that was to better understand the potential of the technologies developed in the "other" work packages and the necessity to prove the interoperability of different technologies.

9 Conclusion

The exploitation of the developed CTC technologies is successful. They are used by all use cases and by many SME projects within THESEUS. Several CTC technologies found their way into international standards. A number of license agreements with third party partners have been concluded and five spin-off companies have been founded. In addition, a number of new R&D projects on national, European and industrial basis have been started to bring the CTC results closer to products.

References

J. Alexandersson, T. Becker, N. Pfleger, Overlay: the basic operation for discourse processing, in *SmartKom: Foundations of Multimodal Dialogue Systems*, ed. by W. Wahlster. Cognitive Technologies (Springer, Berlin/Heidelberg/New York, 7/2006), pp. 255–267

T. Becker, (Multimodale) Dialogsysteme, in *Computerlinguistik und Sprachtechnologie – Eine Einführung*, ed. by K.U. Carstensen, C. Ebert, C. Ebert, S. Jekat, H. Langer, R. Klabunde (Spektrum, Heidelberg, 2010), pp. 624–632

V. Bicer, T. Tran, A. Gossen, Relational Kernel machines for learning from graph-structured RDF data, in *The Semantic Web: Research and Applications*, Heraklion, 29 May–2 June 2011, ed. by G. Antoniou, M. Grobelnik, E. Simperl, B. Parsia, D. Plexousakis, P.D. Leenheer, J. Pan. Volume 6643 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011), pp. 47–62

A. Binder, W. Samek, K.R. Müller, M. Kawanabe, Machine learning for visual concept recognition and ranking for images, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

J. Bock, Map PSO results for OAEI 2010, in *5th International Workshop on Ontology Matching (OM-2010), Collocated with the 9th International Semantic Web Conference (ISWC-2010)*, Shanghai, vol. 689, Nov 2010, pp. 80–186

J. Bock, J. Hettenhausen, Discrete particle swarm optimisation for ontology alignment. Inf. Sci. **192**(1), 152–173 (2012). http://dblp.uni-trier.de/db/journals/isci/isci192.html#BockH12

J. Bock, A. Lenk, C. Dänschel, Ontology alignment in the cloud, in *Proceedings of the 5th International Workshop on Ontology Matching (OM '10), Collocated with the 9th International Semantic Web Conference (ISWC '10)*, Shanghai, vol. 689, ed. by P. Shvaiko, J. Euzenat, F. Giunchiglia, H. Stuckenschmidt, M. Mao, I. Cruz (CEUR Workshop Proceedings, 2010), pp. 73–84 http://ceur-ws.org

J. Bock, T. Tserendorj, Y. Xu, J. Wissmann, S. Grimm, A reasoning broker framework for Protégé, in *11th International Protégé Conference*, Amsterdam, June 2009

M. Bundschus, M. Dejori, M. Stetter, V. Tresp, H.P. Kriegel, Extraction of semantic biomedical relations from text using conditional random fields. BMC Bioinform. **9**(207), (2008) http://dblp.uni-trier.de/db/journals/bmcbi/bmcbi9.html#BundschusDSTK08

M. Bundschus, S. Yu, V. Tresp, A. Rettinger, M. Dejori, H.P. Kriegel, Hierarchical Bayesian models for collaborative tagging systems, in *Proceedings of the IEEE International Conference on Data Mining (ICDM '09)*, Miami, ed. by W. Wang, H. Kargupta, S. Ranka, P.S. Yu, X. Wu (IEEE Computer Society, 2009), pp. 728–733. http://dblp.uni-trier.de/db/conf/icdm/icdm2009.html#BundschusYTRDK09

P. Ciaccia, M. Patella, P. Zezula, M-tree: an efficient access method for similarity search in metric spaces, in *Proceedings of 23rd International Conference on Very Large Data Bases (VLDB '97)*, Athens, 25–29 Aug 1997, ed. by M. Jarke, M.J. Carey, K.R. Dittrich, F.H. Lochovsky, P. Loucopoulos, M.A. Jeusfeld (Morgan Kaufmann, 1997), pp. 426–435

F. Graf, H.P. Kriegel, M. Schubert, S. Pölsterl, A. Cavallaro, 2D image registration in CT images using radial image descriptors, in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Toronto, ed. by G. Fichtinger, A.L. Martel, T.M. Peters. Volume 6892 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011), pp. 607–614. http://dblp.uni-trier.de/db/conf/miccai/miccai2011-2.html#GrafKSPC11

F. Grafa, R. Greila, H.P. Kriegela, M. Schuberta, A. Cavallarob, Enhanced detection of the vertebrae in 2D CT-images, in *Proceedings of the SPIE Medical Imaging 2010: Image Processing*, San Diego, vol. 8314, Feb 2009

S. Grimm, J. Wissmann, Elimination of redundancy in ontologies, in *Proceedings of the 8th Extended Semantic Web Conference (ESWC '11)*, ed. by G. Antoniou, M. Grobelnik, E.P.B. Simperl, B. Parsia, D. Plexousakis, P.D. Leenheer, J. Pan. Volume 6643 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011), pp. 260–274. http://dl.acm.org/citation.cfm?id=2008892.2008916

Y. Huang, V. Tresp, M. Bundschus, A. Rettinger, H.P. Kriegel, Multivariate structured prediction for learning on semantic web, in *Proceedings of the 20th International Conference on Inductive Logic Programming (ILP '10)*, ed. by P. Frasconi, F.A. Lisi. Volume 6489 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2010), pp. 92–104

J. Kleb, A. Abecker, Entity reference resolution via spreading activation on RDF-graphs, in *The Semantic Web: Research and Applications: 7th Extended Semantic Web Conference*, ed. by L. Aroyo, G. Antoniou, E. Hyvönen, A. ten Teije, H. Stuckenschmidt, L. Cabral, T. Tudorache. Volume 6088 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2010), pp. 152–166. http://dblp.uni-trier.de/db/conf/esws/eswc2010-1.html#KlebA10

F. Kuhlmann, J. Hannemann, M. Traub, C. Böhme, S. Zillner, A. Cavallaro, S. Seifert, B. Decker, R. Traphöner, S. Kayser, U. Lindemann, S. Prasse, G. Marczinski, R. Grützner, A. Fasse, D. Oberle, The THESEUS use cases, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

J. Liebetrau, S. Nowak, S. Schneider, Evaluation of image annotation using Amazon mechanical Turk in ImageCLEF, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

M. Löckelt, M. Deru, C.H. Schulz, S. Bergweiler, T. Becker, N. Reithinger, A unified approach for semantic-based multimodal interaction, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

H. Müller, T. Tsikrika, Global pattern recognition: the ImageCLEF benchmark. IAPR Newsl. **32**(1), 3–6 (2010)

National Institute of Standards and Technology (NIST). Secure Hash Signature Standard (SHS) (FIPS PUB 180-2), 2002

K. Nazemi, M. Breyer, D. Burkhardt, D.W. Fellner, Visualization cockpit: orchestration of multiple visualizations for knowledge-exploration. Int. J. Adv. Corp. Learn. **3**(4), 26–34 (2010a). http://dblp.uni-trier.de/db/journals/i-jac/i-jac3.html#NazemiBBF10

K. Nazemi, C. Stab, D.W. Fellner, Interaction analysis for adaptive user interfaces, in *Advanced Intelligent Computing Theories and Applications*, ed. by D.S. Huang, Z. Zhao, V. Bevilacqua, J.C. Figueroa. Volume 6215 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2010b), pp. 362–371

K. Nazemi, C. Stab, A. Kuijper, A reference model for adaptive visualization systems, in *Human-Computer Interaction: Design and Development Approaches*, ed. by J.A. Jacko. Volume 6761 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011), pp. 480–489

N. Pfleger, *Context-Based Multimodal Interpretation: An Integrated Approach to Multimodal Fusion and Discourse Processing*. PhD thesis, Saarland University, 2007

D. Porta, M. Deru, S. Bergweiler, G. Herzog, P. Poller, Building multimodal dialog user interfaces in the context of the Internet of services, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

M.P. Schambach, Recurrent HMMs and cursive handwriting recognition graphs, in *10th International Conference on Document Analysis and Recognition (ICDAR)*, Barcelona (IEEE Computer Society, 2009), pp. 1146–1150. http://dblp.uni-trier.de/db/conf/icdar/icdar2009.html#Schambach09

T. Skopal, J. Pokorný, M. Krátký, V. Snášel, Revisiting M-tree building principles, in *Advances in Databases and Information Systems*, ed. by L. Kalinichenko, R. Manthey, B. Thalheim, U. Wloka. Volume 2798 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2003), pp. 148–162

A.F. Smeaton, P. Over, W. Kraaij, Evaluation campaigns and TRECVid, in *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, Santa Barbara, 26–27 Oct 2006 (ACM, New York, 2006), pp. 321–330

D. Sonntag, N. Reithinger, G. Herzog, T. Becker, A discourse and dialogue infrastructure for industrial dissemination, in *Spoken Dialogue Systems for Ambient Environments: Second International Workshop (IWSDS 2010)*, Gotemba, 1–2 Oct 2010, ed. by G.G. Lee, J. Mariani, W. Minker, S. Nakamura. Volume 6392 of Lecture Notes in Artificial Intelligence (Springer, Berlin/Heidelberg/New York, 2010), pp. 132–143. http://www.springerlink.com/content/5149m52mt5378316/

C. Stab, K. Nazemi, M. Breyer, D. Burkhardt, J. Kohlhammer, Semantics visualization for fostering search result comprehension, in *The Semantic Web: Research and Applications*, Heraklion, ed. by E. Simperl, P. Cimiano, A. Polleres, O. Corcho, V. Presutti. Volume 7295 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2012), pp. 633–646

A. Stoffel, D. Spretke, H. Kinnemann, D.A. Keim, Enhancing document structure analysis using visual analytics, in *Proceedings of the 2010 ACM Symposium on Applied Computing (SAC '10)*, Sierre, Mar 2010, ed. by S.Y. Shin, S. Ossowski, M. Schumacher, M.J. Palakal, C.C. Hung (ACM), pp. 8–12. http://dblp.uni-trier.de/db/conf/sac/sac2010.html#StoffelSKK10

V. Tresp, Y. Huang, M. Bundschus, A. Rettinger, Materializing and querying learned knowledge, in *Proceedings of the First ESWC Workshop on Inductive Reasoning and Machine Learning on the Semantic Web (IRMLeS '09)*, Heraklion, vol. 474 (RWTH Aachen, 2009)

V. Tresp, Y. Huang, M. Nickel, Querying the web with statistical machine learning, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

W. Wahlster (ed.), *SmartKom: Foundations of Multimodal Dialogue Systems*. Cognitive Technologies (Springer, Berlin/Heidelberg/New York, 2006)

W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (eds.), *Towards the Internet of Services: The THESEUS Research Program* (Springer, Berlin/Heidelberg/New York, 2014)

Z. Wang, A.C. Bovik, *Modern Image Quality Assessment*. Synthesis Lectures on Image, Video, and Multimedia Processing (Morgan & Claypool Publishers, San Rafael, California, 2005)

Further Reading on WP2: "Video, Metadata, Platforms: Processing of Multimedia Data"

U. Engelke, H. Kaprykowsky, H.J. Zepernick, P. Ndjiki-Nya, Visual attention in quality assessment. IEEE Signal Process. Mag. **28**(6), 50–59 (2011). http://dblp.uni-trier.de/db/journals/spm/spm28.html#EngelkeKZN11

S. Han, A. Hutter, W. Stechele, A reasoning approach to enable abductive semantic explanation upon collected observations for forensic visual surveillance, in *International Conference on Multimedia and Expo (ICME)* (IEEE, 2011), pp. 1–7 http://dblp.uni-trier.de/db/conf/icmcs/icme2011.html#HanHS11

S. Han, B. Koo, A. Hutter, V. Shet, W. Stechele, Subjective logic based hybrid approach to conditional evidence fusion for forensic visual surveillance, in *Advanced Video and Signal Based Surveillance (AVSS)*, vol. 7 (IEEE Computer Society, 2010), pp. 337–344. http://dblp.uni-trier.de/db/conf/avss/avss2010.html#HanKHSS10

M. Liu, I.V. Konya, J. Nandzik, N. Flores-Herr, S. Eickeler, P. Ndjiki-Nya, A new quality assessment and improvement system for print media. EURASIP J. Adv. Signal Process. **109**, 1–17 (2012). http://dblp.uni-trier.de/db/journals/ejasp/ejasp2012.html#LiuKNFEN12

E. Mbanya, S. Gerke, P. Ndjiki-Nya, Spatial codebooks for image categorization, in *Proceedings of the 1st ACM International Conference on Multimedia Retrieval (ICRM '11)*, ed.

by F.G.B.D. Natale, A.D. Bimbo, A. Hanjalic, B.S. Manjunath, S. Satoh (ACM, New York, 2011). http://dblp.uni-trier.de/db/conf/mir/icmr2011.html#MbanyaGN11

P. Ndjiki-Nya, M. Köppel, D. Doshkov, H. Lakshman, P. Merkle, K. Müller, T. Wiegand, Depth image-based rendering with advanced texture synthesis for 3-D video. IEEE Trans. Multimed. **13**(3), 453–465 (2011). http://dblp.uni-trier.de/db/journals/tmm/tmm13.html#Ndjiki-NyaKDLMMW11

C. Petersohn, Temporal video structuring for preservation and annotation of video content, in *16th IEEE International Conference on Image Processing (ICIP)*, Cairo, Nov 2009 (Institute of Electrical and Electronics Engineers (IEEE), 2009), pp. 93–96. http://dblp.uni-trier.de/db/conf/icip/icip2009.html#Petersohn09

C. Petersohn, *Temporal Video Segmentation*. PhD thesis, Institute of Technology, Berlin, 2010

T. Schwarze, T. Riegel, S. Han, A. Hutter, S. Nowak, S. Ebel, C. Petersohn, P. Ndjiki-Nya, Role-based identity recognition for TV broadcasts. Multimed. Tools Appl. **63**, 501–520 (2013)

F. Stegmaier, M. Döller, H. Kosch, A. Hutter, T. Riegel, AIR: architecture for interoperable retrieval on distributed and heterogeneous multimedia repositories, in *11th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, Desenzano del Garda, Apr 2010 (IEEE, 2010), pp. 1–4 http://dblp.uni-trier.de/db/conf/wiamis/wiamis2010.html#StegmaierDKHR10

F. Stegmaier, M. Döller, K. Schlegel, H. Kosch, S. Seifert, M. Kramer, T. Riegal, A. Hutter, M. Thoma, H.P. Kriegel, A. Cavallaro, Generische Datenintegration zur semantischen Diagnoseunterstützung im Projekt THESEUS MEDICO, in *41. Jahrestagung der Gesellschaft für Informatik: Datenmanagement und Interoperabilität im Gesundheitswesen (DIG)*, Berlin, Oct 2011

Further Reading on WP5: "Innovative User Interfaces and Visualization"

M. Breyer, K. Nazemi, C. Stab, D. Burkhardt, A. Kuijper, A comprehensive reference model for personalized recommender systems, in *Human Interface and the Management of Information. Interacting with Information*, ed. by M.J. Smith, G. Salvendy. Volume 6771 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011), pp. 528–537. http://dblp.uni-trier.de/db/conf/hci/hci2011-11.html#BreyerNSBK11

D. Burkhardt, C.E. Hofmann, K. Nazemi, C. Stab, M. Breyer, D.W. Fellner, Intuitive semantic-editing for regarding needs of domain-experts, in *Proceedings of ED-Media 2010, World Conference on Educational Multimedia, Hypermedia and Telecommunications* (AACE, 2010), pp. 860–869

T. May, J. Kohlhammer, Towards closing the analysis gap: visual generation of decision supporting schemes from raw data. Comput. Graph. Forum **27**(3), 911–918 (2008). http://dblp.uni-trier.de/db/journals/cgf/cgf27.html#MayK08

C. Stab, K. Nazemi, M. Breyer, D. Burkhardt, J. Kohlhammer, Semantics visualization for fostering search result comprehension, in *The Semantic Web: Research and Applications*, ed. by E. Simperl, P. Cimiano, A. Polleres, O. Corcho, V. Presutti. Volume 7295 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2012), pp. 633–646. http://dblp.uni-trier.de/db/conf/esws/eswc2012.html#StabNBBK12

Semantic Concept Identification for Images and Videos

Eugene Mbanya, Sebastian Gerke, Christian Hentschel, Antje Linnemann, and Patrick Ndjiki-Nya

Abstract The easy availability of image and video digitalization technologies as well as low-cost storage media nowadays enable the rapid propagation of digitalized image and video content to various distributed archives and especially the World Wide Web. This allows for easy access to enormous collections of image and video content for everyone. With the continuous growth in the sizes of these collections, effective and reliable management of the data becomes a more and more challenging task. Consequently, there is a need for essential techniques such as indexing, which enables more efficient browsing, searching and manipulation of digital content. In order for indexed digital content to be useful in such application scenarios, the indices must be as rich and as complete as possible.

1 Introduction

Digital media indexing today is mostly still done manually by professionals who assign specific keywords to images and video content. Also, most online photo and video sharing portals allow their users to manually tag media content in order to facilitate search in their collections. Manual tagging of digital content may however lead to inaccuracies when the manually assigned text descriptions misrepresent the media content. Also, in very large unannotated digital media collections, manual tagging becomes very labor intensive, expensive (in corporate settings) and time-consuming (in corporate and private settings) to be feasible. Therefore, automatic approaches for the categorization of images and videos are necessary for mastering the challenge of large-scale multimedia data management.

E. Mbanya (✉) • S. Gerke • C. Hentschel • A. Linnemann • P. Ndjiki-Nya
Fraunhofer Institute for Telecommunications (HHI), Berlin, Germany
e-mail: eugene.mbanya@hhi.fraunhofer.de; sebastian.gerke@hhi.fraunhofer.de; christian.hentschel@hhi.fraunhofer.de; antje.linnemann@hhi.fraunhofer.de; patrick.ndjiki-nya@hhi.fraunhofer.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_7,

Approaches for automatically assigning keywords to images are usually divided into three steps: extraction of low-level metadata, training and classification. In the first step, features (e.g. edges, colors and geometric shapes) suitable for describing image content are automatically extracted from an image. For each category (e.g. "beach") to be learned by an automatic category detection system, a training set of images or videos is required that contains positive and negative samples of the category. In the second step, the system is trained with the features extracted from the training set, in order to be able to distinguish the learned category (positive samples) from other categories (negative samples). The system is then able to classify previously "unseen" images and determine if they belong to the learned category or not, by analyzing their features.

Technologies for automatically assigning semantic keywords to videos usually use approaches for categorizing images as a basis, since a video is made up of many fundamental components called frames (usually at least 25 frames/second), which have similar characteristics as images. Hence, the keywords assigned to a video after analyzing its content will be a combination of the keywords assigned to the individual frames of the video.

2 Semantic Concept Detection System

In order to address this challenge of automatic semantic concept identification in images and videos, the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute (HHI) has developed an automatic system (Mbanya et al. 2011a,b). The system currently uses a concept catalogue of over 90 concepts, covering object-based concepts, e.g., *dog*, *person*, *car*, scene-based concepts e.g. *landscape*, *beach*, representation-based concepts e.g. *portrait*, *art*, sentiment-based concepts e.g. *happy*, *sad*, etc. However, the concept catalogue can be extended to include any desired number of concepts as suits the targeted application scenario.

2.1 System Approach

The approach used by the system to classify images and videos is based on the state-of-the-art *visual vocabulary* (Csurka et al. 2004) model. Visual vocabularies are a technology borrowed from natural language processing and allow for an image to be represented in terms of occurrences of visual words instead of single pixels. The rationale behind this abstraction from pixel level to a more compact representation is twofold: Firstly, it greatly reduces the amount of data used to describe an image; a characteristic which is crucial for the feasibility of automatic concept classification systems. For example, the amount of data in an 800 by 600 pixels sized image is reduced by a factor of at least 360, even when using a relatively large visual vocabulary of 4,000 visual words. Secondly, it abstracts away details which are supposed to be invariant against semantic concepts/objects depicted in an image.

Unlike in natural language, where vocabularies have evolved over centuries, there are no existing visual vocabularies which can be used and must hence be generated.

The visual vocabulary model used for image and video concept classification can be seen as divided into various stages: Identification of interest points using a suitable interest point detection strategy and description of the interest points using selected features. This is followed by creating a codebook (vocabulary) from the descriptors of the sampled points, and finally using a chosen learning algorithm for training and classification.

Interest points used by the developed system are obtained using a dense sampling strategy i.e. sampling the image at a regular interval of six pixels and at a fixed scale. For each sampled point, OpponentSIFT (van de Sande et al. 2010) features are extracted in the neighborhood of the point in the form of a 384-dimensional histogram. OpponentSIFT features are an extension of SIFT (Scale-Invariant Feature Transform) (Lowe 1999) features to the opponent color space. SIFT features are 128-dimensional histograms of gradient directions calculated around each sampling point and aligned to the main gradient direction, i.e. the gradient direction with the largest magnitude. This results in a rotation-invariant representation of the image patch. In this system, however, the features are not scale invariant due to the lack of an image pyramid consisting of the image at multiple scales from which the features are extracted. This leads to increased gains in runtime performance, though at the expense of classification accuracy.

Codebook creation is performed by vector quantizing the descriptors of the interest points sampled from all images of the training set. The method used in the system is the K-means clustering algorithm, with 4,000 clusters. K-means clustering is an unsupervised classification mechanism which partitions a dataset into a number of predefined clusters (groups), by minimizing a distance measure (usually the Euclidean distance) between each data member and its nearest cluster center. The center of each of the 4,000 clusters represents a distinct visual word in the codebook.

Based on the generated codebook, a 4,000-dimensional feature vector is computed for each image of the training and test set, by assigning each gradient histogram of the interest points from the dense grid sampling to a visual word of the codebook, using a nearest neighbor classifier.

In the next step, all 4,000-dimensional vectors computed from the training set are used to train a classifier. The classifier used is the kernel-based Support Vector Machine (SVM) in a two-class setting for binary classification. For each category, a separate SVM is trained, based on a set of manually labeled training images. Instead of using all images that do not represent the respective category as negative samples, the system selects the negative samples by focusing on those samples that are hard to classify. This is done by classifying a subset of the negative samples and choosing those where the feature vector lies within the support margin around the SVM's separating hyperplane.

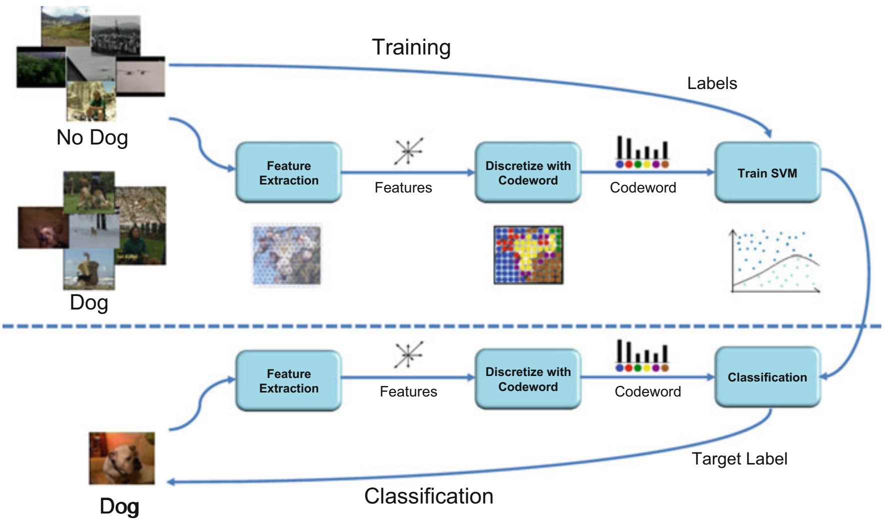


Fig. 1 Workflow of the Fraunhofer HHI semantic concept detection system

The Chi-squared (χ^2) distance, which has been shown to provide good results for comparing distributions (Zhang et al. 2006), is used to calculate the distance between image feature vectors. For two image histograms $H = (h_1, \ldots, h_m)$ and $H = (h'_1, \ldots h'_m)$, the χ^2 is defined as:

$$D(H, H') = \frac{1}{2} \sum_{i=1}^{m} \frac{(h_i - h'_i)^2}{|h_i| + |h'_i|} \,. \tag{1}$$

To incorporate this metric into the Support Vector Machine, a Gaussian kernel is used:

$$K(H, H') = \exp\left(-\frac{1}{\gamma} D(H, H')\right) \,. \tag{2}$$

The normalization parameter γ and the cost parameter C of the SVM are optimized using cross-validation. To speed up the training of the SVM, a precomputed kernel matrix is used. The trained SVM is then used to classify images or video frames of a test set as belonging to the learned category or not. Figure 1 shows the workflow of the semantic concept detection system.

Preserving Spatial Information Representing an image as a 4,000-dimensional vector using the visual vocabulary model results in a loss of spatial information. While this is tolerable for some categories, other categories such as “landscape” or “beach” tend to be more sensitive to the loss in the image’s spatial composition, since the spatial layout plays a significant role in the perception of these categories.

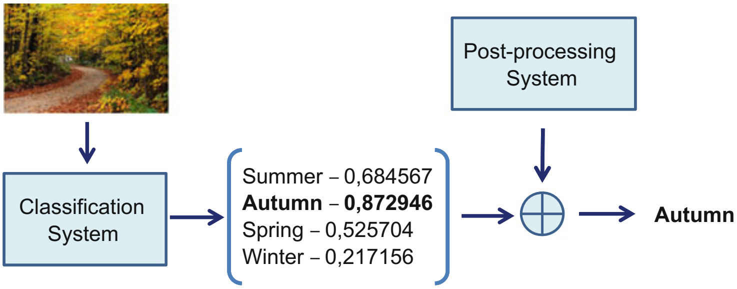


Fig. 2 Example illustration of exclusion rule

To improve classification performance for such categories, the semantic concept detection system can incorporate two different strategies for including spatial information into the model.

The first strategy to incorporate spatial information into the model: *Spatial Pyramids* (Lazebnik et al. 2006) divide an image into the spatial partitions 1×3 (consisting of three vertically stacked regions) and 2×2 (consisting of four image quadrants). For each region of the spatial partitions, a 4,000-dimensional histogram is computed and all the histograms are concatenated. This results in histograms of dimension 4,000 (for the entire image), 12,000 (for the 1×3 spatial partition) and 16,000 (for the 2×2 spatial partition) per image. In the spatial pyramids model, all three histograms are fused and used for training the SVMs.

The second strategy: *Spatial Codebooks* (Mbanya et al. 2011b) extract spatial features from the densely sampled grid by concatenating the computed OpponentSIFT orientation histogram for each sampled interest point and its location vector in the image or video frame. All the computed spatial features from all images or video frames from the training are vector quantized using the K-Means clustering algorithm, resulting in a "spatial codebook". Training the SVM using this model happens analogously as described in the general method (Sect. 2.1).

Category Post-processing The semantic concept detection system further performs a post-classification optimization step to improve the classification results of the different concept classifiers. This optimization step uses a set of rules to exploit dependencies between category sets which are learned from the training set and applied to classification results of the test set. These rules could be exclusion or inference rules. Exclusion rules are applied to category sets such as "winter", "autumn", "summer" and "spring" or "indoor" and "outdoor", ensuring that only one of the mutually exclusive categories from each set is assigned to an image or video frame. Inference rules are applied to sets of categories such as "tree" and "plant" or "cloud" and "sky", in order to boost the probability of existence of the inferred category (e.g. plant) to the inferring category's (e.g. tree) probability of existence, if it is otherwise smaller. Example illustrations of exclusion and inference rules can be seen in Figs. 2 and 3.

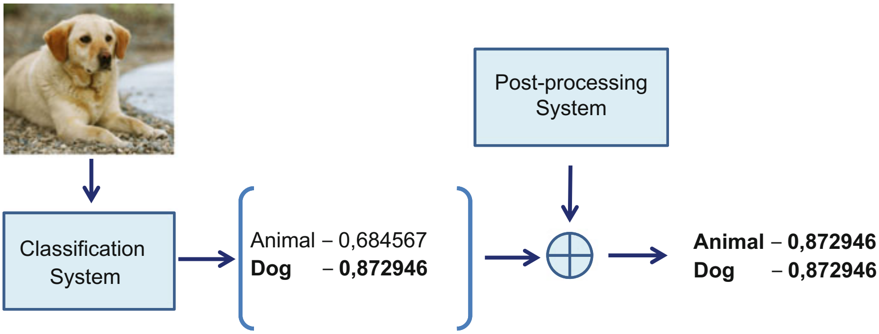


Fig. 3 Example illustration of inference rule

2.2 System and Results

The evaluation of the semantic concept detection system was performed using two well-known datasets; the data collection from the ImageCLEF 2010 & 2011 Photo Annotation Tasks and the PASCAL VOC 2007 data collection.

The PASCAL VOC 2007 is a dataset containing nearly 10,000 images of 20 different concepts with mostly object-based concepts (e.g. airplane, bicycle, etc.). The dataset is divided into a predefined training set (5,011 images) and a test set (4,952 images).

The ImageCLEF dataset consists of images from the publicly available Flickr photo collection and it contains 18,000 images and over 90 different concepts. In addition to object-based concepts, this dataset also contains scene-based concepts, e.g., *landscape*, *beach*, representation-based concepts e.g. *portrait*, *art*, and sentiment-based concepts e.g. *happy*, *sad*, etc. The dataset is divided into a predefined training set (8,000 images) and a test set (10,000 images).

Table 1 shows the gains in Mean Average Precision (MAP) obtained when spatial information is included into the classification model (configurations 2×2, 1×3 and Spatial Codebooks) and when classification is performed without the inclusion of spatial information into the model (configuration 1×1).

Table 2 shows the gains in Ontology Score with Flick Context Similarity (OS-FCS) (Mbanya et al. 2010) obtained when post-classification processing algorithms are applied to the classification results of the concept classifiers.

Table 1 MAPs of evaluations on the PASCAL VOC 2007 dataset for different classifications using the HHI semantic concept detection system, showing gains achieved when spatial information is included in the model (Mbanya et al. 2011b)

Configuration	MAP
1×1	0.447
2×2	0.457
1×3	0.477
Spatial codebooks	0.480

Table 2 MAPs of evaluations on the ImageClef dataset for different classifications using the HHI semantic concept detection system, showing gains achieved when post-processing algorithms are applied on the classification results (Smeulders et al. 2000)

Configuration	OS-FCS
General method without post-processing	0.6318
General method with post-processing	0.6401

In addition to the automatic semantic concept detection system, a web-based demonstrator has been developed which provides an interface for browsing images using the concepts detected by the system. The demonstrator also provides another interface which allows for real-time semantic concept detection in images and video key-frames, i.e. a user can upload an image and the system returns a list of concepts detected in the image with the corresponding confidence values after an analysis of the image's content, as seen in Fig. 4.

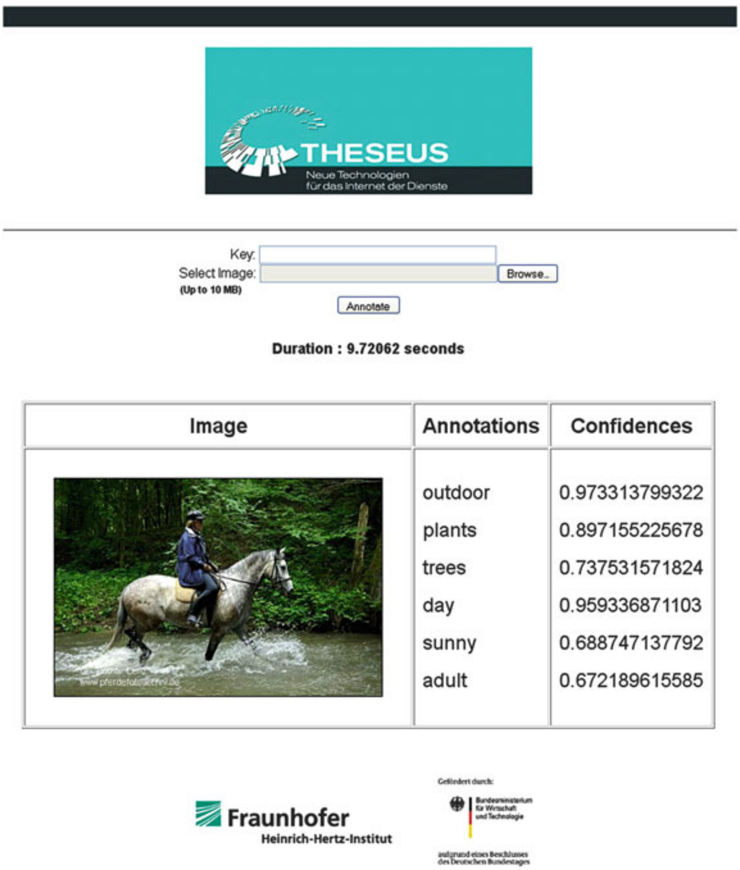


Fig. 4 Screenshot depicting results of real-time concept detection for an image

2.3 Video Specific Adaptations

One main factor which influences the classification accuracy of semantic concept detection systems is the size and the quality of the ground truth (training) data. A large and well balanced collection of positive examples (i.e. the training set consisting of images or video key-frames representative of the concepts present in the target domain (test set)) leads to superior classification results. For videos, such a training set can easily be provided for narrow domains, as they "have a limited and predictable variability in all relevant aspects of their appearance" (Smeulders et al. 2000). Videos of medical and industrial inspection as well as satellite and surveillance videos are examples of such narrow domain sets (Roach et al. 2002), since recording characteristics within these sets are similar. In contrast, general broadcast and most web content belongs to broad domains, which "have an unlimited and unpredictable variability in their appearance even for the same semantic meaning" (Smeulders et al. 2000). Finding appropriate training sets for this kind of content is a challenging task. On the other hand, there exist more extensive and precise annotations for still image datasets, since static image annotation is more common. Consequently, this raises the question of whether images, when used as an additional resource for training, can improve automatic video annotation.

Fusing different digital content domains such as images and videos raises some general challenges. For instance, the variation in resolution within either the video itself or the image domain is uncommon. However, when mixing media content from different sources, datasets are no longer homogeneous in terms of their resolution. Unfortunately, approaches that are scale invariant by definition, e.g. the Harris-Laplace interest point sampling strategy usually used with SIFT features, perform worse than dense interest point sampling strategies (van de Sande et al. 2010). Furthermore, visual characteristics of video key-frames differ from those of images (e.g. blur, image composition) which may be caused by motion, compression or differing natural scene statistics. Using the automatic concept detection system, the Fraunhofer HHI has carried out evaluations to address these issues and has found out that using simple adaptation techniques on the visual features and the content itself can result in higher classification accuracies.

Adaptation Techniques To evaluate the influence of augmenting video datasets with image datasets for semantic concept detection, the following cross-media adaptation techniques were investigated:

- Scaling of images to match the resolution of video frames before extracting the OpponentSIFT features used in the concept detection workflow. The rationale behind this approach was that using heterogeneous image sizes might have a negative effect on the classification results.
- Extracting OpponentSIFT features at multiple (3) scales from the originally sized images and video frames and using the multi-scaled features in the concept detection workflow. The rationale behind this approach was that extracting

Table 3 MAPs of evaluations on the ImageClef and TRECVID datasets for different classification configurations using the HHI semantic concept detection system, showing gains achieved with different cross-media adaption techniques

Adaptation technique	MAP
Original images + original videos	0.210
Resized images + original videos	0.197
Multiscaled images + multiscaled videos	0.229
Normalized images + normalized videos	0.219
Normalized and aligned images + normalized videos	0.215
Replicated images + replicated videos	0.223
Multiscaled normalized replicated images + multiscaled normalized replicated videos	0.243

features at multiple scales could be used as an alternative method to compensate for the difference in resolution between the images and video frames.

- Normalizing the OpponentSIFT features extracted from the originally sized images and video frames separately to zero mean and unit variance. Normalizing features however results in loss of information that is encoded in the mean and variance of the feature dimensions. To compensate for this loss of information, a second normalization approach was investigated whereby an inverse normalization of the normalized image features is performed using the variance $\boldsymbol{\sigma}_V^2$ and the mean $\hat{\boldsymbol{x}}_V$ obtained from the video features as follows:

$$\boldsymbol{x}_i' = \frac{(x_i - \hat{x}_I)}{\sigma_I^2} \times \sigma_V^2 + \hat{x}_V \ . \tag{3}$$

- Augmenting the feature vectors of the images and video key-frames obtained after the K-NN classification step (and used for training and classification in the SVM) in a similar manner to that described in Roach et al. (2002), which has been shown to improve classification performance. The image and video key-frame feature vectors were replicated with the zero vector $\mathbf{0} = (0, 0, \ldots, 0) \in \mathbb{R}^F$ in the respective position as shown below, resulting in a vector dimension three times the original vector size:

$$\hat{\boldsymbol{h}}_V = \langle h_v, h_v, 0 \rangle \ , \quad \hat{\boldsymbol{h}}_I = \langle h_I, 0, h_I \rangle \ . \tag{4}$$

Evaluations Different adaptation techniques were evaluated using two well-known data collections: the image data collection from the ImageCLEF 2011 Photo Annotation Task and the video data collection from the 2008 TRECVID High-level feature extraction task. The TRECVID dataset consists of key-frames extracted from a collection of news and documentary videos from the Netherlands Institute for Sound and Vision. For the evaluation, emphases were laid on the categories which collectively occurred in both datasets, such as dog, bridge, flower, mountain, etc.

Table 3 shows the gains in Mean Average Precision (MAP) for the different cross-media adaptation techniques. All of the adaptation techniques, except where the images were resized, resulted in classification gains as opposed to when no adaptation was performed at all. The highest overall classification gains however were seen to occur from a combination of some of the adaptation techniques: i.e. an initial extraction of OpponentSIFT features from the images and video frames at multiple (3) scales (multiscaled), followed by a feature normalization (to zero mean and unit variance), which were then used to create the codebook and feature vectors as described in the workflow (normalized). This was then followed by a replication of the feature vectors, as in Eq. 4 (replicated). This resulted in a gain of $\sim$3.3 % MAP compared to when no adaptation was performed at all.

3 Conclusion

The Fraunhofer HHI has developed a generic semantic concept detection system which can be used to automatically detect a wide range of varying semantic concepts in images and videos. Through different possibilities to extend the base system with specific modules, the classification performance can be improved by the inclusion of spatial information into the classification model and the application of post-classification processing algorithms to the classification results of the classifiers. In addition, the system can integrate different cross-media adaptation techniques to improve classification performance when augmenting video datasets with image datasets for semantic video concept detection.

References

G. Csurka, C.R. Dance, L. Fan, J. Willamowski, C. Bray, Visual categorization with bags of keypoints, in *Workshop on Statistical Learning in Computer Vision (ECCV)*, Prague, 2004, pp. 1–22

S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, New York, vol. 2 (IEEE Computer Society, 2006), pp. 2169–2178. http://dblp.uni-trier.de/db/conf/cvpr/cvpr2006-2.html#LazebnikSP06

D.G. Lowe, Object recognition from local scale-invariant features, in *Proceedings of the 7th IEEE International Conference on Computer Vision*, Kerkyra, 1999, vol. 2, pp. 1150–1157

E. Mbanya, S. Gerke, C. Hentschel, P. Ndjiki-Nya, Sample selection, category specific features and reasoning, in *Conference on Multilingual and Multimodal Information Access Evaluation*, Amsterdam, ed. by V. Petras, P. Forner, P.D. Clough (Sept 2011a)

E. Mbanya, S. Gerke, P. Ndjiki-Nya, Spatial codebooks for image categorization, in *Proceedings of the 1st ACM International Conference on Multimedia Retrieval (ICRM '11)*, Tsukuba, ed. by F.G.B.D. Natale, A.D. Bimbo, A. Hanjalic, B.S. Manjunath, S. Satoh (ACM, New York, 2011b), http://dblp.uni-trier.de/db/conf/mir/icmr2011.html#MbanyaGN11

E. Mbanya, C. Hentschel, S. Gerke, M. Liu, A. Nürnberger, P. Ndjiki-Nya, Augmenting bag-of-words – category specific features and concept reasoning, in *International Conference of the Cross-Language Evaluation Forum (CLEF)*, Padua, ed. by M. Braschler, D. Harman, E. Pianta, 2010, http://dblp.uni-trier.de/db/conf/clef/clef2010w.html#MbanyaHGLNN10

M. Roach, J.S.D. Mason, N.W.D. Evans, L.Q. Xu, F. Stentiford, Recent trends in video analysis: a taxonomy of video classification problems, in *Proceedings of the International Conference on Internet and Multimedia Systems and Applications (IASTED '02)*, Crete, 2002, pp. 348–353, http://dblp.uni-trier.de/db/conf/imsa/imsa2002.html#RoachMEXS02

A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. **22**(12), 1349–1380 (2000)

K.E.A. van de Sande, T. Gevers, C.G.M. Snoek, Evaluating color descriptors for object and scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. **32**(9), 1582–1596 (2010), http://dblp.uni-trier.de/db/journals/pami/pami32.html#SandeGS10

J. Zhang, M. Marszalek, S. Lazebnik, C. Schmid, Local features and kernels for classification of texture and object categories: a comprehensive study, in *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW '06)*, New York, June 2006, vol. 73(2), pp. 213–238, http://dblp.uni-trier.de/db/journals/ijcv/ijcv73.html#ZhangMLS07

Content Identification and Quality-Based Ranking

Claudia Nickel, Xuebing Zhou, Mohan Liu, and Patrick Ndjiki-Nya

Abstract Copyright infringements on the Internet are a major problem for rights owners. Especially copies with a high visual quality must not be published uncontrolled on free online portals – in the perception of content owners such as TV stations for instance. In the THESEUS CORE TECHNOLOGY CLUSTER (CTC), the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute (HHI) and the Fraunhofer Institute for Computer Graphics Research (IGD) have developed innovative technologies for copyright protection. Perceptual hashing algorithms are used to extract short hash values from the content which allow a robust identification of the same. The identified content can be further analyzed concerning its visual quality and thus its appropriateness for online publishing. For that, automatic quality assessment algorithms are applied. This article describes these two approaches. In addition, the sample application "Detection of copyright infringements" that combines the two developed technologies is described.

1 Introduction

Multimedia content, such as images or videos, undergoes various processing steps after it has been captured. Before transmission via a network, compression is often applied, but also other modifications, such as changes in brightness or size, are common. The challenge when designing perceptual hashing algorithms is that these modified versions have to be assigned to a hash value which is similar to the one of

C. Nickel (✉) • X. Zhou
Fraunhofer Institute for Computer Graphics Research (IGD), Darmstadt, Germany
e-mail: alexander.nouak@igd.fraunhofer.de

M. Liu • P. Ndjiki-Nya
Fraunhofer Institute for Telecommunications (HHI), Berlin, Germany
e-mail: Mohan.Liu@hhi.fraunhofer.de; patrick.ndjiki-nya@hhi.fraunhofer.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_8,

the original content. At the same time, dissimilar hash values have to be assigned to different content. This allows the robust identification of several versions of the same content based on the distance of their hash values.

2 Perceptual Hashing

In THESEUS perceptual hashing algorithms for images and videos have been developed. The algorithm for videos computes the hash value based on block-wise correlation between frames. It has a high robustness against different kinds of distortions such as noise, compression and brightness or contrast modifications.

The algorithm for images divides each color channel into overlapping blocks and creates the hash value based on the color distribution of the blocks. This results in a robust hash value while keeping the computational complexity low. A fast computation and comparison of the hash values is especially important when handling large databases.

In addition to the detection of copyright infringements, perceptual hashes can be used to identify duplicates in databases. In combination with the quality assessment this can be used to ensure compact databases of high quality content. Furthermore, perceptual hashes can be used as link to a metadata database containing information such as copyright regulations or the owner of the image. For perceptual hashing for videos, further application scenarios are possible, e.g. in the area of video restoration. In this case, often several pieces of the original content do exist. By comparing the hash values it is possible to identify overlapping video sections which allow a reconstruction of the original content.

Formal Definition Given two finite alphabets A_1 and A_2, a perceptual hash function is a non-injective function $H\colon A_1^* \rightarrow A_2^k$, with the following properties:

- H is a one-way function, i.e. for all $M \in A_1^*$ there is no efficient method to compute pre-image M from image $h = H(M)$.
- Hash value $h = H(M)$, with $|h| = k$, is easy to compute, given M.
- Perceptually identical or similar data M and M' are mapped to an identical or similar hash value (robustness): $P\{H(M) \approx H(M')\} \approx 1$, $\forall M \approx M'$, M, $M' \in A_1^*$.
- Different data M_1 and M_2 are mapped to different hash values (discrimination): $P\{H(M_1) = H(M_2)\} \approx 0$, $\forall M_1 \neq M_2$, $M_1, M_2 \in A_1^*$.

The third property makes perceptual hashing distinct from cryptographic hashing. Cryptographic hashes are designed to protect integrity of original data (Preneel 1993). Cryptographic hashes of similar data must be different. In addition, it must be difficult to obtain the clear text (the original data) from the plain text (ciphered data) – this is also called pre-image resistance. So cryptographic hashing remains collision free in a more strict way. Cryptographic hashing is absolutely independent of its clear text (input data). Even if only 1 bit is flipped, it will result in different

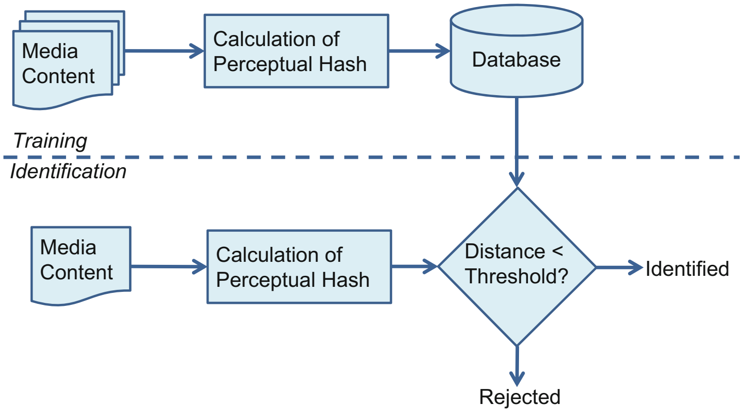


Fig. 1 Training and identification process using perceptual hashing

hash values. By contrast, perceptual hashing needs the similarity of hash values of manipulated data and the original data to realize content identification (Mihçak and Venkatesan 2002).

Identification Process Figure 1 depicts the identification process. During the training phase, the hash values of the existing media content are calculated. They are stored in a database together with relevant metadata, like copyright owner or date of acquisition. In the identification phase, the hash value of the query data is calculated and compared to the hash values stored in the database. The resulting distance values are compared to an algorithm- and application-specific threshold. If the distance is below that threshold, the query data is identified to be contained in the database. The queried content is assumed not to be stored in the database otherwise.

Developed Approach for Video Perceptual Hashing The first step for the calculation of the perceptual hash for videos is a partition of each frame of the video into small blocks. Based on the luminance of the pixels, the linear correlation of temporally adjacent blocks is calculated. This data is used to calculate the spatio-temporal difference between the blocks. Finally, these differences are binarized to obtain the spatio-temporal feature of the frame. The final perceptual hash is a two-dimensional bit-matrix, which contains the spatio-temporal features of each frame of the video. Each frame is compressed to an N-bit string by applying the aforementioned method. This string is denoted as a sub-hash and is one row in the hash matrix. A perceptual hash matrix contains a fixed number of these sub-fingerprints from consecutive frames. Figure 2 shows two binary sample hash values of two different video snippets. The left hash value was computed from a video with slowly varying frames, whereas the video corresponding to the right picture was temporally highly varying.

Producing a binary fingerprint is not a necessary property of perceptual hashing algorithms. However, it reduces the complexity of the database structure and the

Fig. 2 Training and identification process using perceptual hashing

matching process because it allows for a fast calculation of the hash distance. The obtained structure of the perceptual hashes is suitable not only for identifying the complete content but also for localization of a clip inside its original movie. To achieve this, all possible hash blocks are compared. Therefore, the comparison of two videos is started at their maximum matching. Afterwards, one of the hashes is shifted by one sub-hash forward and backward along the temporal direction, until there is no overlapping at all. After each shift the normalized Hamming distance is calculated. The minimum of these distances is considered as the final distance of the hash comparison.

Evaluation The presented algorithm was evaluated by Fraunhofer IDMT, which was responsible for the evaluation work package in THESEUS. To analyze the robustness against various distortions, a video database consisting of 100 original videos with a total length of around 200 min was used. Twenty of these videos were modified in ten different ways to form the query set. These modifications included changes in brightness, change of color and resolution, taking a 5-s excerpt, adding noise and horizontal flip. The evaluation showed that the algorithm is very robust against changes in brightness, addition of noise, taking excerpts and conversion to CIF. After these modifications, 100 % of the videos were identified. A horizontal flip resulted in a much lower recognition rate of 35 %. Due to the fact that the algorithm is based on inter-frame correlation, the videos cannot be recognized anymore after a reduction of the frame rate.

3 Quality Assessment

Compression artifacts are the most typical distortion in online video clips. Their specificities are twofold: blur and block structures. It must be said, however, that the latter specificity depends on the compression method (Rao and Hwang 1996). So-called no-reference (NR) quality measures are required in the given context. They determine the quality of an image based on its intrinsic properties as no reference image is known. As such, NR quality assessment is one of the most

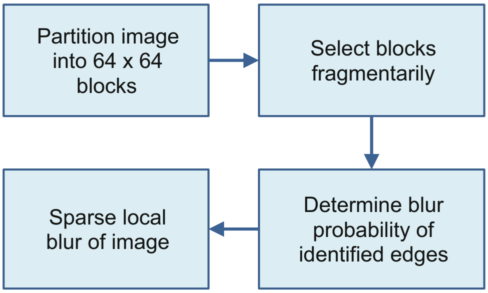


Fig. 3 Workflow of the proposed sharpness metric

difficult tasks in the field and requires prior knowledge to be efficient (Wang and Bovik 2006).

Sharpness Sharpness is one of the most important properties of image quality. Many reasons beyond compression may lead to the distortion of image sharpness such as smoothening/denoising, defocusing, motion, or discoloration of photos. Most of the objective NR sharpness metrics are proposed in the frequency or spatial domain. Although the metric *Just Noticeable Blur* (JNB) (Ferzli and Karam 2009) and its improved version *Cumulative Probability of Blur Detection* (CPBD) (Narvekar 2009) feature a better performance than other state-of-the-art NR metrics, their runtime costs are very high. We therefore propose a two-stage sharpness metric that shows better runtime performance than JNB and CPBD.

Figure 3 illustrates the computation of the local blurriness probability. The image is first partitioned into 64×64 blocks. However, evaluation of dominant structures and textures may not require selecting all blocks. Thus, we select the blocks fragmentarily to increase processing speed. Our experiments have shown that the best performance is achieved if the total area of selected blocks is not less than 25 % of the image area independent of the content. The second step consists in the block selection. The blocks may be selected at $m \,\%2 = 0$ and $n \,\%2 = 0$, where m and n correspond to coordinates at block resolution. Thirdly, a block is classified as an edge block when at least 0.2 % of its pixels are marked as edge pixels (Ferzli and Karam 2009). The blur probability of an edge block is defined as the sum of blur probabilities of all pixels within the edge block (Ferzli and Karam 2009). The overall local blur probability is defined as

$$S_L = \frac{N_b \Gamma_n}{\Gamma}, \quad (1)$$

where N_b is the number of processed edge blocks, Γ is the sum of the blur probabilities of all processed edge blocks and Γ_n is a normalization factor defined as

$$\Gamma_n = \left(\sum_b \left| \frac{64 \times 64 \times 0.2\,\%}{w_{\mathrm{JNB}}(b)} \right|^{\beta} \right)^{\frac{1}{\beta}}, \quad (2)$$

which denotes the sum of the blur probabilities of all edge blocks with ideal sharpness qualities. $w_{\mathrm{JNB}}(b)$ is the edge width of the JNB based on the contrast of the block b. The parameter β controls the curvature of the psychometric probability function and is chosen between 3.4 and 3.8. However, the sparse local sharpness analysis does not measure global image properties. According to the curves and surface theory, edges of objects, which have distinctive shapes, are normally continuous in a small area. Therefore, it can be supposed that the edges in an unassessed edge block are similar to the edges in its analyzed neighbor edge blocks. The proportion between the edge and non-edge blocks is approximately equal to the proportion between the high and low frequency components in an image. We suggest using *Higher Order Statistics* (HOS) to extract the high frequency components from images, as they can suppress Gaussian noise and preserve non-Gaussian information (Gelle et al. 1997). The weight of the high frequency components W, which is estimated as the distribution of edges-of-interest within the image, can be used to extend the local metric S_L to the whole image area. Thus, the proposed sharpness metric based on both local and global sharpness analysis is defined as

$$S_q = W \cdot S_L \ . \tag{3}$$

Blocking Artifacts Blocking artifacts are caused by lossy compression during encoding. The images are partitioned into blocks. A compression based on the discrete cosine transform (DCT) is applied to every block. At low bit rates, the independent quantization of DCT coefficients causes perceivable discontinuities across the block boundaries. These visually apparent artifacts are referred to as blocking artifacts.

Most NR metrics for blocking artifacts are modeled in the spatial domain. They are usually based on the three characteristics of perceivable blocking artifacts, the strength of the blocking boundaries, the discontinuities across block boundaries and the flatness of the image. Compared to other NR blocking metrics, the metric defined in Sheikh et al. (2002) for JPEG and H.26x/MPEG coding artifacts has a good balance between complexity and performance. However, the range of this metric depends on its parameter sets. Thus, we normalize this metric to range $[0, 1]$ to achieve a resolution invariant measure.

Evaluation The CSIQ (Larson and Chandler 2010) database was selected to evaluate the metrics. The *Pearson* (CC) and *Spearman* (SROCC) coefficients (Renaud et al. 2003) were used to measure the correlation of the proposed metrics with *Mean Opinion Scores* (MOS) from subjective ratings. High CC and SROCC scores relate to high accuracy, monotonicity and consistency of the metric under test (Renaud et al. 2003). The best performances are highlighted in the corresponding tables. The complexity of each metric was evaluated by measuring the mean runtime in *seconds per image* (s/img). Two full-reference (FR) objective metrics, PSNR and SSIM (Wang et al. 2004), were also evaluated as benchmarks. The operating system

Table 1 Performances of sharpness measures using the Gaussian blur dataset of the CSIQ database

Metrics	CC	SROCC	Complexity (s/img)
PSNR	0.824	0.862	**0.004**
SSIM (Wang et al. 2004)	0.850	**0.925**	0.083
JNB (Wang and Bovik 2006)	0.588	0.717	1.447
CPBD (Ferzli and Karam 2009)	0.830	0.885	1.479
Proposed	**0.887**	0.872	0.459

used for the experimental environment was Windows 7 64-bit Professional with i7-M620 2.67 GHz CPU and 8.00 GB memory.

The performance of the evaluated sharpness metrics on the Gaussian blur dataset of the CSIQ database are summarized in Table 1. Our sharpness approach yields the best results based on CC. Those even outperform the FR metrics. For SROCC, similar results to CPBD are achieved. Our metric, however, has a speed-up factor greater than three times that of CPBD. The FR metrics show better SROCC scores than our metric. They are, however, not suitable for the given application context.

Table 2 shows the performance of the metrics of blocking artifacts. It is shown that the proposed blocking metric has higher CC and SROCC than other NR and FR metrics with similar runtime complexity.

4 Sample Application: Detection of Copyright Infringements

A combination of the developed technologies for quality analysis and perceptual hashing fulfills the needs of content owners that want to identify their content, e.g. to detect duplicates in databases or to detect copyright infringements. In addition, it is important to determine the quality of the detected data to rank the detected copies accordingly. This supports the content owners in making decisions for further actions, e.g. to determine which of the copies should be deleted.

A possible workflow can be described as follows. Perceptual hashing is used to find copies of videos within a scanned source – e.g. an online video portal or a library. Afterwards, the results list is sorted based on the determined quality score. The first step is to detect copies of a given search video (see Sect. 2). The hash value of the query video is computed and compared to the hash values of the videos

Table 2 Performances of blocking measures using the JPEG dataset of the CSIQ database

Metrics	CC	SROCC	Complexity (s/img)
PSNR	0.728	0.881	**0.004**
SSIM(Wang et al. 2004)	0.904	0.922	0.006
BAVE(Yammine et al. 2010)	0.937	0.902	1.049
Proposed	**0.947**	**0.933**	0.006

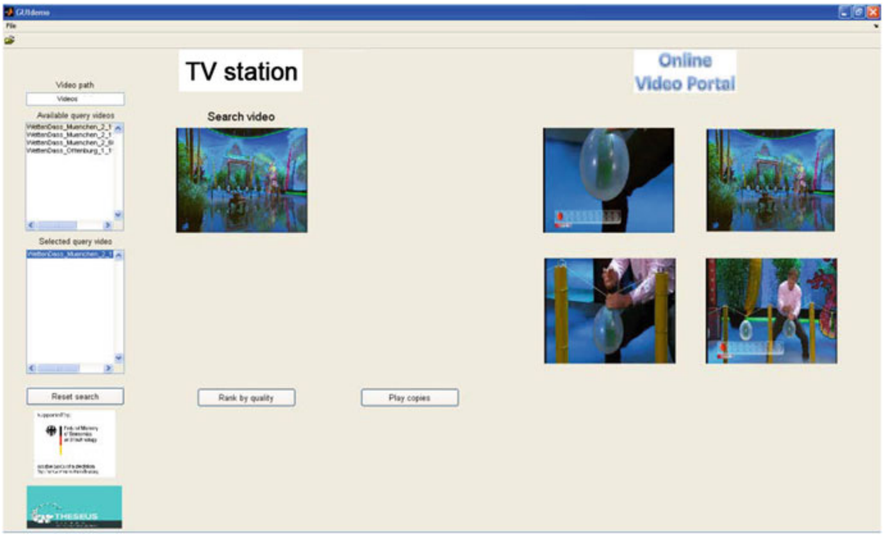


Fig. 4 Result view of the video search

within the database. Based on the distance between the hash values, copies of the search video are detected. Figure 4 shows the result view of the video search. For the given search video, displayed on the left, four copies have been detected which are shown on the right-hand side. These copies have been detected, although they are partly of very low quality and are only extracts of the search video. To sort the videos, a quality score is calculated using a perceptually driven quality analysis method based on the visual quality with a high correlation to human perception (Fig. 5). The high quality video is shown on the left; the detected low quality video is shown on the right. The diagram below the videos displays the quality of the video computed at different frame indices. The score range is 0–100, with 100 corresponding to the best attainable quality. The displayed video obtained a low overall score of 38.82, which correlates to the perceived bad quality of the video. The availability of a quality score for each video enables the possibility to exclude and remove undisturbed, high quality copies from online video portals. Low quality copies disturbed by compression or blurring, however, can be published for free for promotional purposes. Figure 6 again shows the search video on the left-hand side. The demonstrator allows setting a quality threshold to distinguish between low and high quality videos. The identified copy on the right is of high quality and should be removed from the video portal. Ranking by the factor "visual quality" is considered a new approach to sort search results.

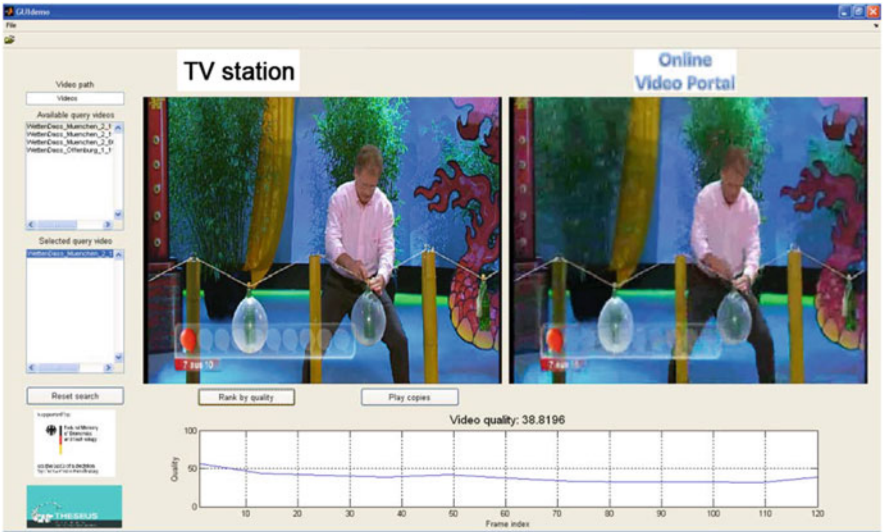


Fig. 5 Calculation of the quality score

Fig. 6 Quality-based sorting

5 Conclusion

This article describes two methods which support the efficient handling of multimedia data. Perceptual hashes are useful for detecting copies of multimedia data in large databases. Because the hash values are calculated based on image or video features they are robust against common manipulations like image or video compression or resizing. For given multimedia data a quality value can be determined which corresponds to the perceived quality of a human observer. This quality value can be used to identify further steps to handle the video. The described sample application ranks videos based on their quality value and enables the content owner to detect those copies of high quality which should not be distributed for free in online web portals.

References

R. Ferzli, L.J. Karam, A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). IEEE Trans. Image Process. **18**(4), 717–728 (2009), http://dblp.uni-trier.de/db/journals/tip/tip18.html#FerzliK09

G. Gelle, M. Colas, G. Delaunay, Higher order statistics for detection and classification of faulty fanbelts using acoustical analysis, in *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics*, Banff, July 1997, pp. 43–46

E.C. Larson, D.M. Chandler, Most apparent distortion: full-reference image quality qssessment and the role of strategy. J. Electron. Imaging **19**, 1–21 (2010), http://dblp.uni-trier.de/db/journals/jei/jei19.html#LarsonC10

M.K. Mihçak, R. Venkatesan, New iterative geometric methods for robust perceptual image hashing, in *Security and Privacy in Digital Rights Management*, ed. by T. Sander. Volume 2320 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2002), pp. 13–21

L. Narvekar, N.D. Karam, A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection, in *International Workshop on Quality of Multimedia Experience (QoMEx)*, San Diego, July 2009, pp. 87–91

B. Preneel, Cryptographic hash functions: an overview, in *Proceedings of the 6th International Computer Security and Virus Conference (ICSVC '93)*, New York, 1993

K.R. Rao, J.J. Hwang, *Techniques and Standards for Image, Video, and Audio Coding* (Prentice Hall, Upper Saddle River, 1996)

R. Renaud, V. Baroncini, A.A. Gusmão, D. Hands, H. Ikeda, C. Lee, G. Morrison, H. Myler, F. Oberti, A.C.F. Pessoa, A.M. Rohaly, M. Vandyke-Lewis, A. Vincent, A.B. Watson, S. Wolf, A. Wörner, Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment. Technical report, VQEG Final Report of FR-TV Phase II Validation Test, 2003

H.R. Sheikh, Z. Wang, A.C. Bovik, No-reference perceptual quality qssessment of JPEG compressed images, in *International Conference on Image Processing*, Rochester, 2002, vol. 1, pp. 477–480, http://dblp.uni-trier.de/db/conf/icip/icip2002-1.html#SheikhWB02

Z. Wang, A.C. Bovik, No-reference image quality assessment, in *Modern Image Quality Assessment* (Morgan & Claypool, San Rafael, 2006)

Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004), http://dblp.uni-trier.de/db/journals/tip/tip13.html#WangBSS04

G. Yammine, E. Wige, A. Kaup, A no-reference blocking artifacts visibility estimator in images, in *7th IEEE International Conference on Image Processing (ICIP)*, Hong Kong (IEEE, 2010), pp. 2497–2500, http://dblp.uni-trier.de/db/conf/icip/icip2010.html#YammineWK10

Semantics in Environmental Search Systems

Ulrich Bügel, Veli Bicer, Jens Wissmann, and Andreas Abecker

Abstract In this article, we sketch three different and complementary, semantics-enabled approaches to search for structured and unstructured environmental data and information. The SUI system is a search broker using ontology-based metadata and background knowledge for query interpretation, query dispatching, and mapping between different systems. The HIPPOLYTOS system also employs ontology-based metadata over query templates for accessing structured relational data in a data warehouse, and ontological background knowledge for query relaxation. The KOIOS system also addresses relational data, but realizes a schema-agnostic search approach that derives distributional semantics from database content in order to better interpret user queries.

1 Introduction

Nowadays, there is a wealth of data and (multimedia) information available containing knowledge about our environment: knowledge about quality of air, quality and quantity of drink and waste water, usage of land, nature protection zones and habitats, and waste management and recycling, as well as urban settlement structures, traffic networks and traffic flows, or suitability of certain places for

U. Bügel (✉)
Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB), Karlsruhe, Germany
e-mail: ulrich.buegel@iosb.fraunhofer.de

V. Bicer • J. Wissmann
FZI, Forschungszentrum Informatik, Karlsruhe, Germany
e-mail: velibice@ie.ibm.com; wissmann@fzi.de

A. Abecker
DISY Informationssysteme GmbH, Karlsruhe, Germany
e-mail: abecker@disy.net

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_9,

renewable energy installations. This knowledge is today the basis for various important activities and decisions of public authorities (e.g. in urban planning, crisis prevention, ...) and it can be the basis for manifold new value-adding services in tomorrow's data economy, in areas such as real-estate management, tourism, health research, rural development, energy planning, etc. Much of this knowledge has been collected by public authorities and is increasingly being opened to the public because of freedom-of-information acts and open data initiatives. However, the large physical and logical distributedness and heterogeneity of environmental information makes it extremely difficult to find and appropriately use all potentially relevant information pieces. Here, semantic technologies can be very valuable. In this article, we sketch several solutions developed within THESEUS and related activities that illustrate the use of semantics for searching environmental information.

2 Example 1: The SUI Semantic Search Broker

The SUI[1] project team Bügel et al. (2011), which worked closely with THESEUS, aimed at the provision of accurate, comprehensive and condensed information contained in *distributed* environmental information systems via public search interfaces. It defined a "minimal-invasive" semantic-search architecture trying to require only a minimum of changes to existing search infrastructures. SUI defined a semantics-enabled search-broker architecture for multi-source search over heterogeneous environmental information systems (see Fig. 1), which has been implemented in a prototype for the next generation Environmental Information System (EIS) of Baden-Wuerttemberg ("Umwelt-BW 3.0").

SUI accepts queries from an end user, pre-processes them semantically, uses its knowledge about connected target systems (of which Google Search Appliance – GSA – is playing a central role for text search) to dispatch queries to appropriate target systems, and then mashes up the partial results in a unified result page. The core functionality of SUI is embedded in the semantics-enabled *Search Broker*, shown in Fig. 2, which acts as a meta-search engine for the environmental portal. The search broker keeps semantic descriptions of the connected target systems, thus being able to (i) send specific queries to the appropriate target systems which may be able to answer them; (ii) create the syntactically and semantically appropriate query for each target system, based on the knowledge of what kinds of parameters they expect; and (iii) process the different result formats correctly. To identify the intended semantics of a user query, the Search Broker employs *specialized plug-ins*

---

[1] SUI = "*S*emantische Suche in *U*mwelt*i*nformationssystemen"/Semantic Search in Environmental Information Systems; the SUI project was run within the German environmental R&D cooperation network KEWA/MAF-UIS and was partially funded by the Baden-Wuerttemberg State Ministry of the Environment, Climate Protection and the Energy Sector (Ministerium für Umwelt, Klima und Energiewirtschaft des Landes Baden-Württemberg).

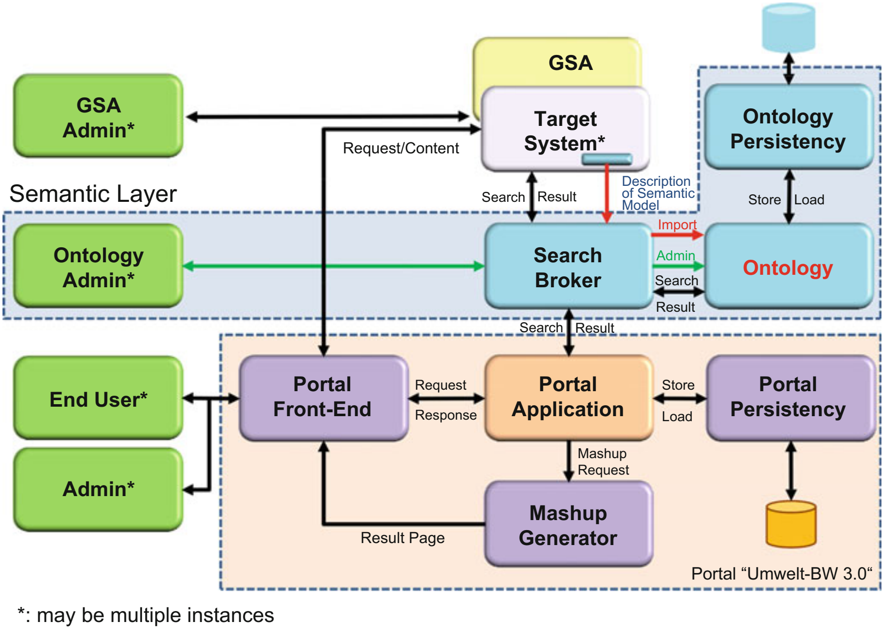


Fig. 1 Architecture of portal "Umwelt-BW 3.0"

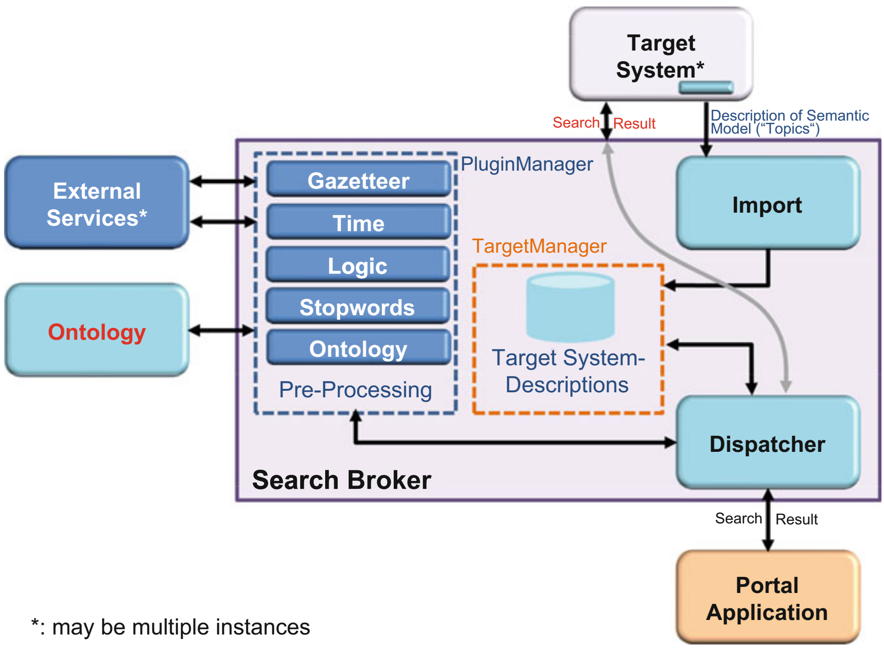


Fig. 2 Internal architecture of SUI semantics-enabled search broker

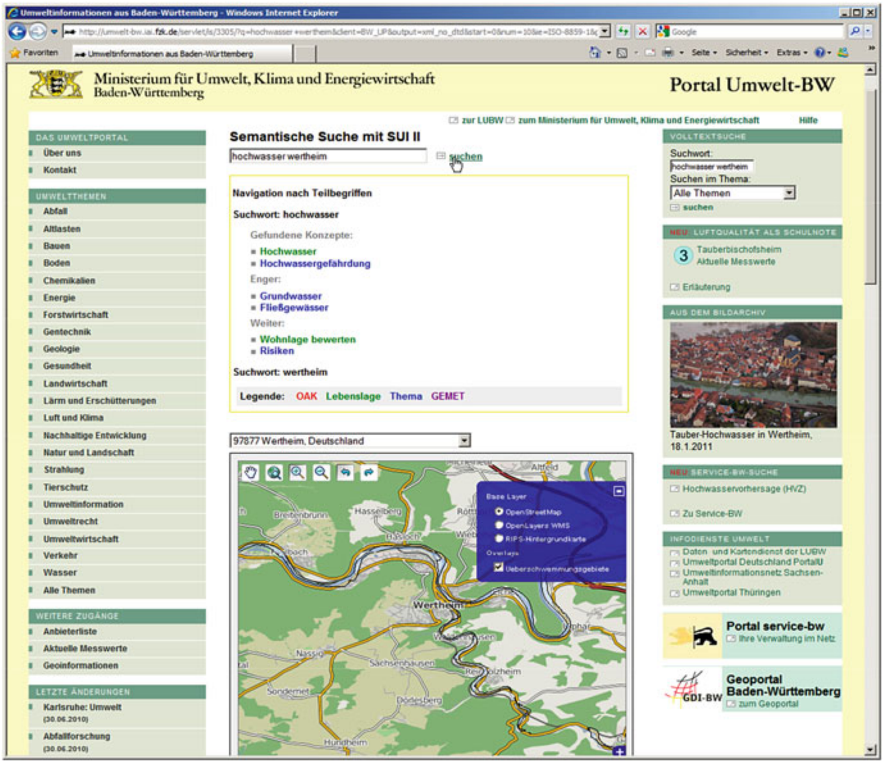


Fig. 3 Result mashup in SUI

which can resolve spatial references (Gazetteer) and temporal terms in a query, and which associate environmental thematic issues contained in a query to the appropriate ontology elements. After pre-processing, the Search Broker queries the target systems for their result data. A mashup component is responsible for the presentation of search results. Depending on the formats supplied, the portal may display them differently, but in an *integrated* results page. A broad variety of content can be represented adequately, e.g., map layers with nature reserves, measuring points as geo-point objects, full-text search results, tables, or multi-media data. Figure 3 shows the result mashup page. Directly below the query-input field there is an information box which shows the ontology concepts identified in the natural-language query, as well as broader and narrower terms for manually refining the query. The rest of the window, except for the navigation area at the left-hand side, presents different kinds of result formats in different ways:

- *Geo data* are presented in a web map client,
- *Hyperlinks*, for instance, can be shown as link lists,
- *Tabular data or diagrams* may be converted to HTML,

- *Text news* (like RSS), e.g., in overview lists,
- *HTML* fragments or micro-formats can be included at fixed places in the GUI,
- Results of *full-text* search can be presented as result lists.

3 Example 2: HIPPOLYTOS Search for Structured Environmental Data

The THESEUS SME project HIPPOLYTOS focused on *structured data* in relational databases or a Data Warehouse (Abecker et al. 2011). It developed a search layer on top of such data repositories implemented using the Cadenza software of disy GmbH.[2] Figure 4 illustrates the HIPPOLYTOS functionality: If the user types in “Eisenschrott Ballungsraum Stuttgart” (“iron junk city region Stuttgart”) in a Google-like query interface, the system reasons as follows:

- “Iron junk” is not a technical term in environmental information systems, but “recyclable fraction FE scrap” is – which is represented in the ontology, with “iron junk” as a *synonymous* wording.
- The ontology also contains the taxonomic knowledge that “potential recyclables” is a *super-concept* of FE scrap and that “metal” is a super-concept of iron/FE whereas “waste” is a super-concept of scrap.
- It also contains in its taxonomy the knowledge that “recyclable fraction Aluminium scrap” and “recyclable fraction glass” may be *siblings* to “recyclable fraction FE scrap” in the taxonomy.
- Furthermore, the lexical part of the ontology knows that “city region” is a *synonym* for “metropolitan region” or for “urban agglomeration”, an informal term that can be mapped to several spatial interpretations, such as the city of Stuttgart, the Stuttgart region constituted by six neighboring administrative districts, or the geographic area within a certain radius around the Stuttgart city center.

Using the lexical and conceptual background knowledge, the system can identify a number of stored, semantically indexed *selectors* – parameterized, predefined query templates, accessing the data sources in the back end. The match between query concepts and annotation concepts of stored selectors can be based on:

[2] disy Cadenza (http://www.disy.net/produkte/cadenza.html) is a system for building search, analysis, and visualization solutions for spatial data. At its core stands a repository system, which manages the back end data sources. Cadenza allows creating so-called *Selectors*, pre-defined query templates for the back end systems which are designed by domain experts for specific query and analysis tasks. Selectors can be described with text metadata. In HIPPOLYTOS, these text metadata for Selectors are extended by semantic metadata referring to environmental domain ontologies.

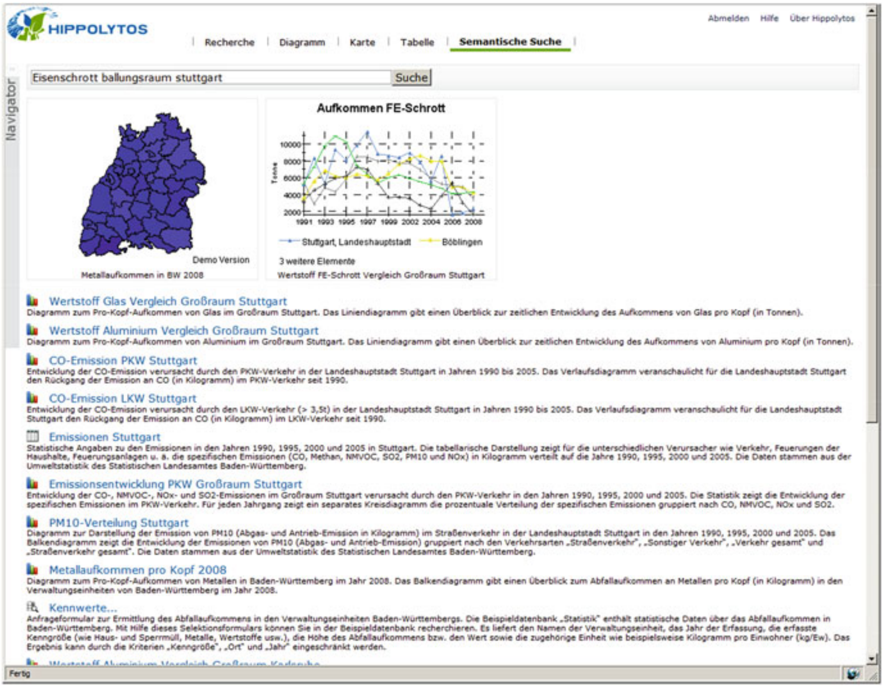


Fig. 4 HIPPOLYTOS Query for "Eisenschrott Ballungsraum Stuttgart"

- The *subject matter* of the selector (e.g., there may be a selector querying for the amount of certain recyclable `materials`, which is a parameter of this selector) in sorted waste of a given `region` (second parameter) in a given `timeframe` (third parameter) – here a proximity match could be made with the "potential recyclables" concept in the set of super-concepts of the query concepts.
- The co-domain of the selector *parameters* (e.g., "FE" could be a parameter value for the first parameter of the example selector above, and "Stuttgart" for the second).
- The *visualization* or presentation type (data value, data series, data table, map-based visualization, specific diagram type, etc.) for the results. For instance, if the query would contain terms like "comparison", "trend", or "distribution", this could give hints about the expected kind of presentation.

Then – for the given query – the most appropriate selectors and parameter settings can be identified and sent to the back end system. The result screen in Fig. 4 shows a ranked list of potential result selectors as well as previews of the visualized results of the two top-ranked ones.

4 Example 3: KOIOS Schema-Agnostic Search

The KOIOS system (Bicer et al. 2011), developed in the THESEUS CORE TECHNOLOGY CLUSTER Work Package 3 in close collaboration with HIPPOLYTOS, applies a so-called *schema-agnostic search*, which takes a set of keywords and heuristically creates a number of potential SQL queries which *might* have been meant by the user when launching his keyword query. These hypotheses are based on the given DB-schema and the statistical distribution of DB values occurring in the concrete, actual DB content. Based on these hypothetical SQL queries, one can then select those Cadenza Selectors which come close to them.

KOIOS operations can be divided into two main stages: *preprocessing* and *search*. *Preprocessing* is an offline stage (not visible to the users) that mainly creates three special search indexes out of the structured data of a given database or RDF store. The core index directly created from the data is called *data index*, which is a graph-based representation of the data implemented as an inverted index, and optimized for efficient access. Based on the data index, two other indexes are also created: the *keyword index* is mainly designed for IR-style access that captures the unstructured part of the data, and the *schema index* is extracted from the data, representing classes and relationships among them. The second stage, *search*, is the actual part in which the system interacts with the user. The user specifies his/her information with a short keyword query, considered as a set of keywords. Based on this query, the system first discovers possible keyword elements using the keyword index to find particular tuples (entities) in the data in which one of the keywords occur. A number of keyword tuple sets are created for each keyword in the query. These sets are then combined with the schema information, resulting in an augmented schema graph, which represents the query space. By exploring this graph, a number of structured query graphs is constructed, each of which can be executed on the data index to find relevant results. This part of the search stage mainly interprets the user's possible information needs in terms of structured query graphs, and computes their corresponding result sets. Based on the outputs of this stage, KOIOS generates a number of *facets* to facilitate further interactions and refinements of query and results. It uses a faceted search interface to present the possible categories (facets) and values generated from the underlying results. This helps the user to refine his/her query.

In practice, with a large DB-schema and often occurring data values, normally many different hypotheses will be possible. In order to quickly deliver probably highly relevant hypotheses to the user, ranking mechanisms have been carefully examined for KOIOS. Furthermore, the user can be offered a *faceted-search GUI* (see Fig. 5): If different potential query interpretations differ in several dimensions, each dimension will be represented by one tab in the left-hand part of the search GUI in Fig. 5. There, a search for "CO emissions in the city of Karlsruhe" yields a number of possible selectors which can be differentiated according to:

- The `year` for which information is sought (only years for which we have actual DB content will be offered).

Fig. 5 KOIOS Faceted-Search Interface for Query "Karlsruhe CO Emissionen". A selection on the left-hand side will immediately refresh the list of possible resulting queries shown on the right-hand side. Clicking one of these results on the right-hand side will then evaluate this selector and deliver the actual query results from the back end database

- The `administrative region` which is examined by the selection (concretely, Karlsruhe city versus Karlsruhe county).
- The actual `measured value` considered (e.g., CO emissions from motorbikes, CO emissions from passenger cars, CO emissions from trucks, etc.).

5 Conclusion

In this article, we briefly sketched three complementary approaches to support the search for heterogeneous and multimedia environmental information through semantic technologies: The *SUI semantic search broker* settles upon existing search interfaces and collects heterogeneous information from manifold existing back end systems. Here, ontologies are the basis for expressing metadata about back end search systems and support query disambiguation. Recent extensions of the system and subject to ongoing work are, for instance, the better automatic metadata creation for information sources and the use of ontology mapping approaches for cross-system search (Bügel et al. 2011). Prototypical implementations have been coupled to the environmental information system of the German federal state of Baden-Wuerttemberg. The *HIPPOLYTOS semantic data search* employs ontologies for providing the background knowledge to query reformulation and query extension to access predefined data selectors. It is closely coupled to disy's commercial data warehouse. At the time of writing this article, parts of the research results have already been integrated in the Cadenza release 2012. The *KOIOS schema-agnostic search* creates an internal representation of distributional semantics in order to cater to scenarios where no explicit metadata or formal ontologies are available. KOIOS has also experimentally been coupled with disy Cadenza. Altogether, the three approaches can be seen to be complementary: HIPPOLYTOS and KOIOS might be added to existing relational data systems and then be integrated as specific back end systems into an overall SUI architecture (Abecker et al. 2011). We consider approaches like those presented in this article particularly important, because more and more public sector information will become open data in the near future, and we need appropriate technology to make sense and draw benefit from this data.

References

A. Abecker, U. Bügel, R. Ebel, T. Schlachter, Integrating semantic search for relational data into environmental information systems, in *25th International Conference on Environmental Informatics (EnviroInfo-2011)*, Ispra, 2011

V. Bicer, T. Tran, A. Abecker, R. Nedkov, KOIOS: utilizing semantic search for easy-access and visualization of structured environmental data, in *The Semantic Web – ISWC 2011*, ed. by L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N.F. Noy, E. Blomqvist. Volume 7032 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011), pp. 1–16, http://dblp.uni-trier.de/db/conf/semweb/iswc2011-2.html#BicerTAN11

U. Bügel, F. Chaves, M. Schmieder, B. Schnebel, C. Düpmeier, T. Schlachter, R. Weidemann, R. Ebel, M. Tauber, B. Ellmenreich, A. Sawade, SUI für Umweltportale – Entwurf und prototypische Implementierung einer Architektur für die semantische Suche im Portal Umwelt-BW, in *F+E Vorhaben KEWA. Kooperative Entwicklung wirtschaftlicher Anwendungen für Umwelt, Verkehr und benachbarte Bereiche in neuen Verwaltungsstrukturen. Phase VI, 2010/11*, ed. by R. Mayer-F öll, R. Ebel, W. Geiger. Volume 7586 of KIT Scientific Reports (KIT Scientific Publishing, Karlsruhe, 2011), pp. 21–32

Reasoning Brokerage: New Reasoning Strategies

Jürgen Bock

Abstract Once a formal representation of data is available an important issue is to infer additional aspects out of this knowledge base. But according to the type of representation scheme chosen, different techniques can be applied or gain better or more accurate results. A reasoning broker system offers the possibility to apply strategies for selecting the best reasoning system, or for letting run different reasoners in parallel. In this article a reasoning broker system enabling the usage and integration of remote reasoners is presented. Additionally, the new reasoning capability of anytime reasoning has been developed and integrated into the reasoning broker.

1 Introduction

One of the key advantages of representing domain knowledge in ontologies is the formal semantics provided by the ontology language's logical underpinning. For instance, the Web Ontology Language (OWL) (Bock et al. 2009a) is based on description logics (Baader et al. 2003), a family of decidable fragments of first-order logic. This formal semantics enables the possibility of conducting logical reasoning in order to infer implicit knowledge from explicitly stated axioms and facts. In order to gain this added value, powerful inference engines (reasoners) are required, which draw conclusions on the provided ontology.

Depending on the use case at hand, the extent to which logically expressive language features are exploited in ontologies differs. Studies show that, in fact, most of the ontologies found on the Web are of low expressiveness in terms of language features used (Bock et al. 2008). Other differences that can be observed are in the

J. Bock (✉)
FZI, Forschungszentrum Informatik, Karlsruhe, Germany
e-mail: bock@fzi.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_10,

size of ontologies with respect to the number of entities referenced by the ontology's axioms. These sizes can range from under a dozen up to several tens of thousands of entities. A significant characteristic when analyzing the size of an ontology is the type of the entities that occur in large numbers. For instance, there are ontologies that contain a large number of classes, thus forming a taxonomically rich knowledge model. On the other hand, there are ontologies with a smaller number of classes and statements about their terminological relations, but a great amount of instances asserted to those classes.

This variety of characteristics that discriminate the ontology landscape impose different challenges on a reasoning engine, which consequently has to cope with these ontologies. In the past decade, a plethora of reasoners has been developed for OWL.[1] Some reasoners, such as HermiT[2] or Pellet[3] can serve as reference implementations for the Description Logic fragment underlying OWL and DL, and thus are capable of processing ontologies that exploit the full logical expressiveness. Other reasoners, such as KAON2,[4] strive for efficient reasoning with ontologies containing a great number of instances, at the cost of slightly reduced logical expressiveness. With the standardization of OWL 2, a range of language profiles (Calvanese et al. 2009) was introduced, which deliberately reduce the expressive power in order to reduce reasoning complexity. Reasoners such as CEL,[5] TrOWL,[6] or ELLY[7] implement such language profiles. Table 1 shows a (non-exhaustive) list of reasoners together with some of their properties.

Apart from the obvious differences in language expressiveness reflected by OWL profiles, there are more subtle characteristics, for instance in the combination of language features, which have significant performance impact on the reasoning calculus. Developers of reasoners typically implement various optimization strategies that are automatically switched on, if the input ontology allows this. Nevertheless, there are major differences in the runtime performance of state-of-the-art reasoning systems depending on the input ontology and the reasoning task at hand. The various strengths and weaknesses of reasoners have been studied extensively in recent years (Bock et al. 2008; Gardiner et al. 2006; Liebig 2006; Luther et al. 2009; Pan 2005), leading to the conclusion that there is no single best reasoning system for all reasoning scenarios. Moreover, this leads to a major inconvenience for the developer of any semantic application intending to utilize reasoning capabilities, since choosing the best reasoner is a nontrivial task. In particular, in the case where either input ontologies or reasoning tasks change, there might be different reasoners suitable for different invocations.

---

[1] http://www.w3.org/2007/OWL/wiki/Implementations

[2] http://www.hermit-reasoner.com/

[3] http://clarkparsia.com/pellet/

[4] http://kaon2.semanticweb.org/

[5] http://lat.inf.tu-dresden.de/systems/cel/

[6] http://trowl.eu/

[7] http://elly.sourceforge.net/

Table 1 Non-exhaustive selection of state-of-the-art reasoning systems (Source: Bock et al. (2012))

CEL	Native Profiles	EL
	Semantics	Direct
	(Non-)conformance	Lacks support for nominals (ObjectHasValue and ObjectOneOf) and data types/values
	Algorithm	Proprietary
	API	OWL API (not for v3.1)
	Authorization	Open source
ELLY	Native profiles	EL, RL
	Semantics	Direct
	(Non-)conformance	OWL profile support under development
	Algorithm	Rule inferencing
	API	OWL API (old version)
	Authorization	Open source
FaCT++	Native profiles	DL
	Semantics	Direct
	(Non-)conformance	Fully conforming except for keys and some data types
	Algorithm	Tableau
	API	OWL API, DIG interface
	Authorization	Open source
HermiT	Native profiles	DL
	Semantics	Direct
	(Non-)conformance	Fully conforming
	Algorithm	Hypertableau
	API	OWL API
	Authorization	Open source
Pellet	Native profiles	DL, EL
	Semantics	Direct
	(Non-)conformance	Fully conforming
	Algorithm	Tableau
	API	OWL API, Jena API, DIG interface
	Authorization	Open source
RacerPro	Native profiles	DL
	Semantics	Direct
	(Non-)conformance	No nominals and RBox
	Algorithm	Tableau
	API	OWL API, DIG interface, OWLlink
	Authorization	Commercial (free for research)
TrOWL	Native profiles	DL, EL, QL
	Semantics	Direct
	(Non-)conformance	
	Algorithm	Various proprietary
	API	OWL API (not for v3.1)
	Authorization	Open-source

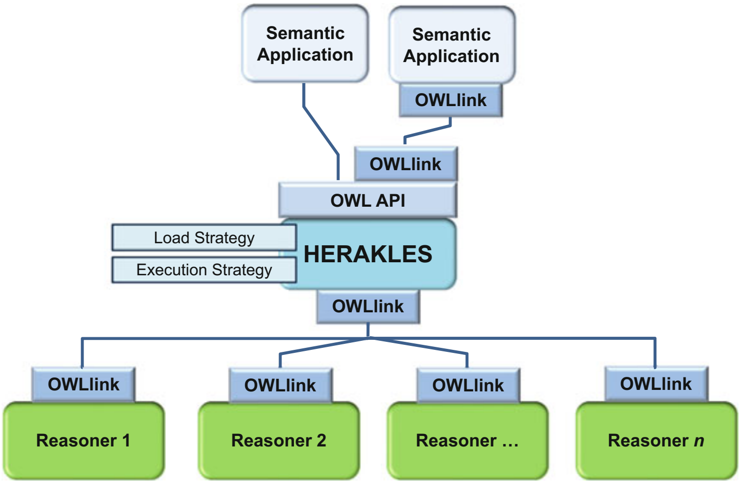


Fig. 1 Architecture of the HERAKLES reasoning broker (Source: Bock et al. (2012))

This article describes how the problem can be addressed using a reasoning broker approach, which serves as a hub between existing reasoners and a client. For the client, this broker appears as a single reasoner but delegates reasoning request intelligently to its (specialized) reasoning engines in the back end. The broker thus utilizes the capabilities of all connected reasoners, and provides additional added value, such as anytime reasoning simulation by approximation.

In the following Sect. 2, the general architecture of the reasoning broker system is introduced, which can be configured by several broker strategies, which are presented in Sect. 3. The special case of anytime behavior by approximation is dealt with in Sect. 4, before the article concludes in Sect. 5.

2 Broker Architecture

A reasoning broker framework for OWL has been implemented as the HERAKLES system[8] (Bock et al. 2009b,c) in the context of the THESEUS research program. The overall architecture of the system is illustrated in Fig. 1.

From an abstract point of view, the broker connects to several external 3rd party reasoning systems in the back end, while serving as a single reasoner endpoint for a semantic client application. The way the broker interacts with the external reasoning systems is driven by two broker strategies: A load strategy and an execution strategy, each of which is exchangeable in order to accommodate the

[8] http://herakles.sourceforge.net

particular usage scenario of the broker (see Sect. 3). Moreover, since the broker is the central component of a distributed reasoning infrastructure, it can implement centralized caching and load balancing mechanisms, in order to provide instant response for reoccurring requests, and efficiently manage the available reasoner resources.

From a more technical point of view, the HERAKLES system is internally based on the OWL API (Horridge and Bechhofer 2011), a popular, Java-based framework for dealing with OWL ontologies and a number of conforming reasoners. Due to this, semantic applications that are based on the OWL API as well can directly use the reasoner interface implementation provided by HERAKLES. On the other hand, HERAKLES implements the OWLlink protocol (Liebig et al. 2010), a standardization effort aiming at setting up a common interface in order to connect different semantic applications in particular reasoners. Adapting to this, HERAKLES can connect with arbitrary OWL reasoners that exhibit themselves as OWLlink servers.[9] Since the OWLlink protocol is HTTP-based, HERAKLES can operate with remote reasoners in a distributed environment. The fact that HERAKLES itself acts as an OWLlink server enables the setup of an intelligent reasoning service in a highly distributed environment.

3 Broker Strategies

The reasoning broker does not implement a reasoning engine or an inference calculus of any kind. Instead, reasoning requests are intelligently delegated to external third party reasoners that might be specialized for efficient processing of particular requests. This intelligent delegation is controlled by two broker strategies that are integrated into the broker as independent modules:

- A *load strategy* is responsible for loading ontologies into the registered external reasoners.
- An *execution strategy* is responsible for invoking the external reasoners for particular reasoning requests.

Both, load and execution strategy are exchangeable in order to accommodate the particular broker usage scenario at hand.

Loading The load strategy manages the initialization of the broker and all (relevant) registered external reasoners with the ontologies in the current reasoning scenario. Different load strategies are possible. A *basic* strategy simply asks all registered reasoners to load the ontologies as requested by the client. An *analyzing* strategy extracts various features from the ontologies in order to use this information for intelligent selection of reasoners for particular reasoning tasks (see Sect. 3).

[9] Reasoners that do not provide an OWLlink server interface have to be wrapped with minor effort.

Parallelization The scalability issue in the context of semantic technologies becomes more and more important. One way to address scalability in terms of ontology reasoning is to utilize parallelization (Bock 2008) techniques where possible. A straightforward way of using a reasoning broker for executing reasoning requests is to delegate the request to all registered (and idling) reasoners and return the result of the reasoner that answers first. However, it has to be ensured that the reasoners do not consume shared computing resources which would slow down the algorithms due to reduced computational power. Thus the best setting for utilizing such a strategy would be a distributed computing infrastructure with independent compute nodes, or at least guaranteed computational resources for each node (e.g. in a virtual environment, such as compute clouds).

Selection In case distributed computing resources are rare, or there is any other reason to save computational resources, it is advisable to avoid blind invocation of reasoning tasks on all available reasoners. Thus, using the features extracted by an analyzing load strategy, a *selection* strategy can determine the best suitable reasoner(s) for a given request. Apart from looking at simple rules based on the known capabilities of the various reasoners, a machine learning approach has been developed (Bock et al. 2012). In this approach, a corpus of real-world ontologies from the Web and a set of generated queries are used to determine the fastest reasoner for each ontology/query combination. The features of ontology and query are used to teach a model that is later used to predict the best suitable reasoner for any new ontology/query combination. Experiments have shown that this selection strategy can predict the best suitable reasoner with an accuracy of up to 77 %.

Fault Recovery Parallelization and the availability of a number of different reasoners can also be exploited to improve the robustness of the whole reasoning process. A *fault tolerant strategy* has been developed, which works similarly to the basic parallelization strategy, but reacts to failures of external reasoners. Such failures can occur if a reasoner does not support a particular reasoning request, or if a reasoner is in a prototypical and thus unstable state. Since a number of reasoners is invoked simultaneously, the failure of one reasoner does not cause the others to stop. An experiment with a series of subsequent queries was carried out in order to test the fault recovery achieved by the said strategy (Bock et al. 2009b). A set of 100 queries was executed using the HERAKLES reasoning broker. In four test runs HERAKLES was configured three times with a single reasoner registered (FaCT++, KAON2, and Pellet), and once with all reasoners registered. As Fig. 2 shows, the configuration with FaCT++ shows the best runtime performance. However, FaCT++ failed in 32 of the 100 queries. Compared to that, Pellet failed in no case, at the cost of higher runtime. (In fact it could be observed that the queries that caused long runtimes for Pellet are the ones that caused FaCT++ to fail rather quickly, which explains the runtime discrepancy.) Using the broker configuration with all reasoners registered and the fault tolerant strategy activated, all queries could be processed with no failures in less time than when using Pellet alone. The case where a series of queries is executed on a fixed set of ontologies is assumed

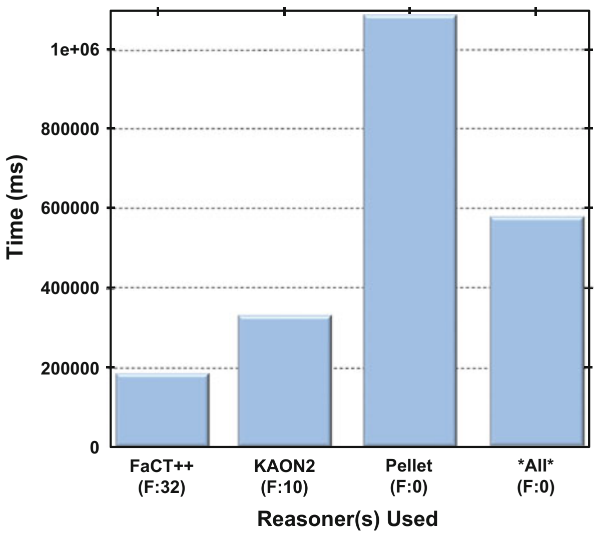


Fig. 2 Runtime comparison of reasoners in the reasoning broker framework: Query stream experiment with 100 queries. The bars denote the total runtimes using the broker configured with a single reasoner, or with all reasoners (*right bar*). Numbers in parentheses denote the number of failures (Sources: Bock et al. (2008, 2009b))

to be a common one. Thus a broker framework such as HERAKLES provides some real added value in terms of fault tolerance and runtime improvement.

4 Anytime Reasoning Through Approximation

Most reasoners are developed with the goal of providing sound and complete answers to queries. For the reasoning task of instance retrieval, this would mean that a request to obtain instances of a class description answers with a set which on the one hand contains only *correct* instances (soundness), and on the other hand *all* instances (completeness) for which it can logically be inferred that they belong to the given class description. These soundness and completeness guarantees are causing the algorithms to invoke complex inference procedures of high computational complexity. Recent approaches developed in the context of the THESEUS research program give up these soundness and completeness guarantees for an improved runtime performance by approximation (Tserendorj 2010). The Screech system developed in THESEUS is based on the KAON2 reasoner[10] and comes in three variants (Tserendorj et al. 2008):

- SCREECH-ALL: This variant ensures completeness but gives up soundness.
- SCREECHNONE: This variant ensures soundness but gives up completeness.
- SCREECH-ONE: This variant gives up both soundness and completeness.

[10] http://kaon2.semanticweb.org/

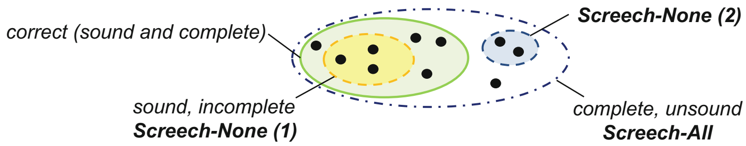


Fig. 3 Combination of variants of the approximate reasoning system SCREECH in order to achieve simulated anytime behavior for instance retrieval requests

It has been analyzed, how the combination of different kinds of anytime algorithms exhibits an anytime behavior that cannot be achieved using a single algorithm (Rudolph et al. 2008). The SCREECH variants introduced above can be arranged in such a setting as illustrated in Fig. 3. For a given class expression, an instantiation of SCREECH-NONE is invoked, which is sound but incomplete, and thus delivers correct instances of this class expression, but the result set might not be complete. Invoked at the same time, an instantiation of SCREECH-ALL, which is complete but unsound, delivers a set that contains all correct instances, but potentially also incorrect ones. A second SCREECH-NONE instance invoked with the *complement* of the original class expression delivers a set of correct negative results, and thus can be used to determine a subset of the result set from SCREECH-ALL that is definitely incorrect. Though the SCREECH instances work as black boxes, they can be invoked in parallel and will probably show different runtimes. After any reasoner returns its results, they can be delivered to the client, provided the soundness and completeness properties are indicated accordingly. Thus, an example run may provide results as follows. First, SCREECH-ALL provides a set of instances which might contain incorrect ones. Second, SCREECH-NONE (positive query) returns with a subset of the results from SCREECH-ALL. These are definitely correct and can be marked accordingly. Third, SCREECH-NONE (negative query) returns with a subset of the results from SCREECH-ALL, which are definitely incorrect and can be removed from the result set. The final result is a better approximation than using any of the SCREECH variants alone. In order to finally get the correct (sound and complete) result set, a standard reasoner may have been invoked in parallel as well, which is expected to have a slower runtime performance, but is able to verify the final result. These steps in result delivery provide the client with approximate intermediate results that are subsequently refined.

An *anytime strategy* was implemented for the reasoning broker HERAKLES, which uses the broker facilities of parallelization. To this end, a novel reasoner interface was implemented to allow for asynchronous reasoning calls – a feature required for anytime result delivery. The novel interface and broker strategy were prototypically integrated into the ontology authoring tool Protégé (Bock et al. 2009c), where a color encoding was used to indicate the “certainty” of instances in the result set.

5 Conclusion

Reasoning brokerage was introduced as a concept that tackles the problem of diversity in reasoner performance, which can be observed for existing reasoning systems. To this end, the broker framework HERAKLES was presented as a solution for semantic application developers, who face the problem of selecting the most appropriate reasoning system for a given task. On the one hand, the broker is flexible in terms of exchangeable broker strategies, while on the other hand it simplifies the usability by providing the standardized programming interfaces of the OWL API and OWLlink to the application developer. A collection of broker strategies has been implemented accommodating parallelization, selection (including machine learning techniques), fault recovery, and anytime simulation through combinations of approximate reasoning systems.

Apart from facilitating the incorporation of existing reasoners in semantic applications, the reasoning broker system can encourage developers of reasoning systems to focus on particular optimizations for specific reasoning problems, without losing a large number of potential clients. HERAKLES is designed to ease the development of additional broker strategies, which might be required in a particular usage scenario. This flexibility supports the growing interest in semantic technologies on the level of powerful and suitable reasoning systems for intelligent applications.

References

F. Baader, D. Calvanese, D. McGuinness, D. Nardi, P. Patel-Schneider, *The Description Logic Handbook: Theory, Implementation and Applications* (Cambridge University Press, New York, 2003)

C. Bock, A. Achille Fokoue, P. Haase, R. Hoekstra, I. Horrocks, A. Alan Ruttenberg, U. Sattler, M. Smith, OWL 2 Web Ontology Language – Structural Specification and Functional-Style syntax (2009a), http://www.w3.org/TR/owl2-syntax/

J. Bock, Parallel computation techniques for ontology reasoning, in *The Semantic Web – ISWC 2008*, ed. by A.P. Sheth, S. Staab, M. Dean, M. Paolucci, D. Maynard, T.W. Finin, K. Thirunarayan. Volume 5318 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2008), pp. 901–906, http://dblp.uni-trier.de/db/conf/semweb/iswc2008.html#Bock08

J. Bock, P. Haase, Q. Ji, R. Volz, Benchmarking OWL reasoners, in *Proceedings of the ARea2008 – Workshop on Advancing Reasoning on the Web: Scalability and Commonsense*, Tenerife, June 2008, ed. by F. van Harmelen, A. Herzig, P. Hitzler, Z. Lin, R. Piskac, G. Qi. CEUR Workshop Proceedings, vol. 350, http://ceur-ws.org

J. Bock, U. Lösch, H. Wang, Automatic reasoner selection using machine learning, in *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics*, Helsingborg, ed. by D.D. Burdescu, R. Akerkar, C. Badica (ACM, New York, 2012), http://dblp.uni-trier.de/db/conf/wims/wims2012.html#BockLW12

J. Bock, T. Tserendorj, Y. Xu, J. Wissmann, S. Grimm, A reasoning broker framework for OWL, in *Proceedings of the 6th International Workshop on OWL: Experiences and Directions (OWLED*

'09), Chantilly, ed. by R. Hoekstra, P.F. Patel-Schneider. CEUR Workshop Proceedings, vol. 529 (2009b), http://www.webont.org/owled/2009/papers/owled2009_submission_13.pdf

J. Bock, T. Tserendorj, Y. Xu, J. Wissmann, S. Grimm, A reasoning broker framework for protégé, in *11th International Protégé Conference*, Amsterdam (June 2009c)

D. Calvanese, J. Carroll, G. De Giacomo, I. Herman, B. Parsia, P.F. Patel-Schneider, A. Ruttenberg, U. Sattler, OWL 2 Web Ontology Language – Profiles (2009), http://www.w3.org/TR/owl2-profiles/

T. Gardiner, I. Horrocks, D. Tsarkov, Automated benchmarking of description logic reasoners, in *Proceedings of the 19th International Workshop on Description Logics*, ed. by B. Parsia, U. Sattler, D. Toman. Volume 189 of CEUR Workshop Proceedings (Windermere, Lake District, UK, 2006), http://dblp.uni-trier.de/db/conf/dlog/dlog2006.html#GardinerHT06

M. Horridge, S. Bechhofer, The OWL API: a Java API for OWL ontologies. Semant. Web **2**(1), 11–21 (2011), http://dblp.uni-trier.de/db/journals/semweb/semweb2.html#HorridgeB11

T. Liebig, Reasoning with OWL – system support and insights. Technical Report 2006-04, Universität Ulm - Fakultät für Informatik, Sept 2006

T. Liebig, M. Luther, O. Noppens, M. Wessel, OWLlink. Semant. Web – Interoper. Usability, Appl. **2**(1), 23–32 (2010)

M. Luther, T. Liebig, S. Böhm, O. Noppens, Who the Heck is the father of Bob? A survey of the OWL reasoning infrastructure for expressive real-world applications, in *The Semantic Web: Research and Applications*, ed. by L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano, T. Heath, E. Hyvönen, R. Mizoguchi, E. Oren, M. Sabou, E. Simperlm. Volume 5554 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2009), pp. 66–80

Z. Pan, Benchmarking DL reasoners using realistic ontologies, in *Proceedings of the OWL: Experiences and Directions Workshop*, Galway, ed. by B.C. Grau, I. Horrocks, B. Parsia, P.F. Patel-Schneider. Volume 188 of CEUR Workshop Proceedings (2005), http://dblp.uni-trier.de/db/conf/owled/owled2005.html#Pan05

S. Rudolph, T. Tserendorj, P. Hitzler, What is approximate reasoning? in *Web Reasoning and Rule Systems*. Volume 5341 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2008), pp. 150–164

T. Tserendorj, Approximate assertional reasoning over expressive ontologies. PhD thesis, Fakultät für Wirtschaftswissenschaften, Karlsruher Institut für Technologie (KIT), Jan 2010, http://digbib.ubka.uni-karlsruhe.de/volltexte/documents/1235410

T. Tserendorj, S. Rudolph, M. Krötzsch, P. Hitzler, Approximate OWL-reasoning with screech, in *Web Reasoning and Rule Systems*, ed. by D. Calvanese, G. Lausen. Volume 5341 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2008), pp. 165–180, http://dblp.uni-trier.de/db/conf/rr/rr2008.html#TserendorjRKH08

A Unified Approach for Semantic-Based Multimodal Interaction

Markus Löckelt, Matthieu Deru, Christian H. Schulz, Simon Bergweiler, Tilman Becker, and Norbert Reithinger

Abstract This article gives an overview of the basic dialog shell for multimodal interaction that was created as the foundation for the software components developed in the CORE TECHNOLOGY CLUSTER of the THESEUS research program. It explains the basic setup of systems based on the dialog shell for specific use cases. The general processing paradigm for interaction in our framework using semantic interface elements as well as several dialog phenomena handled by the dialog shell are described. Finally, a selection of different demonstrator systems that were built for a range of applications is presented briefly. They offer access to quite varied use case scenarios featuring multimodal input and output, including gestures, multi-touch, spoken input and output, and interactions across mobile and stationary devices.

1 Introduction

The purpose of the CORE TECHNOLOGY CLUSTER Work Package 4 (CTC WP4) is to provide generic solutions for interaction tasks that arise in semantics-based multimodal systems. Those solutions should however also allow for a flexible adaptation and configuration with respect to different use case scenarios and system setups. Since many core methods from the tasks can be reused across different use cases, the tasks implement them in generic form, with additional configurations for the concrete scenario when a system is put together. Some requirements, such as

M. Löckelt (✉) • M. Deru • S. Bergweiler • T. Becker
DFKI GmbH, German Research Center for Artificial Intelligence, Saarbrücken, Germany
e-mail: Markus.Loeckelt@dfki.de; Matthieu.Deru@dfki.de; Simon.Bergweiler@dfki.de; Tilman.Becker@dfki.de

C.H. Schulz • N. Reithinger
DFKI GmbH, German Research Center for Artificial Intelligence, Berlin, Germany
e-mail: Christian.Schulz@dfki.de; Norbert.Reithinger@dfki.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_11,

the ability to manage the system lifecycle, messaging and deployment description are independent of the use case. As one of the subtasks, *Semantic User Interface Framework*, an integrated software environment was developed that can deploy and host a situation-aware, semantics-driven dialog shell to cover these requirements. Section 2 describes this environment, called the *Dialog Shell*. The other core software modules that are used in every dialog shell system and their tasks are *Multimodal Semantic Processing*, *Semantic Navigation and Interaction*, *Interactive Service Composition and Query* and *Semantic Output Representation*. Their roles are described in more detail in Sect. 2, too. Software modules for the other tasks in WP4 have a supporting role in the core interaction and were integrated on an as-needed basis for each use case. For example, the *Personalization Module* can be left out for use cases that do not require a user model.

All WP4 tasks use a shared knowledge base. It comprises a hierarchy of types and a fast representation for typed semantic objects called *eTFS* along with some operators useful in dialog processing. There are also conventions on formulating semantic queries for the back end. The semantic representation is described in Sect. 3. In Sect. 4, we explain the overarching interaction paradigm of the Dialog Shell. The basic setup already provides useful mechanisms to deal with some dialog phenomena like cross-modal references; Sect. 5 describes how the context representation is leveraged for this. Finally, Sect. 6 takes a brief look at some of the realized systems that demonstrate different use case aspects.

2 A Generic Dialog Shell

The purpose of the CORE TECHNOLOGY CLUSTER is that base technologies and resources that will be required in more than one use case should only be developed once by a research team and made available to the partners, instead of being re-implemented multiple times by each use case team. This includes pure software modules such as the Joint Service Engine (JSE), but also interface languages such as the presentation markup language (PreML), knowledge sources such as the base ontology, rule sets defining system behavior and many more. The work package aims to deliver a seamless user experience when dealing with semantically represented data and services in multimodal interaction. For the use case integrator, it requires a unified integration and lifecycle management platform for such a set of resources and modules that each is dedicated to specific steps during the processing of a user interaction. For a typical service-oriented use case, this includes analyzing the user input from various modality channels, combining these in a fusion module, managing the interaction, determining and calling appropriate back end services, planning the system response and finally rendering the output for various modalities.

The "Ontology-based Dialog Platform" framework (*ODP*) was designed and implemented in the THESEUS research program specifically to integrate and connect the software modules developed in the different tasks of WP4. To provide an interaction infrastructure for a given use case project, an *ODP* instance is created and

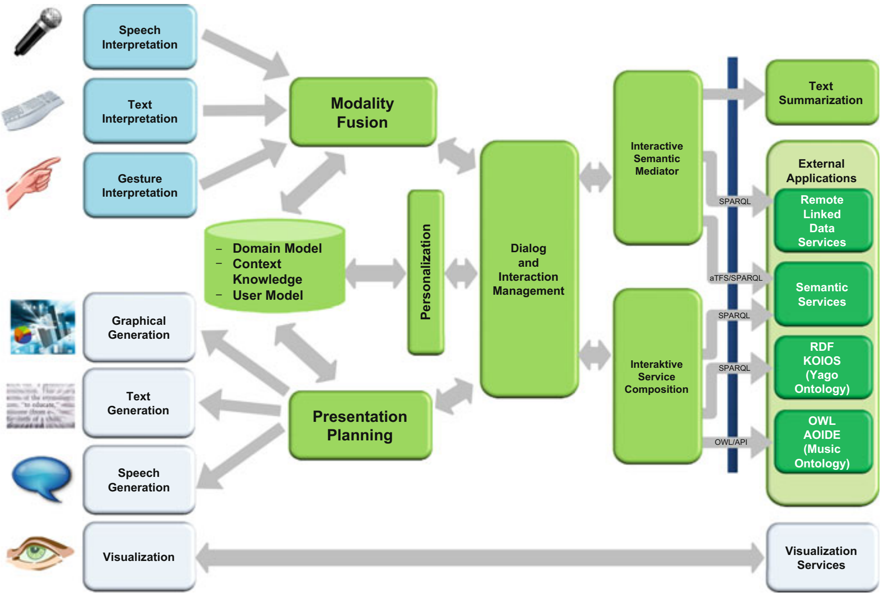


Fig. 1 The architecture of the dialog shell

configured to comprise a system of multiple modules, where each module realizes a specific function in interaction processing, such as spoken input interpretation, dialog modelling, or presentation planning, similarly to the architecture introduced in Maybury and Wahlster (1998). Lifecycle management and message passing for the modules are handled by a central part of *ODP* called the "Ontology-Based Middleware" (OBM), which is a continuation of work that has been started in the *SmartWeb* project (Reithinger et al. 2005).

For the different demonstrators that were built during the project, the combinations of modules vary slightly; however, the standard parts of a system and the corresponding subtasks are as depicted in Fig. 1. A typical user interaction is processed by several core modules that take on roughly the following roles:

- *Multimodal Semantic Processing*: Input interpretation for different modalities, semantic fusion of multimodal input, and interaction context representation,
- *Semantic Navigation and Interaction*: Planning, executing and coordinating the system actions to user input between back end and result presentation,
- *Interactive Service Composition and Query*: Gathering and combining responses to user requests using single or composite semantic web service queries from heterogenous information sources,
- *Semantic Output Representation*: Creating a multimodal result presentation featuring interactive, intuitive and responsive direct user manipulation.

In the concrete system, user inputs – e.g., spoken utterances or gestures – enter the system via hardware sensors connected to one of several *recognizers* for the different modalities of the system. In some cases, the input will also need further analysis; e.g., spoken input has to be interpreted with the help of a grammar as done by the module *Multimodal Semantic Processing*. Several inputs may also be combined to yield a *cross-modal* input representation. For this, an interaction model allows us to resolve cross-modal references and references with respect to the interaction history. Based on the resulting semantic structure representing a user request, the module *Semantic Navigation and Interaction* devises a course of action to fulfill the request. This typically involves one or more queries to the system back end that can be single or composite service invocations or queries. The module *Interactive Service Composition and Query* is responsible for translating back end queries into atomically executable sub-requests and reassembling the results before returning them to the system. *Semantic Navigation and Interaction* then will generate a semantic representation of a system answer for the user and pass it on to the *Semantic Output Representation*, which is responsible for constructing and rendering a concrete presentation in terms of *Semantic Interface Elements (SIEs)* (see Sect. 4).

Apart from this basic core sequence, processing may involve other modules based on the use case, the system configuration, and the user input. Additionally, it does not always have to follow the fixed order; part of the information flow goes "backward", such as the presentation module informing the interaction context module of the information items that were presented to the user.

3 Semantic Representation

Enhanced Typed Feature Structures (eTFSs) One important goal for our systems is to strive for an interaction at or very close to real-time speed. Internal processing is based on a unified semantic representation that is used throughout the entire processing chain and directly supported by the platform. Enhanced Typed Feature Structures (eTFSs) are used as a computationally efficient representation into (and from) which ontologies can be mapped.

Standard ontology-based representations for semantic content, like OWL or RDF(S), in many cases cannot deliver this. While they offer powerful modelling possibilities and comprehensive logic inference, they can be quite unwieldy and in some cases computationally expensive to handle, especially in the case of nontrivial models. This situation has been continually improving in recent years; however, in the development of ODP it was decided that the semantic representation of the objects in the interaction itself – in contrast to the back end data – should be encoded in a simplified formalism optimized for processing speed. This formalism is an *extension* of the "typed feature structure" (TFS) representation widespread in computational linguistics (Carpenter 1992), and is consequently called eTFS. It offers some additional operations, such as overlay and subsumption-based

Fig. 2 Classes from the base ontology (*odp* namespace) and the *CIRIUS*-ontology (*dm* namespace); slots for the *dm#Location* concept

matching (Alexandersson and Becker 2001) that are useful for dialog processing and dialog modelling. An eTFS structure can also be straightforwardly serialized in an XML document; such documents comprise the communication messages that are exchanged between ODP modules. The types of eTFS structures are defined in a hierarchical type system featuring multiple inheritance of slots via sub-typing. The ODP framework provides a tool suite called the *workbench* that can be used to define type hierarchies; for a new dialog system, the type hierarchy of a foundational base ontology is extended with concepts that represent the semantic objects specific to the domain of the given system. Figure 2 shows an example. The base ontology is part of the *ODP* framework and contains a set of concepts that are intended to be used for different use cases. For example, it contains generic representations for graphical presentations (e.g., lists, sets and tables), different forms of interaction (e.g., dialog acts and gestures) and abstract process types (e.g., information retrieval, service invocation, clarification). When defining a system, many of those can (and are meant to) be reused: the system designer defines subclasses that inherit the abstract structure, but are enriched with slots that add the concrete requirements of the use case at hand. Several *ODP* modules such as the discourse modeler *FADE* and the dialog manager *FLINT* are implemented as production rule systems where much of the adapting work for a system instance consists in extending a set of predefined

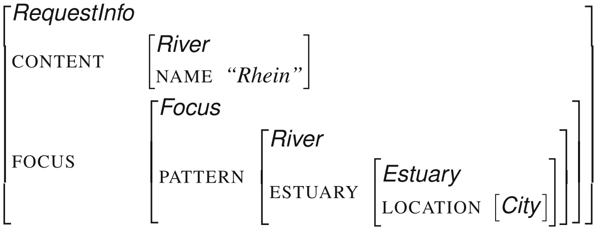


Fig. 3 An example of an eTFS structure representing a user query

rules with specialized rules tailored to the use case. To facilitate this, *ODP* offers support in the form of rule editors and rule matching monitors. Applying rules is quite straightforward, since the necessary subsumption matching is part of the eTFS operator set.

Queries A special case of concepts predefined in the base ontology whose concrete instances can change and be adapted to a specific scenario are `Query` concepts. This message type is used to inquire about information via the back end; however, it is intentionally abstract and does not specify which services in particular should be queried; this decision is left to a *service* module that uses knowledge about the capabilities of the available semantic services to compose an *aggregate query* from the combined user inputs and deliver an answer. The abstract `Query` concept only needs to specify *what* information is wanted, not *how* it should be obtained. For this, it uses two main attributes (or "slots"):

- The CONTENT slot contains a semantic object. It is usually underspecified, i.e., instead of one piece of information as a solution of the query, it describes a *set* of solutions, or a *solution space*.
- The FOCUS slot points to the part of the solution object that is requested. There can be more than one focus slot for one query; in this case, all focused slots are considered as requests as well as constraints.

Note that it is not an error if *more* than the requested information is returned; this is actually quite often the case to allow the presentation manager to create a presentation that overanswers in a way convenient for the user.

Figure 3 depicts a (slightly simplified) eTFS query structure. The structure consists of typed objects that have *slots* giving their relations to other nested typed objects or atomic values like strings or numbers. In the example, the object of type *RequestInfo* contains a slot CONTENT linked to an object of the type *River*, which in turn has a NAME slot with the string value "Rhein". The focus is on estuaries of this content object. Note that the *location* slot of the *Estuary* object is of the type *City*. This means that other types of (correct) answers – such as "The Rhine flows into the North Sea" – are explicitly excluded.

Fig. 4 Types of touch interaction in the *CALISTO* system. The user can use touch gestures to interact with spotlets. Data can be "thrown" (transferred) with the "frisbee gesture" from a mobile device to the stationary display

4 Multimodal Interaction

Semantic Interface Elements Different situation-aware views of data are offered by *Semantic Interface Elements* (SIEs), which are the first-class objects of the interaction and allow for an intuitive access to semantic data provided via the application-specific back end. The interaction with the SIE takes place on a number of levels that stand for different tiers of technical processes, but also for different ways of understanding the presentation and the context in which the objects are addressed: They are graphical as well as abstract conceptual entities that offer ways to manipulate and trigger operations on semantic information. There are also additional non-graphical representation levels. One is the *ontological* representation of the underlying content: the knowledge spotlet is the output component of semantic objects of the type *InformationSearch*. Lastly, the type of *process* that is initiated by receiving an *InformationSearch* object from the user to be resolved corresponds to the knowledge spotlet.

SIEs thus are a further extension of the principle of "no presentation without representation". In addition to the fact that all the data objects presented by the system have an underlying semantic representation, the presentation objects themselves – called *Spotlets* – are associated with specific *types* of semantic data defining the processing and available interactions for them (see Fig. 4). They can be interfaces to maps or media content or to entirely application-specific content; however, their affordances and controls should be reasonably clear to the user without needing an introduction.

Fig. 5 HTML5-based spotlets

If the user moves and releases an object onto an area associated with an operation – a so-called *drop zone* – of a SIE, it is triggered; depending on the context, this can be a service invocation, a search, or any other operation. Which process will be triggered is determined by the combination of the semantic object and the Semantic Interface Element. There can be also more than one process associated with one spotlet, e.g., corresponding to different drop zones or contexts. One advantage of the SIE abstraction is that the underlying functionality can be ported to different interaction platforms (conventional computer screens, large touch screens, phone interfaces), while the messaging on the *ODP* framework can remain unchanged. Figure 5 shows some spotlets from touch-screen based systems ported to HTML5 to be rendered in a plain web browser.

Pattern-Based Interaction Management As mentioned before, the fusion and discourse engine FADE and the dialog management module FLINT are both based on a system of production rules. The general principle for processing in this case is that there is a set of semantic objects, called the *working memory*, that at each point represents the interaction context. A processing module runs with a set of production rules, which consist of a *condition* (a pattern that can be matched by semantic objects) and an *action* part (a sequence of operations to be executed). If an input arrives during system execution (e.g., an interpreted user input), it is added to the working memory. The production rules system then loops over the rules and the working memory elements (WME). If the condition part of a rule matches a WME, the rule "fires", i.e., its *action* part is executed. Actions essentially can (1) create, delete and update WMEs, and (2) create and send messages for other modules, such as the JSE or the presentation manager. For operations that cannot be realized in terms of manipulating WMEs, such as direct low-level computations, the system designers can define and call plugins written in Java. The loop continues until the system re-enters a stable state in which no further rules can be applied.[1] The purpose of the two modules is, respectively:

- *FADE*
  - Maintaining a dynamic information state that contains the interaction context from the contributions of the user and the system;
  - Using this context to resolve referring expressions (see Sect. 5).
- *FLINT*
  - Matching the user input against a set of defined processes to be able to determine actions that can fulfill the user's requests;
  - Triggering and coordinating a matching (possibly multi-step) process. Such an action plan generally consists of back end queries, but may also involve other actions, e.g., subdialogs to clarify the user's intentions;
  - Constructing an abstract presentation response for the user from the results.

Several other, optional modules also use the production rule system (e.g., the TTS generation module). Since the underlying lean semantic representation is the same for all rule-based modules, there is no need for resource-intensive conversion or copying operations between modules. Additionally, the same editor tools can be used for all production rule-based knowledge sources. There are several additional features such as rule weights which cannot be described here for space reasons.

5 Default Coverage of Dialog Phenomena

For state-of-the-art spoken dialog systems, it is a widespread view that in order to model and interpret the communicative intention transmitted by the user, we need to consider findings coming from human factor experiments suggesting incorporating natural multimodal interaction patterns by humans. A key element in multimodal systems towards a more natural human-computer interaction is the integration of all input modalities offered to the user into one single semantic interpretation. After all, the sound interpretation of mixed input modalities through the user's action allows us to handle different dialog phenomena such as the interpretation of ambiguous, under-specified and partial utterances.

In order to manage the resolution of elliptical and anaphoric expressions, we have used a discourse modeling module covering mechanisms including:

[1]Generally, an essential action of any rule is to remove the WME that triggered it, to avoid endless re-firing.

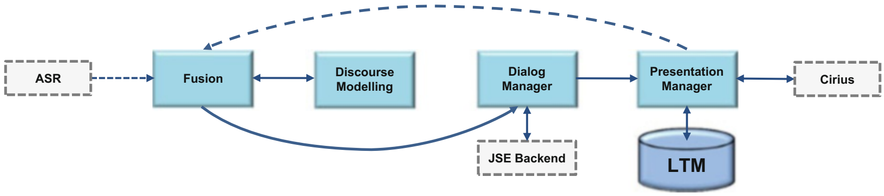


Fig. 6 The pipeline of resolving multimodal expressions

1. Fusion – integration of user's turn regardless of the used modality, i.e., interpreting and transforming the incoming input actions into a homogeneous semantic representation,
2. Discourse Modelling – the storage of discourse objects and moreover the process of identifying referring expressions, detecting the missing information within the discourse history and recomputing the eTFS in a sufficiently informative manner as to be processed by the successive components of the dialog system.

Figure 6 shows the incoming and outgoing dependencies of the fusion component and the discourse model with respect to the other components of the dialog system.

The approach to multimodal integration is centered around the idea that each input event needs to be interpreted with respect to its local turn context, embracing all previously recognized input events together with the dialog state assigned to the same user turn. As for the input event delivered by the speech recognizer, the fusion component integrates the interpretation of a hypothesis only if the input event outperforms a confidence value. This ensures that only speech input based on good recognition quality obtains the authorization to potentially contribute to the integrated meaning of a multimodal utterance. The dotted lines in Fig. 6 indicate the flow of context information that is fed into the discourse model. In addition to speech input, the framework allows us to channel context information into the discourse model stemming from user actions based on non-verbal inputs, i.e., pointing gestures. Thus, each time the user touches a graphical object, the presentation manager will receive a message and retrieve the corresponding ontological concept from the Long Term Memory (LTM) that eventually will be propagated to the discourse model (for more detail, see Sect. 5).

One major principle of our framework is the demand to be on one hand flexible and generic enough to deal with a variety of phenomena, and on the other hand to reduce the work load when instantiating a dialog system for a new application: The developer of the dialog system does not need to take care of the synchronization of input modalities. In case the speech recognizer is activated and ready to process some audio input, the pending mode of the speech recognizer prevents the access of the rule engine to rules of classes that may cause the corruption of the dialog state. Moreover, the framework comprises already the symbolic representation inside the ontology modeling the core concepts to deploy the discourse processing module.

Hereafter, we will show how to make use of the ontological concepts for the purpose of homogeneous user turn integration and the resolution of referring expressions.

The Interaction Context Tracking the semantic content of the interaction in terms of the user's "dialog turns" allows a system to deal with some common dialog phenomena; here is an example of a multimodal dialog involving some of them (↗ signifies a pointing gesture):

1. **U**: "What are the neighboring countries of France?"
2. **S**: "I found eight results, for example Spain, Andorra and Switzerland." (*Shows a list of neighboring countries*)
3. **U**: "How many people live [↗ there]?"
4. **S**: "At last count, Germany had 81,799,600 inhabitants." (*Focuses on the selected country and shows the most recent population count*)
5. **U**: "Show me information on its capital city."
6. **S**: "Here is some information on Berlin." (*Shows Berlin on the map and presents further relevant information*)

The interpretation modules can fuse contributions in different modalities, yielding a combined input representation to interpret cross-modal references, or to resolve ellipsis against the interaction history. Here, we give an outline to the processing within the framework that controls the underlying operations to tackle:

- *Completion of elliptic intention hypotheses*: Ellipsis completion is computed for utterances like "And the Rhine?". This sentence would not make any sense if information of the preceding discourse were not taken into account. In this example the user might have asked in the previous turn for the estuary of some river.
- *Resolution of deictic expressions*: This applies in utterance like "And this river?". Besides the lack of explicit focus inside the question, this turn implies a non-verbal source of input. For the purpose of comprehending the full communicative intention of the user, it is crucial to interpret as well non-verbal actions. In this case, the Discourse Model must be fed by the object denoted by the gesture that typically accompanies referring expressions, i.e., "this river".

Ellipsis Processing Discourse objects are containers for concepts that were directly mentioned during the preceding discourse. They comprise a unified representation of the semantic information gathered so far. Our notion is that each entity may serve as referent for referring to an expression (e.g. objects, events, states or collections of the same) and introducing a new instance of a discourse object. A so-called local focus stack bears pointers to discourse objects which, in case the users utterance contains referrings expressions, will be extended with all mentioned concepts available as potential referents along the course of the succeeding dialog session. Discourse objects stemming from speech are identified by analyzing the semantic interpretation, i.e., the reference to a sub-object of the semantic representation of the best scored intention hypotheses. The representation of referring expressions is defined as an ontological instance of the concept `RefProp`. The adequate semantic

$$\begin{bmatrix} \textit{River} & \\ \text{ESTUARY} & \begin{bmatrix}\textit{Estuary} & \\ \text{BODYOFWATER} & \begin{bmatrix}\textit{River}\end{bmatrix}\end{bmatrix} \\ \text{REFPROP} & \begin{bmatrix}\textit{RefProp} & \\ \text{TYPE} & \textit{“def”}\end{bmatrix} \\ \text{RESOLVED} & \textit{true}\end{bmatrix}$$

Fig. 7 Semantic interpretation of the referent (*River*)

$$\begin{bmatrix} \textit{River} & \\ \text{ID} & \textit{“id\_4711”} \\ \text{NAME} & \textit{“Rhine”}\end{bmatrix} \quad \begin{bmatrix} \textit{River} & \\ \text{REFPROP} & \begin{bmatrix}\textit{RefProp} & \\ \text{TYPE} & \textit{“deictic”}\end{bmatrix}\end{bmatrix}$$

Fig. 8 (*left*) An instance of *River* located in the LTM. (*right*) The semantic interpretation of a deictic reference

interpretation of the substructure that serves as referent for the partial utterance "And the Rhine?" is as shown in Fig. 7. Here the instance of *River* is marked with the *RefProp* object; hence it will be integrated as a discourse object within the discourse model. The presence of the slot RESOLVED indicates if a referring expression has been already resolved or not, which basically determines the distinction if we deal with a referring expression or a referent. It follows that the partial utterance "And the Rhine?", which omits the semantic interpretation of the estuary information by the absence of the slot ESTUARY, can be recovered by traversing the local focus stack in historical order until a successful unification with the discourse object is returned.

Cross-Modal References The technical demonstrator *CIRIUS* provides the modality to place deixis with an accompanying gesture. A pointing gesture typically refers to entities of the type `PhysicalObject` representing objects that can be perceived in the visual environment. If a `PhysicalObject` is explicitly activated through the mentioning of a discourse object, it can serve as a referent for a referring expression. The presentation manager allocates a unique identifier to all ontological objects that will be displayed to the user and stores them inside the LTM. The eTFS on the left side of Fig. 8 represents the instance within the dialog system that is offered as a graphical entity to the user. Thereafter, the presentation manager listens to events triggered by a pointing gesture that conveys merely the identifier designated to the graphical object. By this means, only graphical objects that are contained by the LTM will be integrated into the discourse model. This step will enable the discourse processing module to resolve deictic expressions in the same manner as the reconstruction of missing parts within an utterance by a search over a list consisting of potential candidates inside the discourse model. The eTFS structure that represents the corresponding semantic interpretation, which invokes the resolution of deixis, is displayed on the right in Fig. 8.

Fig. 9 Interaction with a multi-touch system installed on a stele (*CALISTO*)

6 Conclusion

During the course of the THESEUS research program, several systems were designed and constructed to showcase and exemplify different aspects and capabilities of the work package sub-tasks and the dialog shell as a whole. In many cases, they were also used as first "proof-of-concept" instantiations for newly developed features. In the following, short descriptions of some demonstrators and their scope are given while pointing out some highlights and special aspects:

COMET: This early instance of an *ODP*-based system implements a semantic-based multimodal interaction with a music database. The system interfaces directly with a portable iPod music player, extracting the title descriptions dynamically and using them to generate parts of the resources on the fly, such as the recognizer grammar. *COMET* demonstrated the feasability of the dialog shell concept. It established that the combined spoken and gestural input can operate in real time over a nontrivially sized dataset. The system implements multimedia operations over the semantic back end data (such as music and video playback) and an intuitive user interface with a compact set of features controlled by spontaneous and natural interaction.

The *ISIS* system demonstrates how to connect the dialog shell with very large semantic knowledge bases. It combines access to *DBPedia* and the semantic database of the ALEXANDRIA use case. This demonstrator is a combined effort of several work packages. The WP3 provided an additional abstraction layer ("KOIOS") to mediate the differences between the semantic representations of dialog shell and back end representation so that they could be treated by the rest of the system in a uniform way. WP5 contributed new visualization methods for the front end presentation.

CALISTO realizes an interactive and collaborative platform with some focus on intuitive cross-device interaction (Bergweiler et al. 2010). The system is a media terminal that also supports collaboration and data exchange between multiple users and devices (see Fig. 9). One feature is a gesture-based intuitive media transfer between mobile devices and the *CALISTO* demonstrator via the so-called “frisbee gesture” enabling the users to figuratively “throw” data from their cell phones, to subsequently appear on *CALISTO*’s surface. In this case, it becomes apparent that creative application of rather conventional technical means (via Wi-Fi connection) can lead to a big usability gain through well-applied intuitive and “fun” interaction metaphors.

CIRIUS is a multimodal interactive system that enables access to information, multimedia content, and services from a semantic database by means of spoken language and gestures. The demonstrator uses semantically annotated content to help users to discover interesting facts about rivers, cities, and points of interest in Germany. The *CIRIUS* system also interprets various input modes like spoken language and pointing gestures separately or in combination. In addition, it enables rapid access to semantic services as well as the ontological representation of the extracted information. *CIRIUS* uses web services to find additional photos, texts, and videos. The focus of the *CIRIUS* system is on speech interactions and multimodal fusion, where users can easily click on an item and utter a command like “Give me more information about this!” and have the input be resolved from the combined modalities in the context.

References

J. Alexandersson, T. Becker, Overlay as the basic operation for discourse processing in a multimodal dialogue system, in *Proceedings of the IJCAI Workshop Knowledge and Reasoning in Practical Dialogue Systems*, Seattle, 2001

S. Bergweiler, M. Deru, D. Porta, Integrating a multitouch Kiosk system with mobile devices and multimodal interaction, in *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS ’10)*, Saarbrücken (ACM, New York, 2010), pp. 245–246, http://doi.acm.org/10.1145/1936652.1936698

B. Carpenter, *The Logic of Typed Feature Structures* (Cambridge University Press, Cambridge, UK, 1992)

M.T. Maybury, W. Wahlster, *Readings in Intelligent User Interfaces* (Morgan Kaufmann Publishers, San Francisco, 1998)

N. Reithinger, S. Bergweiler, R. Engel, G. Herzog, N. Pfleger, M. Romanelli, D. Sonntag, A look under the hood – design and development of the first smartweb system demonstrator, in *Proceedings of the 7th International Conference on Multimodal Interfaces (ICMI ’05)*, Trento (ACM, New York, 2005), pp. 159–166

Building Multimodal Dialog User Interfaces in the Context of the Internet of Services

Daniel Porta, Matthieu Deru, Simon Bergweiler, Gerd Herzog, and Peter Poller

Abstract We will show how to build innovative multimodal dialog user interfaces that integrate multiple heterogeneous web services as data sources on the basis of the Ontology-based Dialog Platform (*ODP*). More specifically, we will describe how to exploit *ODP*'s well-defined extension points and how generic *ODP* processing modules can be adopted, in order to support a rapid dialog system engineering process. By means of the latest *ODP*-based educational information system *CIRIUS* and the *ODP* workbench, a set of Eclipse-based editors and tools, we demonstrate step-by-step along the generic multimodal dialog processing chain what has to be done for developing a new multimodal dialog user interface for a specific application domain.

1 Introduction

The ultimate goal for an advanced user interface (UI), which employs multimodality for enhanced human-computer interaction, is to provide a natural and intuitive multimodal dialog. Knowledge-based processing and adequate semantic models constitute key ingredients to achieve this. Compared with conventional graphical UIs, however, application-specific tailoring of multimodal dialog systems requires additional engineering effort. Our Ontology-based Dialog Platform (*ODP*) considered here provides a generic framework for multimodal dialog UIs, which is designed to ease the development of such interfaces for diverse usage scenarios in varying application domains. The complexity of building new UIs is reduced significantly by encapsulating recurring tasks and functionalities.

D. Porta (✉) • M. Deru • S. Bergweiler • G. Herzog • P. Poller
DFKI GmbH, German Research Center for Artificial Intelligence, Saarbrücken, Germany
e-mail: Daniel.Porta@dfki.de; Matthieu.Deru@dfki.de; Simon.Bergweiler@dfki.de; Gerd.Herzog@dfki.de; Peter.Poller@dfki.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_12,

In the context of the Internet of Services, business logic is executed in and data is delivered from heterogeneous and complex web services. We focus on user interaction with the Internet of Services and optimize the *ODP* framework as a means for implementing natural UIs for services. This results in architectural decisions and a streamlined and more formal development process. In this article, we concentrate on the latter topic while the former is discussed in-depth in Löckelt et al. (2014). The aim is that developers be able to easily design and implement homogeneous *ODP*-based service UIs that hide the complexity of the back end infrastructure from the user. But it is also clear that domain-specific adaptations and extensions for new UIs will still be necessary and possible. Typically, domain-specific customizations may require a substantial amount of custom programming and thus development time. The *ODP* framework, which evolved as a full-fledged multimodal dialog shell from the former SmartKom (Herzog and Ndiaye 2006) and SmartWeb (Sonntag et al. 2007) multimodal dialog systems, accounts for domain adaptations and it is especially designed for that. At its core, a reusable and consistent base model and generic processes, which handle most of the interaction-specific tasks out of the box, are complemented by transparent scenario- and domain-specific customization points.

We realized this by applying a model-driven development approach. Multimodal dialog UIs involve a large number of models that incorporate or depend on each other. Model-driven development is a software development approach for automatically generating running applications from formal representations of a domain and system. A formal description of system parts by models pays off in better readable documentation, improved reusability and easier adaptation, and hence in the reduction of development time.

This article focusses on the underlying development method of our *ODP* approach. The following Sect. 2 first provides a short overview of related work regarding development methods for natural UIs. Our own method is then discussed in Sect. 3. The conception and refinement of the proposed development method builds upon the implementation of various multimodal dialog service UIs that have already been built using the evolving *ODP* framework. Section 4 then illustrates how the *ODP* approach has been applied for the realization of *CIRIUS*, a collaborative kiosk information system with multimodal dialog interaction. The concluding Sect. 5 summarizes obtained results and further work.

2 Related Work

Building natural UIs, i.e. UIs that offer a maximum user experience, is a difficult task because user experience is influenced by many factors such as the user's current usage context. Here, a formal development method that accounts for several, possibly orthogonal layers of abstraction helps to manage the complexity by separating functional requirements and properties.

In the research area of Model-Based User Interface Design (MBUID[1]), the conceptual CAMELEON Reference Framework (Calvary et al. 2003) defines a standard architecture for modeling UIs. Starting from a formal user-centric task description, model transformations iteratively derive platform-independent and then platform-specific UI model fragments on different levels of abstraction. Based on the formal UI models, the concrete UI can finally be generated in a subsequent code generation step. Over the last few years, several instantiations of the framework were developed (Limbourg et al. 2005; Mori et al. 2004). Usually, UIs resulting from these transformation and generation processes are self-contained and proprietary. With the ongoing success of the Internet of Services and the accompanied shift of business logic and data into cloud infrastructures, some modern approaches also consider browser-based UIs (Stanciulescu et al. 2005) as well as UIs for services (Paterno et al. 2009). In parallel, the inclusion of additional modalities beyond point-and-click or touch raises interest (Ertl 2009; Heinrich et al. 2007; Paterno et al. 2008). However, true multimodality, i.e. the simultaneous and cross-referencing use of several modalities as well as support for natural language discourse phenomena cannot be achieved so far with these model-based approaches. This is because MBUID concentrates on formal methods for describing UIs but neglects the psychological and linguistic background of human communication and interaction metaphors. Thus, adequate models capturing these metaphors and respective processing components, e.g., for modality fusion, as available in the *ODP* framework, are still missing in the MBUID approaches.

In contrast, Gandhe et al. (2009) describe an integrated authoring tool for the rapid prototyping of tactical questioning dialog systems by non-experts. The resulting system is used in a training scenario, where the user has to interrogate virtual characters. Authors have to first model relevant domain concepts and virtual characters that participate in the questioning. Characters can have different domain knowledge, attitudes, social behavior and roles. In a second step, a complete set of domain-specific dialog acts (Bunt et al. 2010) are derived, based on the character-specific domain knowledge and a set of predefined and thus domain-independent dialog acts. This ensures completeness, i.e. all dialog acts that are relevant for the domain can be generated, which in turn means that all information can be queried by the user and (dis-)confirmed by the characters. The last step is to specify valid utterances for the user (as speech input) and the characters (speech output) and link them to the respective dialog acts. Hence, e.g., speech input is directly mapped to a subset of the derived dialog acts transporting a certain amount of information. This ensures consistency, because only valid dialog acts can be generated at runtime. That means they can be correctly handled by the dialog manager. *ODP*, as a multimodal dialog framework, pursues a similar approach in that not only speech input but in general all input modalities are directly mapped

[1]There is also a W3C working group examining this area; see http://www.w3.org/2011/mbui/

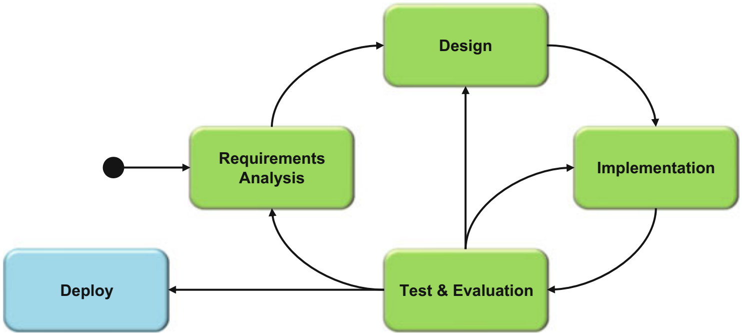


Fig. 1 The software development cycle

to (partial[2]) dialog acts. On the output side, dialog acts generated from the dialog manager are processed by a presentation planning component.

The WAMI toolkit (Gruenstein et al. 2008) enables developers to develop, deploy, and evaluate browser-based multimodal UIs without the need for heavy modeling formalisms. A rapid prototyping approach combines web programming techniques for the graphical UI part with a low-level speech grammar format for the auditive part of the UI. AT&T's Speech Mashup Framework (Di Fabbrizio et al. 2009) takes a similar rapid prototyping approach but relies more on W3C standards for defining speech grammars (SRGS, Hunt and McGlashan 2004), speech synthesis (SSML, Burnett et al. 2004) and multimodal interaction annotations (EMMA, Baggia et al. 2009). *ODP* also adopts these standards for speech input and output. But developers are not forced to touch these low-level formats directly; rather they can work on a more abstract level.

In the TEXO use case, new services are developed in a holistic and model-driven approach according to the Integrated Service Engineering (ISE) methodology discussed in Kett et al. (2014). The ISE methodology accounts for different service facets at different abstraction layers and hence maps to the typical workflow for developing an electronic service in larger organizations. It also includes preceding requirements, analysis, and design stages before the actual implementation stage (see Fig. 1). The distinction between a method and a methodology is controversially discussed in science and both terms are often used synonymously. In short, a methodology also considers coordination strategies of development teams as well as different roles and skills inside a team whereas a method "only" focuses on a prescribed workflow of doing something. In this context, we refined the ISE workflow for modeling service UIs in compliance with the conceptional CAMELEON framework with a special focus on multimodal dialog service UIs and our own development method (Porta 2010).

[2]Partial dialog acts have to be resolved by the input fusion component.

3 The ODP Development Method

In this section, we focus on the *ODP* development method. First, we discuss how a development method is designed in principle (Sect. 3.1). Then, we report on how the *ODP* development method has evolved based on three major requirements (Sect. 3.2). Finally, we describe the method in detail in Sect. 3.3.

3.1 Conception of a Development Method

In general, developing software is an iterative process that is often described by a cycle such as the one shown in Fig. 1. Starting from a *Requirements Analysis*, first concepts are developed in the *Design* phase that serve as a basis for the actual *Implementation*. The *Implementation* needs to be tested in regular intervals in order to discover functional and technical deficiencies as early as possible. In addition, a running prototype needs to be evaluated in a user study. Deficiencies and usability issues discovered in the *Test & Evaluation* phase trigger an iteration through the cycle in order to fix them and finally *deploy* the software.

Interestingly, this development cycle is on a meta-level also applicable to the conception of a development method implemented by a technical framework. In this section, we first want to reflect on that theoretically and then apply those thoughts to the conception of the *ODP* framework's development method.

Given an initial description of what should be developed by means of a method, you gather some initial requirements of how that should be done. In the design phase, you start thinking of how to conceptually fulfill the requirements in terms of concrete techniques and tools. Moreover, it is important to identify relevant development tasks as well as dependencies between them in order to define an initial workflow. In the implementation phase, you work on the technical realization of the design concepts. That implies adaptations and extensions to the underlying framework that ought to implement the method, if it already exists. To point that out, such a method has a defining impact on the implementing framework. Or the other way around, a framework for whatever purpose is just as good as the method of how to use it. In the testing phase, you have to know the expected results of the tests in advance. Leaving technical errors aside, you usually test against the actual functional specification derived from the identified requirements. Thus, you can answer the question of whether the software or the method meets the requirements or not. So, testing a method, i.e., deciding whether it in principle allows you to do the things in the intended way, is clearly possible.

However, testing a method also depends on the implementing technical framework used during the tests as a proof-of-concept. The evaluation of a development method, i.e., testing it with users or developers, is more difficult. Assume there are two frameworks that both implement the same development method, e.g., the CAMELEON implementations. Then you can evaluate which one better implements the method. Although in this case, you did not evaluate the method itself but the frameworks, this information is very valuable for deciding on several competitive frameworks or, more applicable to our situation, measuring progress in the development of a single framework by comparing the latest version of the framework with a previous version. In order to evaluate a method, it basically has to be compared to another method for the same purpose. This, however, requires us to establish comparable or better equivalent test conditions. Unfortunately, this is hard to achieve because two methods most likely depend on two different implementing frameworks. Applying pure virtual evaluation methods, e.g., the cognitive walkthrough method, for comparing two development methods without a grounding to technical implementations might also deliver bad results because identified issues could perhaps be resolved by an intelligent implementation. But, what can be done is comparing two iterations of the same method and hence measuring progress in the development of that method since the two versions still rely on the same (perhaps refined due to the refined method) technical framework. The method has evolved from gained experience in applying it and thus from discovered deficiencies.

3.2 Developing the ODP Development Method

During the THESEUS research program, we adopted our framework for implementing several multimodal dialog UIs for different usage scenarios and application domains. We want to briefly point to three of them.

In the TEXO context, Porta et al. (2009) describe a mobile business application that enables decision-makers on the go to still participate in important business processes. The mobile user can handle purchase order requisitions in an enterprise resource planning system and can search for alternative products. Found alternatives can be sorted according to different criteria and visualized in three-dimensional space. In the MEDICO context, Sonntag and Möller (2010) describe a stationary medical system that supports a radiologist in finding a diagnosis, asking for a second opinion, and deciding on an appropriate medical treatment. The system allows semantic image annotations and presents former patients with similar findings. The CALISTO collaborative kiosk infotainment system (Bergweiler et al. 2010) combines mobile and stationary interaction scenarios. Typically, CALISTO instantiations can be deployed in museums or exhibitions. Visitors can ask for relevant information by interacting with a large tabletop surface and their mobile devices, which can also be used for synchronizing and sharing media assets. The necessary data is retrieved by accessing a large heterogeneous service back end which also comprises semantic web services.

By using the *ODP* framework for implementing these UIs, we not only constantly refined the technical framework but also its underlying development method. Before we describe the latest iteration of the development method and show how the method is applied to a new UI, we first want to discuss the three principle requirements that directed the development of both the *ODP* framework and its development method. However, the technical framework has its own requirements not discussed here, e.g., real-time processing for responsive interaction. Fulfilling them is in our opinion also essential for a successful commercial exploitation.

First, as the acronym implies, everything in the *ODP* world must be semantically modeled. This is important for robust multimodal processing and dialog management. Second, building multimodal dialog service UIs is an inter-disciplinary task that requires technical but also, if applicable, psychological know-how in the fields of graphical UIs, vocal UIs, multimodal processing, dialog management, as well as knowledge representation, service technologies, and last but not least the actual domain of interest. Here, the method must support a clear separation of concerns or models. Third, recurring work has to be done in a straightforward and uniform way. Even though the *ODP* framework contains generic components for multimodal dialog processing and back end retrieval on top of a uniform base modeling, domain-specific customizations are necessary, e.g., modeling relevant domain concepts and integrating domain-specific services. These should be realized as easily as possible.

These requirements can be further decomposed into more concrete corollary ones. So, against the background of responsive multimodal dialog UIs (a framework requirement), the first (method) requirement contradicts with the notion of fast processing. Thus, a lightweight, i.e., a fast and sufficiently expressive, semantic data modeling language is necessary, and we chose extended typed feature structures (eTFSs) (Carpenter 1992). The second requirement implies on a technical level also a clear separation of concerns in terms of a minimum-dependency architecture and transparent interfaces for the generic components. A transparent interface "allows the connection and operation of a system, subsystem, or equipment with another without modification of system characteristics or operational procedures on either side of the interface" (Weik 2001). But this separation of concerns must also be reflected in the tooling as this is the most visible point for developers working with the *ODP* framework. The third requirement motivates decisions regarding how to realize recurring but domain-specific work conceptually and technically. Conceptually, the *ODP* framework pursues a declarative approach of "programming" wherever possible as this seems more suitable for UI experts who are not necessarily experts in the Java programming language. Furthermore, as already mentioned in the related work section, developers are not forced to work on the most concrete level of abstraction, e.g., write SRGS or SSML documents. Rather, the *ODP* framework offers more abstract and comprehensible work levels. Technically, the *ODP* framework supports developers with a complete workbench for the rapid development of multimodal dialog service UIs (Sonntag et al. 2010a). It contains specialized editors supporting auto-completion as well as syntax and consistency checks and integrates life debugging facilities based on the Eclipse Platform. Hence, the *ODP* framework consists on the one hand of a runtime environment for executing multimodal service UIs and on the other hand of a design-time workbench for the integrated and model-driven development of these UIs.

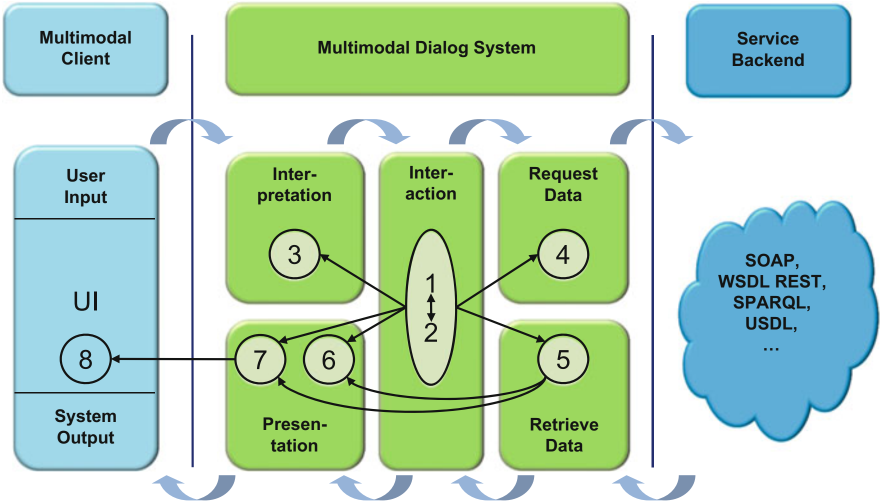


Fig. 2 The dialog processing chain and relevant tasks for building multimodal dialog UIs

3.3 How to Build an ODP-Based Multimodal Dialog Service UI

In this section, we discuss the *ODP* development method and concentrate on the *Implementation* step in the software development cycle (see Fig. 1). That means, the *Requirements' Analysis* as well as the conceptual *Design* of the UI are already available. For obtaining these, we apply standard techniques adapted specifically to multimodal dialog UIs as discussed in Sonntag et al. (2010b). Figure 2 shows the major steps in the processing chain of the *ODP* runtime environment from *User Input* to *System Output*. These steps also correspond to generic components discussed in Löckelt et al. (2014). Figure 2 further assigns development steps (as numbered circles) to the processing steps and highlights the dependencies between the development steps. Work in such a development step has to be carried out by the responsible expert(s) for the respective generic runtime component.

1. *Task Modeling*. The tasks the user should be able to perform are derived from the *Requirements' Analysis* and conceptually specified in the *Design* phase of the software development cycle (see Fig. 1). So, they depend on the application scenario but not on a specific input or output modality. As one of the first things, the user tasks have to be modelled by means of the eTFS meta-model and derived from the generic `odp#Task` concept. Figure 3 shows the `odp#Task` concept in its inheritance hierarchy. All further concepts mentioned in the following are also highlighted in Fig. 3.
2. *Domain Modeling*. As user tasks carry domain-specific data, modeling relevant concepts naturally depends on the supported user tasks and vice versa. These

Fig. 3 The *ODP* base model

are also modeled by means of the eTFS meta-model and are inherited from the concept `odp#Resource`. For exploiting generic processes, some concepts might need to be annotated with dialog-specific annotations. Annotations are realized by deriving the domain concept from multiple specializations of the `odp#Annotation` concept, e.g., `odp#NamedEntity`. As task and domain modeling are closely related, both actually take place at the same time. As a consequence, subsequent steps depend on both, the task and the domain model.

3. *Speech Recognition Grammar*. Writing low-level grammars, e.g., by means of the SRGS standard, is a tedious task. The *ODP* framework already contains domain-independent grammars, e.g., for social obligations dialog, but also algorithms for dynamically generating and updating named entity grammars. Hence, only domain-specific grammars must be written by means of a dedicated editor of the *ODP* workbench. Here, in a more convenient approach, which can be compared to the approach pursued by Gandhe et al. (2009) (see Sect. 2), a developer can specify complete utterances that consist of several alternative phrases. An utterance is then directly mapped by the developer to a semantic interpretation represented as an eTFS structure based on the underlying ontology, i.e., a `dm#CommunicativeAct`. It can also be marked as partial in order to trigger the multimodal fusion component at runtime. The fusion component then applies a generic process to resolve missing information from either the discourse (elliptic reference) or other input modalities (cross-modal reference). At runtime,

the speech recognition grammar is compiled into a valid SRGS grammar that is compatible with most off-the-shelf speech recognizers.

4. *Service Integration*. Domain-specific services have to be integrated into the *ODP* runtime environment, and tasks to be performed by the user have to be decomposed and finally grounded to service calls. Most work has to be accomplished in terms of mapping tasks to (a chain of) service calls and writing declarative rules for transforming eTFS-based data objects into XML-based business objects that can be processed by the invoked services.
5. *Data Mediation*. When a service call is completed, retrieved data has to be mediated back into the internal eTFS-based domain model. This also requires declarative transformation rules for lifting the data from a syntactic XML notation to an eTFS-based semantic level suitable for multimodal dialog processing. In principle, this is the opposite direction of the transformation rules written for *Service Integration*.
6. *Speech Synthesis Templates*. We pursue a generic output fission strategy according to symmetric multimodality (Wahlster 2003). That means, whenever a user interacts by speech, the *ODP* also generates a synthesized speech output as a response. To ease the development, the *ODP* framework also contains generic templates for speech output, e.g., for error recovery, social obligations, and named entities. Domain-specific auditive system feedback must be specified in terms of declarative rule-based speech synthesis templates that transform eTFS-based data objects into text. Here, the same rule-based transformation mechanism is used as for *Service Integration* and *Data Mediation*. At runtime, this text-based speech output is finally transformed into valid SSML markup that also contains optional grapheme-to-phoneme transcriptions for the correct pronunciation of foreign language words as described in Schehl et al. (2008).
7. *GUI Model*. According to the "no presentation without representation" paradigm, it is necessary to semantically represent interaction-relevant graphical components, because a user might refer to them or the displayed content either by speech or deictic gestures. Such a GUI model is composed of `odp#ViewComponents` that are managed and updated by the server-side *ODP* infrastructure. It contains information about the position, dimension, and visibility of each `odp#ViewComponent` as well as the displayed content. A special kind of `odp#ViewComponents` are Semantic Interface Elements, aka `dm#Spotlets`. Conceptually, spotlets (Sonntag et al. 2009) are intelligent presentation and interaction agents with independent internal logic that know how to present and interact with certain types of objects. Following the approach described in Neßelrath and Porta (2011), a spotlet defines a visual template for an arbitrary data model containing relevant triggers and actions. A data model can be connected to (and thus eventually presented in) a spotlet by using data bindings. This is to some extent comparable to the Model-View-ViewModel pattern used in Microsoft's WPF framework. However, the GUI model is kept on the server-side. It is thus not directly rendered and needs to be synchronized with the actual client-side UI implementation.

8. *UI Implementation*. First of all, no work has to be done for the auditive part of the UI. We already implemented ready-to-use client-side libraries for audio streaming (receiving and sending). These are embedded in lightweight browser-based multimodal clients for mobile devices (Porta 2010) as well as generic desktop-based implementations of the Spotlet GUI concept in HTML5 or Adobe AIR/Flash. These implementations also allow for client-side style customizations. For the graphical presentation, declarative transformation rules have to be written to transform domain-specific eTFS-based content back into XML data structures that can be consumed and processed by the concrete presentation components.

4 Building the CIRIUS System

CIRIUS is an interactive kiosk system on a large tabletop surface that provides users access to heterogeneous information sources using spoken language, pointing gestures and the combination of both within the same user inquiry. Users can explore geographical information about countries in Europe as well as rivers, federal states, cities and points of interest in Germany. A sample screenshot is provided in Fig. 4. A typical multimodal dialog interaction between user and system is as follows (↗ signifies a pointing gesture):

1. **USER:** "What are the neighboring countries of France?"
2. **CIRIUS:** "I found eight countries, for example Germany, Switzerland and Italy." (*Shows a list of neighboring countries, see Fig. 4*)
3. **USER:** "How many people live [↗ there]?" (pointing gesture to the list item displaying Germany)
4. **CIRIUS:** "At last count, Germany had 81,799,600 inhabitants." (*Focuses on the selected country and shows the most recent population count*)
5. **USER:** "Show me information on its capital city."
6. **CIRIUS:** "Here is some information on Berlin." (*Shows Berlin on the map and presents further relevant information*)

In the following, we describe the relevant details of each of the aforementioned eight general development and implementation steps that were conducted to build up the *ODP*-based multimodal dialog system *CIRIUS*. To illustrate these steps more precisely, we pick the first user interaction with the system and discuss the relevant artifacts that are necessary to make this specific interaction run:

1. *Task Modeling*. Since the main purpose of the *CIRIUS* system is to allow a user to explore and browse geographic information, most user tasks can be seen as `dm#AttributeRetrievalTasks`. Task instances represent the abstract request for certain attributes of a given (geographic) entity. Tasks for managing individual spotlets like opening, closing, or moving a spotlet on the surface or playing or pausing displayed media are inherited from the concept `dm#SpotletSpeechInteraction` and can be reused.

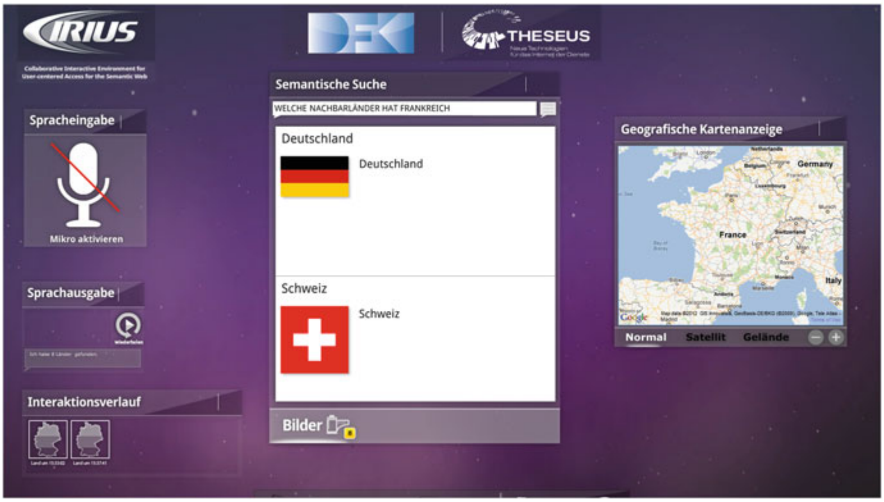


Fig. 4 The *CIRIUS* GUI

2. *Domain Modeling*. The *CIRIUS* domain is geographic information in Europe with a focus on Germany. Thus, new domain-specific concepts for locations are needed. These are subdivided into logical, physical, and political locations. Logical location concepts are `dm#Border` and `dm#Estuary`. Physical locations are `dm#River` and `dm#Sea`. Political locations are `dm#Country`, `dm#FederalState`, and `cm#City`. These concepts are further derived from `odp#NamedEntity`. For the representation of the different result media, existing concepts like `dm#Image`, `dm#Text`, and `dm#Video` can be reused.
3. *Speech Recognition Grammar*. For the focused interaction, a speech grammar rule is necessary. A respective excerpt is shown in Fig. 5. Here, a monomodal, non-referencing speech input "What are the neighboring countries of France?" is mapped to a full interpretation in terms of a `dm#TaskRequest` (derived from `dm#CommunicativeAct`). This object contains the actual task, i.e., `dm#AttributeRetrievalTask`. In order to allow the user to address previously mentioned or currently displayed countries, a slight modification of this rule is necessary. The definite `dm#Country` object inside the `odp#InfoRequest` object needs to be marked as reference. Then, the user can also ask questions like "What are the neighboring countries of [this ↗] country?". The missing information of the resulting partial interpretation, i.e., of which country neighboring countries have to be retrieved, has then to be inferred at runtime in the input fusion step (see Fig. 2) either from a pointing gesture or from the current discourse.
4. *Service Integration*. For providing (not only) geographic information, the *CIRIUS* system integrates numerous heterogeneous (semantic) web services such as BBC-Review, DBpedia, Google API (Images, Maps, and Videos), YouTube,

```
<utterance name="S_GET_NEIGHBOURING_COUNTRIES">
  <phrases>
    <phrase>?PLZ WELCHE NEIGHBOURING_COUNTRIES [(hat)
      (grenzen ?DENN an)] ?DENN COUNTRY ?PLZ</phrase>
    <phrase>?PLZ WELCHE (sind ?DENN die)
      NEIGHBOURING_COUNTRIES (von) COUNTRY ?PLZ</phrase>
    <phrase> ... </phrase>
  </phrases>
  <semantic-interpretation>
    <object type="odp#TaskRequest">
      <slot name="odp#hasContent">
        <object type="dm#AttributeRetrievalTask">
          <slot name="dm#hasRequest">
            <object type="odp#InfoRequest">
              <slot name="odp#hasContent">
                <variable name="COUNTRY" />
              </slot>
              <slot name="odp#hasFocus">
                <object type="odp#Focus">
                  <slot name="odp#hasContent">
                    <object type="dm#Country">
                      <slot name="dm#hasNeighbor">
                        <object type="dm#Country"/>
                      </slot>
                    </object>
              ...
      </slot>
    </object>
  </semantic-interpretation>
</utterance>
```

Fig. 5 Speech recognition rule for speech-only queries about neighboring countries. Please note that the phrases are in German

Freebase, Yahoo Weather, and a data service provided by the ALEXANDRIA use case. How the integration as well as the subsequent processing are accomplished is discussed in Bergweiler (2014). Here, it is sufficient to treat the service integration component as a black box which receives fully specified `odp#InfoRequests` like the one contained in Fig. 6.

5. *Data Mediation*. After the respective information sources have been queried, the result is mediated back from the service-specific data representation formalism into a valid eTFS representation and integrated into the initial request object (see Fig. 6). Finally, this structure is packed into an `odp#PresentationTask` object (derived from `odp#ApplicationTask`) and sent to the presentation planning component. This is also covered in Bergweiler (2014).
6. *Speech Synthesis Templates*. As the user interacts by speech, a voice prompt needs to be generated. In order to verbalize the result, i.e., the neighboring countries of France, a set of transformation rules is necessary. For the focused interaction, three domain-specific rules have to be provided, which recursively

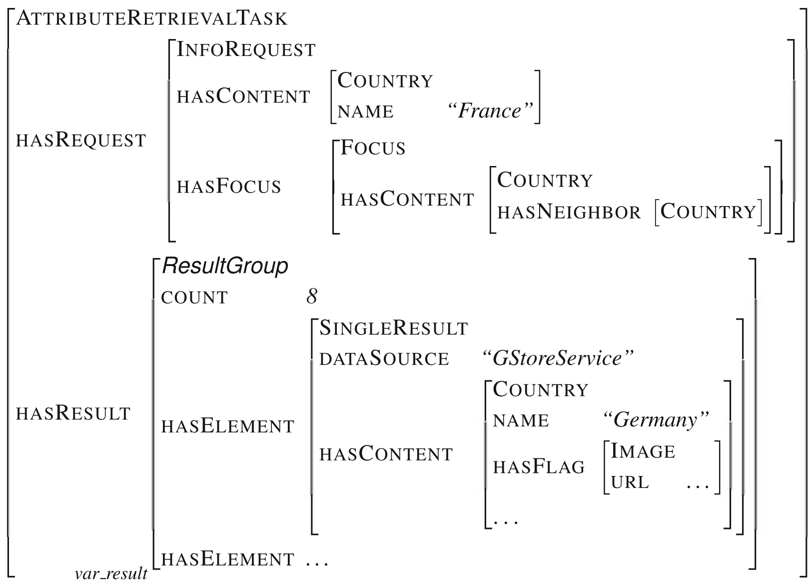


Fig. 6 An example of an eTFS structure representing a user query for neighboring countries

invoke generic rules for list management and named entities. The domain-specific rules distinguish whether zero, one or more than one neighboring country was found. The rule for multiple results is depicted in Fig. 7. It introduces an additional variable binding *var_result* in the condition part that is used to access the result structure in the mapping part. The binding is also shown in Fig. 6.

7. *GUI Model*. Basically, the *CIRIUS* GUI model is composed of different spotlets. Figure 4 shows the knowledge spotlet in the center and the map spotlet on the right. For the knowledge spotlet, a developer needs to provide templates for the presentation of list items of different types, e.g., `dm#Country` or `dm#River`. Then, at runtime, the respective template is selected, based on the item type of the list that has to be presented. In the focused interaction, a list of `dm#Country` objects is displayed in the knowledge spotlet. The respective template consists of an image for displaying the flag and a label for the county name, both nested in a list item view component. Tapping on a list item puts the underlying country in the focus of the discourse. But beforehand, the UI implementation needs to receive an update. This is created in a rule-based transformation which transforms the result list to the XML structure shown in Fig. 8.
8. *UI Implementation*. The client-side *CIRIUS* UI implementation is designed for large tabletop surfaces and implemented in the Adobe Flash/AIR technology (see Fig. 4). The main reason for choosing Flash/AIR by the fact that a high number of high definition videos and multimedia contents have to be displayed and played back simultaneously. The synchronization with the server-side GUI model is achieved by HTTP-based REST (Fielding 2000) communication in combination

```
<generate name="neighboring_countries:multiple"
  weighting="1.5">
  <condition name="this">
    <object type="odp#PresentationTask">
     <slot name="hasContent">
          [Structure of Fig. 7]
     </slot>
    </object>
  </condition>
  <mapping>
    I found \${var_result.count} countries, for example
    <generate name="var_result">
      <parameters>
        <param key="maxcount">3</param>
        <param key="list">conjunction</param>
      </parameters>
    </generate>
    .
  </mapping>
</generate>
```

Fig. 7 Rule-based speech synthesis template for uttering multiple neighboring countries

```
<presentation target="KnowledgeSpotlet">
  <query>what are the neighboring countries of france</query>
  <list type="country">
    <item id="49180ede-536c-4b08-886e-744241e3b1ee">
      <label>Germany</label>
      <image src="http://..."/>
    </item>
    ...
  </list>
</presentation>
```

Fig. 8 GUI update for the knowledge spotlet

with long-polling AJAX requests. Thus, the client implementation receives on the one hand GUI updates from the *CIRIUS* dialog system. On the other hand, each interactive GUI element on the client-side can, e.g., when tapped, transmit its unique identifier to the dialog system in order to trigger actions.

5 Conclusion

The Ontology-based Dialog Platform provides a generic conceptual architecture suitable for THESEUS use case scenarios and beyond. The integrated *ODP* workbench with its integrated set of Eclipse-based editors and tools in combination with the underlying development method simplifies multimodal dialog engineering significantly. Taking the educational information system *CIRIUS* as an example for a recent *ODP*-based dialog application, we illustrated how to build innovative multimodal dialog UIs that integrate multiple heterogeneous web services as data sources. Our rapid dialog system engineering process relies on well-defined extension points within the *ODP* framework, and the adaptation of the provided generic processing modules.

The important transition from research prototype to production use has been achieved as well. The commercial *ODP* version has been made available through our DFKI spin-off SemVox,[3] where a broad range of operational multimodal dialog applications and products has been realized.

The engineering approach presented here constitutes a viable solution and offers additional research potential. So far, we consider a development method that already covers some methodological aspects. It would be beneficial to extend the method to a holistic methodology by integrating agile team coordination and management methods.

References

P. Baggia, D.C. Burnett, J. Carter, D.A. Dahl, G. McCobb, D. Raggett, EMMA: Extensible MultiModal Annotation Markup Language – W3C Recommendation (Feb 2009), http://www.w3.org/TR/emma/

S. Bergweiler, Interactive service composition and query, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

S. Bergweiler, M. Deru, D. Porta, Integrating a multitouch Kiosk system with mobile devices and multimodal interaction, in *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*, Saarbrücken (ACM, New York, 2010), pp. 245–246, http://doi.acm.org/10.1145/1936652.1936698

H. Bunt, J. Alexandersson, J. Carletta, J.W. Choe, A.C. Fang, K. Hasida, K. Lee, V. Petukhova, A. Popescu-Belis, L. Romary, C. Soria, D. Traum, Towards an ISO standard for dialogue act annotation, in *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC '10)*, Valletta, ed. by N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias (European Language Resources Association (ELRA), 2010), http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.178.9209

D.C. Burnett, M.R. Walker, A. Hunt, Speech synthesis markup language (SSML) version 1.0 – W3C recommendation (Sept 2004), http://www.w3.org/TR/speech-synthesis/

G. Calvary, J. Coutaz, D. Thevenin, Q. Limbourg, L. Bouillon, J. Vanderdonckt, A unifying reference framework for multi-target user interfaces. Interact. Comput. **15**(3), 289–308 (2003), http://dblp.uni-trier.de/db/journals/iwc/iwc15.html#CalvaryCTLBV03

R.L. Carpenter, *The Logic of Typed Feature Structures: With Applications to Unification Grammars, Logic Programs and Constraint Resolution*. Volume 32 of Cambridge Tracts in Theoretical Computer Science (Cambridge University Press, Cambridge, UK, 1992)

G. Di Fabbrizio, T. Okken, J.G. Wilpon, A speech mashup framework for multimodal mobile services, in *Proceedings of the 11th International Conference on Multimodal Interfaces*

[3] http://semvox.de

and the 6th Workshop on Machine Learning for Multimodal Interfaces (ICMI-MLMI '09), Cambridge, ed. by J.L. Crowley, Y. Ivanov, C.R. Wren, D. Gatica-Perez, M. Johnston, R. Stiefelhagen (ACM, New York, 2009), pp. 71–78, http://dblp.uni-trier.de/db/conf/icmi/icmi2009.html#FabbrizioOW09

D. Ertl, Semi-automatic multimodal user interface generation, in *Proceedings of the 1st ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '09)*, Pittsburgh, ed. by T.C.N. Graham, G. Calvary, P.D. Gray (ACM, New York, 2009), pp. 321–324, http://dblp.uni-trier.de/db/conf/eics/eics2009.html#Ertl09

R.T. Fielding, Architectural styles and the design of network-based software architectures. Doctoral dissertation, University of California, 2000, http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm

S. Gandhe, N. Whitman, D. Traum, R. Artstein, An integrated authoring tool for tactical questioning dialogue systems, in *Proceedings of the 6th IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems* (Association for the Advancement of Artificial Intelligence (AAAI), Pasadena, California, 2009), http://www.ida.liu.se/~arnjo/Ijcai09ws/

A. Gruenstein, I. McGraw, I. Badr, The WAMI toolkit for developing, deploying, and evaluating web-accessible multimodal interfaces, in *Proceedings of the 10th International Conference on Multimodal Interfaces (ICMI '08)*, Chania, ed. by V. Digalakis, A. Potamianos, M. Turk, R. Pieraccini, Y. Ivanov (ACM, New York, 2008), pp. 141–148, http://dblp.uni-trier.de/db/conf/icmi/icmi2008.html#GruensteinMB08

M. Heinrich, M. Winkler, H. Steidelmüller, M. Zabelt, A. Behring, R. Neumerkel, A. Strunk, MDA applied: a task-model driven tool chain for multimodal applications, in *Task Models and Diagrams for User Interface Design*, ed. by M. Winckler, H. Johnson, P. Palanque. Volume 4849 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2007), pp. 15–27, http://portal.acm.org/citation.cfm?id=1782434.1782439

G. Herzog, A. Ndiaye, Building multimodal dialogue applications: system integration in smartkom, in *SmartKom: Foundations of Multimodal Dialogue Systems*, ed. by W. Wahlster. Cognitive Technologies (Springer, Berlin/Heidelberg/New York, 2006), pp. 439–452, http://dblp.uni-trier.de/db/series/cogtech/54023732.html#HerzogN06

A. Hunt, S. McGlashan, Speech recognition grammar specification version 1.0 – W3C recommendation (Mar 2004), http://www.w3.org/TR/speech-grammar/

H. Kett, M. Winkler, K. Kadner, Integrated service engineering (ISE), in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

Q. Limbourg, J. Vanderdonckt, B. Michotte, L. Bouillon, V. López-Jaquero, USIXML: a language supporting multi-path development of user interfaces, in *Engineering Human Computer Interaction and Interactive Systems*, ed. by R. Bastide, P.A. Palanque, J. Roth. Volume 3425 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2005), pp. 200–220

M. Löckelt, M. Deru, C.H. Schulz, S. Bergweiler, T. Becker, N. Reithinger, A unified approach for semantic-based multimodal interaction, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

G. Mori, F. Paterno, C. Santoro, Design and development of multidevice user interfaces through multiple logical descriptions. IEEE Trans. Softw. Eng. **30**, 507–520 (2004), http://portal.acm.org/citation.cfm?id=1018383

R. Neßelrath, D. Porta, Rapid development of multimodal dialogue applications with semantic models, in *Proceedings of the 7th IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems (KRPD '11)*, Barcelona, July 2011

F. Paterno, C. Santoro, J. Mantyjarvi, G. Mori, S. Sansone, Authoring pervasive multimodal user interfaces. Int. J. Web Eng. Technol. **4**, 235–261 (2008), http://portal.acm.org/citation.cfm?id=1366965.1366970

F. Paterno, C. Santoro, L.D. Spano, MARIA: a universal, declarative, multiple abstraction-level language for service-oriented applications in ubiquitous environments, in *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 16 (ACM, New York, 2009), pp. 1–30

D. Porta, Towards model-driven development of mobile multimodal user interfaces for services, in *Informatik 2010: Service Science – Neue Perspektiven für die Informatik, Beiträge der 40. Jahrestagung der Gesellschaft für Informatik*, ed. by K.P. Fähnrich, B. Franczyk. Volume 175 of Lecture Notes in Informatics (2010), pp. 497–502, http://dblp.uni-trier.de/db/conf/gi/gi2010-1.html#Porta10

D. Porta, D. Sonntag, R. Nesselrath, A multimodal mobile B2B dialogue interface on the iPhone, in *Proceedings of the 4th Workshop on Speech in Mobile and Pervasive Environments (SiMPLE '09) in Conjunction with Mobile (HCI '09)*. (ACM, Bonn, Germany, 2009), http://www.dfki.de/web/forschung/publikationen?pubid=4177

J. Schehl, A. Pfalzgraf, N. Pfleger, J. Steigner, The BabbleTunes system: talk to your iPod! in *Proceedings of the 10th International Conference on Multimodal Interfaces (ICMI '08)*, Chania (ACM, New York, 2008), pp. 77–80, http://doi.acm.org/10.1145/1452392.1452408

D. Sonntag, M. Deru, S. Bergweiler, Design and implementation of combined mobile and touchscreen-based multimodal web 3.0 interfaces, in *Proceedings of the International Conference on Artificial Intelligence (ICAI '09)* (Pasadena, California, 2009)

D. Sonntag, R. Engel, G. Herzog, A. Pfalzgraf, N. Pfleger, M. Romanelli, N. Reithinger, SmartWeb handheld – multimodal interaction with ontological knowledge bases and semantic web services, in *Artifical Intelligence for Human Computing*, ed. by T.S. Huang, A. Nijholt, M. Pantic, A. Pentland. Volume 4451 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2007), pp. 272–295

D. Sonntag, M. Möller, A multimodal dialogue mashup for medical image semantics, in *Proceedings of the 15th International Conference on Intelligent User Interfaces (IUI '10)*, Hong Kong (ACM, New York, 2010), pp. 381–384, http://doi.acm.org/10.1145/1719970.1720036

D. Sonntag, N. Reithinger, G. Herzog, T. Becker, A discourse and dialogue infrastructure for industrial dissemination, in *Spoken Dialogue Systems for Ambient Environments: Second International Workshop (IWSDS 2010)*, Gotemba, 1–2 Oct 2010, ed. by G.G. Lee, J. Mariani, W. Minker, S. Nakamura. Volume 6392 of Lecture Notes in Artificial Intelligence (Springer, Berlin/Heidelberg/New York, 2010a), pp. 132–143, http://www.springerlink.com/content/5149m52mt5378316/

D. Sonntag, C. Weihrauch, O. Jacobs, D. Porta, THESEUS CTC-WP4 usability guidelines for use case applications, Technical report, DFKI GmbH, BMWi (Apr 2010b), http://www.dfki.de/web/forschung/publikationen?pubid=4788

A. Stanciulescu, Q. Limbourg, J. Vanderdonckt, B. Michotte, F. Montero, A transformational approach for multimodal web user interfaces based on UsiXML, in *Proceedings of the 7th International Conference on Multimodal interfaces (ICMI '05)*, Trento (ACM, New York, 2005), pp. 259–266, http://dblp.uni-trier.de/db/conf/icmi/icmi2005.html#StanciulescuLVMM05

W. Wahlster, Towards symmetric multimodality: fusion and fission of speech, gesture, and facial expression, in *KI 2003: Advances in Artificial Intelligence*, ed. by A. Günter, R. Kruse, B. Neumann. Volume 2821 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2003), pp. 1–18

M.H. Weik, *Computer Science and Communications Dictionary* (Kluwer, Boston, 2001), http://www.springer.com/computer/swe/book/978-0-7923-8425-0

Interactive Service Composition and Query

Simon Bergweiler

Abstract The various modern services on the Internet offer a wealth of information. Therefore, the provision of specific credible information becomes more important. Nowadays, different lightweight application architectures based on web services establish themselves as de facto standards for data access and data publishing via so-called REST-based interfaces on the Internet. Within these approaches a flexible integration of services data is becoming a reality, and for the Internet of the future, these activities, designed as Linked Open Data, are becoming increasingly important. In this paper we describe the development of a framework named Joint Service Engine (JSE) for the discovery, integration, processing, and fusion of "semantically" annotated web services. The focus is on the discovery of appropriate service information and the fusion of all gathered information in real time. The processing chain includes the harmonization of different data structures obtained from heterogeneous web services that have to be mapped to specific predefined domain ontologies. In a multimodal dialog shell, JSE supports the system to answer spoken queries like "Give me the neighboring countries of Germany, their populations and the geo-coordinates of the respective capitals". This requires several service calls and the filtering and combination of their results to provide an adequate answer.

1 Introduction

Today, web services are an important feature in the areas of information provision, process management, and Enterprise Application Integration (EAI), supported by the popular paradigm of service-oriented architectures (SOAs). In fact, on the

---

S. Bergweiler (✉)
DFKI GmbH, German Research Center for Artificial Intelligence, Saarbrücken, Germany
e-mail: Simon.Bergweiler@dfki.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_13,

one hand, web services offer advantages, like loose coupling of components, and consequently their easy replacement or combination, but, on the other hand, lead to further technical challenges in terms of discovery, management, reuse, and composition. Due to the rapid and uncontrolled growth of the Internet, the process of finding and processing relevant information turns out to be very problematic.

Although not all the efforts of information sources of the service web can be taken into account, the most common approaches are all considered in this investigation. The techniques of discovery, orchestration and combination of web services provide a powerful new dimension of information access for end users. Rich new information spaces can be realized through the individual combination of information currently residing in separated web services. But for end users, it is a difficult task to find and access such information sources by traditional web search engines.

The goal of our framework, named Joint Service Engine (JSE), has been the implementation of a multimodal module for interactive access to complex web services, as well as the provision of methods and tools for annotation and processing of web services for semantic interaction. Furthermore, communication protocols have been established, which permit the information exchange of the developed component with the bundled interaction systems on an interaction-specific level.

While semantic technologies and the connected interaction system have been described in Löckelt et al. (2014) and Porta et al. (2014), the following section about web services and semantics is dedicated to the use of these technologies in real applications. In this context, the next section introduces the integration and harmonization of the obtained information. This is followed by an overall description of the architecture of the developed framework and a more detailed explanation of the individual core modules and their integration and tasks.

2 Web Services and Semantics/Annotations

Today's web services and their corresponding infrastructure are based on the first generation of industry standards for service description, discovery, and composition, which do not address semantic representations yet. The process of discovery and composition of web services is a more complex task, particularly with regard to the potential number of available heterogeneous information sources.

Web services are gaining particular importance within enterprise applications supported by the popular paradigm of service-oriented architectures (SOAs) and issues of enterprise application integration (EAI). In our scenario, the considered web services provide interfaces to access structured data, stored in large relational databases or NoSQL databases. On the one hand common client applications can access these databases via web services, but on the other hand, these data-based web services can in turn act as service consumers to ensure immediate access to needed additional information of remote sources. In order to implement the exchange of information, vendors of relational databases such as Microsoft, Oracle and IBM rely on traditional web service technologies and use their own middleware frameworks. The most prominent examples currently used are WSDL for service description, UDDI for service publication and discovery and, currently evolving, WS-BPEL for process definition and execution.

In our consideration, ordinary web service descriptions are supplemented by annotations which describe the functionality of a service in more detail, e.g., with technical description languages like WSDL or WADL (Chinnici et al. 2007; Hadley 2009). The added annotations extend the service description to a machine-readable specification of service functionality. With reference to the efficient and detailed service description it is possible to quickly find and invoke suitable services compliant with the present use case. User participation in such a scenario demands highly flexible and interactive approaches of dynamic service discovery and composition. This can only be achieved by exploiting semantic descriptions of web service capabilities specified in a machine-understandable way. WSDL 2.0, which now supports RDF, constitutes only a first initial step towards semantic annotations.

Consequently, the scientific community promotes approaches like OWL-S or WSMO for full semantic annotation of web services (Lausen et al. 2005; Martin et al. 2004). Oberle (2006) and Fensel et al. (2007) give a good detailed description regarding these approaches for the current state of the art in this area. However, these approaches lack the proof of suitability in practical use, and in particular do not involve the user in the process of information retrieval. If some information is missing in the composition process, the access fails or delivers unsatisfactory answers. If the user is included in the composition loop, missing information like authorization, preferences or context data for certain services can be added interactively. Thus, the user participates in the service composition and query, and has the power to steer the search and control the results. Service execution results can be adaptively clustered or filtered according to the user's needs.

A first generic approach for interactive composition was realized in the SmartWeb project (Sonntag et al. 2007). The service composer of SmartWeb comprised a plan-based approach combined with semantic matching techniques for automatic service composition (Sirin and Parsia 2004). It realized a mapping of the semantic structure of the user's input to an initial state and goal description for the planning process. Additionally semantic service descriptions, based on the WSDL specification of the services, are mapped to planning operators. During the composition process, the module provides information about missing data that must be communicated to the user and initiates a replanning process on an appropriate response, thus increasing accuracy and robustness of the answering process.

Recently, different lightweight web service application architectures have been established that make information available via RESTful interfaces (REpresentational State Transfer). But the usual problems of interoperability, data type mapping and interconnection go hand in hand with the development of web services in the context of the Semantic Web (Fensel et al. 2003). For this reason, in 2006 Tim Berners-Lee motivated the development of the Linked Open Data Initiative, which was initiated to make machine-readable data available. Linked data is structured data that can be used in traditional web applications, but it is additionally enriched by links or references, pointing to several semantically relevant databases on the Web (Krummenacher et al. 2010). Particularly, academic institutions, archives, and libraries rely on the dissemination of information from research and teaching and provide persistent and reliable services. First link services link key words and information to people with publicly available information from DBpedia.[1]

3 Integration and Harmonization of Information

In the digital age of high-speed broadband Internet access, the retrieval of information through web services and simple cellular phone applications moves further and further into the focus of public interest. Different web service interfaces offer an unprecedented ability to access concrete content of different information spaces at the same time. But for users the question arises about how reliable, composable, and reusable this information that is selectively obtained from various sources is. For a more detailed consideration of these questions, we try to approach each issue through a use case. To answer user queries of different contexts, a simple web service is no longer sufficient as the sole data source for reliable information, because it is usually designed for the representation of data in a particular context. Rather interesting is the question about how services can be combined or orchestrated to satisfy a need for information that actually exists and how to handle the data formats of the individual sources. More specifically, web services operate according to different domain knowledge with various data exchange formats, such as XML, SPARQL,[2] JSON[3] or RDFs. Due to this heterogeneity, for a conventional system it is difficult to extract and interpret concrete factual knowledge on a particular topic from several different data sources.

This article describes the JSE approach, i.e., the development of a framework that collects information from multiple heterogeneous web services and facilitates the transfer of data between extraction and harmonization. It provides access to contents of different databases encapsulated by web service interfaces. All gathered data will be prepared in the context of the given use case and returned to the user. The developed framework supports the user of an interactive, web-based interaction system in the process of annotation, integration and aggregation of heterogeneous information that is accessible through web services.

---

[1] http://dbpedia.org/

[2] http://www.w3.org/TR/rdf-sparql-query/

[3] JavaScript Object Notation – http://www.json.org/

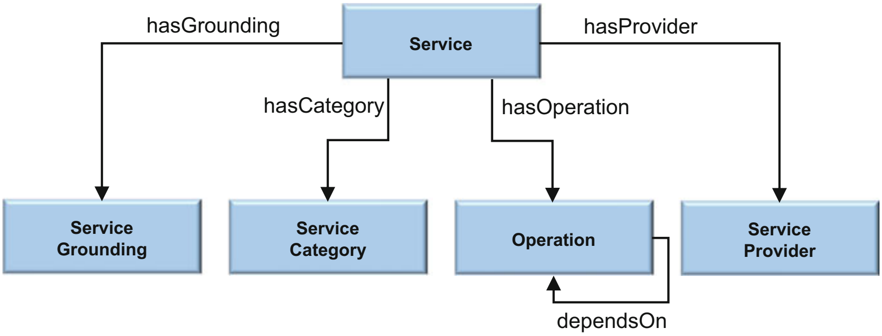


Fig. 1 Representation of web services

4 Knowledge Representation

The chosen approach describes the access to specific and complex knowledge by combining multiple knowledge sources. But in this context the question arises about how information can be represented in order to be reusable and interchangeable. To this thematic issue Gruber in 1993 applied the term "ontology", which is widely accepted in the Semantic Web community: "An ontology is an explicit specification of a conceptualization" (Gruber 1993). In his remarks, the ontology describes the basic vocabulary for detailed description of a domain of interest and specifies the meaning of terms defined within the vocabulary.

4.1 Service Ontology

A service ontology was developed as a means of domain-independent representation of web services and their properties. The purpose is the representation of web services and their properties using the definition of anchor points that link specific concepts of the service ontology with those from the domain ontology and their associated use cases. The functional description of a service is based on the operations and their functional properties, such as input and output parameters and pre- and post-conditions. Therefore, the modeling and semantic annotation of the used services and their inputs and outputs follows the specified top-level ontology OWL-S; see Martin et al. (2007) for more details. But in this implementation, the semantic annotation process is made proactively and manually, based on the central vocabulary of the service ontology. In the following, the basic concepts of the service ontology and the anchor points are introduced.

The service concept is the main concept of the service ontology, which describes a web service by relations to other concepts. An overview of these concepts is shown in Fig. 1. The semantic classification of a web service, which essentially serves to discover a service from a category, is domain-specific. This can be done by defining the sub-concept *Service Category*. The concept *Service Provider* encapsulates all information about the provider of the service, and the *Service Grounding* defines the mapping of a semantically described service on its technical implementation (e.g. WSDL, WADL, and REST). The grounding is used in the execution process of a service, when a service must be instantiated. A semantically described service is basically independent of its realization, and for this reason, it is possible to define multiple groundings.

In general it is necessary to divide operations into two categories: atomic operations that relate directly to the technical realization of the service, and composite operations, a composition of various operations. The composition can be defined recursively and results finally in a tree structure of compound operations, wherein the leaves of this tree are atomic operations that can be executed. In addition to this hierarchical decomposition of a complex operation for each operation, a dependency relation can be defined that specifies the execution order as a partial order.

4.2 Framework Ontology

The semantic representation of services and their content is used for the integration of heterogeneous information, whereas ontologies act as inter-lingua between different data formats and communication protocols. These metadata annotations are the basis for the aggregation of semantic web services through formal rules and automated reasoning (*Knowledge Matching*), similar to innovative types of content-ecosystems and collaborative platforms.

The developed framework attempts to harmonize all extracted relevant facts and relations by mapping the heterogeneous result structures to an internal domain ontology. This ontology is used for the central storage of the entire extracted contents and implicit relationships. At the same time, it serves as source and allows centralized access to the gained content in order to perform further processing steps or to generate the final result structures for the user.

5 Joint Service Engine

The "Joint Service Engine" (JSE) is an approach to querying and integrating factual knowledge of heterogeneous web services and passing it to an interaction system, such as the implemented multi-modal dialog shell (Löckelt et al. 2014). According to a user query, multimedia data and text documents are retrieved from different connected web services. In the current realization, the framework provides access to different types of services, such as BBC-Review, DBpedia, Google API (Images, Maps, and Videos), YouTube, Freebase, Yahoo Weather and the Neofonie data service. All listed services encapsulate broad databases, and provide access to a large pool of stored information with a high degree of reliability. In terms of the THESEUS research program JSE provides the integration of factual knowledge of the individual use cases such as ALEXANDRIA or ORDO. JSE aggregates the individual back end results of all connected data sources and transforms them into the representation format needed by connected client systems.

As outlined above, the JSE framework is based on a distributed component architecture whose independent core modules encapsulate functionality of individual processing steps, such as the *Answer Module*, *Semantic Web Service Repository*, *Mapping Core*, *Execution* and the *Planning Module*. The following section describes the core modules and their interaction in detail. As JSE is implemented as a multi-user system, queries can be executed in parallel.

6 Core Modules and Architectural Overview

In combination with an interaction system, like the developed dialog shell, JSE can handle and process complex spoken natural language queries. Such a query could be formulated as follows: "Give me the neighboring countries of Germany, their populations and the geo-coordinates of the respective capitals". This natural language query is evaluated by the dialog shell and transferred as utterance in a domain-specific query structure, which serves as input to the JSE. The dialog shell ontology uses extended typed feature structures (eTFSs) for semantic data representation. The information flow is depicted in the overview of the individual JSE components in Figs. 2 and 3, and explained below.

On the JSE side the *query decomposition* module serves as the central interface for retrieving information. In the query module, the transmitted request structure will be evaluated and decomposed into several sub-query parts. The query decomposition process is carried out, based on the users' ontology that defines the vocabulary of the domain of interest, and the resulting query constructs are passed to the *planning module*.

Corresponding to each part of the query, the *planning* module analyzes and interprets what to do and which services appear suitable to answer the query. In this process, particular attention is paid to the aspect of service matchmaking.

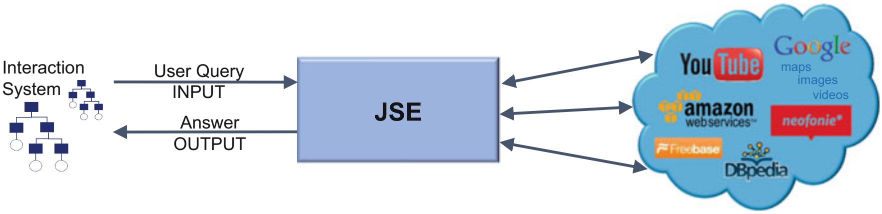


Fig. 2 Architectural overview

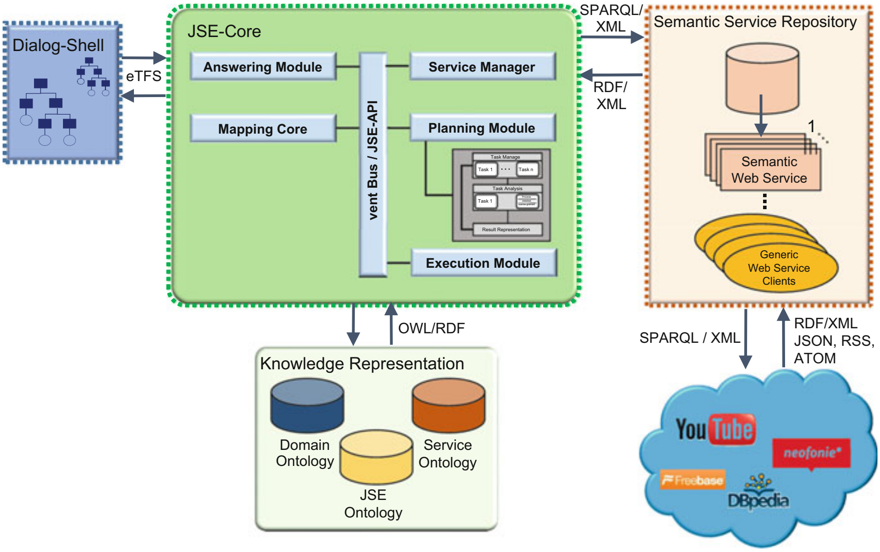


Fig. 3 Detailed architectural overview of the Joint Service Engine

For the matchmaking process of the *semantic service repository*, the extracted domain information of the query analyzing process is used as a basis for determining appropriate service descriptions. As a result of the matchmaking process, a list of matching service descriptions is created and returned to the planning module.

If a query is not sufficiently specified, because of ambiguity, unknown terminology or missing information, the planning process fails. This occurs when a query does not provide any indication or relation to a concrete concept. In this case several possibilities for interpretation arise, e.g., for the query input string “Berlin” it is not obvious whether it refers to the city or the federal state. In the context of the modeled domain, relevant interpretations are handled by predefined rules that refer to specific types and concepts. In order to set a process priority, all rules are weighted by confidence values. A clarification task tries to add these associated specific types and concepts to the query and once again pass them to the answering module.

This component passes the list of matching service descriptions to the *execution module* and monitors the execution process in order to reschedule in case of failure, e.g., for non-availability of a service. In the execution process, the module involves the respective client applications associated with their service, and returns the result structures for analysis to the planning module. This module uses rules defined in the mapping core and maps the result structures to specific instances of the JSE ontology. If there are additional planning steps available, the harmonized contents in turn serve as input for the next request. In a final processing step the overall result structure will be generated. Therefore the previously created instances of the JSE ontology must be mapped to the domain ontology of the dialog shell, by using predefined rules for parameter mapping of the mapping core.

The JSE framework was developed as a service-oriented architecture (SOA) and is integrated on an application server. The semantic service repository provides further standalone SOAP-based services in addition to the actual JSE application.

However, it must also be acknowledged that the user's query is underspecified, so conventional planning fails. In that case the *what-if-planning* procedure creates a hypothesis query by enriching the original query with heuristical information to force the query analysis and interpretation process.

6.1 Answering Module

Following the model of context-driven information retrieval, the specific information needs of the user are taken into account in order to create a better basis for decision making. One objective of query processing, the disclosure of hidden information, is based on the information specified in the query, such as action, time, location, urgency and the user's knowledge. But, depending on the specified situation, there is room for interpretation, which can be considered context-sensitive and refers to different results. In our approach, we draw our own conclusions and form hypotheses adequately for a specific request pattern in case of underrepresented or ambiguous query structures.

The first module of the JSE processing chain is the *Answering Module* that retrieves and understands user query expressions according to a vocabulary defined by an interaction ontology which is discussed in detail in Porta et al. (2014). Based on this interaction ontology the users are able to add additional constraints, in the form of metadata, to the semantic request that refines the precision of the selection results, and increases the ability to find the right information sources. That query, originally formulated using the eTFS format, is transformed semi-automatically into an ontology-based query representation, based on our internal query representation structure. The eTFS format is basically the semantic communication standard of the speech and gesture recognizing in dialog systems. Mapping standard OWL and RDF(S) ontologies to lightweight eTFS variants is possible semiautomatically. Although a full OWL/RDF(S) expressiveness cannot be maintained in the eTFS version, this is not needed for the task of query representation.

The answering module decomposes the incoming query into several sub-query parts and transmits each part for further processing to the planning component. At the end of the processing chain of the JSE framework the final result structure must be produced. Therefore just the relevant knowledge facts of all gathered information must be selected, and transferred to the appropriate output format. A special topic has been deposited in the internal query structure that refers to the mapping rules, defined in the *Mapping Core* component, described further below.

Within the JSE processing, a query is transformed multiple times, and services are selected in accordance with the requirements of this query. An abstract view for illustrating the various query transformation steps is shown in Fig. 4. Compositing steps and service instances have been omitted for clarity.

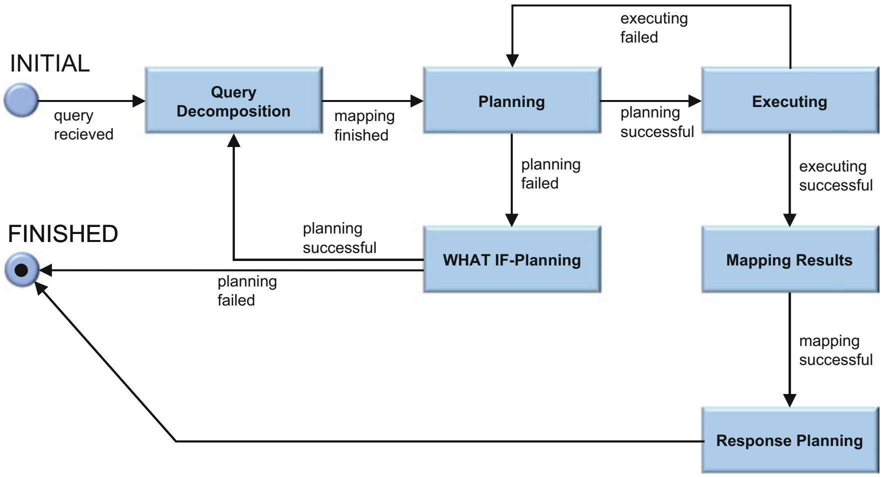


Fig. 4 Information flow in the Joint Service Engine

The input query (A) for this component, illustrated in Fig. 5, could be formulated in eTFS or in a special XML-based query format. This eTFS query is transformed by a JSE API to the internal query representation structure (XML), which has a direct binding in Java objects. In the second step (B), the information need (requested data) and the input available are clearly specified. Thus, the proper service needed to answer the request can be chosen using the available service grounding descriptions.

Depending on the service that matches the information need, different execution paths can be pursued from then on. For SPARQL-based services (*left path*), data type mapping consists of transforming the query into a corresponding SPARQL query (C), and mapping constants as necessary. This is necessary, since, for example, entity references in the internal ontology representation (B) are typically not equal to the ontology used by the service we want to access later. Then the query is sent to the back end SPARQL endpoint. Its results get transformed into the internal JSE-ontology by mapping rules (E).

For other services (e.g., SOAP/REST-based web services; see *right path* in Fig. 5), entity resolution and data type mapping is done in the step from B to C. Then, the web service is called, resulting in (E), a representation of results in a format set by and specific to the web service called. The final result representation is created as the harmonized internal results (F).

6.2 *Planning Module*

Each planning step utilizes a JSE ontology, which contains a priori no knowledge. The body of factual knowledge is always built for the particular context and discarded after production of the response structure. Accordingly, in the current

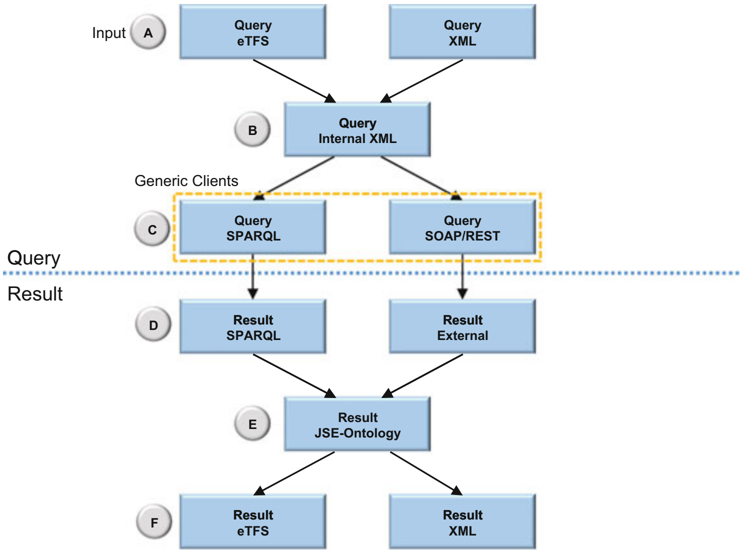


Fig. 5 Representations of queries and results

implementation, the module behaves stateless and the process of analysis and interpretation relies solely upon context information from manually predefined plans. These plans define pre- and post-conditions to identify relevant context-related information that is required to initiate the matchmaking process of the semantic service repository. Based on this information as input, the semantic matchmaking service is looking for adequate service descriptions stored in the appropriate directory and all resulting service descriptions are returned as a result set to the planning module.

In a composition process, proactive individual tasks can be combined and stored as a plan, but implicit and explicit composition tasks must be distinguished. An implicit task, such as executing the transformation or harmonization of data by an additional service, can be used several times and is essential for the combination of services. Only such services can be combined, where the result of the first service may in turn serve as input for the following service. Thus, for example, the country code is represented internally in the JSE framework in a standardized format (ISO 3166), but the several connected services provide different representations of country codes. Before the retrieved content can serve as input for the next service in the planning chain, it must be implicitly harmonized. In the harmonization process, the newly obtained information is merged with existing instances represented in the ontology, leading to an expansion of the existing body of knowledge.

This query from the context of the final demonstrator serves as an example for service composition: “Give me the neighboring countries of Germany, their populations and the geo-coordinates of the respective capitals.”

Each defined plan also specifies the execution order of services. Referring to the above example, the capitals of neighboring countries should be obtained only if the countries are known. The planning module thus performs the following operational steps to answer this query and web services are combined or aggregated if necessary:

1. Find services that can deliver the neighboring states in Europe.
2. Call these services and add the results to the JSE ontology.
3. Find services to determine the population of countries in Europe.
4. Call these services and use the found states as input. Add the results to the existing instances in the JSE ontology.
5. Find services that provide the capitals of countries in Europe.
6. Call these services and use the found states as input. Add the results to the existing instances in the JSE ontology.
7. Find services that provide geographical coordinates for cities.
8. Call these services and use the found cities as input. Add the results to the existing instances in the JSE ontology.
9. Create the final result for the requesting interaction system.

Depending on the complexity of the connected web services, queries can be formulated in a combined form, like RDF triple stores by SPARQL query language.

The planning module uses a monitoring component to control the execution of the associated planning steps which initiate calls to specific remote web services. By combining different types of content and functionality of individual services, the JSE framework offers new information sources and innovative value-added services to the user.

6.3 Execution Module

The *Execution Module* instantiates and invokes the determined web services based on the specific defined service grounding information. The service grounding contains the reference to the particular query structure to call the service via a generic client. The results of all involved clients are transferred to a common vocabulary, based on the JSE ontology. At several points in the execution process, the mapping core is needed to transform data structures. For this harmonization of the service results, the module uses the *Mapping Core* with its predefined mapping rules for each service and client. If multiple services cover the same topic, their clients are accessed simultaneously and the results are transformed according to the mapping rules.

6.4 Mapping Core

As mentioned in the section on query decomposition, during the processing chain, content must be repeatedly transformed from one data format to another. Within the *Mapping Core*, formal mapping rules are used that allow a higher quality data type mapping on a more abstract level, like the mapping of an external ontology or taxonomy of objects to the internal ontology. In a general declarative element-based mapping process, input and output parameters are mapped to and from the internal data structure. The mapping core module is realized as a generic component and is used in the following core modules:

- *Answering module*: To process a user query, it must first be decomposed into sub-queries. For this purpose a transfer of the incoming data on to the internal data representation structure is necessary. To carry out this transfer, this component depends on the ontological infrastructure and the defined data formats of a use case. Based on the incoming query a topic is set, which relates to the format of the result structure (XML, OWL or eTFS). According to this format a result structure with all relevant obtained results is generated and returned to the user.
- *Service repository*: The planning component creates a query to the service repository to find a suitable service. Therefore the internal query format must be mapped to a SPARQL or XML query. Similarly, the response of the services repository must be mapped to the internal representation format. This is needed to associate the semantic service description with the respective generic client representation. This procedure achieves a proper execution and invocation of all identified suitable services.
- *Service execution*: A central problem for the integration of information consists in the individual design of different REST-based services, which use mostly their own internal data representation that does not follow any standard. For example, many linked data sources offer a format for exchanging data such as RDF/XML, XML (atom) or JSON. Therefore, in some cases, the individual result structures of the external web services must be first transferred in an intermediate step to an internal XML meta-format before they can be mapped to JSE's ontological data structure. All relevant input and output parameters of the retrieved external result structures, e.g. XML, are associated with a data type in the JSE's ontology.

6.5 Semantic Service Repository

The dynamic discovery and binding of suitable services is a characteristic for service-oriented architectures. In our approach the planning component is the service consumer that discovers a service by defining certain criteria in a request sent to a service registry, which manages a potentially large service repository. The response of this service registry is a result set of matching services. The consumer gets the result set of matching services and the service descriptions in OWL-S from the service registry. On the basis of these descriptions the consumer is able to bind and call the services dynamically at runtime.

Service Repository Broker (Hosting of Services) The service registry broker manages all semantically annotated services in the service repository. For each service there exists a corresponding semantically annotated service wrapper, which is created at design time (predefined). This technique allows integrating different service types from SOAP to REST, and their different description languages, such as WSDL or WADL. The entire semantic service modeling is based on the service ontology as described above. With its semantic expressiveness it represents the specific domain of the included information, the defined operations and the associated message elements.

As described in the section on the service ontology, it is possible to define a semantic service description in the semantic service repository, which already provides a composition of atomic transformation operations. Depending on the use case of the connected application, the user can individually decide whether to deposit a composed operation as a semantically described web service, or to concatenate single tasks in the planning module.

Matchmaking (Discovery of Services) The matchmaking process of finding the best suitable web service based on the characteristics of query or customer-supplied information is facilitated by using a semantic service registry. This registry is one crucial part of the components architecture and can be accessed as a typical web service search engine.

The matchmaking engine serves as part of the semantic service repository and matches the incoming semantic requests of connected components. The goal of this matchmaking component is to search for services with a semantic service description using semantic search techniques. That means the discovery of a service is made by the matching of keywords and of detailed semantic descriptions; in particular, the contextual meaning of terms refers to the searchable data space. The successful search will return a set of any services which match the specified condition.

If the matchmaking process identifies a service, the service broker delivers the service representation. This service description contains a reference to a generic client, which can call and invoke the functionality of the service.

Service Grounding – Generic Service Clients The service grounding refers to finding a concrete service implementation for a given abstract service specification. This typically involves technical descriptions of service data type mappings and calling details, e.g. protocol and request, and response types of the service. In this implementation the service grounding is done by matching a specified identifier to a description of a concrete generic client implementation and a special predefined more or less complex query structure.

7 Conclusion

We have outlined the general design and implementation of the Joint Service Engine, especially with regard to the core modules and their functionality. Our solution addresses different data sources and integrates their knowledge in spite of different query languages and different domain ontologies. The merging and linking of knowledge based on user-centered context contributes to a refinement of information delivered by semantically annotated web services.

The JSE engine combines the functionality that allows accessing SOAP/REST as well as SPARQL services. The most important feature is the functionality, which enables the system to connect with SPARQL endpoints, allowing access to large RDF triple stores. In order to facilitate efficient processing of domain knowledge, a mechanism for importing lightweight variants of domain ontologies has been created. In the future, further adjustments of data formats and query representations will be carried out, making the component more customizable while keeping standards compliance.

For the immediate future, it is planned to switch from the current simple ontology-based template query representation to a representation that allows arbitrary queries and complex query patterns. This representation will be based on a variant of SPIN[4] (*SPARQL Inferencing Notation*), which is, in turn, a representation of the SPARQL query language in RDF(S).

JSE has been employed successfully in various technology demonstrators in different application domains, including public presentations at the CeBIT fairs in 2011 and 2012 in Hanover, Germany.

References

R. Chinnici, J.J. Moreau, A. Ryman, S. Weerawarana, Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language – W3C Recommendation, June 2007, http://www.w3.org/TR/wsdl20/

D. Fensel, J.A. Hendler, H. Lieberman, W. Wahlster (eds.), *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential* (MIT, Cambridge, 2003)

D. Fensel, H. Lausen, J. de Bruijn, M. Stollberg, D. Roman, A. Polleres, J. Domingue, Web services, in *Enabling Semantic Web Services – The Web Service Modeling Ontology* (Springer, Berlin/Heidelberg/New York, 2007), pp. 37–54

T.R. Gruber, A translation approach to portable ontology specifications. Knowl. Acquis. Spec. Issue: Curr. Issues Knowl. Model. **5**(2), 199–220 (1993)

M. Hadley, Web Application Description Language – W3C Member Submission, Aug 2009, http://www.w3.org/Submission/wadl/

R. Krummenacher, B. Norton, A. Marte, Towards linked open services and processes, in *Future Internet – FIS 2010*, ed. by A.J. Berre, A. Gómez-Pérez, K. Tutschku, D. Fensel, vol. 6369 of

[4] http://spinrdf.org/sp.html

Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2010), pp. 68–77, http://dx.doi.org/10.1007/978-3-642-15877-3_8

H. Lausen, A. Polleres, D. Roman, Web Service Modeling Ontology (WSMO) – W3C Member Submission, June 2005, http://www.w3.org/Submission/WSMO/

M. Löckelt, M. Deru, C.H. Schulz, S. Bergweiler, T. Becker, N. Reithinger, A unified approach for semantic-based multimodal interaction, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

D. Martin, M. Burstein, J. Hobbs, O. Lassila, D. McDermott, S. McIlraith, S. Narayanan, M. Paolucci, B. Parsia, T. Payne, E. Sirin, N. Srinivasan, K. Sycara, OWL-S: Semantic Markup for Web Services – W3C Member Submission, Nov 2004, http://www.w3.org/Submission/OWL-S/

D. Martin, M. Burstein, D. McDermott, S. McIlraith, M. Paolucci, K. Sycara, D.L. McGuinness, E. Sirin, N. Srinivasan, Bringing semantics to web services with OWL-S. World Wide Web **10**(3), 243–277 (2007)

D. Oberle, *Semantic Management of Middleware* (Springer, Berlin/Heidelberg/New York, 2006)

D. Porta, M. Deru, S. Bergweiler, G. Herzog, P. Poller, Building multimodal dialog user interfaces in the context of the internet of services, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

E. Sirin, B. Parsia, Planning for semantic web services, in *In Semantic Web Services Workshop at 3rd International Semantic Web Conference (ISWC 2004)*, 2004, http://www.mindswap.org/papers/SWS-ISWC04.pdf

D. Sonntag, R. Engel, G. Herzog, A. Pfalzgraf, N. Pfleger, M. Romanelli, N. Reithinger, SmartWeb handheld – multimodal interaction with ontological knowledge bases and semantic web services, in *Artificial Intelligence for Human Computing*, ed. by T.S. Huang, A. Nijholt, M. Pantic, A. Pentland. Volume 4451 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2007), pp. 272–295

Intelligent Semantic Mediation, Knowledge Acquisition and User Interaction

Daniel Sonntag and Daniel Porta

Abstract The design and implementation of combined mobile and touchscreen-based multimodal Web 3.0 interfaces should include new approaches of intelligent semantic mediation, knowledge acquisition and user interaction when dealing with a semantic-based digitalization of mostly unstructured textual or image-based source information. In this article, we propose a semantic-based model for those three tasks. The technical components rely on semantic web data structures in order to, first, transcend the traditional keyboard and mouse interaction metaphors, and second, provide the representation structures for more complex, collaborative interaction scenarios that may combine mobile with terminal-based interaction to accommodate the growing need to store, organize, and retrieve all these data. Interactive knowledge acquisition plays a major role in increasing the quality of automatic annotations as well as the usability of different intelligent user interfaces to control, correct, and add annotations to unstructured text and image sources. Examples are provided in the context of the MEDICO and TEXO use cases.

1 Introduction

In virtually all business and social community domains, we increasingly rely on digitized information. There is a growing need to store and organize all this data, including business records, folksonomy data (creating and managing tags to annotate and categorize content), medical laboratory reports, Internet pages, and images of all kinds. In addition to pure Web 2.0 interfaces, such as folksonomies and wikis, Web 3.0 interfaces bring together the social and community aspects of user-generated content of folksonomies with the semantic web structures of the

D. Sonntag (✉) • D. Porta
DFKI GmbH, German Research Center for Artificial Intelligence, Saarbrücken, Germany
e-mail: Daniel.Sonntag@dfki.de; Daniel.Porta@dfki.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_14,

THESEUS research program. One of the core intentions of the Semantic Web is providing machine understandable data for effective retrieval of commercial, scientific, and cultural data in a universal medium. The intuitive access to the semantic data requested via an application-specific back end demands a unified approach for semantic-based multimodal interaction (see Porta et al. 2014) and a discourse and dialog infrastructure for industrial dissemination (Sonntag 2010). However, the actual design and implementation of combined mobile and touchscreen-based multimodal Web 3.0 interfaces must include new approaches of intelligent semantic mediation, knowledge acquisition and interaction in general when dealing with a semantic-based digitalization of mostly unstructured textual or image-based source information.

Most interestingly, the idea of the Semantic Web provides new opportunities for semantically enabled user interfaces. For example, in the context of Human Computing (Huang et al. 2007), anticipatory user interfaces should be human-centered and require humanlike interactive functions (which include the understanding of certain human behavior) for intelligent semantic mediation, knowledge acquisition, and interaction. Over the last several years, we have explored technical components that rely on semantic web data structures in order to (1) transcend the traditional keyboard and mouse interaction metaphors, and (2) provide the representation structures for more complex, collaborative interaction scenarios that, most interesting, may combine mobile with terminal-based interaction, or the physical with the virtual world in order to accommodate the growing need to store, organize, and retrieve all this data. For this purpose, we introduced the notion of *collaborative multimodality*, which serves as the missing link between traditional Human-Computer Interaction (HCI) and intuitive human-centered designs for our particular purpose of storing, organizing, and retrieving digitized information in the THESEUS use cases (Sonntag 2012).

2 Background of Collaborative Multimodality

Many systems are available that translate natural language input into structured ontological representations (e.g., AquaLog (Lopez et al. 2005)), port the language to specific domains, e.g., ORAKEL (Cimiano et al. 2007), or use reformulated semantic structures, e.g., NLION (Ramachandran and Krishnamurthi 2009). AquaLog presents a solution for a rapid customization of the system for a particular ontology; with ORAKEL a system engineer can adapt the natural language understanding (NLU) component (Engel 2006) in several cycles, thereby customizing the interface to a certain knowledge domain. The system NLION uses shallow natural language processing techniques (i.e., spell checking, stemming, and compound detection) to instantiate an ontology object. However, if one looks closer at actual industrial projects' requirements for data management, such as in the THESEUS use cases, this idealistic vision of dealing with ontologies and language-based user input begins to blur, mainly because of software infrastructure or usability issues. Previous research in intelligent interaction systems (see for example Bunt and Beun 2001) has emphasized the important relationship between cooperation and multimodal communication. Especially in industrial, semantic-based application domains of collaborative multimodality, the extended concept of *task-based* cooperation should be one of the major driving forces in order to store, organize, and retrieve digitized information of a particular THESEUS use case domain. That means we deal with semantic service tasks such as booking flights, answering specific questions about a particular domain (e.g., medical conditions), or helping a user acquire the data in a semantic machine-readable form in order to increase the data quality and interoperability for future tasks, in which the semantically enriched data sources can be used. An important side condition of our multimodal interaction principles (based on discourse theory) is that dialog utterances are also treated as actions of the multimodal dialog system manager. We claim that the artificial separation of dialog and action into distinct user interface modes is a symptom of a deeper dichotomy between the technologies used to implement speech-based interaction system actions in combination with traditional desktop-based HCI methods.

AI techniques (Russel and Norvig 2003) can be used to model this complex interaction behavior, and machine learning (e.g., see Mitchell 1997) plays a significant role in the modeling of discourse information, collaborative goals, and content information over time. We are almost certain that multimodal dialog-based communication with machines will become one of the most influential AI applications of the future. Apple shipped the first voice speech input software Siri, which takes the application context (e.g., email client or web search) into account to the iPhones, in 2011. Only three years ago, multimodal interfaces for coherent dialog were among the main achievements of the SmartWeb system (funded by the German Federal Ministry of Education and Research with grants totaling 14 million €) (Sonntag 2010; Wahlster 2008). In SmartWeb, questions and commands are (additionally) interpreted according to the context of the previous conversation (Apple has just begun to assimilate the basic technologies, for example, the disambiguation of a context is not yet integrated). We will take this multimodal dialog infrastructure as an example of how applications like Siri can be implemented (see the range and similarity of voice applications to the prior research projects). The full list of collaborative multimodality principles for THESEUS use cases can be summarized as follows:

1. HCI systems must not artificially separate dialog and action,
2. HCI dialogs should be planned with modality independent ontology-based systems,
3. HCI systems must be multimodal and use AI technology,
4. HCI systems must employ mixed initiative dialogs.

In this contribution, we will focus on the fourth principle; HCI systems must employ mixed initiative dialogs, and bring this into the context of the main three collaborative THESEUS interaction tasks of interactive semantic mediation (Sect. 3), interactive knowledge acquisition (Sect. 4), and applying suitable usability rules in the context of specific industrial THESEUS use case tasks (Sect. 5). The major challenge is the so-called *knowledge acquisition bottleneck*. We cannot acquire the necessary use case knowledge that ought to be used in the software application easily as it is hidden in the heads of the use case experts. Mixed initiative user interaction should allow us to rely on both automatic and manual procedures for semantic mediation and the annotation of unstructured data.

3 Interactive Semantic Mediation

Interactive semantic mediation has two aspects: (1) the mediation with the processing back end, the so-called dynamic knowledge base layer (i.e., the heterogeneous data repositories), and (2) the mediation with the user (advanced user interface functionality for the adaptation to new industrial use case scenarios). We developed a semantic mediation component to mediate between the query interpretation created in the dialog shell and the semantic background services prevalent in the industry application context. The mediator can be used to collect answers from external information sources. Since we deal with ontology-based information in heterogeneous terminology, *ontology matching* has become one of the major requirements. The ontology matching task can be addressed by several string-based or structure-based techniques (see Euzenat and Shvaiko 2007). As a new contribution in the context of large-scale speech-based AI systems, we think of ontology matching as a dialog-based *interactive* mediation process (cognitive support frameworks for ontology mapping involve users). The overall approach is depicted in Fig. 1. A dialog-based approach can make more use of partial, unsure mappings. The user's expertise is included for difficult cases of semantic interoperability. Furthermore, our approach increases the usability in dialog scenarios where the primary task is different from the matching task itself (see many industrial usability requirements). So as not to annoy the user, he/she is presented only the difficult cases for disambiguation feedback; thus we use the dialog shell to basically confirm or reject preconsidered alignments (Sonntag 2008).

4 Interactive Knowledge Acquisition

We cannot acquire the necessary knowledge that ought to be used in the software application easily, as it is hidden in the heads of experts. In MEDICO, we provided an example of how an incremental knowledge acquisition process for radiology images can be implemented to solve this problem. Thereby, we integrated semantic web technologies with a variety of automatic and manual annotation tools for radiology images. According to the complex medical finding process, the different annotation tools should be used for very specific purposes. This divide-and-conquer strategy turns out to be very effective in the radiology domain.

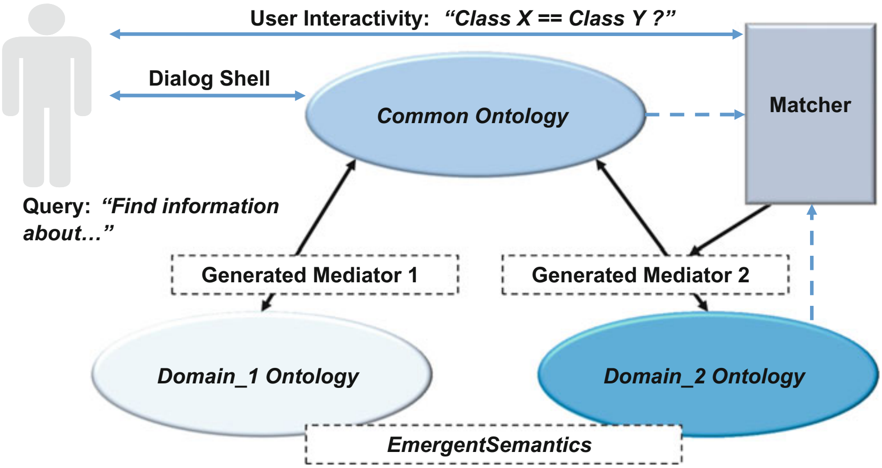


Fig. 1 Interactive semantic mediation for industry applications

The incremental knowledge acquisition process (Fig. 2), which is a universal, use case independent process, relies on the structured ontological knowledge of the particular domain (domain ontologies). Based on this prerequisite, we have been trying to formulate the process of automatic and manual data annotation. Hereby, two factors play a major role: the quality of automatic annotations and the usability of different intelligent user interfaces to control, correct, and add annotations.

For us, usability means, that people can use an Artificial Intelligence (AI) prototype easily and efficiently to accomplish their tasks. Prototypes that are usable enable domain experts to concentrate on their tasks rather than pay attention to the tools they use to perform their tasks. The prevalent interaction design issue that follows this definition is that the intelligent interfaces are:

- Efficient to use;
- Quick to recover from errors; and
- Visually pleasing.

To achieve all three of these, a careful selection of involved components for manual annotation is vital. In the following, we will exemplify this in the medical imaging domain of the MEDICO use case. It should be noted that all concepts and issues of the proposed incremental knowledge acquisition process can be easily transferred to any other use case domain that deals with unstructured information repositories such as texts and images.

This can be substantiated in MEDICO by the current developments in clinical practice, where *structured reporting* should be introduced. This means that the radiologists fill in special standardized forms. Radiologists feel restricted by these standardized forms and fear a decrease in focus and eye dwell time on the images (Jameson et al. 2009; Weiss and Langlotz 2008). As a result, the acceptance for

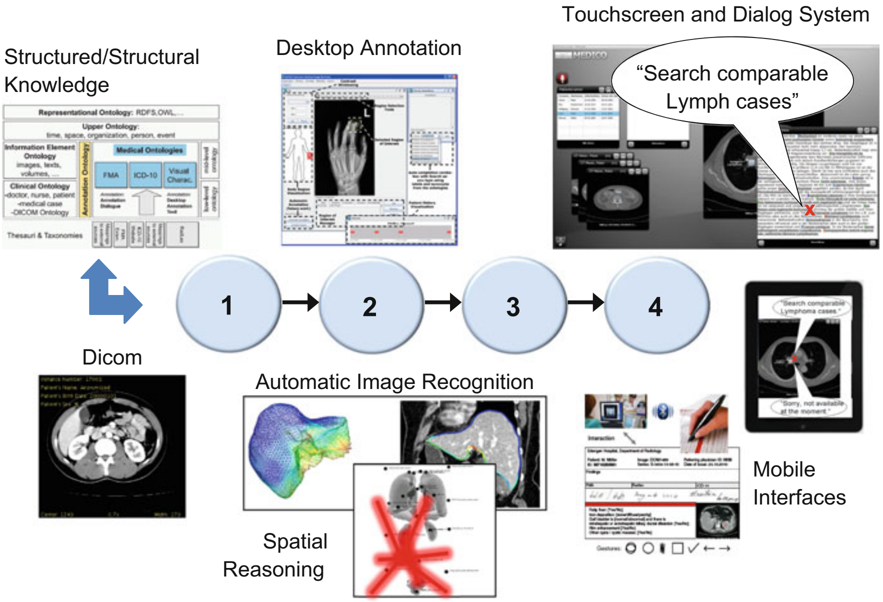


Fig. 2 Incremental knowledge acquisition process by example (MEDICO)

structured reporting is still low among radiologists, while referring physicians and hospital administrative staff are generally supportive of structured standardized reporting, since it can be used more easily for further processing. As a matter of fact, image semantics with RDF is a further step in this direction. These issues are explained in the context of industrial usability and our basic process steps for industrial dissemination. The incremental knowledge acquisition process (Fig. 2) has four steps:

First, the automatic metadata are extracted from the DICOM images and instantiated according to the structured/structural knowledge model. After that, a direct access to the RDF statements is possible while using the query language SPARQL.

Second, the automatic image recognition software runs over the images to produce anatomical annotations according to the structural knowledge model. According to the spatio-anatomical ontology, automatic spatial plausibility checks can be executed. Hereby, the spatial reasoning process runs completely automatically and only the outlier configurations are presented to the medical experts.

Third, the experts can then use the manual annotation tool (desktop-based) to correct or extend these configurations. At this stage, a very comprehensive set of image semantics, namely the study, patient, and low-level image feature information in combination with the automatically detected anatomical concepts and manual annotations with the desktop tool are available. These image model instances are not accurate enough for a treatment plan, but accurate enough to be used in a semantic search and annotation system, the dialog shell, which the senior radiologist can use. For this purpose, the image annotations are accurate enough. The point is that the automatic annotations need not be supervised or revised by any experts at this stage.

Fourth, only when the images are retrieved and considered for a medical treatment plan, can accurate disease annotations be added by the senior radiologist while using the dialog system which displays the image and patient data on a large touchscreen. It is even possible to search for similar disease annotations in other patients' contexts for a comparative study. In THESEUS, we also extended the high-level process of patient findings and image annotations to a mobile scenario, where we can use a special pen to recognize annotations on normal paper and/or used the iPad as a mobile dialog system and touchscreen device for the senior radiologist (also see the project RadSpeech[1]).

Our hope is that the resulting process successfully supports the complex health-care process in which radiology images are used. The development of automatic processing applications is as essential as the design and implementation of intelligent user interfaces for specific purposes. In our view, only such a combination will produce successful decision support systems for industrial dissemination of specialized applications in industry domains. Towards this goal, our interactive, semi-automatic approach should be connected with a standardized extensible framework for building search solutions to access unstructured information in an enterprise. SMILA[2] may provide such an architecture.

5 Usability of Collaborative Multimodality

Usability is quite simple to define; it means that people can use a product easily and efficiently to accomplish their tasks. Products that are usable enable workers to concentrate on their tasks and do "real work", rather than pay attention to the tools they use to perform their tasks.

As Jameson et al. (2009) point out, many research prototypes that use technically advanced but unimportant or unrealistic functionality for the specific domain or personal activities do not provide the AI support that users would appreciate most. This can make a complex speech dialog system languish as an infertile research prototype on demonstration computers which cannot be used in the context of industrial prototypes or real-world industrial dissemination. Accordingly, the binocular view of intelligent interfaces for industrial dissemination should study not only the suitability of a single algorithm and a component performance for a given user task, but also the industry user's interaction requirement in which the interaction will be used. For example in the MEDICO use case, our specific radiology case, the feature that only a senior radiologist is responsible for the

[1] http://www.dfki.de/RadSpeech/

[2] http://www.eclipse.org/smila/

treatment plan, implicates that his or her interaction with the annotation system must be designed to be very effective. Although it is widely reductive to put it this way, a senior radiologist has three main goals: (1) access the images and image (region) annotations (a summary can also be synthesized), (2) complete them, and (3) refine existing annotations. These tasks can best be fulfilled while using a multimodal dialog system. In contrast, less demanding manual annotation tasks, such as the correction of organ detection algorithms of image region selection, can be done by a first-year resident with the help of our desktop-based annotation tool. This tool can also easily be installed on virtually any computer in a hospital, whereas a speech dialog system requires a specific hardware infrastructure.

In order to come up with the different THESEUS testing scenarios, each of which has specific design and testing recommendations, we first extracted some substantial information from the different use cases in different user settings: we discerned between those who will use the system, where they will use the system, and what they will do with the system. After considering the results, we determined that the THESEUS research program works with seven different scenarios. We provide a decision tree that leads to specific recommendations for designing and testing with prototypes for each of the different scenarios and user settings (Sonntag et al. 2010).

Based on the introduction of a new scenario in the TEXO use case, which is based on car insurance services, we recently developed a browser-based demonstrator for multimodal mobile claims notifications. Typically, claims notifications adhere to a strict business process specified by the insurance provider. In our scenario (mobile device and data visualization), the insurance provider requires a policy holder to pass through the following process steps in order to successfully submit a claims notification (Fig. 3 shows the respective screenshots). All screens display a process view on top. It indicates the steps to follow during the claim notification process. After five minutes of free exploration time with the TEXO application (Fig. 3), and additional hints from the instructor, users had two attempts to successfully perform a complex task. This task includes a user authentication, the validation of contact information, the specification of the kind/cause of damage, the damaged car parts, and the location of the incident and images.

(1) This screen in the mobile claims notification application (mobile Schadenmeldung) allows the user to input and check personal information (Persoenliche Daten). He can be identified through an identification number (nPA). (2) The next screen "Damage Appraisal" (Schadenaufnahme) asks the users which type of claim they wish to make. The options are: accident, theft, vehicle damage without further property damage, vehicle damage with further property damage. Please note that "Damage Appraisal" (Schadenaufnahme) appears under the round blue icons and above the back button and helps show the user where he is in the process. (3) The user has selected the third option, vehicle damage without further property damage, and a list appears of additional required information. From top to bottom these are: cause of damage, damaged area(s), vehicle location, pictures of the vehicle, and synchronize data. (4) The user begins at the top and decides to give a cause of damage. The system requests, "Please describe the cause of damage." The user can now describe the damage via natural speech, and the dialog system extracts

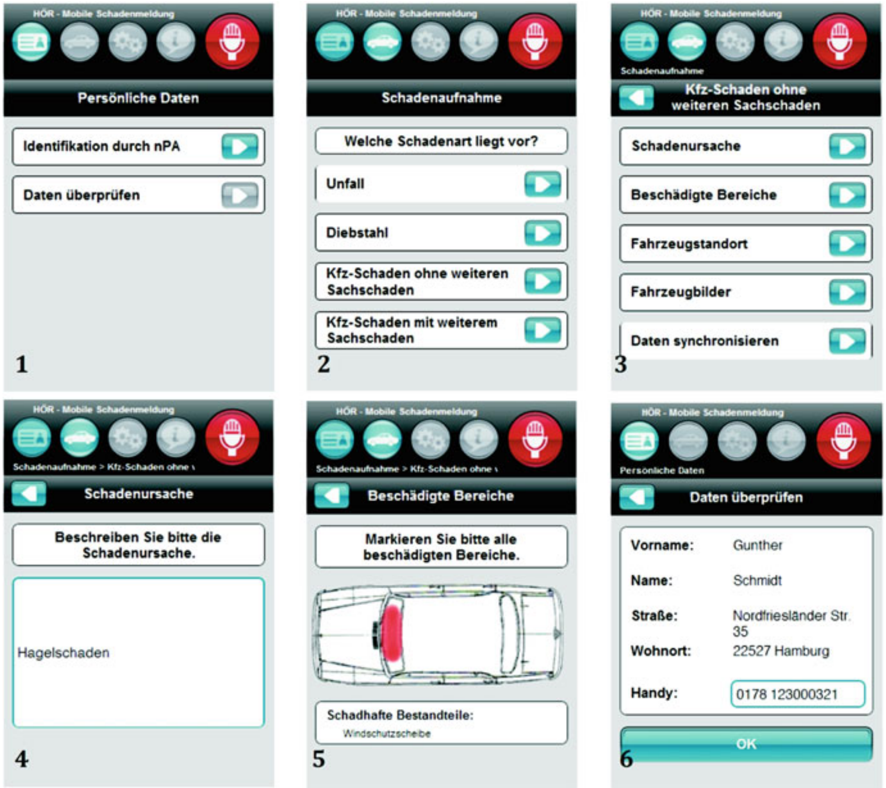


Fig. 3 TEXO sequence: authentication, validation of contact information, specification of the kind of damage, the cause of damage, damaged car parts, the location of the incident and images

"Damage by hail" from his utterance. The user clicks on the back button to return to the option screen and chooses the next option, damaged area(s). (5) In the screen for damaged areas, the system asks the user to please mark all damaged areas. He marks the windshield which appears on the screen in red. Underneath the picture of a vehicle, a list of damaged areas (Schadhafte Bestandteile) is created: windshield (Windschutzscheibe). (6) In this final screenshot, the user wants to verify his personal information. His first name (Gunther), last name (Schmidt), street (Nordfriedlaender Str. 35), city and zipcode (22527 Hamburg), and cell phone number (0178 123000321) are displayed. The user may click on the "OK" button to verify the information or select any of the categories to change them.

6 Conclusion

Multimodal interaction offers great opportunities when designing mobile human computer interfaces. In THESEUS, we implemented several collaborative user scenarios for data management purposes. In our view, collaborative interaction scenarios that combine mobile with terminal-based interaction accommodate the growing need to store, organize, and retrieve data. The usage of automatic detectors of unstructured content is seen as a great opportunity for dealing with the knowledge acquisition bottleneck – in combination with a clever manual annotation step according to our incremental procedure for semantic mediation and knowledge acquisition. Semantic interpretations of detector results and dialog utterances may become the key advancements in semantic search and dialog-based interaction for industrial applications, thereby mediating and addressing dynamic, business-relevant information sources. Future directions include the usage of reasoning procedures. For example, the goal of diagnostic quality control as envisioned in MEDICO is to use medical image content information for the automated staging of lymphoma patients, i.e., the knowledge about the number of enlarged lymph nodes indicated by automatic and/or manual annotation. This information relates to the patient degree and can be inferred automatically.

References

H. Bunt, R.J. Beun (eds.), *Cooperative Multimodal Communication*. Volume 2155 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2001), http://dblp.uni-trier.de/db/conf/cmc/cmc1998.html

P. Cimiano, P. Haase, J. Heizmann, Porting natural language interfaces between domains: an experimental user study with the ORAKEL system, in *Proceedings of the 12th International Conference on Intelligent User Interfaces (IUI '07)*, Honolulu (ACM, New York, 2007), pp. 180–189

R. Engel, SPIN: a semantic parser for spoken dialog systems, in *Proceedings of the 5th Slovenian and First International Language Technology Conference (IS-LTC '06)* (Ljubljana, Slovenia, 2006)

J. Euzenat, P. Shvaiko, *Ontology Matching* (Springer, Berlin/Heidelberg/New York, 2007)

T.S. Huang, A. Nijholt, M. Pantic, A. Pentland (eds.), *Artifical Intelligence for Human Computing*. Volume 4451 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2007)

A. Jameson, A. Spaulding, N. Yorke-Smith, Introduction to the special issue on usable AI. AI Mag. **30**(4), 11–15 (2009), http://www.aaai.org/ojs/index.php/aimagazine/article/view/2273

V. Lopez, M. Pasin, E. Motta, AquaLog: an ontology-portable question answering system for the semantic web, in *The Semantic Web: Research and Applications*, ed. by A. Gómez-Pérez, J. Euzenat. Volume 3532 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2005), pp. 546–562

T.M. Mitchell, *Machine Learning* (Mcgraw-Hill New York, 1997)

D. Porta, M. Deru, S. Bergweiler, G. Herzog, P. Poller, Building multimodal dialog user interfaces in the context of the internet of services, in *Towards the Internet of Services: The THESEUS*

Research Program, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

V.A. Ramachandran, I. Krishnamurthi, NLION: natural language interface for querying ONtologies, in *Proceedings of the 2nd Bangalore Annual Compute Conference (COMPUTE '09)*, Bangalore, ed. by R.K. Shyamasundar (ACM, New York, 2009), http://dblp.uni-trier.de/db/conf/compute/compute2009.html#RamachandranK09

S. Russel, P. Norvig, *Artificial Intelligence: A Modern Approach*, vol. 2 (Pearson, Upper Saddle River, 2003)

D. Sonntag, Towards dialogue-based interactive semantic mediation in the medical domain, in *Proceedings of the Third International Workshop on Ontology Matching (OM) collocated with the 7th International Semantic Web Conference (ISWC '08)*, Karlsruhe (2008)

D. Sonntag, *Ontologies and Adaptivity in Dialogue for Question Answering*. Volume 4 of Studies on the Semantic Web (AKA and IOS Press, Heidelberg, 2010)

D. Sonntag, Collaborative Multimodality. KI – Künstliche Intelligenz **26**(2), 161–168 (2012), http://dblp.uni-trier.de/db/journals/ki/ki26.html#Sonntag12

D. Sonntag, C. Weihrauch, O. Jacobs, D. Porta, THESEUS CTC-WP4 usability guidelines for use case applications, Technical report, DFKI GmbH, BMWi, April 2010, http://www.dfki.de/web/forschung/publikationen?pubid=4788

W. Wahlster, SmartWeb: Ein multimodales Dialogsystem für das semantische Web, in *Informatikforschung in Deutschland*, ed. by B. Reuse, R. Vollmar (Springer, Berlin/Heidelberg/New York, 2008), pp. 300–311

D.L. Weiss, C.P. Langlotz, Structured reporting: patient care enhancement or productivity nightmare? Radiology **249**, 739–747 (2008)

SemaVis: A New Approach for Visualizing Semantic Information

Kawa Nazemi, Matthias Breyer, Dirk Burkhardt, Christian Stab, and Jörn Kohlhammer

Abstract Information is an indispensable resource today. Access to and interaction with information play more and more a key role, whereas the amount of accessible information increases. Semantic technologies provide new solutions to structure this important property. One promising way to access the complex semantic structures and the huge amount of data is visualization. Today's Semantic Visualization systems offer primarily proprietary solutions for predefined and known users and usage scenarios. The adaptation to other scenarios and users is often cost- and time-consuming. This article presents a novel model for a fully adaptable and adaptive Semantics Visualization framework. Starting with the introduction of a new visualization model, the implementation of this model will be described. The article concludes with selected advantages of the described visualization technology.

1 Introduction

Information constitutes a crucial resource of enterprises. An efficient way for accessing the "enterprise information", e.g. personnel expertise, contact persons or projects and resources, is important to reduce costs and time. Besides the important role of information in several industrial and knowledge working areas, the everyday knowledge adoption process is mainly based on information acquisition from heterogeneous computer-based multimedia systems.

Semantic technologies provide different methods for annotating the underlying information with an appropriate "meaning". With the characteristics of building sentences in natural languages by subject, predicate and object, information entities

K. Nazemi (✉) • M. Breyer • D. Burkhardt • C. Stab • J. Kohlhammer
Fraunhofer Institute for Computer Graphics Research (IGD), Darmstadt, Germany
e-mail: kawa.nazemi@igd.fraunhofer.de; matthias.breyer@igd.fraunhofer.de; dirk.burkhardt@igd.fraunhofer.de; christian.stab@igd.fraunhofer.de; joern.kohlhammer@igd.fraunhofer.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_15,

are set in relation to other information, attributes or resources. Furthermore, the information entities are abstracted in appropriate concepts and categories respectively. The semantic structure allows searching, finding and gathering the requested information in an efficient way, whereas the result processing can be enhanced with rule-based and machine-learning methods to provide more insights into the given domain of information. The existing semantic technologies are able to provide the answer of known questions in an efficient way, provided the questions and the goals of the given search are known. But there are several usage scenarios and cases, where searching has more an exploratory character, and the formulation of the question or the search term is not known explicitly. Further semantic technologies provide us with their underlying structure to capture a whole information or knowledge domain and gather information about the context of the information.

The described cases are only examples for the exigency of investigating a complementary discipline for acquiring knowledge by considering and exploiting the semantic structure of the given knowledge domain, namely *Information Visualization*. Different research groups and institutions have worked on accomplishing visualization with the semantic structures. Herewith different terms have arisen for the merged technologies, e.g. Ontology Visualization or Semantic Visualization. Since these technologies are mostly focusing on visualizing semantic structures, the term Semantics Visualization is appropriate in our opinion.

The THESEUS CORE TECHNOLOGY CLUSTER *Innovative User Interfaces and Visualizations* (CTC WP5) investigated techniques and concepts to visualize semantic data for heterogeneous users, usage scenarios and semantic characteristics. The challenge of the research work was the conceptualization and development of a technology that is user-oriented on the one hand and features the characteristics of a core technology on the other hand. Therefore the adaptable and adaptive SemaVis Technology was developed, which covers both the features of a user, use case and context adaptable core technology and a user-centered and adaptive Semantics Visualization.

In this article the *SemaVis* technology will be described, starting with the abstract model of visualization, which subdivides the visualization of semantic data in three abstracted layers. In the following, the implementation of the described model as a framework will be introduced. The article concludes with some advances of the *SemaVis* technology and an application example.

2 The Model of Semantics Visualization

Semantic technologies process data to enrich them with meaning and structure and provide information that can be further processed by machines for a better understanding of the underlying domain. With this enrichment not only is the search process more efficient, the acquisition of implicit information is provided in a simple way. Semantic transitions allow us by using various techniques to get information about data relationships that are not explicitly modeled. The transitivity of the relations assumes knowledge about a domain and has to be formalized in a machine readable way. The efficiency of the semantic search and information acquisition is constrained by both the knowledge of the user (he knows what he is searching for) and the formalization of the semantics. But not in every case does the user really know what the main target of his search is, and which question has to be answered. *Semantics Visualizations* provide the ability to explore in a semantic space and acquire information in answers and questions, because "There is nothing better than a picture for making you think of questions you had forgotten to ask (even mentally)" (Tukey and Tukey 1985).

To explore information and provide the ability of verbalizing questions was one of the main tasks given by the heterogeneous use cases. The challenge was to develop user-centered visualizations as core technologies to support the heterogeneous users, usage scenarios and especially the use cases in the THESEUS research program. To achieve this goal, a fully adaptable visualization technology was necessary to meet the requirements of a core technology. Therefore the whole process of the information visualization was investigated and the visualizations and the main user interface were subdivided into different layers of abstraction. Visualizations can be described according to the following characteristics: *what is displayed*, *where is it displayed*, and *how is it displayed*?

To ensure a fine-granular adaptation of the whole visualization process, the model for *Semantics Visualization* has to adapt each piece of the given visual information separately. Therefore we separated the visual information pipeline into three different layers of abstraction (Nazemi et al. 2011b). The basic concept of our model was the reference model for information visualization (Card et al. 1999), which processes the data in three sequential but iterative steps of *Data Transformation*, *Visual Mappings* and *View Transformations* from raw data to visualization. According to this model, we defined the visualization model as a transformation pipeline for semantic data with *Semantics*, *Layout* and *Presentation* (Nazemi et al. 2011b).

The main goal of this model is to provide in each level of abstraction the ability to adapt the visual appearance and provide a user-, use case-, and task-centered visualization model implementable as core-technology. The main difference between this model and existing models is the fact that human perception and the way humans process visual information played a key role in the design of the model. We separated the visual information based on the findings in the area of human visual processing (Treisman 1985, 1986; Wolfe 2007). The *Semantics* layer in our model considers the amount and complexity of data to be visualized and transforms the required data to geometric structures. These geometries are analyzed in the *layout* layer. The visual placement (on the plane) (Bertin 1983) and the closeness to other visual entities are defined in this layer. For each of the defined visual placement or layout algorithms a visual appearance is defined in the *presentation* layer, the highest level of the model. Then each of these layers can be adapted according to the requirements of the given scenario or users. Figure 1 displays the abstracted model of *SemaVis* according to Nazemi et al. (2011b).

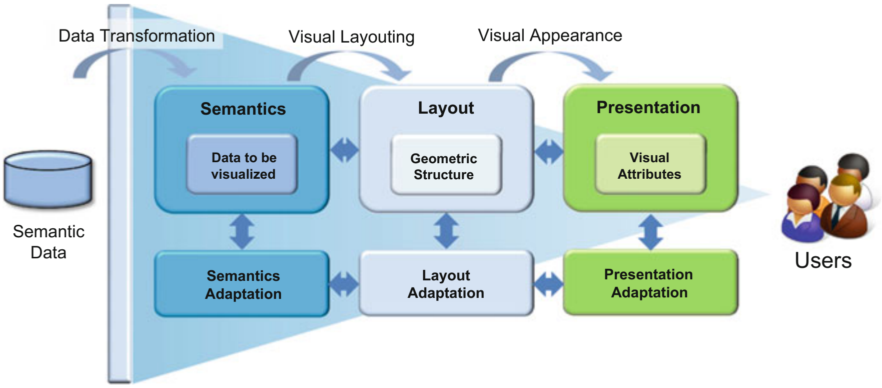


Fig. 1 The abstracted model of *SemaVis*: *Semantics Visualization*

The generation of visualization is processed as follows: First the *Semantics* layer analyzes the semantic structure, properties and data. Thereafter appropriate layout algorithms are chosen for the data or for certain parts of the data. Finally the chosen layouts are visualized by the *presentation* layer. Each of these layers can be manipulated by a simple visualization modeling language and can be adapted to several users, usage scenarios or corporate designs. The main user interface, which may include more than one visualization type, is generated using the same model.

Semantics – Modeling Data for Visualization Semantics defines which data is visualized. It contains information about the data (what is the data about), its structure (e.g. hierarchy, incoming and outgoing relations) and implicit and explicit properties. The main task of this layer is to parameterize the data and provide the opportunity to adapt visualization in the lowest level. To visualize semantic data, the formalized network structure of semantic entities, attributes and information have to be transformed in a visualizable model. Additionally, the terminological schema level has to be considered in the visualization if semantics formalisms are given. The semantics formalisms intersect the data artifacts to be modeled for visualization. Therefore *SemaVis* applied them as abstracted conceptualizations in the *Semantics* layer, covering semantic networks, frame-based logics, and description logics (Fensel et al. 2003; Hitzler et al. 2008). The semantics data model in *SemaVis* is compliant with *lightweight semantics*, consisting of concepts, concept taxonomies, relationships or roles between concepts, and properties describing concepts (Gómez-Pérez et al. 2004). The *Semantics* layer models formal axioms, functions, rules, procedures and constraints to model heavyweight formal semantics formalisms (Gómez-Pérez et al. 2004) for visualization. This enables us to visualize the full scope of formal semantics with SemaVis.

Layout – Structuring Visual Information The *Layout* layer is responsible for structuring and transforming the semantic information to produce a geometric structure of the data for visualization. The semantic entities identified and extracted in the *Semantics* layer are systematically analyzed and extended with layout-specific (e.g. spatial) information. *Layout* defines where the semantics will be visualized. Several graphical metaphors, e.g. graph-layout algorithms or space-filling visualizations are chosen, based on the data characteristics derived in the previous step. The analyzed data structures are weighted and classified for visualization. Further adequate layout and placement algorithms (e.g. hierarchical, matrix, sequential of organic layout structure) represent the semantic structure in this layer. Additionally, further impact factors like users' preferences and current tasks are considered in the layout process step for deriving an appropriate geometric representation of the semantic information. It should be noted that only the geometrical layout is defined on this layer, but not yet the visual appearance.

Presentation – Representing Semantics with Visual Attributes *Presentation* is the highest level of the *SemaVis* model and defines the visual appearance of the geometric data model. Based on the outcome of the layout layer, where the positioning of the visualization entities are computed, this layer adopts the most important visual attributes to describe the entities visually. The visual attributes used in this layer are based on the pioneering work of Bertin (1983) enhanced by findings of human visual processing abilities (Treisman 1985; Wolfe 2007). The visual attributes or features adopted in *Presentation* are based on the pre-attentive information processing. Studies on human perception could show, that several visual features drive our attention within a very short time (below 250 ms) to a target (target detection) (Treisman 1985, 1986; Wolfe 2007). The *SemaVis Presentation* layer implements the most valuable visual features (e.g. color, size, texture, shape, orientation, position, transparency) to visualize the geometric model and support users in their heterogeneous tasks. Like the other layer of the model, the *Presentation* layer is adaptable to different scenarios and supports the idea of a core technology for visualizations.

3 The Model of Semantics Visualization as Framework

The above described model for *Semantics Visualization* is a theoretical concept of the *SemaVis* framework. The following section gives an overview about the system architecture and the modular implementation of the model.

System-Architecture The foundation of the *SemaVis* framework is a modular and expandable system architecture that includes three different layers, (1) data layer, (2) logic layer and (3) model layer (see Fig. 2). The data layer contains on the one hand the semantics information that is loaded and generalized into an internal data representation for visualization and on the other hand script-based files that allow the configuration of the framework for certain application scenarios. In the logic layer the data and the configuration files based on the *Semantics Visualization Markup Language* (SVML) are processed and interpreted for further

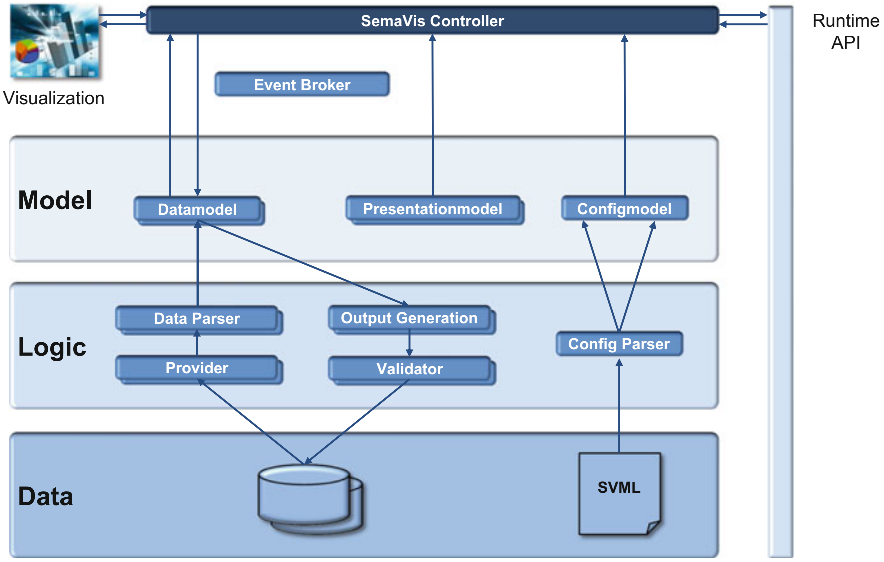


Fig. 2 System architecture of the *SemaVis* framework with *arrows* representing the information flow between modules

utilization in the *Semantics Visualization* model. The model layer includes the data model that stores semantic data and offers certain data access methods that provide uniform access to semantics information. Additionally the model layer contains the configuration model holding configuration data for certain scenarios and the *Presentation* layer as a model.

The data model is responsible for visualizing specific data management of semantic knowledge. It provides a generalization for different semantic data types and thus allows uniform access for the semantics visualizations to the semantic structures. The data provider is responsible for gathering data from a knowledge repository and transfers the semantics information to the data parser, which converts the data into the internal representation in the data model. If data entities are changed or edited in the visualization, the output generator and the validator transfer the changes to the underlying knowledge repository. The *Presentation* model provides a unification of visual properties (e.g. color, size, etc.) initially loaded from the configuration file. The event broker constitutes the central hub for each event that controls the state of the visualization and observes interaction events. All described modules are initialized and managed by the SemaVis-Controller, which provides all needed models to the visualization components and manages the central processes in the framework and the in- and outgoing commands from the runtime application programming interface (API).

Modular Integration of Visualization Components The depicted processing steps from the system architecture represent logical modules with a specific task.

These modules contain the logic to process and transform the input data to an output. To control the processing within the modules, the system can be configured statically by a configuration file and dynamically by using the system implemented API. Moreover, we define *modules* and *sub-modules*. Modules represent the essential processing steps from data to the final graphical user interface (GUI). Each module acts as a standalone application with just the defined input parameters and outputs as result. Furthermore, every processed result of a module will be transmitted through each following module. This allows, for instance, the consideration of extracted data characteristics from the *Semantics* layer to the *Presentation* layer to set the final visual features. This modular architecture allows the flexible replacement and extension of modules with additional functionality without affecting the general processing pipeline. Due to the standalone conceptualization, it is not required to perform always the whole visualization pipeline. A change needs just the replacement or reconfiguration of the affected module. Among others, a very dynamic influence factor during the visualization process is the interaction of users, which can require a couple of analyses to ensure an efficient user support to extract the semantics and intension of the interaction. For these steps we defined additional *sub-modules*, which deal with just a single task and are strongly related to just one main module. In contrast to modules, they are not formally described; thus the heterogenous usage has a supporting character.

Data Integration *SemaVis* is intended and designed as a core-technology, thus *SemaVis* is neither domain dependent nor developed for a specific user group. Particularly important is the loosely coupled software design paradigm pursued for the data integration layer, which is responsible for the communication with external data and ontology management systems. For integrating *SemaVis* into various use case scenarios this layer can be adapted, extended, and configured by framework extensions or by settings in the SVML-Markup configuration.

SemaVis supports the integration of different data sources, offering different data formats simultaneously. Further, loading-on-demand, editing, and annotation functionalities are supported. Even gathering data from a specific data source in a given data format and storing changes or annotations to another data management system in a different data format is supported due to the internal representation of the data in the *Semantics* layer. To provide these scopes and functionalities, the data integration concept differentiates five responsibilities: *data provider*, *data parser*, *data model*, *output generator* and *validator*. The data provider implements the connection between *SemaVis* and the data or ontology management system. The most prominent connections like *file-based*, *SparQL* or *Web Services* offered by most of the existing management systems are implemented. Even proprietary connections can be easily implemented with modular characteristics of *SemaVis*. The data parser converts and transforms the provided data and inserts the information into the internal data model. Due to the announcement of the Semantic Web community and standardization formalisms like the *Resource Description Framework* (RDF), the accordant schema *RDFS* and *Web Ontology Language* (OWL) are predominant and thus supported. Additionally *SemaVis* supports graph languages like GraphML or the technically lower syntactic levels (e.g. *XML* or *JSON*).

Gathered data is passed to the internal data model which allocates and manages it. Other modules can query specific data artifacts to the model. Further, semantic analysis algorithms of the *Semantics* layer identify semantics which may be relevant for the user or the current task. This requires methods for retrieving special artifacts of the formal semantics, which are managed and provided within the internal data model. For transmitting edited and annotated information, the output generator preprocesses this data to the sinks formats. Preconditioned, the data or ontology management system supports data evolution and the changes can be inserted into the original database. Otherwise a different management system can be triggered to store it.

To guarantee only feasible changes, the validator verifies the syntactic correctness. In most application scenarios the validator is closely coupled with the verification mechanisms of the back end system. This computation distribution allows better performance results, especially if a large amount of semantic data has to be investigated in the verification process. Due to this fact user demand for rapid feedback as part of the verification process is performed on the client side. So, the user can continue her workflow of editing or annotating the semantics data while the system checks iteratively the integrity of the changes.

4 Advances in Semantics Visualization

Reducing Complexity with Complementary Visualizations Semantically annotated data contains heterogenous and complementary attributes for describing the knowledge domain. Existing visualizations cover either one of these attributes or visualize the whole spectrum. Both variations have advantages and disadvantages. While visualizing the whole semantic spectrum is often complex to understand and not always necessary, the reduction to certain semantic characteristics might reduce the information content. Therefore it is necessary to provide a solution where the different semantic characteristics can be visualized in an easy and understandable way.

SemaVis implements a multiple-visualization user interface as a visualization cockpit (Nazemi et al. 2010a). The visualization cockpit separates the information units from each other and visualizes this information in separate visualization units. The advantage of the separation of complex information units is obvious: The user of a cockpit is able to perceive the required information very fast and react to the perceived information very fast. *SemaVis* has transformed this idea for visualizing semantics. A set of different visualization techniques was developed; each visualization focuses on a certain characteristic of semantic data, e.g. a spatial space-filling visualization was developed to visualize the concept hierarchies (Nazemi et al. 2009), graph visualization methods are used for visualizing relation, and timeline visualizations are used for visualizing explicit properties

(Stab et al. 2010). The visualization cockpit metaphor is more than the combination of multiple visualizations. It provides the separation of the semantic information in several modes:

1. Aspect-oriented mode: visualizes the same data with different visualizations. The user is able to combine visualizations to see the same information from different perspectives. The visualizations are linked with each other. Every interaction with one of the visualizations changes the view in the combined one.
2. Visual Comparative mode: In this mode the visualizations are not necessarily linked with each other. The user is able to compare certain parts of the same semantics data by using different visualization techniques.
3. Semantics Comparative mode: In this mode the same visualization combinations are used for two different semantics datasets. The user is able to compare two different semantics databases with the same visualizations.
4. Level-of-Detail mode: In this mode the same visualization technique is used twice on the screen, whereas the presentation layer differs. The user is able to zoom in (semantically or visually) one of the visualizations and keep the overview in the other one.

Adaptive Semantics Visualization The layer-based adaptability of *SemaVis* provides an adequate framework for generating heterogeneous user interfaces by manipulating the entire spectrum of the visualization characteristics. This ability is used to further provide an automatic adaptation to several impact factors, e.g. user interaction or data structure. The goal of the automatic adaptation of the visualization is to support the users in their navigation process during the exploration of information.

Therefore a new approach was conceptualized, where the impact factors are captured and enriched with semantics with a direct link to the semantics of the data. A user interaction, for instance, is determined as a three-dimensional conceptualization of the interaction taxonomy. Thereby the semantic taxonomy of the data is used and enriched with a hierarchy of the interaction device and a hierarchical representation of the visualization. These three dimensions, the creation of unique interaction semantics, provide a meaning of the users' natural consequence of the visualization operation (Nazemi et al. 2011b). Further, the characteristics of the data, especially the semantic structure and the explicit and implicit properties, are considered in a similar way. Therefore a data semantics is generated that provides information about the current visible data (Nazemi et al. 2011a). The implicit gathered information is used to personalize both the different layers of the visualizations and the main user interface. While the different layers of the visualization aim to support the navigation and exploration process of the user, the main user interfaces recommend several visualizations for a more efficient way of acquiring information. The transformation of the determined "context semantics" is orchestrated by a probabilistic algorithm (Nazemi et al. 2010b) to both the different graphical layers for adapting the visualization characteristics and the main user interface for recommending a set of adequate visualizations. Therefore both aspects are considered, the current context (given data and recognized activity)

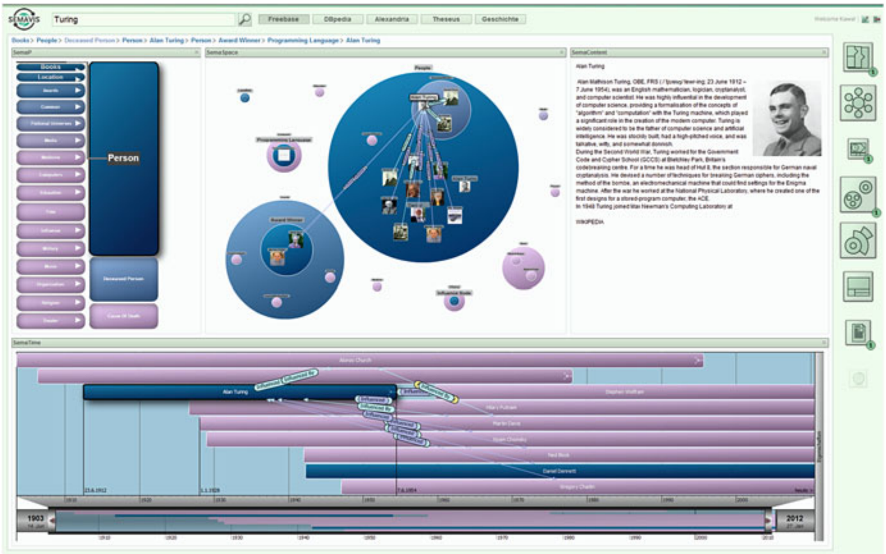


Fig. 3 User-centered adaptation of *SemaVis* for the search term *Turing*

of usage and the preference of the user. We try to comprehend the exploratory adaptation with the following example: A user wants to know which persons were influenced by the work of the computer scientist *Turing*. He searches in the *SemaVis* search interface for the term “Turing” and enters the exploration space. Because the term *Turing* appears in different knowledge categories, the main user interface recommends a spatial visualization for the concept hierarchy; hereby the concept uses the presentation layer for highlighting concepts of interest.

With the cockpit metaphor, a composition of different visualizations is automatically created, which shows the relevant information for the user separated in several views. A further visualization pictures the relations of the search terms to other data entities, whereas the relation of interest is highlighted and the other relations are put in the background using the transparency feature of the presentation layer. Since the work of Turing influenced persons even after his lifetime, a further visualization is chosen, which visualizes explicitly the time and semantic relationships. The main display further recommends several visualizations with probabilistic quantifiers for the given exploratory search issue. The user is able to choose other visualizations recommended by the systems or reduce the information space by removing visualizations from the main user interface. Figure 3 displays the described adaptation process for the term *Turing*.

5 Conclusion

Semantic technologies provide a promising way to access information and complex data structures. Today's technologies and research focus on the "machine readability", whereas the human factor is often neglected. *Semantics Visualizations* are promising approaches for providing human-centered solutions for interacting with semantic data. This article presented the *SemaVis* technology based on a novel model for a fully adaptable and adaptive Semantics Visualization. The model subdivides the visualization pipeline in three visual characteristics and provides consequently a fine-granular adaptation to heterogeneous users and usage scenarios. Further, the implementation of the described model as a modular framework technology was described. The modularity allows integrating not only visualizations but also recommendation and analysis algorithms into the framework. The advantages of the framework were described with an example of exploratory search.

References

J. Bertin, *Semiology of Graphics* (University of Wisconsin Press, Madison, 1983)

S.K. Card, J.D. Mackinlay, B. Shneiderman, *Readings in Information Visualization – Using Vision to Think* (Academic, San Diego, 1999)

D. Fensel, J.A. Hendler, H. Lieberman, W. Wahlster (eds.), *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential* (MIT, Cambridge, 2003)

A. Gómez-Pérez, M. Fernández-López, O. Corcho, *Ontological Engineering* (Springer, Berlin/Heidelberg/New York, 2004)

P. Hitzler, M. Krötzsch, S. Rudolph, Y. Sure, *Semantic Web - Grundlagen*, eXamen.press (Springer, Berlin/Heidelberg/New York, 2008)

K. Nazemi, M. Breyer, C. Hornung, SeMap: a concept for the visualization of semantics as maps, in *Universal Access in Human-Computer Interaction. Applications and Services*. ed. by C. Stephanidis. Volume 5616 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2009), pp. 83–91, http://dblp.uni-trier.de/db/conf/hci/hci2009-7.html#NazemiBH09

K. Nazemi, M. Breyer, D. Burkhardt, D.W. Fellner, Visualization cockpit: orchestration of multiple visualizations for knowledge-exploration. Int. J. Adv. Corp. Learn. (IJAC) **3**(4), 26–34 (2010a), http://dblp.uni-trier.de/db/journals/i-jac/i-jac3.html#NazemiBBF10

K. Nazemi, C. Stab, D.W. Fellner, Interaction analysis for adaptive user interfaces, in *Advanced Intelligent Computing Theories and Applications*, ed. by D.S. Huang, Z. Zhao, V. Bevilacqua, J.C. Figueroa. Volume 6215 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2010b), pp. 362–371

K. Nazemi, D. Burkhardt, A. Praetorius, M. Breyer, A. Kuijper, Adapting user interfaces by analyzing data characteristics for determining adequate visualizations, in *Human Centered Design*, ed. by M. Kurosu. Volume 6776 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011a), pp. 566–575, http://dblp.uni-trier.de/db/conf/hci/hci2011-16.html#NazemiBPBK11

K. Nazemi, C. Stab, A. Kuijper, A reference model for adaptive visualization systems, in *Human-Computer Interaction: Design and Development Approaches*, ed. by J.A. Jacko. Volume 6761 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011b), pp. 480–489

C. Stab, K. Nazemi, D.W. Fellner, SemaTime – timeline visualization of time-dependent relations and semantics, in *Advances in Visual Computing*, ed. by G. Bebis, R.D. Boyle, B. Parvin, D. Koracin, R. Chung, R.I. Hammoud, M. Hussain, K.H. Tan, R. Crawfis, D. Thalmann, D. Kao, L. Avila. Volume 6455 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2010), pp. 514–523, http://dblp.uni-trier.de/db/conf/isvc/isvc2010-3.html#StabNF10

A.M. Treisman, Preattentive processing in vision. Comput. Vis. Graph. Image Process. **31**(2), 156–177 (1985), http://dblp.uni-trier.de/db/journals/cvgip/cvgip31.html#Treisman85

A.M. Treisman, Features and objects in visual processing. Sci. Am. **255**(5), 114–125 (1986)

J.W. Tukey, P.A. Tukey, Computer graphics and exploratory data analysis: an introduction, in *Proceedings of the Sixth Annual Conference and Exposition: Computer Graphics ’85*, vol. 3 (Dallas, Texas, 1985), pp. 773–785

J.M. Wolfe, Guided search 4.0: current progress with a model of visual search, in *Integrated Models of Cognitive Systems*, ed. by W.D. Gray (Oxford University Press, New York, 2007), pp. 99–119

Further Reading

I. Torre, Adaptive systems in the era of the semantic and social web, a survey. User Model. User-Adapt. Interact. **19**(5):433–486 (2009)

From Raw Data to Rich Visualization: Combining Visual Search with Data Analysis

Thorsten May, Kawa Nazemi, and Jörn Kohlhammer

Abstract Visual analytics is an interdisciplinary field of research at the boundary between data mining, statistics and visualization. Patterns and relations in the data complement a semantic representation of knowledge on a lower level of abstraction. One important goal of visual analytics is to find relations hidden in vast amounts of data, which can be turned into useful knowledge. Analysis needs to be "visual", because human's visual cognitive abilities are important for the identification and refinement of the analytical process. Further the results of the analysis have to be presented in a way to match the user's perspective on the proposed task. However, typical users are not experts in statistics or data mining. The challenge of visual analytics is to keep domain experts in charge of the analytical process while reducing the workload due to the complexity of the techniques. While search and analysis usually are mentioned in different contexts, they are highly interdependent processes. In fact, every exploratory analysis is a search for new knowledge. In turn, this knowledge can be used to refine future searches by introducing new concepts or relations to draw from. This article will show how automated and visual methods can be combined to connect knowledge artifacts on multiple levels of abstraction.

1 Introduction

Vast amounts of information are available in today's digital data repositories. As the available information provides a good opportunity to face the needs in our society, the benefits of existing information management systems are not fully exploited. One promising way to face the information overload problem is by the embracement of the analytical and visual approaches (May 2011). The interdisciplinary field

T. May (✉) • K. Nazemi • J. Kohlhammer
Fraunhofer Institute for Computer Graphics Research (IGD), Darmstadt, Germany
e-mail: thorsten.may@igd.fraunhofer.de; kawa.nazemi@igd.fraunhofer.de; joern.kohlhammer@igd.fraunhofer.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_16,

incorporating these two areas is visual analytics (Keim et al. 2010). The ultimate goal of visual analytics is to provide the best possible way for gathering knowledge from complex and mass data using visual methods. The visual representation of hyper-dimensional data enables solutions for heterogeneous visual tasks. Especially, the visual analysis and decision-making tasks are supported with visual analytics methods in a suitable way. In many application areas, success depends on the availability of *right information*, *at the right time*, and presented *in the right way*. Visual analytics designs and develops techniques to meet these requirements. In many important decisions, the connections between cause effects and goals are not trivial. In order to make choices, we refer to our own experience and knowledge, but we also want to use of other available sources of knowledge. In fact, this is the reason why huge amounts of data have been collected in the past decades in every field of science or business.

However, *data* is not knowledge; the acquisition of data is no longer the driving problem. Drowning in data, we need to separate the relevant from the irrelevant for every new application or problem statement. Visual analytics enables people to find hidden relations and to turn the data into useful and defensible knowledge. The goal of research is the creation of techniques to discover knowledge and to foster its constant revision and refinement. In the end, this knowledge grounds future decisions on sound empirical arguments.

Furthermore, visual analytics is about making information available at the right time. Distilling the right information is a long-term goal based on historical data and experiences. In comparison, making information available is a short term goal which consists of applying existing knowledge to the current situation. Because the amount of data to be considered may still be overwhelming even after the selection of relevant data, this process needs to be supported by methods, which automatically transform the data into useful information.

Finally, visual analytics is about presenting information in the right way: Even the simplest information may be difficult to understand if the way of presentation is inadequate. In the past, research has focused on faithful visual representation of data. In contrast, recent research increasingly focuses on user-dependent variables like role, task and mental models. Evaluation of visualization techniques is still a challenge, because the user's prior knowledge remains an important uncontrolled variable and it is still impossible to measure "understanding" or "knowledge gained" objectively. Understanding user-dependent variables allows us to adapt the presentation and to seamlessly integrate it into the workflow of the user. One main research topic of the THESEUS *Core Technology for Innovative User Interfaces and Visualizations* was the investigation of analytical and visualization methods. The main goal was to provide an enhanced model for coupling methods from both areas and therewith support the main idea and goals of visual analytics (Keim et al. 2010).

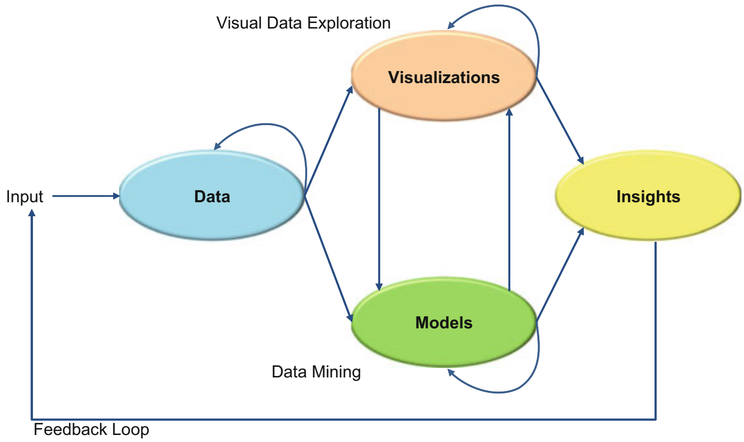


Fig. 1 The analytical reasoning process (Adapted from Thomas and Cook (2005))

2 Visual Analytics and the Analytical Reasoning Process

Thomas and Cook (2005) define visual analytics as the "Science of analytical reasoning supported by interactive, visual interfaces". Analytical reasoning is a process which connects raw data with applicable knowledge via a number of steps of abstraction and refinement. Thomas and Cook call these steps *analytical artifacts* and distinguish four different levels in the analytical reasoning process (see Fig. 1). We present the steps to locate our technological contribution in the subsequent sections of this article.

These steps have been established for data analysis, but in turn they are relevant for the design and refinement of (visual) search as well. One end represents all these sources of data and information of potential relevance to the human user. The other end represents the human user, his tasks and mental models used to solve this task. Most likely, there is a mismatch between these representations, which means that the data representation does not fit the way the user works with the data. In a typical search scenario, it is likely that the amount of data might be impossible to handle for a user. Just as often, there might be a gap between the intention of the user – expressed as a query – and the result of the search. Hence, on the grand scale, the analytical reasoning process establishes a connection between human understanding and data representation. However, there is not – and there will never be – a one-size-fits-all recipe of how to establish this connection. To dissect the process, Thomas and Cook (2005) distinguish the following artifacts:

- *Elemental Artifacts*, commonly referred to as *data*, represent a number of virtually isolated facts or observations which are considered relevant for the question at hand. In the most general perspective to analysis, the type of data is not prescribed – it can be structured or unstructured, textual, numerical or hybrid data.

- *Patterns*, the first step of refinement, is the aggregation of facts to meaningful and useful sets, which may be later refined to build concepts. Patterns are the "currency of analysis": Single isolated facts do not allow for defensible generalizations to create sound arguments and decisions. Finding patterns in the data is the domain of data mining and information visualization research. However, a formal definition of what a pattern actually distinguishes from noise cannot be stated. It remains the responsibility of the human analyst to choose methods which guarantee the relevance of patterns.
- *Concepts*, *Models*, *Arguments*, which are introduced as higher-order artifacts. They represent formal descriptions and relations in mathematical language or even close to natural language. Taxonomies and Ontologies are examples of these artifacts. Inference is a typical example of operations on these artifacts.
- *Hypothesis and Stories* are those higher-order artifacts which most closely match the analytical question at hand.

At the first glance, it appears that the analytical reasoning process is a chain of transformations to turn data into understandable knowledge. Yet, this is only half of the truth. The second important goal is the development and verification of these transformations, which also requires human decisions: The selection of data, the choice of analytical methods and models from a large zoo of techniques and the parametrization of these methods are not trivial. Yet, all of these may have severe effects on the outcome of the analysis (May 2011). When searching for new knowledge, we need to bear in mind that the analyst did not make the best choices a priori. Analysis is an iterative, non-linear process and it requires the constant comparison, refinement and possibly the revision of these decisions (Keim et al. 2008).

Keeping the "human in the loop" of the analytical process is a major claim of visual analytics research. *Analysis must not be alchemy*; professionals in this field – like professionals in any other business – can be expected to defend their choice of methods and parameters – especially if the results turn out to be useless or even harmful. The problem of dealing with purely automated methods is not their chance of failing. The problem of automated analysis is – how do we know, if we fail?

3 Distilling the Right Information: Combining Strengths of Man and Machine

Why *visualize* analysis? The obvious strengths of automated techniques are the abilities to store, process and retrieve virtually infinite amounts of data. A more subtle nonetheless important advantage of automated methods is that their results are deterministic, repeatable and comparable. This is an indispensable requirement to all scientific and business applications. Traditionally, data analysis has been the domain

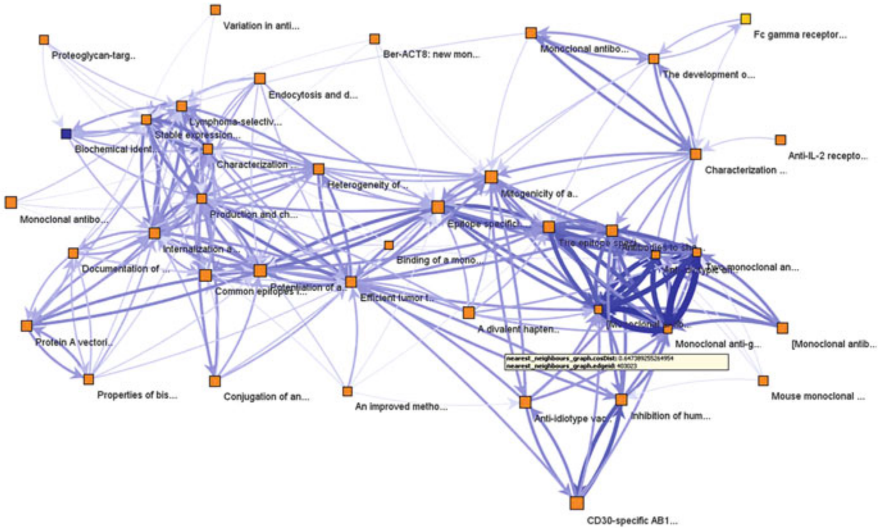


Fig. 2 Large graphs in *RelaNet* (Adapted from von Landesberger (2010))

of statisticians or data mining and machine-learning experts. Ever since computers have existed, they have been used by them to strengthen or refute hypotheses based on empirical data, to identify patterns or to create prediction models from the data. Why do we need to include a new technology to the repertoire of methods?

Visualization introduces human strengths into the analysis, which mitigates some fundamental flaws with automated methods. One flaw is that automated methods will never be able to discover something *unexpected* – by definition. Consider, for example, a technique which is designed to identify all linear correlations between different word count statistics from a large text corpus. We may expect to find linear correlations with this technique because we specifically choose to do so. Selecting another technique will allow us to find another class of patterns or relations, but no single automated technique can claim to find *everything* that might be relevant. In the end, it remains the responsibility of the human to specify what to search for. To make informed decisions, visualization helps the user to get familiar with the data. Another flaw is that automated methods will *always* present a result, regardless of its relevance or quality. Most of them are optimization techniques, which search for the best option from a large but very well-specified set of candidates. To illustrate the problem, imagine you have to pick the best tool from your toolbox to screw something to the wall. If your toolbox contains a large set of candidate screwdrivers, you most likely end up with a good choice. If your toolbox contains a set of candidate hammers, you still might find a "best" one, but still encounter problems in its application.

Traditionally data analysis has been the domain of statisticians or data mining and machine-learning experts. Their methodologies do not explicitly include visual-

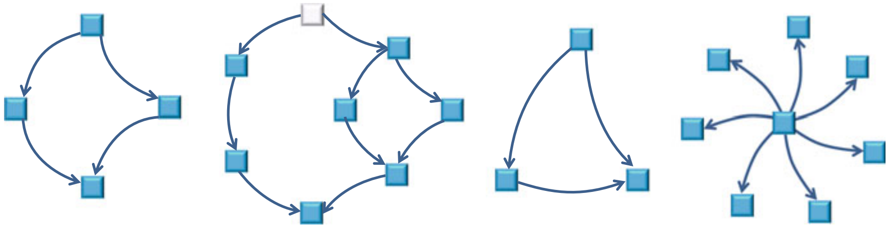


Fig. 3 Motifs in *RelaNet* (Adapted from von Landesberger (2010))

ization. Statisticians typically do *confirmatory analysis*. Starting with a hypothesis, statistical methods define criteria which can be applied to test data to strengthen or refute the hypothesis. The counterpart of confirmatory analysis is *exploratory analysis* as introduced by Tukey and Tukey (1985) in the 1970s. Exploratory analysis starts with raw data. The rationale of exploratory analysis is that there might be more interesting patterns and relations in a data collection than a statistician might be able to think of. Its goal is to identify new and useful statements about the data which are worth further investigation. Techniques from data mining and machine-learning are devised to create patterns and prediction models from the data.

4 Closing the Analytic Gap: An Example

One example for the coupling of knowledge discovery (Fayyad et al. 1996) methods with visual representations is *RelaNet*, a visual analytics system developed in THESEUS CTC WP5 (von Landesberger 2010; von Landesberger et al. 2011). *RelaNet* visualizes relationships and correlation of data entities as directed graphs. Dependencies within abstract data objects, e.g. chemical substances, documents, people, or any combination of data entities, can be visualized without the usage of explicitly modeled semantics. The visualization of such graphs, which are using not the metadata but the data entities themselves, may grow very fast (see Fig. 2). To handle the problem of huge graphs and provide a comprehensible view on the given data entities, *RelaNet* incorporates analytical methods to allow the search for *motifs*.

One main aspect of *RelaNet* is the *motif search* in the various graphs. *Motifs* are sub-graphs describing specific relationship constellations. The search for meaningful constellations, which appear frequently in ideal cases, is one of the methods for the abstraction of complex graphs. The distribution of certain *motifs* in the underlying network provides useful information for a better search in large data structures (von Landesberger 2010). Figure 3 displays a set of *motifs*, which can be searched in the complex graph visualization of *RelaNet*. The *motif search* uses statistic-analytical methods to find similar correlations between data entities.

The search for and examination of complex relations is a strategy for the simplification of complex graphs, in which the information of the individual elements is transferred to the "meaning" the complex structures.

5 Conclusion

The THESEUS *Core Technology for Innovative User Interfaces and Visualizations* investigated methods for coupling analytical and visual methods to face the problem of information access. With new methods on *motif search*, the research in THESEUS CTC WP5 focused on large graphs and the detection of sub-graphs. A contribution was made to the area of visual analytics, an interdisciplinary field incorporating data mining, statistics and visualization to visually represent *right information*, *at the right time*, and *in the right way*.

References

U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, From data mining to knowledge discovery in databases, AI Mag. **17**(3), 37–54 (1996)

D. Keim, G. Andrienko, J.D. Fekete, C. Görg, J. Kohlhammer, G. Melançon, Visual analytics: definition, process, and challenges, in *Information Visualization: Human-Centered Issues and Perspectives*, ed. by A. Kerren, J.T. Stasko, J.D. Fekete, C. North. Volume 4950 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2008), pp. 154–175

D.A. Keim, J. Kohlhammer, F. Mansmann, T. May, F. Wanner, Visual analytics, in *Mastering the Information Age – Solving Problems with Visual Analytics*, ed. by D.A. Keim, J. Kohlhammer, G. Ellis, F. Mansmann (Florian Mansmann, Goslar, Germany, 2010), pp. 7–18

T. May, *Modelle und Methoden für die Kopplung automatischer und visuell-interaktiver Verfahren für die Datenanalyse*, PhD thesis, Technische Universität Darmstadt, 2011

J.J. Thomas, K.A. Cook (eds.), *Illuminating the Path: The Research and Development Agenda for Visual Analytics* (IEEE Computer Society Press, Los Alamitos, California, 2005)

J.W. Tukey, P.A. Tukey, Computer graphics and exploratory data analysis: an introduction, in *Proceedings of the Sixth Annual Conference and Exposition: Computer Graphics '85*, Dallas, vol. 3, (Dallas, Texas, 1985), pp. 773–785

T. von Landesberger, *Visual analytics of large weighted directed graphs and two-dimensional time-dependent data*, PhD thesis, Technische Universität Darmstadt, 2010, http://tuprints.ulb.tu-darmstadt.de/2242/

T. von Landesberger, A. Kuijper, T. Schreck, J. Kohlhammer, J.J. van Wijk, J.D. Fekete, D.W. Fellner, Visual analysis of large graphs: state-of-the-art and future research challenges. Comput. Graph. Forum **30**(6), 1719–1749 (2011), http://dblp.uni-trier.de/db/journals/cgf/cgf30.html#LandesbergerKSKWFF11

Machine Learning for Visual Concept Recognition and Ranking for Images

Alexander Binder, Wojciech Samek, Klaus-Robert Müller, and Motoaki Kawanabe

Abstract Recognition of a large set of generic visual concepts in images and ranking of images based on visual semantics is one of the unsolved tasks for future multimedia and scientific applications based on image collections. From that perspective, improvements of the quality of semantic annotations for image data are well matched to the goals of the THESEUS research program with respect to multimedia and scientific services. We will introduce the data-driven and algorithmic challenges inherent in such tasks from a perspective of statistical data analysis and machine learning and discuss approaches relying on kernel-based similarities and discriminative methods which are capable of processing large-scale datasets.

A. Binder (✉) • W. Samek
Fraunhofer Institute for Computer Architecture and Software Technology (FIRST), Berlin, Germany

Machine Learning Group, TU Berlin, Berlin, Germany
e-mail: alexander.binder@tu-berlin.de; wojciech.samek@tu-berlin.de

K.-R. Müller
Machine Learning Group, TU Berlin, Berlin, Germany

Department of Brain and Cognitive Engineering, Korea University, Seoul, Korea
e-mail: klaus-robert.mueller@tu-berlin.de

M. Kawanabe
ATR Brain Information Communication Research Laboratory, Kyoto, Japan
e-mail: kawanabe@atr.jp

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_17,

1 Introduction

Visual concept recognition in its broadest sense is about the identification of semantic content in images, based on the visual information in the pixels of an image. This idea is as old as research on Artificial Intelligence itself and has been pursued since the availability of digitized images and computing machines, yet it started to flourish just in the mid-1990s when computing hardware became affordable and its computing power expanded rapidly. The identification of visual content is one of the unsolved tasks for future applications based on digital image collections. With the advent of the Internet and the use digital cameras, digital images have become widely available. This trend is enhanced by the evolution of embedded digital imaging systems in consumer products such as mobile phones and their increased availability in medical applications such as radiology. The collection of image databases, like digital archives in medial or medical facilities, is facilitated by falling prices of storage devices. All this fuels the demand for systems which are capable of identifying in one or another sense visual content in a large number of images. We will focus here on semantic concept recognition in images.

Formally, a visual concept can be represented as an indicator function $\mathbb{I}_C$ on the space of all images $\mathscr{X}$ such that $\mathbb{I}_C(x) = 1$ denotes the presence of concept C in an image $x \in \mathscr{X}$:

$$\mathbb{I}_C : \mathscr{X} \longrightarrow \{0, 1\} . \tag{1}$$

For ambiguous semantic concepts this definition is extended by assigning an image x a score $f_C(x)$ in a bounded interval (e.g. $[0, 1]$) which represents a numerical value for the strength of the presence of a semantic concept in an image:

$$f_C : \mathscr{X} \longrightarrow [0, 1] . \tag{2}$$

This numerical value can be interpreted in a probabilistic manner as the agreement of a set of human annotators with respect to the question of whether an image belongs to a semantic concept or not. In the context of statistical methods this is known as label noise. Such ambiguities arise naturally for concepts denoting the emotional impression of an image or concepts related to aesthetic quality.

We set the aim of Semantic Concept Recognition as the prediction of all semantic concepts present in an image. In contrast to multi-class approaches we aim at recognizing multiple concepts in an image, e.g. at finding the set of concepts *Outdoor*, *Plants*, *Day*, *NeutralIllumination*, *PartlyBlurred*, *ParkGarden*, *Toy*, *Natural*, *Cute* and *Calm* in one image. Thus, we are interested in predicting a large set of generic semantic concepts in contrast to a small set of highly specialized concepts as is the aim of face recognition, for example. One image may show multiple concepts. The prediction output is desired to be a continuous score instead of a binary decision. The continuous score allows us to provide hints about uncertainty of the prediction. Such information is highly useful for the common search scenario in which a user is interested in finding the top K most likely images for a selected concept. The continuous score may be used for ranking in such a context. Such a ranking-based search scenario has proven to be practical and represents the state of the art for finding images according to their provided file names by common Internet search engines. The notion of concept recognition may be adapted to more specific concepts: if one considers visual objects with a clearly determinable area, then one can pursue object detection, which aims at finding a tight bounding box around the object or segmentation which assigns each pixel in the image a label corresponding to one of the objects or the background.

Why Is Concept Recognition in Images a Difficult Task? One may ask why common Internet search engines use the image search based on file names as the standard tool while search based on visual content appears to be in the beta phase at best. In this section we discuss some issues and challenges of visual concept recognition in images.

Generally speaking one may observe a large variability in the *semantic* structure of visual concepts. This presents a challenge for algorithms capable of predicting semantic concepts and ranking images according to them.

Key factors for the variance in the structure of a semantic concept are the presence and absence of a wide range of all kinds of visual cues, their composition and their contribution to the classification of an image in a non-deterministic manner. The following examples will exemplify the above.

Concepts are defined by the presence of several visual cues in the image. The difference from classical object recognition is that the visual cues may vary highly and may not be classified into one simple object class. Consider the concept *concert*. Photos showing a small group of people known to be famous music artists on stage are likely to belong to such a concept. At the same time a large group of hobby artists playing in an orchestra also shows a *concert*.

The composition of cues beyond their mere presence may play an important role: A person holding a guitar in a certain pose may contribute to the classification as a *concert*. However another pose with a guitar on his back may depict rather a travelling person. Similarly, music at a funeral scene is less likely called a *concert*. One can think of many setups of musical instruments and people which are more or less likely to be a *concert*.

This reveals that general semantic concepts are more difficult to recognize compared to classic single object recognition. Another reason besides the wide range of possible cues is that cues contribute in a non-deterministic way (which can be modeled in a probabilistic fashion) to the rating for belonging to semantic concepts. Consider the concept *streetscene*: the presence of roads and buildings are cues for such a concept; however, the density and height of buildings, the density of roads and the density of parked cars are important for judging whether this is a *streetscene* or just a lonely road outside a town with some buildings near it. This probabilistic contribution of cues and their composition becomes obvious for concepts related to aesthetic quality or emotional impact such as *funny* or *scary*.

We can identify some special cases of the variability of cues which are mentioned briefly:

- *Varying positions and sizes of regions in an image relevant for a semantic concept:* When limited to objects one will note that an object can fill a large fraction of the image or a very small region. A smaller object may have a highly varying position within the image. Similarly, the appearance of an object may vary with its viewpoint. The same holds for cues contributing to a semantic concept instead of objects.
- *Occlusion of regions in an image relevant for a semantic concept:* Regions of an image relevant for the recognition of a semantic concept can be occluded.
- *Clutter and complex scene compositions:* Images can have large areas which are irrelevant for the recognition of a visual concept. Consider the example of a living room which contains somewhere an object of the class *bottle* which is to be found.

The Role of Label Noise The informal points discussed above may be interpreted in more formal statistical terms: they have two effects on increasing the difficulty of the classification problem. The first effect in a probabilistic classification setting is an increased irregularity of the optimal decision boundary. In a probabilistic setting the decision boundary for a visual concept would be the set of images x with maximal uncertainty with respect to the question of whether a visual concept is present or not: $P(\mathbb{I}_C(x) = 1) = 0.5$.

The second effect is increased label noise. Label noise can be measured as the uncertainty of human annotators in assigning an image to belong to a semantic concept. Mathematically it can be modeled as the probability of an image belonging to a concept $P(\mathbb{I}_C(x) = 1)$.

Label noise has intuitively a deteriorating impact on classification accuracy, and more importantly on model selection. Learning of a support vector machine corresponds to the choice of a function from a class of functions via selecting its parameters when solving the optimization problem.

Theorem 6 in Massart and Nédélec (2006) provides lower bounds for the expected risk in empirical risk minimization depending on a uniform bound for the label noise which is applicable to support vector machines with Gaussian kernels on bounded domains for distributions with smoothly differentiable Bayes boundaries.

Theorem 1 (Theorem 6 from Massart and Nédélec 2006). *Let μ be a probability measure on $\mathscr{X}$ and S be some class of classifiers on $\mathscr{X}$ such that for some positive constants K_1, K_2, ϵ_0 and r*

$$K_2\epsilon^{-r} \leq H_1(\epsilon, S, \mu) \leq K_1\epsilon^{-r}$$

for all $0 < \epsilon \leq \epsilon_0$, where $H_1(\epsilon, S, \mu)$ denotes the $\ell_1(\mu)$-metric entropy of S. Then, there exists a positive constant K depending on K_1, K_2, ϵ_0 and r such that the bound

$$R_n(h, S, \mu) = \inf_{\hat{s} \in S} \sup_{P \in \mathscr{P}(h,S,\mu)} \mathbb{E}[P(Y \neq \hat{s}(X)) - P(Y \neq s^*(X))]$$
$$\geq K(1-h)^{\frac{1}{1+r}} \max \left(h^{-\frac{1-r}{1+r}} n^{-\frac{1}{1+r}}, n^{-\frac{1}{2}}\right) \quad (3)$$

holds whenever $n \geq 2$.

For understanding the theorem, note that $\mathscr{P}(h, S, \mu)$ is the set of distributions on the input-label product space $\mathscr{X} \times \mathscr{Y}$ such that the input space distribution is μ and the label noise is in each point of $\mathscr{X}$ bounded by $1/2 - h/2$: $|P(Y = 1|X = x) - 0.5| \geq h/2$ and s^* is the Bayes classifier. $\mathbb{E}[P(Y \neq \hat{s}(X)) - P(Y \neq s^*(X))]$ is the deviation between the expected errors of the classifier $\hat{s}$ and the a posteriori optimal Bayes classifier s^*. The supremum is taken over a class of distributions followed by selection of the optimal empirical classifier $\hat{s}$ given knowledge of the distribution. Since the underlying distribution of images and their concept labels is unknown, this implies that the lower bound is an optimistic formulation and results will be worse in practice.

An increase in the overall label noise corresponds to a decrease of the value of h which yields an increased lower bound in Theorem 1 for the expected deviation between the expected error of an optimistically selected classifier and the statistically best possible classifier within a function class. *The qualitative message is that label noise does have a highly deteriorating influence on model selection.*

A Machine Learning Approach As seen in the preceding section, the methodological challenge consists in the ability to deal with a large set of differing visual concepts, ranging from localized objects to overall emotional impressions, the large variability of image appearance present within general visual concepts beyond simple objects like *partylife*, *beachholiday*, *mountains*, *indoor*, *euphoric*, *aestheticimpression* and the disagreement of human annotators in labelling images for such concepts.

Statistical learning methods are constructed to be robust against annotator disagreement, labelling errors, varying image qualities and scales of visible cues, differing lighting conditions, occlusion and clutter in images.

They are able to extract relevant structures based on *implicit and statistical* definitions as when labelling example images as belonging to the same concept even when there is no way to define deterministic rules for what a concept should be. This makes statistical learning methods a valuable complement to *explicit and deterministic* knowledge modelling via ontologies.

The principle of the imaging solutions is based on countering the large variability of visual concepts by a large set of varying image representations being computed for each image and learning the optimal weighting of these representations for each visual concept based on criteria derived from state-of-the-art statistical machine learning as depicted in Fig. 1.

Two key components of the approach will be described in the following: Bag of Words Features and the learning of kernel combinations.

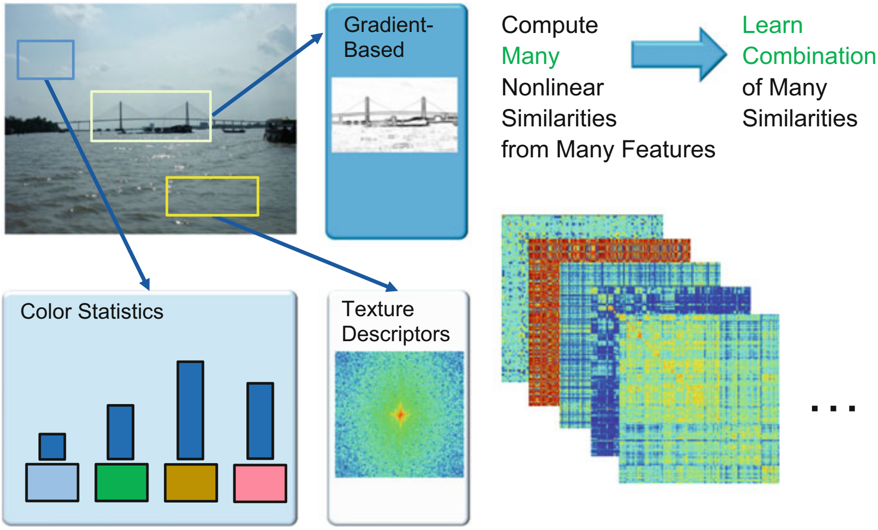


Fig. 1 Graphical depiction of the multiple feature extraction and multiple kernel learning-based machine learning approach

2 Bag of Words Features

In the last decade bag of words features (Csurka et al. 2004) have evolved into the dominating feature extraction approach for visual concept recognition and ranking. They deliver consistently the top-ranked solutions in this field, as shown over the past few years in state-of-the-art benchmark challenges like Pascal Visual object categorization (Everingham et al. 2010) and ImageCLEF PhotoAnnotation (Nowak et al. 2011), which deal with images with high visual variance. For image recognition problems on highly specialized narrow domains, other feature extraction methodologies may deliver better results.

The bag of words feature is a three-stage feature extraction principle which is not linked to the SIFT (Lowe 2004) feature itself. In the first stage a set of local features is computed from an image. Formally, a local feature is a vector computed over a localized subset of the image. The SIFT descriptor (Lowe 2004) is the most popular choice. Besides the choice of the local feature, regions for its computation have to be chosen. Typically, local features are computed on small overlapping regions taken from the whole image. Apart from grid sampling, biased random sampling (Binder et al. 2011; Yang et al. 2009b) may serve for the computation of the corresponding descriptor regions. The number of local features may vary across images, for example by adaptation to image size. The second stage, the computation of the set of visual words, is done once during training time. Formally, a visual word is a point in the space of local features. One possibility to compute the visual words is by discretization of the local feature density using k-means. Practically proven alternatives are radius-based clustering (Jurie and Triggs 2005), Bayesian methods like pLSA (Hofmann 1999) and more commonly Fisher vectors based on Gaussian mixture models (Csurka et al. 2010) and sparse coding (Yang et al. 2009a). There has been considerable research on improvements for visual word generation, such as hierarchical clustering (Nistér and Stewénius 2006), class-wise clustering (Wojcikiewicz et al. 2010a), weakly supervised clustering (Binder et al. 2010) or optimization of information-theoretic criteria (Yu et al. 2009). Finally the last stage is the aggregation of local features into the global bag of words feature. The local features are mapped on the visual words, usually with weights based on the distances between the local feature and all the prototypes. Examples are soft codebooks (van Gemert et al. 2008) and fast local linear coding (Wang et al. 2010).

The strength of the bag of words feature lies in its robustness, which comes from the following factors:

- The absence of modelling spatial relations between parts, unlike earlier approaches, which are susceptible to noise in images with complex sceneries.
- The aggregation of local features into a global feature, which implies denoising via averaging of contributions of many local features.
- The choice of robust local features such as SIFT (Lowe 2004) or SURF (Bay et al. 2008) which are known to be invariant against many changes in lighting conditions. See van de Sande et al. (2010) for an overview from a color-theoretic point of view.

Another advantage of bag of words features is their computational scalability. This is an advantage over intuitively more appealing Bayesian approaches which often need to rely on restricted probability models or inference approximations in practice. Computation of bag of words features in real time is demonstrated in Uijlings et al. (2010). van de Sande et al. (2011a) demonstrate their efficient computation on GPUs.

The work (Parikh 2011) shows by comparing against human performance that bag of words features yield a performance similar to that of humans on jumbled images which were cut into square parts and then piecewise randomly permutated and rejoined. The human advantage is our ability to extract spatial relations between parts which took us years of learning in childhood from millions of examples and some hundred thousand years of brain evolution for the base learning system to be operational.

The bag of words method is also applied with superior results in competitions in related domains such as semantic indexing for videos in TRECVID (Inoue et al. 2011) or, as the winning entry, in the ILSVRC2011 large-scale object detection challenge (van de Sande et al. 2011b).

Despite their robustness for domains with highly variable images, bag of words features are also applied to narrow domains such as concept recognition for medical images (André et al. 2011; Cruz-Roa et al. 2011; Xu et al. 2011).

In conclusion, bag of words features are very well suited for image annotation tasks.

3 Learning Kernel Combinations

Once there are many possible feature representations at hand for each image, the question arises about how to combine them. The answer from the point of statistical machine learning is to learn a combination from image features and labels, trying to separate the images belonging to a visual concept from the rest in a high dimensional feature space. The typical dimensionality of single features of the order of several thousands limits the usefulness of classification trees and simple nearest neighbor algorithms. Several principled machine learning algorithms have been developed, the classic Boosting (Freund and Schapire 1995) the most famous one.

We will focus on kernel-based algorithms (Müller et al. 2001). A kernel can be said to be a positive definite matrix of similarity values between data samples. These similarity values have been computed from the features available for each image. Given a loss function, the similarity values contained in a kernel matrix can be used to learn a classifier which predicts the annotation. A well established method for learning such a classifier is support vector machines (Cortes and Vapnik 1995) which have been used with success in a variety of fields such as genome analysis, computational chemistry, finance, brain-computer interfacing, and image and natural language processing.

The learning of kernel combinations relies on multiple kernel learning (Kloft et al. 2011), which extends support vector machines (see formula (4)). It learns weights for a linear combination of kernels in addition to the parameters of an ordinary support vector machine for each visual concept. The need for learning kernel combinations in image annotation tasks has been widely recognized in the international community, as seen by the application of sparse multiple kernel learning to object recognition (Gehler and Nowozin 2009; Kumar and Sminchisescu 2007), object detection (Lampert and Blaschko 2008), or the development of alternative non-sparse Kernel learning algorithms (Yan et al. 2009) influenced by multiple kernel learning:

$$\begin{aligned} \min_{\beta,w,b,\xi} \quad & \frac{1}{2}\sum_{j=1}^{m} \beta_j \boldsymbol{w}_j' \boldsymbol{w}_j + C\|\boldsymbol{\xi}\|_1 \\ \text{s.t.} \quad & \forall i : \; y_i \left(\sum_{j=1}^{m} \beta_j \boldsymbol{w}_j' \psi_j(\boldsymbol{x}_i) + b \right) \geq 1 - \xi_i \\ & \boldsymbol{\xi} \geq \mathbf{0}\, ; \quad \boldsymbol{\beta} \geq \mathbf{0}\, ; \quad \|\boldsymbol{\beta}\|_p \leq 1\, . \end{aligned} \tag{4}$$

The advantage of non-sparse multiple kernel learning lies in the combination of the robustness of the support vector machine with adjustable regularization for the learning of kernel weights. The latter constitutes the essential difference with its predecessor, sparse multiple kernel learning (Bach et al. 2004; Lanckriet et al. 2004; Sonnenburg et al. 2006).

The image annotation system developed by the Fraunhofer FIRST and the TU Berlin has been tested in several international benchmark competitions with undisclosed ground truth for the test data. The joint efforts resulted in its being the third best group at the ImageCLEF2009 Photo Challenge, the fourth best group at the Pascal VOC 2009 Object Classification Challenge, and finally, the winning entry (Binder et al. 2011) for two categories of ImageCLEF2011 PhotoAnnotation Challenge (Nowak et al. 2011), namely multi-modal and visual ranking by the mean average precision measure, for which multiple kernel learning was employed.

4 Conclusion

This article has shown, on the basis of selected scientific contributions, that machine learning has proved its high usefulness in practical image annotation (Kawanabe et al. 2011; Samek et al. 2011). This was – along with excellent publications (Binder et al. 2012a,b, 2013, 2010; Wojcikiewicz et al. 2010a,b) – also demonstrated in successful submissions to international annotation competitions (Binder and Kawanabe 2009; Everingham et al. 2010; Nowak et al. 2011) and most recently as the best visual and multi-modal submission entry in the ImageCLEF2011 PhotoAnnotation Challenge (Binder et al. 2011).

When striving for intelligent machine vision a number of hard challenges still lie ahead. Multimodal sources of information and their respective hierarchical structures have to be fused on firm theoretical grounds; here a lot is still to be done. Machine-learning-based visual concept annotation and ranking needs to be combined with semantic knowledge such as taxonomies in order to express semantic relations between visual concept classes (Binder et al. 2012a; Deng et al. 2011). Semantic classes are far from being independent, thus a challenge will be to model dependency structure appropriately; Multi-task learning will only be a first step in this important direction. Finally, the real world is a complex, structured and highly non-stationary environment. A statistical structure that may hold for an image ensemble in 2000 may not be up to date in 2012. Thus it will be important to establish high performing learning methods that work stably despite the non-stationarity of the world around us and all its structural changes (Sugiyama et al. 2007; von Bünau et al. 2009). The emerging fields of human action classification and anomaly detection in surveillance videos require solutions for such problems.

Acknowledgements This work was primarily supported by the German Federal Ministry for Economic Affairs and Energy (BMWi) under the THESEUS research program (Grant 01MQ07018). Furthermore it was in part supported by the World Class University Program through the National Research Foundation of Korea funded by the Korean Ministry of Education, Science, and Technology, under Grant R31-10008. We express our thanks to Volker Tresp, the work package leader at CTC WP6, Ralf Schäfer from the Fraunhofer HHI and Shinichi Nakajima from the Nikon corporation for the fruitful collaboration.

References

B. André, T. Vercauteren, A.M. Buchner, M.B. Wallace, N. Ayache, Retrieval evaluation and distance learning from perceived similarity between endomicroscopy videos, in *Proceedings of the 14th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'11)*, Toronto, ed. by G. Fichtinger, A.L. Martel, T.M. Peters. Volume 6893 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011), pp. 297–304. http://dblp.uni-trier.de/db/conf/miccai/miccai2011-3.html#AndreVBWA11

F.R. Bach, G.R.G. Lanckriet, M.I. Jordan, Multiple kernel learning, conic duality, and the SMO algorithm, in *Proceedings of the 21st International Conference on Machine Learning (ICML'04)*, Banff, ed. by C.E. Brodley. Volume 69 of ACM International Conference Proceeding Series (ACM, 2004). http://dblp.uni-trier.de/db/conf/icml/icml2004.html#BachLJ04

H. Bay, A. Ess, T. Tuytelaars, L.V. Gool, SURF: speeded up robust features. Comput. Vis. Image Underst. (CVIU) **110**, 346–359 (2008)

A. Binder, M. Kawanabe, Enhancing recognition of visual concepts with primitive color histograms via non-sparse multiple kernel learning, in *Proceedings of the 10th Workshop of the Cross-Language Evaluation Forum (CLEF'09)*, Corfu, Greece, ed. by C. Peters, B. Caputo, J. Gonzalo, G.J.F. Jones, J. Kalpathy-Cramer, H. Müller, T. Tsikrika. Volume 6242 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2009), pp. 269–276. http://dblp.uni-trier.de/db/conf/clef/clef2009-2.html#BinderK09

A. Binder, K.R. Müller, M. Kawanabe, On taxonomies for multi-class image categorization. Int. J. Comput. Vis. **99**(3), 281–301 (2012a). http://dblp.uni-trier.de/db/journals/ijcv/ijcv99.html#BinderMK12

A. Binder, S. Nakajima, M. Kloft, C. Müller, W. Samek, U. Brefeld, K.R. Müller, M. Kawanabe, Insights from classifying visual concepts with multiple kernel learning. PLoS One **7**(8) (2012b). http://dblp.uni-trier.de/db/journals/corr/corr1112.html#abs-1112-3697

A. Binder, W. Samek, M. Kloft, C. Müller, K.R. Müller, M. Kawanabe, The joint submission of the TU Berlin and Fraunhofer FIRST (TUBFI) to the ImageCLEF2011 photo annotation task, in *CLEF (Notebook Papers/Labs/Workshop)*, Amsterdam, ed. by V. Petras, P. Forner, P.D. Clough, 2011. http://dblp.uni-trier.de/db/conf/clef/clef2011w.html#BinderSKMMK11

A. Binder, W. Samek, K.R. Müller, M. Kawanabe, Enhanced representation and multi-task learning for image annotation. Comput. Vis. Image Underst. **117**(5), 466–478 (2013). http://dblp.uni-trier.de/db/journals/cviu/cviu117.html#BinderSMK13

A. Binder, W. Wojcikiewicz, C. Müller, M. Kawanabe, A hybrid supervised-unsupervised vocabulary generation algorithm for visual concept recognition, in *Proceedings of the 10th Asian Conference on Computer Vision (ACCV'10)*, Queenstown, ed. by R. Kimmel, R. Klette, A. Sugimoto. Volume 6494 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2010), pp. 95–108. http://dblp.uni-trier.de/db/conf/accv/accv2010-3.html#BinderWMK10

C. Cortes, V. Vapnik, Support vector networks. Mach. Learn. **20**(3), 273–297 (1995)

A. Cruz-Roa, J.C. Caicedo, F.A. Gonzáez, Visual pattern mining in histology image collections using bag of features. Artif. Intell. Med. **52**(2), 91–106 (2011). http://dblp.uni-trier.de/db/journals/artmed/artmed52.html#Cruz-RoaCG11

G. Csurka, C.R. Dance, L. Fan, J. Willamowski, C. Bray, Visual categorization with bags of keypoints. *Workshop on Statistical Learning in Computer Vision (ECCV)*, Prague, 2004, pp. 1–22

G. Csurka, F. Perronnin, L. Marchesotti, S. Clinchant, J. Ah-Pine, Fisher kernel representation of images and some of its successful applications, in *Proceedings of the International Conference on Computer Vision and Theory Applications (VISAPP'10)*, Angers, ed. by P. Richard, J. Braz (INSTICC, 2010), pp. 21–25. http://dblp.uni-trier.de/db/conf/visapp/visapp2010-1.html#CsurkaPMCA10

J. Deng, A.C. Berg, F.F. Li, Hierarchical semantic indexing for large scale image retrieval, in *Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR'11)*, Colorado Springs (IEEE, 2011), pp. 785–792. http://dblp.uni-trier.de/db/conf/cvpr/cvpr2011.html#DengBL11

M. Everingham, L.J.V. Gool, C.K.I. Williams, J.M. Winn, A. Zisserman, The pascal visual object classes (VOC) challenge. Int. J. Comput. Vis. **88**(2), 303–338 (2010). http://dblp.uni-trier.de/db/journals/ijcv/ijcv88.html#EveringhamGWWZ10

Y. Freund, R.E. Schapire, A decision theoretic generalization of on-line learning and an application to boosting, in *Second European Conference on Computational Learning Theory (EuroCOLT-95)*, Barcelona, ed. by P.M.B. Vitányi (Springer, Berlin/Heidelberg/New York, 1995), pp. 23–37. citeseer.nj.nec.com/freund95decisiontheoretic.html

P.V. Gehler, S. Nowozin, On feature combination for multiclass object classification, in *Proceedings of the 12th IEEE International Conference on Computer Vision (ICCV'09)*, Kyoto (IEEE, 2009), pp. 221–228. http://dblp.uni-trier.de/db/conf/iccv/iccv2009.html#GehlerN09

T. Hofmann, Probabilistic latent semantic analysis, in *Proceedings of the Uncertainty in Artificial Intelligence (UAI'99)*, Stockholm, 1999, pp. 289–296

N. Inoue, Y. Kamishima, T. Wada, K. Shinoda, S. Sato, TokyoTech+Canon at TRECVID 2011, in *TREC Video Retrieval Evaluation*, Gaithersburg, Maryland, 2011

F. Jurie, B. Triggs, Creating efficient codebooks for visual recognition, in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV'05)*, Beijing (IEEE Computer Society, 2005), pp. 604–610. http://dblp.uni-trier.de/db/conf/iccv/iccv2005-1.html#JurieT05

M. Kawanabe, A. Binder, C. Müller, W. Wojcikiewicz, Multi-modal visual concept classification of images via Markov random walk over tags, in *Proceedings of the IEEE Workshop on Applications of Computer Vision (WACV'11)*, Kona (IEEE Computer Society, 2011), pp. 396–401. http://dblp.uni-trier.de/db/conf/wacv/wacv2011.html#KawanabeBMW11

M. Kloft, U. Brefeld, S. Sonnenburg, A. Zien, Lp-norm multiple kernel learning. J. Mach. Learn. Res. **12**, 953–997 (2011). http://dblp.uni-trier.de/db/journals/jmlr/jmlr12.html#KloftBSZ11

A. Kumar, C. Sminchisescu, Support kernel machines for object recognition, in *Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV'07)*, Rio de Janeiro (IEEE, 2007), pp. 1–8. http://dblp.uni-trier.de/db/conf/iccv/iccv2007.html#KumarS07

C.H. Lampert, M.B. Blaschko, A multiple kernel learning approach to joint multi-class object detection, in *Proceedings of the 30th DAGM Symposium*, Munich, ed. by G. Rigoll. Volume 5096 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2008), pp. 31–40. http://dblp.uni-trier.de/db/conf/dagm/dagm2008.html#LampertB08

G.R.G. Lanckriet, N. Cristianini, P.L. Bartlett, L.E. Ghaoui, M.I. Jordan, Learning the kernel matrix with semidefinite programming. J. Mach. Learn. Res. **5**, 27–72 (2004). http://dblp.uni-trier.de/db/journals/jmlr/jmlr5.html#LanckrietCBGJ03

D.G. Lowe, Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)

P. Massart, É. Nédélec, Risk bounds for statistical learning. Ann. Stat. **34**(5), 2326–2366 (2006)

K.R. Müller, S. Mika, G. Rätsch, S. Tsuda, B. Schölkopf, An introduction to kernel-based learning algorithms. IEEE Trans. Neural Netw. **12**(2), 181–202 (2001). http://www.ist.temple.edu/~vucetic/cis526fall2003/SVMintro.pdf

D. Nistér, H. Stewénius, Scalable recognition with a vocabulary tree, in *Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, New York (IEEE Computer Society, 2006), pp. 2161–2168. http://dblp.uni-trier.de/db/conf/cvpr/cvpr2006-2.html#NisterS06

S. Nowak, K. Nagel, J. Liebetrau, The CLEF 2011 photo annotation and concept-based retrieval tasks, in *CLEF (Notebook Papers/Labs/Workshop)*, Amsterdam, ed. by V. Petras, P. Forner, P.D. Clough, 2011. http://dblp.uni-trier.de/db/conf/clef/clef2011w.html#NowakNL11

D. Parikh, Recognizing jumbled images: the role of local and global information in image classification, in ed. by D.N. Metaxas, L. Quan, A. Sanfeliu, L.J.V. Gool, *Proceedings of the International Conference on Computer Vision (ICCV'11)*, Barcelona (IEEE, 2011), pp. 519–526. http://dblp.uni-trier.de/db/conf/iccv/iccv2011.html#Parikh11

W. Samek, A. Binder, M. Kawanabe, Multi-task learning via non-sparse multiple kernel learning, in *Proceedings of the 14th International Conference on Computer Analysis of Images and Patterns (CAIP'11)*, Seville, ed. by P. Real, D. Díaz-Pernil, H. Molina-Abril, A. Berciano, W.G. Kropatsch. Volume 6854 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011), pp. 335–342. http://dblp.uni-trier.de/db/conf/caip/caip2011-1.html#SamekBK11

S. Sonnenburg, G. Rätsch, C. Schäfer, B. Schölkopf, Large scale multiple kernel learning. J. Mach. Learn. Res. **7**, 1531–1565 (2006). http://jmlr.org/papers/volume7/sonnenburg06a/sonnenburg06a.pdf

M. Sugiyama, M. Krauledat, K.R. Müller, Covariate shift adaptation by importance weighted cross validation. J. Mach. Learn. Res. **8**, 985–1005 (2007). http://dblp.uni-trier.de/db/journals/jmlr/jmlr8.html#SugiyamaKM07

J.R.R. Uijlings, A.W.M. Smeulders, R.J.H. Scha, Real-time visual concept classification. IEEE Trans. Multimed. **12**(7), 665–681 (2010). http://dblp.uni-trier.de/db/journals/tmm/tmm12.html#UijlingsSS10

K.E.A. van de Sande, T. Gevers, C.G.M. Snoek, Evaluating color descriptors for object and scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. **32**(9), 1582–1596 (2010). http://dblp.uni-trier.de/db/journals/pami/pami32.html#SandeGS10

K.E.A. van de Sande, T. Gevers, C.G.M. Snoek, Empowering visual categorization with the GPU. IEEE Trans. Multimed. **13**(1), 60–70 (2011a). http://dblp.uni-trier.de/db/journals/tmm/tmm13.html#SandeGS11

K.E.A. van de Sande, J.R.R. Uijlings, T. Gevers, A.W.M. Smeulders, Segmentation as selective search for object recognition, in *Proceedings of the 13th International Conference on Computer Vision (ICCV'11)*, Barcelona, ed. by D.N. Metaxas, L. Quan, A. Sanfeliu, L.J.V. Gool (IEEE, 2011b), pp. 1879–1886. http://dblp.uni-trier.de/db/conf/iccv/iccv2011.html#SandeUGS11

J. van Gemert, J.M. Geusebroek, C.J. Veenman, A.W.M. Smeulders, Kernel codebooks for scene categorization, in *Proceedings of the 10th European Conference on Computer Vision (ECCV'08)*, Marseille, 2008, pp. 696–709. http://dblp.uni-trier.de/db/conf/eccv/eccv2008-3.html#GemertGVS08

P. von Bünau, F.C. Meinecke, F.C. Király, K.R. Müller, Finding stationary subspaces in multivariate time series. Phys. Rev. Lett. **103**(21), 214101 (2009)

J. Wang, J. Yang, K. Yu, F. Lv, T.S. Huang, Y. Gong, Locality-constrained linear coding for image classification, in *Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR'10)*, San Francisco (IEEE, 2010), pp. 3360–3367. http://dblp.uni-trier.de/db/conf/cvpr/cvpr2010.html#WangYYLHG10

W. Wojcikiewicz, A. Binder, M. Kawanabe, Enhancing image classification with class-wise clustered vocabularies, in *Proceedings of the 20th International Conference on Pattern Recognition (ICPR'10)*, Istanbul (IEEE, 2010a), pp. 1060–1063. http://dblp.uni-trier.de/db/conf/icpr/icpr2010.html#WojcikiewiczBK10

W. Wojcikiewicz, A. Binder, M. Kawanabe, Shrinking large visual vocabularies using multi-label agglomerative information bottleneck, in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP'10)*, Hong Kong, 2010b, pp. 3849–3852

R. Xu, Y. Hirano, R. Tachibana, S. Kido, Classification of diffuse lung disease patterns on high-resolution computed tomography by a bag of words approach, in *Proceedings of the 14th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'11)*, Toronto, ed. by G. Fichtinger, A.L. Martel, T.M. Peters. Volume 6893 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011), pp. 183–190. http://dblp.uni-trier.de/db/conf/miccai/miccai2011-3.html#XuHTK11

F. Yan, J. Kittler, K. Mikolajczyk, M.A. Tahir, Non-sparse multiple kernel learning for fisher discriminant analysis, in *Proceedings of the 2009 IEEE International Conference on Data Mining (ICDM'09)*, Miami, ed. by W. Wei, H. Kargupta, S. Ranka, P.S. Yu, X. Wu (IEEE Computer Society, 2009), pp. 1064–1069. http://dblp.uni-trier.de/db/conf/icdm/icdm2009.html#YanKMT09

J. Yang, K. Yu, Y. Gong, T.S. Huang, Linear spatial pyramid matching using sparse coding for image classification, in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'09)*, Miami (IEEE, 2009a), pp. 1794–1801. http://dblp.uni-trier.de/db/conf/cvpr/cvpr2009.html#YangYGH09

L. Yang, N. Zheng, Y. Jie, M. Chen, H. Chen, A biased sampling strategy for object categorization, in *Proceedings of the 12th International Conference on Computer Vision (ICCV'09)*, Kyoto (IEEE, 2009b), pp. 1141–1148. http://dblp.uni-trier.de/db/conf/iccv/iccv2009.html#YangZYCC09

K. Yu, T. Zhang, Y. Gong, Nonlinear learning using local coordinate coding, in *Proceedings of the 24th Annual Conference on Neural Information Processing Systems (NIPS'09)*, Vancouver, ed. by Y. Bengio, D. Schuurmans, J.D. Lafferty, C.K.I. Williams, A. Culotta (Curran Associates, Inc., 2009), pp. 2223–2231. http://dblp.uni-trier.de/db/conf/nips/nips2009.html#YuZG09

Querying the Web with Statistical Machine Learning

Volker Tresp, Yi Huang, and Maximilian Nickel

Abstract The traditional means of extracting information from the Web are keyword-based search and browsing. The Semantic Web adds structured information (i.e., semantic annotations and references) supporting both activities. One of the most interesting recent developments is Linked Open Data (LOD), where information is presented in the form of facts – often originating from published domain-specific databases – that can be accessed both by a human and a machine via specific query endpoints. In this article, we argue that machine learning provides a new way to query web data, in particular LOD, by analyzing and exploiting statistical regularities. We discuss challenges when applying machine learning to the Web and discuss the particular learning approaches we have been pursuing in THESEUS. We discuss a number of applications where the Web is queried via machine learning and describe several extensions to our approaches.

1 Introduction

The traditional means of extracting information from the Web are keyword-based search and browsing. In a search, the user enters query terms and, if lucky, can read off the required information from the returned pages. In browsing, the user follows hyperlinks to gain deeper information on an issue. The Semantic Web adds structured information (i.e., semantic annotations and references) supporting both keyword-based search and browsing. One of the most interesting recent developments here is Linked Open Data (LOD) (Bizer et al. 2009), where information

V. Tresp (✉) • Yi Huang
Siemens AG, Munich, Germany
e-mail: Volker.Tresp@Siemens.com; YiHuang@Siemens.com

M. Nickel
Ludwig Maximilian University Munich, Munich, Germany
e-mail: nickel@dbs.ifi.lmu.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_18,

is presented in form of facts – often originating from published domain-specific databases – that can be accessed both by a human and a machine via specific query endpoints. Thus, one can query for the “10 largest German companies whose CEOs were born in the US” or a list of “genes associated with a given disease”. LOD does not just *reference* information, it *represents* information in the form of simple subject-predicate-object triples. With this novel representation of information, new opportunities for accessing information emerge that explore and exploit regularities in the represented data. In recent years, mostly deterministic regularities, which can be formulated as logical expressions, have been explored. Thus, deductive reasoning might conclude that an author born in Landshut would also be recognized as an author born in Bavaria. Deterministic regularities originate, for example, from natural laws (e.g., law of gravity), from human definitions and conventions (e.g.,“dogs are mammals”), from design (e.g.,“the car only starts when the key is turned”), and from human imposed laws and regulations (e.g.,“work begins at 9 am”). In addition to deterministic or close-to-deterministic regularities, the world also contains statistical regularities. One might debate whether the world is inherently deterministic or probabilistic, but at the abstract level of representation, which is typically available for decision making, the world certainly appears uncertain.[1] Young males typically like action movies but whether young Jack will buy “Iron Man 2” might depend more on the availability of illegal downloads, the opinions of his peers and Jack’s financial situation. A system recommending a movie to Jack must work with the available information, maybe a list of movies that Jack has bought before, and can only make statistical recommendations. Similarly, the interactions between genes and diseases might or might not be inherently deterministic in nature; at the level of current knowledge the relationships are only partially known.

Machine learning is a basis for extracting statistical patterns and in this article we will describe our work on statistical machine learning for the Web as pursued in THESEUS and in the EU FP7 project LarKC (Fensel et al. 2008). In this work, we have proposed, that statistical patterns extracted from the Web via machine learning should be integrated into queries (Tresp et al. 2009). Thus a search for diseases associated with a particular gene can be done in three ways: First, one can study the documents returned via keyword-based search. Second, one can obtain a list of diseases known to be, or suspected to be, associated via a structured query on LOD. Finally, one can use machine learning to extract diseases likely to be related to the gene based on disease and gene attributes and established gene-disease patterns. Note that machine learning depends on repeatable statistical patterns: thus machine learning cannot help to give you the first name of the wife of the US President (a one-to-one relationship), but it can predict the probability of his re-election, his income, the party of his vice president and the number of his expected grandchildren.

[1] Although the world might be governed by scientific laws and logical constraints in general, at the level of abstraction that we and our applications have to function, the world partially appears to be governed by probabilities and statistical patterns.

In the next section we discuss some of the challenges encountered when applying machine learning to LOD. In Sect. 3 we motivate and describe our particular approaches. Section 4 describes a number of applications. One of them is BOTTARI, the winning entry in the ISWC 2011 Semantic Web Challenge.[2] In Sect. 5 we describe extensions and future work. Section 6 contains our conclusions.

2 Challenges for Machine Learning

Machine learning is not only of interest to the Web, the Web also poses interesting research challenges to machine learning. First of all, web data typically does not represent an i.i.d. (independent and identically distributed) statistical sample of some sort but might have been collected and published for any reason, often not systematically. For similar reasons, the data, in general, is incomplete, e.g., from the fact that if a social network lists no friends of Jack one cannot conclude that Jack does not have any friends. In general, negation is very rare on web data, thus one might find information that two persons are friends but rarely that two persons are not friends. This needs to be considered to avoid biased predictions. Another interesting property of web data is that relationships between entities are often more informative than entity attributes, an effect exploited in collective learning: It might be easier to predict Jane's wealth from the wealth of her friends than from known properties of Jane. As in this example, nonlocal aggregated information is often informative for a prediction task and machine learning needs to take this into account. Sometimes, as in the examples mentioned in the introduction, relationships themselves are of interest, e.g., item preferences, friendships, relationships between genes and diseases. Since the number of potential relationships can be very large, the output of a learning system will often be a ranked list of candidate relationships, e.g., a ranked list of recommended items, instead of a single answer. As a particular feature of web data, there is often textual information available that describes entities (e.g., Wikipedia articles), events (e.g., news stories) or topics (e.g., blogs), and this information can often be very useful for the machine learning task. Finally, a machine learning system has to be able to handle the large scale of the Web, its dynamical nature and its noisiness.

3 Predicting Facts with Factorization

In LOD, basic facts are represented as subject-predicate-object triples (s, p, o). In our work, we have been addressing the challenge of using machine learning to predict the likelihood of triples that are not explicitly given in the data. Since triples can

[2] http://challenge.semanticWeb.org/2011/

describe class membership (Jane, rdf:type, Student), entity attributes (Jane, income, High) and relationships (Jane, likes, Jack), triple prediction is a quite general task. Equivalently, we might look at the LOD graph where the nodes represent the entities and a directed link represents (s, p, o) triple, labeled by the predicate. In this view the learning tasks consist of predicting the existence of labeled links not explicitly given in the graph.

Some of the most powerful learning approaches, effective for predicting links in a graph with properties as discussed in the last section, rely on a description of an entity in a latent space. This means that each entity is described by a number of features which might or might not have a real-world meaning and which are abstracted from the information describing the entity. One example would be a cluster assignment, e.g., a student might belong to the cluster of "good students". Another abstract representation can be obtained via factorization approaches. Here an entity is described by degrees of agreement with certain factors. As an example, a student might agree with the factor "good students" to some degree, and with the factor "popular students" to some other degree. It turns out that latent factors are not simply helpful for understanding a domain but are also very effective in link prediction. In a factorization approach, the likelihood of the existence of a link is determined by the scalar product of the latent factors describing the associated entities.[3] Tresp et al. (2009) and Huang et al. (2010) describe the SUNS framework, which is the particular factorization approach developed in our work.

Factorization approaches provide high-quality predictions and are robust to the challenges described in Sect. 2. In particular they are highly scalable by exploiting the sparsity in the data and are suitable for making use of relationship information and for predicting relationships. Extensions towards deductive reasoning, for the inclusion of textual information, and for addressing the dynamical nature of data, will be discussed in Sect. 5.

---

[3] In particular, the probability that a relationship between two entities exists given the knowledge base *KB* is estimated as

$$\hat{P}((Jane, likes, Jack)|KB) = \sum_{i=1}^{L} f_i^{Jane} f_i^{likes,Jack}$$

where $\{f_i^{Jane}\}_{i=1}^{L}$ are the L factors describing Jane, and $\{f_i^{likes,Jack}\}_{i=1}^{L}$ are the L factors describing Jack in his role as an object of the predicate "likes". There are a number of approaches for calculating the factors. In our work in the SUNS framework (Huang et al. 2010; Tresp et al. 2009), we have employed regularized factorization of the associated data matrices. In our three-way tensor approach RESCAL (Nickel et al. 2011), we estimate

$$\hat{P}((Jane, likes, Jack)|KB) = \sum_{i=1}^{L} f_i^{Jane} R^{likes} f_i^{Jack} .$$

Each entity has a unique latent representation, here $\{f_i^{Jane}\}_{i=1}^{L}$ and $\{f_i^{Jack}\}_{i=1}^{L}$, and the relation type specific interaction is modeled by the matrix R^{likes}.

```
PREFIX ya: http://blogs.yandex.ru/schema/foaf/
PREFIX foaf: http://xmlns.com/foaf/0.1/
PREFIX dc: http://purl.org/dc/elements/1.1/
SELECT DISTINCT ?person
WHERE {
  ?person ya:located ?city .
  ?person foaf:knows <http://trelana.livejournal.com/trelana>
  WITH PROB ?prob .
  FILTER REGEX(?city, Munich ) .
}
ORDER BY DESC(?prob)
```

Fig. 1 The query includes the predicted *knows* triples for Trelena and rates them by predicted probability

4 Querying the Web with Machine Learning

In this section we describe some applications of machine learning to LOD. For more details, see the respective references.

Querying Social Networks The experiments presented in this section are based on friend-of-a-friend (FOAF) data, which is part of LOD. The purpose of the FOAF project (Brickley and Miller) is to create a web of machine-readable pages describing people, their relationships, their activities and their interests using W3C's RDF technology. The population is defined by the 32,062 persons in our FOAF subset. 14,425 features are formed by potential friends in the data. Furthermore, 781 attributes refer to general information about age, location, number of blog posts, attended school, online chat accounts and interests. The task is to predict potential friends of a person, i.e., *knows* statements, and the performance is evaluated using a test set of known friendships. In a comparison with competing methods, the factorization approach gave the best performance in predicting new friendships (Huang et al. 2010). The following SPARQL expression illustrates a query for LiveJournal[4] users who live in Munich and might want to be Trelena's friend (see Fig. 1).

The query contains an extended clause WITH PROB that returns the estimated probabilities of a friendship relationship for Trelena, modeled by the *knows* relation. Figure 2 presents a typical query response.

Querying Linked Life Data Life science data forms a significant part of the LOD cloud. To a large extent, this data has been extracted from well-maintained databases such that this portion of LOD is of high quality. We applied our approach to an important problem in the life sciences, i.e., the prediction of gene-disease relationships, and demonstrated that we obtained results competitive those of state-of-the-art solutions.

[4]http://www.livejournal.com/

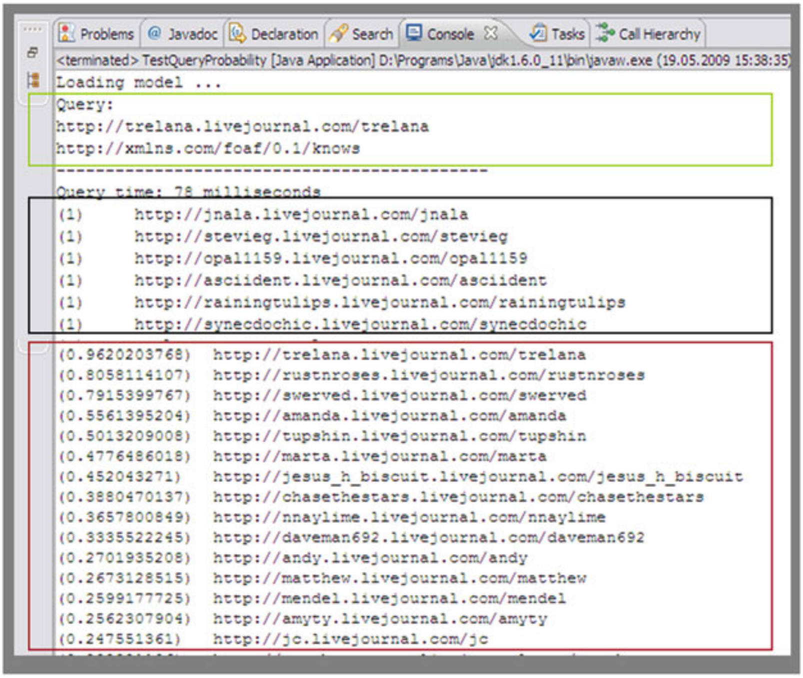


Fig. 2 Query: Who wants to be Trelena's friend. Her actual friends are predicted first with confidence values. Then, interestingly, it is predicted that she should be her own friend, followed by a ranked list of predicted friends

Disease genes are genes involved in the causation of, or associated with, a particular disease. At this stage, more than 2,500 disease genes have been discovered. Unfortunately, the relationship between genes and diseases is far from simple since most diseases are polygenic and exhibit different clinical phenotypes. High-throughput genome-wide studies like linkage analysis and gene expression profiling typically result in hundreds of potential candidate genes and it is still a challenge to identify the disease genes among them. One reason is that genes can often perform several functions and a mutational analysis of a particular gene reveals dozens of mutation sites that lead to different phenotype associations to diseases like cancer (Kann 2010). Analysis is further complicated when environmental and physiological factors come into play as well as by exogenous agents such as viruses and bacteria.

Despite this complexity, it is quite important to be able to rank genes in terms of their predicted relevance for a given disease. Such a ranking can not only be a valuable tool for researchers but also has applications in medical diagnosis, prognosis, and personalized treatment of diseases.

Gene properties differentiate disease genes and have been used as the bases for computational tools to prioritize disease gene candidates. All of the current approaches are based on the integration of different properties such as: gene function (disease genes are expected to share common functional properties), pathways (disease genes are most likely to share common pathways), gene expression (disease genes are expected to be co-expressed), gene regulation (genes within the same gene-regulation network are expected to affect similar diseases), sequence properties, and protein interaction (disease genes are often highly connected with other genes from the same disease). These attributes have been used in our approach as well. In addition, our approach exploits gene-disease interaction patterns. The solution was integrated into the THESEUS MEDICO use case. In seven out of 12 experiments the machine learning approach using the SUNS framework was superior to a leading competing approach based on a fuzzy-based similarity measure. See Huang et al. (2013) for a detailed description of experimental results.

BOTTARI: Personalized and Location-Based Recommendations BOTTARI is an augmented reality application that permits the personalized and localized recommendation of points of interest (POIs) based on the temporally weighted opinions of the community. Opinions on POIs (here: restaurants) were extracted from Twitter[5] microposts with natural language processing and a background ontology. The task of the machine learning module was to rank different restaurants based on the extracted information and based on individual preference profiles of users. A successful evaluation of BOTTARI was carried out using a 3 year collection of tweets of about 319 restaurants located in the 2 km^2 district of Insadong, a popular tourist area of the South Korean city of Seoul.

The technological basis of BOTTARI is the highly scalable LarKC platform for the rapid prototyping and development of Semantic Web applications. BOTTARI is the winner of the ninth Semantic Web Challenge, co-located with the 2011 International Semantic Web Conference. BOTTARI is currently being field tested in Korea by Saltlux (Balduini et al. 2013).

More Examples *DBpedia* (Auer et al. 2008) is part of LOD and contains structured information extracted from Wikipedia. At the time of this writing, it describes more than 3.4 million concepts, including, e.g., 312,000 persons, 413,000 places and 94,000 music albums, DBpedia does not only serve as a “nucleus for the web of data”, but also holds great potential for being used in conjunction with machine learning approaches. DBpedia already provides great value and is useful for accessing facts by answering queries for the famous citizens and the most spectacular sights of a large number of cities. DBpedia data is increasingly getting interesting for machine learning purposes as new and richer relationships are added. In our experiments, we used a population consisting of all members of the German Bundestag to evaluate our approach (Huang et al. 2013). The task was to predict the party membership based on age, state of birth, and keywords from the Wikipedia pages of the Bundestag members as input information. Most informative were the latter two sources, in particular the state information, which can be explained by the

[5]https://twitter.com/

peculiarities of German politics. Overall, we obtained a system that achieved high accuracy. Although there might not be any economic value in this experiment, it provides a first insight into how DBpedia can be queried in the future using machine learning.

The SUNS approach was used in a prototype for analyzing the usage service patterns and for recommending services in a web service platform as part of the THESEUS TEXO use case and in a prototype application for airline claim managements, presented at the CeBIT 2011.

5 Extensions

The factorization approaches briefly described in Sect. 3 are surprisingly general and powerful. In this section we described some important extensions. As mentioned, we used machine learning to predict triples, or labeled links, based on statistical patterns in the data whereas deductive reasoning derives triples using facts and logical expressions, such as rules. Machine learning can easily benefit from deductive reasoning by including the derived triples in training and prediction.

Deductive reasoning can be helpful for aggregation as well. In many applications, information that is not local might become relevant for machine learning. As an example, in the DBpedia experiment party membership is more easily predicted from a politician's state of birth than from his city of birth; however only the latter is explicitly stated in DBpedia. The state can be derived by using geo-reasoning from the city prior to learning. Materialization of knowledge by making implicit knowledge explicit via computing the inductive closure is a highly scalable approach to reasoning.[6]

Important sources of information are often documents describing the involved entities or relations between entities, as shown in the DBpedia experiments in Sect. 4, and textual information can support triple prediction in general. The combination of information extraction from text, deductive reasoning and machine learning to improve triple prediction is described in a probabilistic extension of the factorization approach in Jiang et al. (2012a,c).

Another interesting aspect concerns sequence and temporal information. Often a series like "Star Wars" will be watched in order. Similarly, medical procedures are given in a sensible sequential order. In Tresp et al. (2011) an extension to the factorization model is described that can model both sequential information and absolute time.

In the SUNS approach, triples were mapped to one or several matrices and the latent factors were calculated via a factorization of these matrices. It might be argued that a more natural representation for LOD's triple structure is given by a three-way tensor. Whereas in a matrix an element is addressed by two indices, in a three-way

[6] http://www.larkc.eu/

tensor an entry is addressed by three indices. Nickel et al. (2011) describes an approach where all entities of a domain are mapped to two modes of a three-way tensor and the predicates are mapped to the third mode. In this tensor, an element equal to one indicates that the corresponding (s, p, o) triple is known to be true. For this representation of triples, a particular three-way factorization was developed which permits us to exploit nonlocal information without explicit aggregation by collective learning. Furthermore, in Nickel et al. (2012) this approach was applied to the sizable YAGO-2 ontology, demonstrating its scalability. In Jiang et al. (2012b) an additive model is described that attempts to combine the simplicity of the SUNS approach with some of the powerful features of a tensor model.

6 Conclusion

In this article we have argued that machine learning might become a third important way to access web information, in addition to keyword-based search and structured querying. We have provided examples that illustrate for what kind of queries machine learning might be effective. We have discussed the machine learning approaches pursued in the work in THESEUS and LarKC, which are based on the factorization of matrices and tensors. We expect that machine learning researchers will increasingly consider the Web as a great data source for learning applications. In particular the joint exploitation of different knowledge sources like web documents, Wikipedia, published databases as part of LOD, and other background information poses new interesting research challenges.

References

S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives, DBpedia: a nucleus for a web of open data, in *Proceedings of the 6th International Semantic Web Conference (ISWC'08)*, Karlsruhe. Volume 4825 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2008), pp. 722–735

M. Balduini, I. Celino, D. Dell'Aglio, E.D. Valle, Y. Huang, T. Lee, S.H. Kim, V. Tresp, Reality mining on micropost streams: deductive and inductive reasoning for personalized and location-based recommendations. Semant. Web Interoperability Usability Applicability **2**, 1–16 (2013)

C. Bizer, T. Heath, T. Berners-Lee, Linked data – the story so far. Int. J. Semant. Web Inf. Syst. (IJSWIS) **5**(3), 1–22 (2009)

D. Brickley, L. Miller, The Friend of a Friend (FOAF) project, http://www.foaf-project.org/

D. Fensel, F. van Harmelen, B. Andersson, P. Brennan, H. Cunningham, E.D. Valle, F. Fischer, Z. Huang, A. Kiryakov, T.K. il Lee, L. Schooler, V. Tresp, S. Wesner, M. Witbrock, N. Zhong, Towards LarKC: a platform for web-scale reasoning, in *Proceedings of the IEEE International Conference on Semantic Computing*, Santa Clara, Aug 2008, pp. 524–529

Y. Huang, V. Tresp, M. Bundschus, A. Rettinger, H.P. Kriegel, Multivariate structured prediction for learning on semantic web, in *Proceedings of the 20th International Conference on Inductive Logic Programming (ILP'10)*, Florence, ed. by P. Frasconi, F.A. Lisi. Volume 6489 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2010), pp. 92–104

Y. Huang, V. Tresp, M. Nickel, A. Rettinger, H.P. Kriegel, A scalable approach for statistical learning in semantic graphs. Semant. Web Interoperability Usability Applicability **1**, 1–18 (2013)

X. Jiang, Y. Huang, M. Nickel, V. Tresp, Combining information extraction, deductive reasoning and machine learning for relation prediction, in *Proceedings of the 9th Extended Semantic Web Conference (ESWC'12)*, Heraklion, ed. by E. Simperl, P. Cimiano, A. Polleres, O. Corcho, V. Presutti. Volume 7295 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2012a), pp. 164–178. http://dblp.uni-trier.de/db/conf/esws/eswc2012.html#JiangHNT12

X. Jiang, V. Tresp, Y. Huang, M. Nickel, Link prediction in multi-relational graphs using additive models, in *Proceedings of the 11th International Workshop on Semantic Technologies Meet Recommender Systems & Big Data*, ed. by M. de Gemmis, T.D. Noia, P. Lops, T. Lukasiewicz, G. Semeraro. Volume 919 of CEUR Workshop Proceedings, 2012b, pp. 1–12, CEUR-WS.org. http://dblp.uni-trier.de/db/conf/semweb/sersy2012.html#JiangTHN12.

X. Jiang, V. Tresp, Y. Huang, M. Nickel, H.P. Kriegel, Scalable relation prediction exploiting both intrarelational correlation and contextual information, in *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD'12)*, Bristol, ed. by P.A. Flach, T.D. Bie, N. Cristianini. Volume 7523 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2012c), pp. 601–616. http://dblp.uni-trier.de/db/conf/pkdd/pkdd2012-1.html#JiangTHNK12

M.G. Kann, Advances in translational bioinformatics: computational approaches for the hunting of disease genes. Brief. Bioinform. **11**(1), 96–110 (2010). http://dblp.uni-trier.de/db/journals/bib/bib11.html#Kann10

M. Nickel, H.P. Kriegel, V. Tresp, A three-way model for collective learning on multi-relational data, in *Proceedings of the 28th International Conference on Machine Learning (ICML'11)*, Bellevue, 2011

M. Nickel, V. Tresp, H.P. Kriegel, Factorizing YAGO: scalable machine learning for linked data, in *Proceedings of the 21st International World Wide Web Conference*, Lyon, ed. by A. Mille, F.L. Gandon, J. Misselis, M. Rabinovich, S. Staab (ACM, 2012), pp. 271–280. http://dblp.uni-trier.de/db/conf/www/www2012.html#NickelTK12

V. Tresp, Y. Huang, M. Bundschus, A. Rettinger, Materializing and querying learned knowledge, in *Proceedings of the First ESWC Workshop on Inductive Reasoning and Machine Learning on the Semantic Web (IRMLeS'09)*, Heraklion, vol. 474 (RWTH Aachen, 2009)

V. Tresp, Y. Huang, X. Jiang, A. Rettinger, Graphical models for relations – modeling relational context, in *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (KDIR'11)*, Paris, Oct 2011

Automatic Assessment of Image Quality

Thomas Sporer, Kristina Kunze, and Judith Liebetrau

Abstract Huge amounts of film and image content are in archives stored on analog media, slowly fading away. In professional environments new content is captured electronically. There has always been the wish to capture or reformat, to store and transmit images electronically in very high quality, but only today have the sensors been developed in a way that the capturing can be done at reasonable cost for everybody. These high quality levels for still and moving pictures come with high amounts of data. Therefore encoding is necessary for storage and transmission. In general, the encoding process is not lossless, but the aim is to preserve the perceived quality. Until recently, the only way to assess the quality of encoding schemes used to be visual tests, but meanwhile measurement schemes have been standardized in ITU-R and ITU-T which can estimate the perceived quality if some conditions are met: the measurement scheme must have access to the raw, non-coded, digital content, and the encoding scheme must be in a family of well-known coding schemes. Non-reference (NR) measurement and measurement of innovative coding schemes are an open issue. Within THESEUS, two new NR predictors were developed and their performance was evaluated. In this paper, the procedure for and the results of the assessment of a measurement scheme which estimates quality without access to the reference are presented.

T. Sporer (✉) • J. Liebetrau
Fraunhofer Institute for Digital Media Technology (IDMT), Ilmenau, Germany
e-mail: spo@idmt.fraunhofer.de; ltu@idmt.fraunhofer.de

K. Kunze
Ilmenau University of Technology, Illmenau, Germany
e-mail: kristina.kunze@tu-ilmenau.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_19,

1 Introduction

Digital cameras either as standalone devices or included in mobile phones enable the capturing of high resolution pictures and films. The amount of digital data created that way can only be handled if compression is applied. Some examples might illustrate this: A (still) picture resolution of $2{,}048 \times 1{,}536$ at 24 bit per pixel would result in about 9 MB. A video clip in HDTV format (resolution of $1{,}920 \times 1{,}080$, frame rate 50 Hz, 2 : 1) would result in a SATA rate of about 800 Mbit/s. On the other hand, mobile transmission channels can provide up to 220 kbit/s (GSM/EDGE), 84 Mbit/s (HDSPA+), or 300 Mbit/s (LTE). All these rates are best case for a complete cell, and in a normal situation the rates are much lower.

To reduce the amount of data, picture and video coding are necessary. Lossless coding reduces the redundancy in the content. Sections of an image containing random structures contain hardly any redundancy and cannot be compressed at all. Such random structures might be related to the content (e.g. the structure of sand) or to the capturing device (noise of photo sensor at low lighting). The compression performance of lossless coding is therefore not constant. Therefore, for most applications lossy[1] coding is used. The key requirement for lossy coding is: The coded image should provide almost the same *perceived quality* as the raw image.

The final instances for judging the perceived quality of pictures are visual (subjective) tests, but such tests are time consuming and expensive and not possible in all situations. In such cases, objective measurement schemes, which mimic the rating of test subjects as closely as possible, are necessary. Simple measures of image quality like Peak-Signal-to-Noise-Ratio (PSNR) and Mean Square Error (MSE) do not sufficiently respect the properties of the human perception. It is necessary to consider different approaches for edges in the image ("high precision"), noisy areas ("noise with (statistically) the same structure"), and to look especially at typical coding artifacts like "blocking".

Measurement Principles Measurements of the quality of still pictures and moving pictures share the same principal concepts. A key problem of all measurement schemes is that the measurement system has to know what the image should look like ("what is good quality"). Humans have learned this in their childhood, and have stored expectations in their brain. For measurement schemes different approaches are possible for making assessment of quality reasonable.

In the *Full Reference* (FR) approach (Fig. 1) the non-coded original image is available to the measurement scheme and the measurement scheme compares this so-called reference to the output of a transmission.

The major advantage of this approach is that any input can be used (even an input picture which already has coding artifacts). Unfortunately, this measurement method is only applicable when a reference is available; in several situations it is not possible to access the reference picture for measurement.

---

[1]The term lossy coding is used for codes where there is a difference between the input and output.

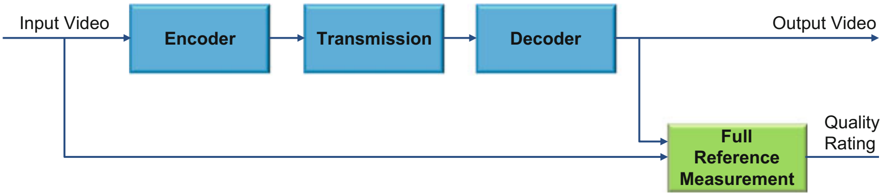


Fig. 1 Full reference management

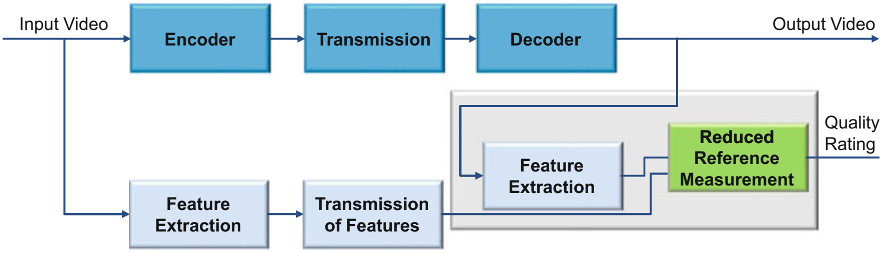


Fig. 2 Reduced reference management

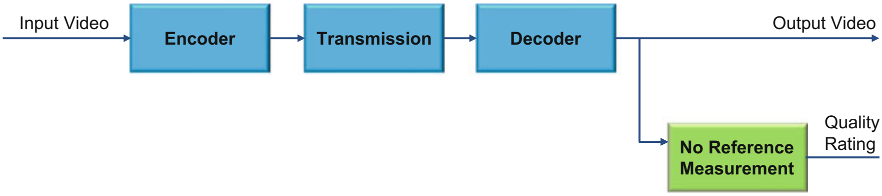


Fig. 3 No reference management

An approach to measuring the quality, even in absence of a reference picture, is the usage of a *Reduced Reference* (RR) approach (Fig. 2): A set of features of the input picture is extracted and transmitted to the measurement system. The measurement system extracts the same features from the output picture and compares these feature sets. Usually the data rate necessary to transmit the features is much smaller than the data rate of the original picture and even smaller than the data rate necessary for the encoded picture. However, this approach is only applicable if an original reference (raw data) exists. Examples of applications where no reference exists are the digitalization and restoration of old films, but also the processing of (digital) archives in which only the compressed content has survived. In this last application the information about quality of the image is an important component in the metadata set describing the content: Journalists often have to choose between different clips describing the same historical event and the selection of the clip also has to consider the (different) picture quality.

Table 1 Full reference measurement

ITU-T	ITU-R	Principle	Resolution	Year
J.144	BT.1683	FR	SDTV	2004
J.247	BT.1866	FR	Multimedia	2008/2010
J.249	BT.1885	RR	SDTV	2011
J.341	BT.1907	FR	HDTV	2011/2012
J.342	BT.1908	RR	HDTV	2011/2012

Measurement schemes using the *No Reference* (NR) approach (Fig. 3) can provide this information. Having no access to the original, such measurement schemes have to make certain assumptions about the content to process. Very often such schemes also make assumptions about the artifacts created by coding schemes and try to "concentrate" on the detection of such artifacts.

International Activities on Quality Assessment In radio and telecommunication the selection of coding schemes important aspect for interoperability and quality of service. Therefore two sectors of the International Telecommunication Union (ITU), namely the Telecommunication Sector (ITU-T) and the Radio Communication Sector (ITU-R), have joined resources with external parties and formed the *Video Quality Expert Group* (*VQEG*).[2] Started in 1997, this group covers subjective and objective assessment.

Table 1 lists the standards for measurement resulting from this work. Some of these recommendations have been published in different years in ITU-R and ITU-T. Therefore 2 years are given in the publication column. The column "resolution" denotes the image resolution of the image: SDTV is conventional TV ("standard definition", 720 × 576@50 Hz, 720 × 480@60 Hz) and HDTV ("high definition", 1,920 × 1,080 and 1,280 × 720). Recommendation J.247 specifies the measurement for lower resolutions like CQIF, CIF and VGA and is intended for mobile devices.

Limitations of Standardized Measurement Schemes The major limitation of all these recommendations can easily be shown just by citing from the text of the recommendations:

> The video quality estimation model described in this Recommendation *cannot be used to replace subjective testing*. Correlation values between two carefully designed and executed subjective tests (i.e. in two different laboratories) normally fall within the range 0.95 to 0.98.
> This Recommendation *cannot be used* to make video system comparisons (e.g. *comparing two codecs*, comparing two different implementations of the same compression algorithm).
> The performance of the video quality estimation model described in this Recommendation is *not statistically better than* PSNR.
> [. . .]
> It should be noted that in case of *new coding* and transmission technologies producing artefacts which were not included in this evaluation, the objective model may produce erroneous results. Here, a *subjective evaluation* is required.
> (Source: International Telecommunication Union (ITU) 2012a,b)

[2] http://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx

Similar caveats can be found in all these recommendations. To summarize: It is believed that these methods only work reliably if exactly the same codices are tested which had been used in the verification of the recommendations.

Application Areas Not Covered An application which is not addressed by these recommendations is the quality assessment for the digitalization and restoration process for films. Today, the quality control is done by human observers, which limits the processing speed dramatically. At the same time, human observers are prone to fatigue and loss of concentration. As mentioned earlier, a measurement system which could replace this human quality control step must be a "no reference" system. Currently no standardized "no reference" schemes exist. THESEUS CORE TECHNOLOGY CLUSTER Work Package 2 developed a new NR measurement scheme and Work Package 8 performed evaluations on the performance of this scheme. The outline of the test procedure and some results are given in the next sections.

2 Test Procedure

General Requirements for Validation of Perceptual Measurement Schemes A major step for the evaluation of the performance of a measurement scheme is the definition of quality. In this context the "gold standard" is the visual experiment with human observers. This subjective evaluation had to follow a formal test procedure (see next section). The usage of the scale in subjective tests in general is dependent on the user group, the content and the spread of quality levels of stimuli within each test. Therefore the comparison of the subjective and the objective data is not an easy task. Within VQEG several metrics for the comparison have been used. They either are looking at the correlation between the two datasets or at different ways to weight the difference between these (prediction error).

In general the assessment of a measurement scheme has to follow a five-step approach:

1. A first set of images with different quality levels is evaluated by human observers. In the following this dataset is called the training dataset (DB1).
2. Images and subjective results for DB1 are used in a supervised training of the measurement schemes which should be evaluated.
3. A second set of images with different quality levels is evaluated by human observers. This dataset is referred to as the verification dataset (DB2).
4. The measurement scheme to be evaluated uses the image data of DB2 and predicts the quality without knowledge of the subjective results from step 3.
5. The results of the subjective tests (step 3) and the objective tests (step 4) are compared.

To get sufficiently reliable information about the performance of the measurement system it is essential that the images used in DB1 and DB2 be different. The incorporated classes of artifacts and their degree have to be the same for both training and dataset. The test subjects used in step 1 and step 3 must be from the same population. The test results obtained in stage 3 must be kept secret before the end of step 4.

Subjective Evaluation As mentioned earlier, the subjective test must follow a formal procedure. In Recommendations ITU-R BT.500 (International Telecommunication Union (ITU) 2012c) and ITU-T P.910 (International Telecommunication Union (ITU) 1999) several test methods are standardized.

For the creation of databases there are several requirements: The database must be sufficiently large and must cover all kinds of artifacts which should be detected by the measurement scheme. In the scientific community, the comparison of results of different measurement schemes is often done by using the same databases. On the other hand these databases are not unknown and cannot be used for verification. To enable comparison with evaluations conducted by other groups a combination of (internationally) known and unknown databases was used here.

Measurement Scheme to be Evaluated This section describes the objective evaluation of NR measurement schemes developed in THESEUS during the last few years of the research program. The work on assessment of FR and RR methods done in the first years of THESEUS is not reported here.

As mentioned in Sect. 1 NR methods must be based on assumptions of the properties of the images and/or assumptions on the structure of typical coding artifacts. The method developed in THESEUS followed both approaches: Two descriptors have been developed, namely a descriptor for "blocking artifacts" (a typical artifact of picture coding algorithms) and a descriptor for "blur" in the image (an artifact which can be related to digitization or post-processing of images). The combination of these two descriptors in one measurement value which uses the same scale as the subjective tests was not done in THESEUS.

Databases, State-of-the-Art Algorithms, and Metric Four databases have been used:

- IDMT II is a database especially for THESEUS,
- CSIQ is a database from Larson and Chandler (2010),
- LiveLabs is database available from the Internet (Sheikh et al.),
- TID2008 is database from Ponomarenko et al. (2009).

Due to the fact that LiveLabs and TID2008 had been used during the development of the NR measurement scheme within the THESEUS research program, they have to be regarded as "known databases". Although the CSIQ dataset was publicly available, it was not used for training of the algorithm, and can hence be called "unknown". IMDT II was always kept secret.

To estimate the performance of the new quality descriptors, two state-of-the-art measurement schemes have been used for comparison: The peak signal to noise ratio (PSNR) is easy to implement and widely used by developers of image coding schemes. The origin of this method dates back to the 1980s (Dicus 1981). The Structural similarity (SSIM) (Wang et al. 2004) takes into account that humans

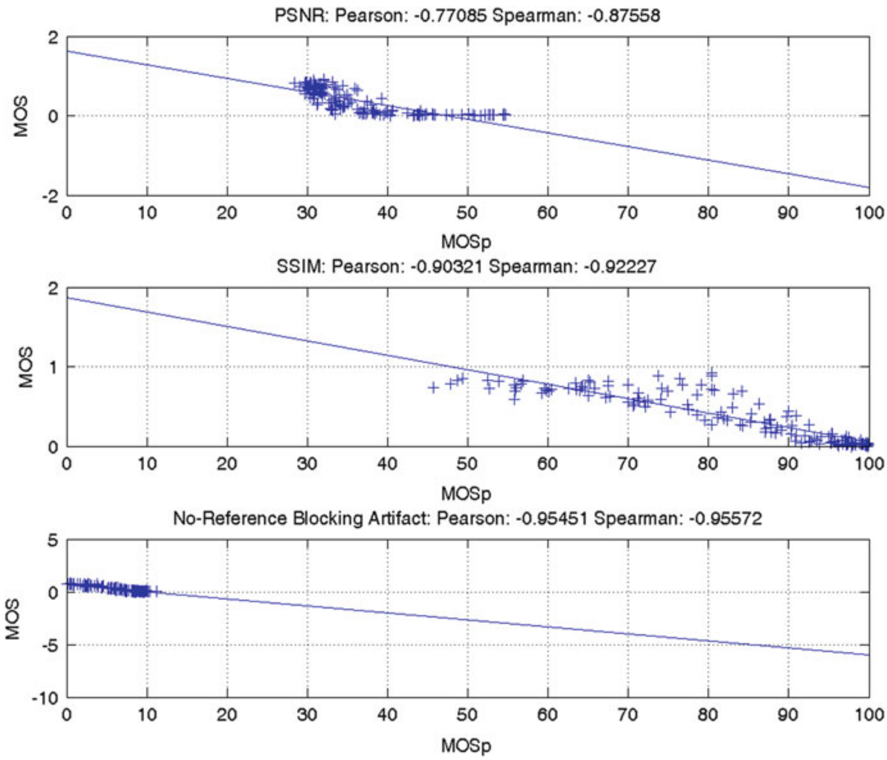


Fig. 4 Results CSIQ – blocking artifact/JPEG

perceive amongst others the structure of images and therefore are sensitive to changes of the structures. A comparison of PSNR and SSIM can be found in Horé and Ziou (2010).

All methods were used to retrieve objective estimations of the perceived visual quality of the pictures. To determine the accuracy of the measurement schemes, scatter plots and correlation of subjective and objective results have been used. To respect different (non-linear) usage of the scale by observers both the Pearson's correlation and the Spearman Rank-Order correlation was calculated. Due to the fact that the raw descriptors are evaluated, the output of the measurement is on a different scale than the subjective data. Therefore no error measures were calculated.

3 Results

Figures 4 and 5 show results for all measurement schemes obtained with the CSIQ database. The subjective dimension is labeled MOS (Mean Opinion Score), the objective dimension is labeled MOSp (Predicted MOS). The database had been split

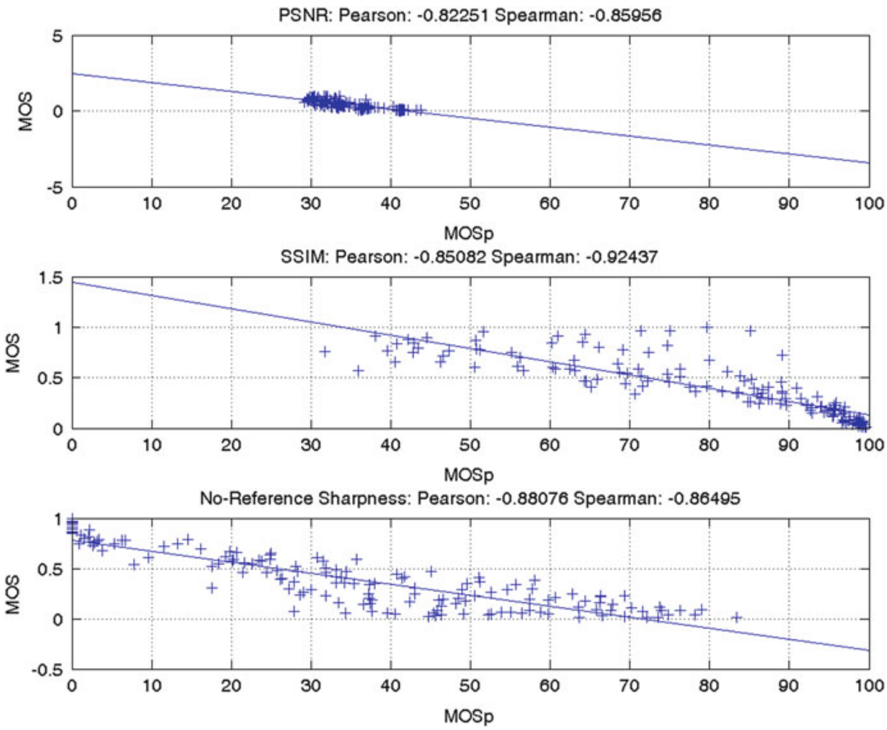


Fig. 5 Results CSIQ – Blur

into two parts showing either blocking or blurring artifacts only. Figure 4 shows the results for blocking and Fig. 5 for the blurring.

At a first glance the performance of the three methods seems to be similar, but it has to be mentioned that PSNR and SSIM are FR methods, while the two descriptors from THESEUS are NR methods: Even without knowledge of the reference they often outperform the other two methods.

Table 2 shows a summary of the performance for the databases CSIQ, LiveLabs, and TID2008. For all these databases the performance of the THESEUS descriptors are very good. It has to be mentioned that in the CSIQ dataset, lower MOSscores denote higher quality. This is contrary to other sets, where the highest MOSscore denotes the highest quality. This leads to negative correlations of all metrics with this dataset.

The results for database IDMT II are not shown here due to several reasons: This database consistently shows conspicuous results for all three measurement schemes. This might be caused by obvious properties of this database: The artifacts in IDMT II are much smaller than those in the other databases: Only a very small number of items show distinct blocking and blurring artifacts, and originally the majority of items used in the subjective tests showed “real-world artifacts” (=mixture of different artifacts). Looking at the differences between the quality judgments of the subjects, a huge variance could be observed.

Table 2 Reduced reference measurement

	Pearson's correlation					
	Blocking			Blur		
	PSNR	SSIM	THESEUS	PSNR	SSIM	THESEUS
LiveLabs	0.88	0.87	0.84	0.80	0.87	0.89
TID2008	0.83	0.90	0.93	0.87	0.91	0.75
CSIQ	−0.77	−0.90	−0.95	−0.82	−0.85	−40.88
	Spearman's Rank Order correlation					
	Blocking			Blur		
	PSNR	SSIM	THESEUS	PSNR	SSIM	THESEUS
LiveLabs	0.93	0.96	0.93	0.84	0.94	0.89
TID2008	0.86	0.90	0.92	0.86	0.94	0.79
CSIQ	−0.88	−0.92	−0.96	−0.86	−0.92	−0.86

However, the results for this database point to another important research topic: The subjective test procedures have to be improved to get reliable data in the presence of only small impairments.

4 Conclusion

Several full and reduced reference methods for the measurement of picture quality are standardized. All these standards have limitations in performance and reliability. A generalized usage of the methods for different content and coding schemes is not possible. In addition these methods require the knowledge of a reference picture or at least the key features of the picture. In the targeted application, scenarios for Internet-based knowledge infrastructure, a measurement scheme for picture quality without known reference, is required. Such "no reference" (NR) methods have not been standardized yet. A first attempt to develop individual NR quality descriptors within THESEUS has been made. The evaluation of these gives evidence that they are working well. They predict the picture quality reliably but with some constraints; especially the prediction of small impairments is insufficient. Also the combination of these descriptors in one measurement value is still missing.

Before measurement schemes for small impairments can be developed and verified, it is necessary to create reasonable training datasets. Therefore, also the procedure for subjective tests for generating the datasets has to be improved.

References

R.L. Dicus, Impulse response estimation with underwater explosive charge acoustic signals. J. Acoust. Soc. Am. **70**(1), 122–133 (1981)

A. Horé, D. Ziou, Image quality metrics: PSNR vs. SSIM, in *20th International Conference on Pattern Recognition (ICPR'10)*, Istanbul (IEEE, 2010), pp. 2366–2369. http://dblp.uni-trier.de/db/conf/icpr/icpr2010.html#HoreZ10

International Telecommunication Union (ITU), Recommendation P.910: subjective video quality assessment methods for multimedia applications. Technical report, 1999

International Telecommunication Union (ITU), Recommendation ITU-R BT.1907: objective perceptual video quality measurement techniques for broadcasting applications using HDTV in the presence of a full reference signal. Technical report, 2012a

International Telecommunication Union (ITU), Recommendation ITU-R BT.1908: objective video quality measurement techniques for broadcasting applications using HDTV in the presence of a reduced reference signal. Technical report, 2012b

International Telecommunication Union (ITU), Recommendation ITU-R BT.500-13: methodology for the subjective assessment of the quality of television pictures. Technical report, 2012c

E.C. Larson, D.M. Chandler, Most apparent distortion: full-reference image quality qssessment and the role of strategy. J. Electron. Imaging **19**, 1–21 (2010). http://dblp.uni-trier.de/db/journals/jei/jei19.html#LarsonC10

N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, V. Lukin, Metrics performance comparison for color image database, in *Proceedings of the 4th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM'09)*, Scottsdale, Jan 2009

H. Sheikh, Z. Wang, L. Cormack, A. Bovik, LIVE image quality assessment database release 2, http://live.ece.utexas.edu/research/quality/

Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004). http://dblp.uni-trier.de/db/journals/tip/tip13.html#WangBSS04

Evaluation of Image Annotation Using Amazon Mechanical Turk in ImageCLEF

Judith Liebetrau, Stefanie Nowak, and Sebastian Schneider

Abstract With the increasing amount of digital information in the Web and on personal computers, the need for systems that are capable of automated indexing, searching, and organizing multimedia documents is incessantly growing. Automated systems have to retrieve information with high performance in order to be accepted by industry and end users. Multimedia retrieval systems are often evaluated on different test collections with different performance measures, which makes the comparison of retrieval performance impossible and limits the benefits of the approaches. Benchmarking campaigns counteract these tendencies and establish an objective comparison among the performance of different approaches by posing challenging tasks and by pushing the availability of test collections, topics, and performance measures. As part of the THESEUS research program, Fraunhofer IDMT organized the "Visual Concept Detection and Annotation Task" (VCDT) of the international benchmark ImageCLEF, with the goal of enabling the comparison of technologies developed within THESEUS CTC to international developments. While the test collection in 2009 was assessed with expert knowledge, the relevance assessments for the task have been acquired in a crowdsourcing approach since 2010 by using the platform of Amazon Mechanical Turk (MTurk). In this article the evaluation of THESEUS core technologies within ImageCLEF is explained in detail. A special focus lies on the acquisition of ground truth data using MTurk. Advantages and disadvantages of this approach are discussed and best practices are shared.

J. Liebetrau (✉) • S. Schneider
Fraunhofer Institute for Digital Media Technology (IDMT), Ilmenau, Germany
e-mail: ltu@idmt.fraunhofer.de; schnsn@idmt.fraunhofer.de

S. Nowak
Hella Aglaia Mobile Vision GmbH, Berlin, Germany
e-mail: research@stefanie-nowak.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_20,

1 Introduction

Within the THESEUS research program, basic technologies are developed and applied in six application scenarios. These scenarios show how the technologies can be used for innovative tools, services, and business models.[1] Experts evaluate the quality of these basic technologies according to requirements of the use cases.

For the majority of scenario applications, special interest lies in semantic search and therefore in the semantic analysis and interpretation of multimedia content. In particular, the categorization and recognition of the content of a picture are put in one focus of the program. One part of semantic analysis is image annotation. Image annotation refers to the task of automatically assigning words or attributes to images that describe characteristics of the image content. These so-called "concepts" include, for example, objects, scene descriptions, quality issues, representation of content, and sentiments. Image annotation is not to be confused with categorization tasks: While the latter describes images with one of several categories, the former applies multiple concepts to one image. In literature, image annotation is also referred to as: multi-label image classification, concept detection, automated photo tagging, high-level feature indexing, image auto-annotation, or image indexing.

Two technologies for automated image annotation were developed and evaluated within the THESEUS research program. The following sections describe the requirements for the assessment, the test methodology, and test data generation in detail.

2 Test Methodology

The performance assessment is conducted by comparing the results of the retrieval algorithm to human judgment: Automatically retrieved annotations of a test collection, specifically created for this purpose, are compared to annotations of this collection provided by humans. A comparison of the retrieval performances of different technologies is only possible when the same test collection is used. Multiple benchmarking initiatives focus on the performance assessment of multimedia systems with standardized test collections for different technologies (usually called "tasks"), such as ImageCLEF (Müller et al. 2010), PASCAL Visual Object Classes (VOC) (Everingham et al. 2010), and TREC Video retrieval evaluation (TRECVid) (Smeaton et al. 2006).

The performance of the THESEUS core technologies was assessed within ImageCLEF in the Visual Concept Detection and Annotation task. Existing test collections are often tailored to their respective challenges and cannot be easily applied to other tasks. As a result, a new test collection had to be built for each task, and several aspects concerning the design and setup of evaluation methodologies

[1] http://theseus.pt-dlr.de/en/about.php

needed to be taken into account: The acquisition of a realistic test collection has to be carried out with respect to the evaluation goal, including the definition of topics and the assessment of relevance. The definition of user preferences for query topics and concepts plays a key role. It must be ensured that systems are capable of indexing concepts the users are interested in, to add value to an application. The dataset and query tasks must reflect the user requirements on multimedia retrieval systems (Nowak 2011). The performance measurement needs to be unbiased and stable against random annotations. Lastly, the evaluation measures must fit the purpose of the assessment. The traditional performance measures Precision and Recall, F-Measure, Average Precision, Area-Under-Curve, or Equal Error Rate have emerged from research on text retrieval and are commonly applied to image retrieval and annotation systems. As part of the THESEUS research program, common performance measures were analyzed and several novel performance measures were proposed that directly adhere to the requirements of the use case scenarios (Nowak 2011).

Based on the requirements of the application scenarios, which include requirements on the diversity of test data, and an international comparison of results, Fraunhofer IDMT decided to evaluate the THESEUS core technologies as systems in an international benchmark. The evaluation of the photo annotation and classification performance of the technologies developed by Fraunhofer FIRST and HHI is performed with the test collection used in the "Visual Concept Detection and Annotation Task" (VCDT) of the international benchmark ImageCLEF. ImageCLEF is an initiative within the Cross Language Evaluation campaign and focuses on the evaluation of image retrieval and annotation algorithms (Müller et al. 2010).

Fraunhofer IDMT has been organizing the VCDT since 2009. The task of the participants is to automatically annotate a number of photos with visual concepts in a multi-label scenario. Fraunhofer IDMT provides the test collection including ground truth for the algorithm training, as well as the photos for the test, and performs the evaluation of all submitted runs. Every interested research group, whether from university or industry, can participate by signing an agreement and registering. The focus of VCDT lies on the automated detection and annotation of concepts in large consumer photo collections. It poses the challenge of classifiers having to be able to scale with large and unbalanced amounts of concepts and data. Hence a large test collection has to be generated, which fits the needs of the evaluation and adheres to real user needs. These requirements are derived from the focus of the THESEUS research program and its application scenarios.

Four application scenarios within THESEUS pose requirements regarding automated annotation of images: CONTENTUS uses semantic technologies to digitalize cultural heritage in the form of texts, images, audio, and video recordings. In ALEXANDRIA, a consumer-orientated knowledge database is created. ORDO uses semantic technologies to organize digital information. The goal of PROCESSUS is the optimization of business processes, including a basic semantic platform, which is able to manage the digital content of business processes. CONTENTUS requests the classification of images with "segment-wise extraction of texture descriptors for detection of texture structure and texture classes, like grass, stone, wood,

Table 1 The 53 concepts annotated in 2009

Number	Concept name	Number	Concept name
00	Partylife	22	Water
01	Family Friends	23	Lake
02	Beach Holidays	24	River
03	Building Sights	25	Sea
04	Snow	26	Mountains
05	Citylife	27	Day
06	Landscape Nature	28	Night
07	Sports	29	No Visual Time
08	Desert	30	Sunny
09	Spring	31	Sunset Sunrise
10	Summer	32	Canvas
11	Autumn	33	Still Life
12	Winter	34	Macro
13	No Visual Season	35	Portrait
14	Indoor	36	Overexposed
15	Outdoor	37	Underexposed
16	No Visual Place	38	Neutral Illumination
17	Plants	39	Motion Blur
18	Flowers	40	Out of Focus
19	Trees	50	Aesthetic Impression
20	Sky	51	Overall Quality
21	Clouds	52	Fancy

water, etc.". CONTENTUS further asks for "extraction of segment and descriptor distribution within the image for complex analysis methods for scene classification [...]". "Combining this information on a high semantic level can result in the identification of a landscape picture with sky, forest, and lake". ALEXANDRIA requests, that buildings, landscapes, or persons be detected. ORDO and PROCESSUS request the recognition of motives and faces in images.

3 Concept Annotation 2009

From these requirements, as well as additional input from user studies (Nowak 2011), a list of 53 concepts for visual annotation was developed as shown in Table 1.

The photos of the MIR Flickr 25,000 image dataset (Huiskes and Lew 2008) were chosen as test data in the task. This collection consists of 25,000 photos from Flickr published under the Creative Commons license. The ground truth assessment was realized in three steps. First, 43 Fraunhofer IDMT employees performed a manual annotation of all photos. All annotators were provided with a definition of the concepts and example images, with the goal of achieving consistent annotation among them. Second, a validation of these annotations was conducted. Due to the large number of people, the number of photos, and the ambiguity of some image content, the annotations were not consistent throughout the database. Three employees performed a screening and re-annotation of those images with obviously wrongly assigned or missing concepts (Nowak and Dunker 2010). In a third step, an agreement between different annotators for the same concepts and photos was calculated. Eleven different persons annotated a subset of 100 photos. For each photo and each concept, a majority vote of annotators was regarded as correct and the percentage of annotators that annotated correctly was utilized as an agreement factor. Regrettably, it was impossible to have each photo annotated by two or three persons to get a validation and derive an agreement on concepts over the whole set.

4 Concept Annotation 2010

The amount of manual annotation labor described in the section "Concept annotation 2009" was already exceptionally high with "merely" 53 concepts. In 2010, the automated concept detection task was extended by 41 new concepts to a total of 93 (one of the concepts of 2009, "*canvas*", was dropped). Manual expert relevance assessment was regarded to be too inefficient, time consuming, and expensive. An alternative was searched for, and the question arose, about whether the relevance assessment for image annotation might be reliably obtained with a crowdsourcing approach.

Among other things, crowdsourcing can be used to outsource simple repetitive tasks to a (preferably) vast network of people, i.e., the crowd. For the concept annotation 2010, the Amazon Mechanical Turk (MTurk)[2] platform was used. At MTurk, mini-jobs (called Human Intelligence Tasks or HITs) are distributed via an online marketplace to registered workers (called turkers) all over the world. Terms such as the prize paid per HIT, completion time, minimum number of different workers needed, special skills, etc. are specified by the requester up front. Workers can choose whether at all or for how many HITs they want to participate.

Design of HIT Template In total, four different tasks (HITs) were prepared to be annotated with MTurk. The turkers were supposed to allocate more specific concepts to each image. For the annotation process, pre-knowledge from the annotations in 2009 was used. Therefore, the new 41 concepts were structured into four groups:

1. *Vehicles*: 1,183 pictures known to include this general concept were further annotated with the more specific concepts *car*, *bicycle*, *ship*, *train*, *airplane*, and *skateboard*. A textbox offered the possibility to input further categories. A checkbox saying that no vehicle is depicted in the photo was included in case of falsely classified pictures. The corresponding survey with guidelines can be found in Fig. 1.

[2] https://www.mturk.com

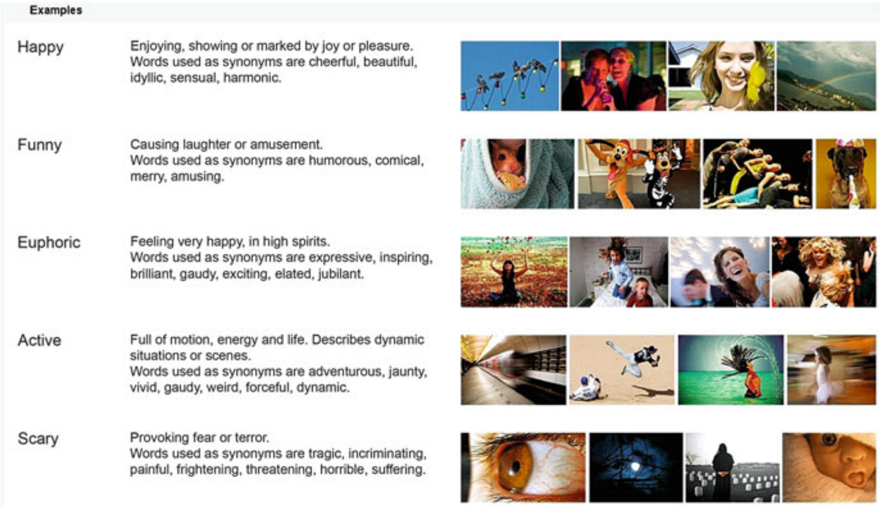


Fig. 1 Example images with definition of sentiments

2. *Animals*: The turkers at Amazon were asked to further classify 1,535 photos into the categories *dog*, *cat*, *bird*, *horse*, *fish*, and *insect*. Again, a textbox offered additional input possibilities.
3. Persons: 5,722 photos of humans were classified with attributes like *female*, *male*, *baby*, *child*, *teenager*, *adult*, and *old person*.
4. *General annotations*: For the following 22 concepts, no prior information could be used and they were hence assessed in all 18,000 photos. The HIT was designed as a survey with six questions aiming to annotate the categories "content elements" (*architecture*, *street*, *church*, *bridge*, *park garden*, *rain*, *toy*, *musical instrument*, and *shadow*), "persons" (*bodypart*), "events" (*travel*, *work*, *birthday*), "image representation" (*visual arts*, *graffiti*, *painting*), "impression" (*artificial*, *natural*, *technical*, *abstract*) and "feelings" (*boring*, *cute*).

For the first three groups each HIT was rewarded with \$0.01. Due to the increasing workload for the fourth group, a reward of \$0.03 was paid for the completion of each HIT. For all concepts, the annotations per photo were obtained three times. The final annotations were built from the majority vote of these three opinions.

5 Concept Annotation 2011

Experiences with MTurk from ImageCLEF 2010 and a pre-study (Nowak and Rüger 2010) showed the applicability of crowdsourcing for human annotation of images. In 2011, nine novel sentiment concepts were added to the test collection and annotated with the MTurk approach. To enrich the quality of the annotations, additional quality assurance mechanisms were incorporated to reduce the impact of spammers on the annotations.

Design of the HIT Template Russell's circumplex model of affect (Russell 1980) was used for the definition of sentiments. It proposes an emotional space with two dimensions: arousal and valence. The x-axis represents the dimension of valence, which ranges from highly negative emotions on the left to highly positive emotions on the right. The y-axis is called arousal (sometimes "activation") and ranges from active to inactive. Emotional adjectives/sentiments are grouped into eight affect concepts and placed in a circular order. A concept named "funny", which is not part of the original model, was included. The sentiment concepts were annotated by asking the turkers what sentiments an image conveys.

The judgment of sentiments is an extremely subjective task, hence the HIT template included a definition of sentiments, synonym sentiments, and example photos (see Fig. 1). This was to help the turkers to become familiar with the meaning of each of the sentiments. The definitions were derived from WordNet 3.0[3] and The Free Dictionary.[4] All sentiments were annotated on a one by one image-based level. Each HIT consisted of ten images. The image was depicted on the left, while on the right the circumplex model was visualized (see Fig. 2). The eight sentiments are structured according to their degrees in the circle as proposed by Russell. The wording was partially changed to better fit the sentiments to describe images (e.g., an image cannot be excited or astonished, but it may look exciting to a human being). Starting with happy at 0°, the circle is composed by funny (about 30°), euphoric (70°), active (90°), scary (150°), unpleasant (180°), melancholic (210°), inactive (270°), and calm/comforting (330°). In the middle of the circle, the concept "no sentiment" is added. This option had to be chosen when no sentiment fit the image. If turkers selected this checkbox, they were asked to give a reason for why no sentiment fits. This question was mandatory to answer, and was included in order to prevent turkers from just clicking this checkbox without thinking about the task. For all sentiments, except "no sentiment", multiple choices could be selected at the same time. Additionally, the turkers were asked on what basis they made their decision: the motif or the overall impression of the image. They could choose on a discrete five-point scale labelled with "motif", "mostly motif", "both equally", "mostly overall impression", and "overall impression". An example for one image is shown in Fig. 2. The HIT template included a verification procedure. For all ten images belonging to one HIT, it was verified that every image was annotated before the turkers were enabled to submit them. In the case of missing answers, the turkers saw the questions marked in red. This procedure ensures that it is not too easy to answer randomly and submit spam, and it helps reduce the work of filtering out incomplete answers that need to be republished. While it does not ensure that all

[3] http://wordnet.princeton.edu/

[4] http://www.thefreedictionary.com/

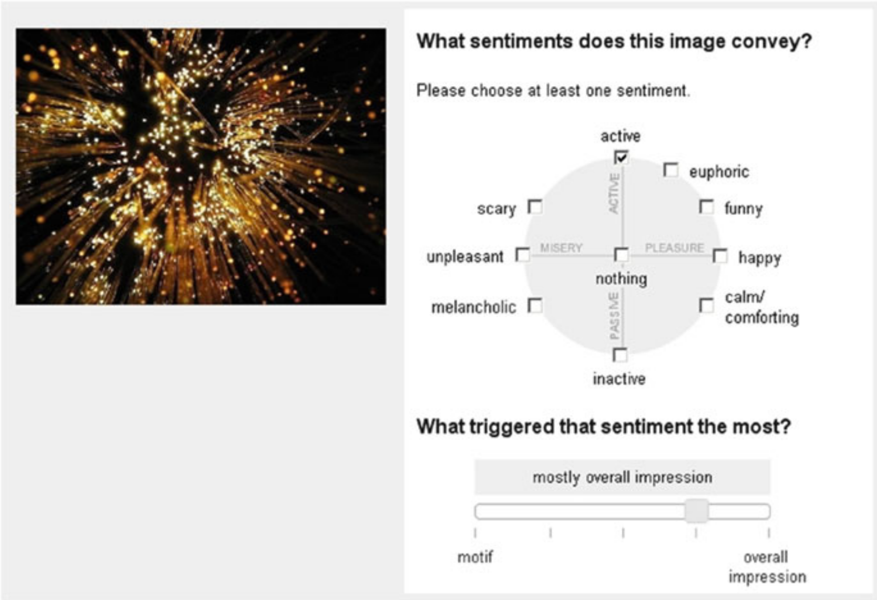


Fig. 2 Example HIT for correct annotation

random annotators are excluded (as they still can randomly answer each question), it raises the costs of cheating compared to the time that is needed to answer honestly.

Annotation Statistics The ground truth was acquired in different annotation batches. First, a pretest was conducted in which 400 images of the training set were annotated. Each survey contained ten images and was annotated three times. This made a total of 120 HITs. Each HIT was rewarded with $0.05. The HITs of the pretest were annotated by a total of 22 turkers with an average annotation time of 3 min and 12 s. The purpose of the pretest was to determine whether the template design and the task were understandable and if the turkers were able to solve the task. Results show, that in about 50 % of the images the turkers agree on a sentiment (or choose neighboring sentiments), while in the other 50 %, they chose opposite sentiments (like *happy* and *melancholic*).

The rest of the training set of the photo annotation task was annotated in 4,225 HITs. Each HIT contained nine photos of the training set and one photo of the pretest. The photos with a clear annotation agreement from the pretest were included as a so-called "gold standard" to verify the answers of turkers in order to reduce spam. For 271 images of the pretest, a sentiment could be assigned by at least two annotators (out of three), excluding the "no sentiment" concept. These 271 were used in the gold standard test of the training set. On average, the same gold standard image was used in three different HITs (excluding repeated HITs). Therefore, there was a possibility that one turker got the same gold standard image assigned up to three times. Each HIT of the training set was rewarded with $0.07 and the HITs were completed on average in 2 min and 36 s with a total of 258 turkers working on the batch.

The test set was assessed in 5,560 HITs, which each included nine images and one gold standard image of the training set. Each image of the test set was annotated five times. A total of 1,112 annotated images of the training set were used as gold standards, so that each HIT contained a different gold standard image (excluding repeated HITs). The images were chosen by a majority consensus on a sentiment. The HITs of the test set were divided into two batches in order not to pose too many HITs at the same time, and annotated by 156 distinct turkers. Each HIT was again rewarded with $0.07. For the first batch of 2,745 HITs, each HIT was annotated on average in 2 min and 8 s, while the 2,815 HITs of the second batch were annotated on average in 1 min and 44 s.

The verification of the work of the turkers was difficult, as the task of sentiment annotation is very subjective. Therefore, we followed several strategies of comparing annotations. The verification of the HITs of the training and test set with the gold standard led to a direct acceptance of 3,204 and 4,358 HITs, respectively, when allowing for deviations of at most 120° on the affect circle. For HITs, that did not pass the gold standard test, we compared the results of the HIT to the four answers of other turkers for the same HIT. For all images of the HIT, the deviation from the annotations of the other HITs was computed and the HITs were accepted when the deviation per image was equal to or less than 120° on the affect circle. A total of seven out of the ten images had to fit. With this procedure, all remaining HITs were accepted.

The final construction of the ground truth considered the majority votes for each image. In the case that no clear answer was given, we decided to discard any sentiment information for that image. In total, about 15 % of the training set images and 14 % of the test set images have no sentiment information. Interestingly, the "nothing" option was chosen rarely by the turkers. None of the images of the training set and only one image of the test was annotated with "nothing" by a majority of turkers.

6 Best Practices

Crowdsourcing is a rather new technology for which no standardization of approaches yet exists. The main challenges lie in the HIT design, in how to balance payments (Horton and Zeckhauser 2010; Mason and Watts 2009), in ethics regarding the possible exploitation of workers in third-party countries (Fort et al. 2011), in legal issues (Felstiner 2011), and in quality control of the obtained answers (Alonso et al. 2008). The following practical guidelines address some of the main challenges.

Hit Design Many crowdsourcing experiments adopt the online marketplace Amazon Mechanical Turk, where mini-jobs (HIT) are distributed to a crowd of people.

Among others, the requester has to define a title, description, and keywords for the work to distribute. It should be kept in mind that the description of the task should be as inviting as possible. The task itself should be formulated to be as easy as possible, due to differing and unknown knowledge levels and linguistic abilities of the turkers. Extensive use of technical language should be avoided; the turkers should be able to successfully complete HITs with common knowledge. To clarify the task, it is advisable to give examples about what is in the scope of the task and what is not. Sometimes the context of the task improves the people's willingness to work (Chandler and Kapelner 2010).

Another crucial element is the usability of the HIT template or interface design. The interface should be easy to handle and prevent incorrect entries. There are two common possibilities to test the usability and functionality of the interface: The use of the developer sandbox and the conduction of pretest with a small dataset.

Lastly, it is important to give the turkers the opportunity to provide feedback on the HIT and analyze it. This can help to increase the description of the task and finally the test results.

Balance Payments The balance of the payment is a crucial point. Although "learning new skills" and "having fun" are types of motivation for turkers, the biggest motivation is earning money.[5] If the reward is too low, no one will work on the HIT. If it is too high, it will attract spammers.

One way to avoid over- or underpayment is to ask turkers what amount of money would be sufficient for the given amount of work. An additional bonus can be paid to turkers who solve many HITs, solve them very fast, or solve them especially well.

Amazon demands fees for every completed HIT; usually 10 % of the reward but at least $0.005. The fee depends on the knowledge of turkers: For HITs exclusively provided to Photo Moderation or Categorization Masters, an additional 20 % fee applies.

Quality Control MTurk offers several ways of assuring quality. Optionally, the turkers can be asked to pass a qualification test before working on HITs, multiple workers can be assigned the same HIT, and requesters can reject work in case the HITs were not finished correctly. The HIT approval rate each turker achieves by completing HITs can be used as a threshold for authorization to work. The time allotted to each HIT can also be defined and used as some kind of quality control: The time should be chosen in such a way that the average turker is able to solve the task in the given time. Too much granted time could tempt the turker to additionally work on other things than the HIT with the result that the task is blocked for other turkers meanwhile. Too short a time period might put stress on the turker and could lead to random answers due to frustration. Moreover, the completion times provide information on the reliability of the answers. The inclusion of gold answers is an appropriate way to ensure reliability of turkers. The gold answer, already known

[5] http://crowdflower.com/blog/2010/02/why-people-participate-on-mechanical-turk-now-as-a-mosaic-plot/

to the requester, is compared to the answer of the turker. If the answer is in line with the gold answer, the data provided by the turker is considered to be reliable. It is recommended to repeat HITs with several turkers and take a combined vote of answers, e.g., a majority vote, to filter out "noise".

As described above, quality estimation is the basis for rejection of work. However, a reliable rejection mechanism is difficult to set up, as the task of image annotation entails ambiguities and subjectivity (Nowak 2011).

7 Conclusion

Although there are some mechanisms to enable quality control of data generated with a crowdsourcing approach, attention has to be paid to noise. Noise is brought into the evaluation process by inconsistencies in relevance judgments among turkers. Its effect on system rankings has to be known. In Nowak (2011) and Nowak and Rüger (2010) reliability of crowdsourced approaches for image annotations and their influence on system rankings in comparison to expert judgments is investigated. The annotations of 11 experts were compared among each other and with the annotations of non-expert annotators at MTurk. The experts had a very high inter-annotator agreement. The accuracy of the turkers showed a high agreement. When comparing the averaged ground truth of the non-experts with that of the experts, the inter-annotator agreement in terms of Cohen's kappa is 0.84 on average. Additionally, different ranking experiments assessed that relevance provided by experts and by turkers is comparable regarding the ability to differentiate among the quality of annotation approaches. These results indicate that there are differences in annotations, but that they do not significantly influence the ranking of different systems in a benchmark scenario. A majority vote of several turkers' opinions is able to filter noise in judgments of non-experts to some extent. The resulting annotation set is of comparable quality to that of the annotations of experts.

The results prove that the process of relevance assessment in image annotation can be outsourced to the crowd. The crowdsourcing approach exhibits a fast and cheap possibility of obtaining relevance judgments while retaining sufficient reliability. This makes crowdsourcing an interesting alternative or addition to expert annotations and pooling methods. Fraunhofer IDMT made the test collections of the ImageCLEF Visual Concept Detection and Annotation Task from 2009 to 2011 publicly available. The photos used in the benchmark as well as the ground truth, Flickr user tags, and EXIF information are freely available.[6]

[6]http://www.idmt.fraunhofer.de/de/projects/expired_publicly_financed_research_projects/photo_annotation.html

References

O. Alonso, D.E. Rose, B. Stewart, Crowdsourcing for relevance evaluation. ACM SIGIR Forum **42**(2), 9–15 (2008). http://doi.acm.org/10.1145/1480506.1480508

D. Chandler, A. Kapelner, Breaking monotony with meaning: motivation in crowdsourcing markets. Technical report, University of Chicago, 2010

M. Everingham, L.J.V. Gool, C.K.I. Williams, J.M. Winn, A. Zisserman, The pascal visual object classes (VOC) challenge. Int. J. Comput. Vis. **88**(2), 303–338 (2010). http://dblp.uni-trier.de/db/journals/ijcv/ijcv88.html#EveringhamGWWZ10

A. Felstiner, Working the crowd: employment and labor law in the crowdsourcing industry. Berkeley J. Employ. Labor Law **32**(1), 143–204 (2011)

K. Fort, G. Adda, K.B. Cohen, Amazon mechanical turk: gold mine or coal mine? Comput. Linguist. **37**(2), 413–420 (2011)

J.J. Horton, R.J. Zeckhauser, Algorithmic wage negotiations: applications to paid crowdsourcing, in *Proceedings of CrowdConf*, San Francisco, California, 2010

M.J. Huiskes, M.S. Lew, The MIR flickr retrieval evaluation, in *Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval (MIR'08)*, Vancouver, ed. by M.S. Lew, A.D. Bimbo, E.M. Bakker (ACM, New York, 2008), pp. 39–43. http://dblp.uni-trier.de/db/conf/mir/mir2008.html#HuiskesL08

W. Mason, D.J. Watts, Financial incentives and the "performance of crowds", in *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP'09)*, Paris (ACM, New York, 2009), pp. 77–85

H. Müller, P. Clough, T. Deselaers, B. Caputo (eds.), *ImageCLEF: Experimental Evaluation in Visual Information Retrieval*. Volume 32 of The Information Retrieval Series (Springer, Berlin/Heidelberg/New York, 2010)

S. Nowak, Evaluation methodologies for visual information retrieval and annotation. PhD thesis, Technische Universität Ilmenau, 2011

S. Nowak, P. Dunker, Overview of the CLEF 2009 large-scale visual concept detection and annotation task, in *Multilingual Information Access Evaluation II. Multimedia Experiments*, ed. by C. Peters, B. Caputo, J. Gonzalo, G.J.F. Jones, J. Kalpathy-Cramer, H. Múller, T. Tsikrika. Volume 6242 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2010), pp. 94–109

S. Nowak, S. Rüger, How reliable are annotations via crowdsourcing? A study about inter-annotator agreement for multi-label image annotation, in *Proceedings of the International Conference on Multimedia Information Retrieval (MIR'10)*, Firenze (ACM, New York), 2010, pp. 557–566

J.A. Russell, A circumplex model of affect. J. Personal. Soc. Psychol. **39**(6), 1161–1178 (1980)

A.F. Smeaton, P. Over, W. Kraaij, Evaluation campaigns and TRECVid, in *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, Santa Barbara, 26–27 Oct 2006 (ACM, New York, 2006), pp. 321–330

Part III
Use Cases

The THESEUS Use Cases

Florian Kuhlmann, Jan Hannemann, Myriam Traub, Christoph Böhme, Sonja Zillner, Alexander Cavallaro, Sascha Seifert, Björn Decker, Ralph Traphöner, Sven Kayser, Udo Lindemann, Stefan Prasse, Götz Marczinski, Ralf Grützner, Axel Fasse, and Daniel Oberle

F. Kuhlmann (✉)
Neofonie GmbH, Berlin, Germany
e-mail: florian.kuhlmann@neofonie.de

J. Hannemann • M. Traub • C. Böhme
German National Library, Frankfurt am Main, Germany
e-mail: j.hannemann@dnb.de; m.traub@dnb.de; c.boehme@dnb.de

S. Zillner
Siemens AG, Munich, Germany
e-mail: sonja.zillner@siemens.com

A. Cavallaro
University Hospital Erlangen, Erlangen, Germany
e-mail: alexander.cavallaro@uk-erlangen.de

S. Seifert
Siemens AG, Erlangen, Germany
e-mail: sascha.seifert@siemens.com

B. Decker • R. Traphöner
Empolis Information Management GmbH, Bielefeld, Germany
e-mail: bjoern.decker@gmail.com; ralph.traphoener@empolis.com

S. Kayser
Attensity Europe GmbH, Saarbrücken, Germany
e-mail: sven.kayser@attensity.com

U. Lindemann
Technische Universität Munich – Product Development, Garching, Germany
e-mail: udo.lindemann@pe.mw.tum.de

S. Prasse
VDMA Verlag GmbH, Frankfurt am Main, Germany
e-mail: stefan.prasse@vdma.org

G. Marczinski
CIM Aachen GmbH, Aachen, Germany
e-mail: mc@cim-aachen.de

R. Grützner
Infoman AG , Stuttgart, Germany
e-mail: ralf.gruetzner@infoman.de

A. Fasse • D. Oberle
SAP AG, SAP Research, Karlsruhe, Germany
e-mail: Axel.Fasse; D.Oberle@sap.com

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_21,

Abstract The THESEUS research program assembled key companies with market power from all types of sectors to jointly develop the innovative products that will enable the knowledge society. There have been carried out six use cases to demonstrate applications based on the developements of the THESEUS CORE TECHNOLOGY CLUSTER. In this article, we will give a short overview of selected results of each use case.

1 Introduction

The THESEUS CORE TECHNOLOGY CLUSTER (CTC, see Becker et al. 2014) and the industry driven application scenarios (use cases) are the two major building blocks for the THESEUS research program. The success of THESEUS was based on the transfer of technology from the CTC to the use cases and on the functional and business requirements from the use cases being covered properly by the underlying technology which was delivered by the CTC.

Within the THESEUS research program the following six use cases have been successfully worked on:

- ALEXANDRIA (leading partner: neofonie GmbH)
- CONTENTUS (leading partner: German National Library)
- MEDICO (leading partner: Siemens AG)
- ORDO (leading partner: Empolis Information Management GmbH)
- PROCESSUS (leading partner: Empolis Information Management GmbH)
- TEXO (leading partner: SAP AG)

Each use case was led by an industrial partner who was committed to carrying out the project and bringing the results to the market, gathering other partners contributing to it. This formed a sub-consortium which may have had specific agreements to deal with intellectual property (IP) and commercial exploitation issues.

It is understood that commercial exploitation itself was not funded. The work plan of a use case and its associated budgets were designed to lead to one or several demonstrators, which at most can be real-scale prototypes close to the market, but not directly marketable products.

The typical time horizon for a use case was 3–5 years, with intermediate shorter term goals every few months. The initial precise work plan was defined within that time frame, but the flexible structure of the THESEUS research program allows refining it if necessary.

Each use case was structured into several work packages, led by one industrial partner. Work packages represent smaller pieces of work, such as demonstrators, platforms and other building blocks for the use case.

The following sections are designed to give a first look under the hood of the THESEUS use cases with some of the major achievements described briefly. For a deeper insight, further results are referenced.

2 ALEXANDRIA: A Consumer-Oriented Knowledge Database

Like the ancient world's most famous library in the city of the same name, the THESEUS use case ALEXANDRIA is seeking ways to structure the exponentially growing quantities of knowledge, reaching from handwritten books from the far past, to collaboratively written documents of today in the Web 2.0. The key idea is to automatically describe document content by meaningful annotations, i.e., semantic annotations, using a common vocabulary. The annotations, i.e., the structured data, allow more efficient information access and new ways of analysis. The particular focus in ALEXANDRIA was the integration of functionalities used by social media platforms in a portal built on top of semantic technologies. One of the key challenges was to elaborate techniques that would allow users without deeper knowledge of semantic technologies to collaboratively use a semantic network (a graph of knowledge, or more abstractly, of things) and extend the containing knowledge.

The heart of the platform is a complex knowledge base that was manually designed and later filled with data from the web of Linked Open Data (LOD), a worldwide initiative to provide structured datasets to the public and link these datasets with each other, so that data entries in one dataset are linked to the corresponding entry in another set. As domains of general interest, the knowledge base was representing information about famous people, places and organizations and relationships between those entities. The design principles of the ALEXANDRIA knowledge base are described in more detail in Wendt et al. (2014).

To give users access to the knowledge, different kinds of visual navigation were designed and implemented. Depending on the concept of the actual viewed entity (e.g. person, place, etc.), a particular kind of visualization is automatically chosen as a default. If the user is, e.g., viewing the entity "Germany", a geographic map and a timeline are displayed. Both navigation elements are connected so that the user is enabled to browse all semantically related data either by scrolling through the timeline (and seeing connected events appearing and disappearing on the map), or by moving and zooming the map; thus the user can visually explore the history of a country. For different concepts other visualization forms are used, e.g. when viewing a person, the corresponding network connections to related persons are graphically displayed.

The knowledge base can also support the automatic extraction of new information about the containing entities out of unstructured documents on the World Wide Web. In a multi-stage pipeline, a big part of German online news publications are gathered and analyzed. The first step in this analytic process is the identification of entities. This step involves linguistic pre-processing, usage of statistical models to gather entity candidates and the exploitation of network information from the knowledge base to automatically resolve ambiguous entities (there are for example several persons in the knowledge base that would match the name "Peter Müller", so the system has to decide which one to choose).

The next step of information extraction is to detect relations either between two entities or between an entity and an attributive value (e.g., a date, an amount of money, etc.) in the document. This step is described in more detail in Xu et al. (2014).

In the information extraction pipeline, several annotators were built to extract semantic information from text documents. Some examples are extractors for time values (relative mentions like "last month" and absolute ones like "February 2012"), direct and indirect speech, events with deeper information about time and location of the event, opinions from persons about other persons or organizations, co-reference analysis, and extractors for the most specific places in a document to automatically display a map according to the content. Statistics are built upon the relative number of occurrences in German media. These statistics are used to calculate a value for the "prominence" of entities for each of the last 30 days in a sliding time frame. This measurement is combined with a popularity score calculated from the semantic network. The latter is calculated, based on the number of incoming links (the relations between two more entities), to the entity in the semantic graph.

Information generated in the information extraction process is written back into the knowledge base. Automatically detected relations are gathered in a separate graph, because precision of the extracted information is still too low to be acceptable by the end users of the platform. Concepts for a manual evaluation of this data by users in a playful way were developed and designed, but a concrete implementation was not part of the project.

Using mainly the same technologies as in the information extraction described before, an interactive question answering system was built, giving the users of the platform another way to explore the data of the knowledge base. Users can type in questions in natural language (at the moment only German is provided). A system called "Semantic Type Ahead" gives the user suggestions of how his input is interpreted by the engine and permits the possibility of manually selecting an entity in an ambiguous situation. E. g., if the user asks a question like "Which politicians were born in Berlin between 1950 and 1970?", the engine displays various options for the interpretation of Berlin (East-Berlin, West-Berlin, Berlin in Texas) the user can choose from. By default, the most popular or "prominent" entry of an entity is used for calculating the answers; if the user actively chooses another entity, the answers are changed instantly.

As a fall-back in case an answer cannot be provided by the knowledge base (because either the relationships are not modelled yet, or the data is too sparse to give an answer to a specific question), search results from German news are displayed to the user. This functionality is based on a more traditional, keyword-based search technique combined with the search in the already disambiguated (by the user on the search side, by the system on the news side) named entities.

Besides the two ways mentioned before to populate the knowledge base (by using Linked Open Data and by using information extraction on unstructured data) a third alternative is to extend the semantic network. Users of the ALEXANDRIA platform are enabled to actively edit and extend data by using form-based input masks. A quality management process is introduced to ensure the correctness of

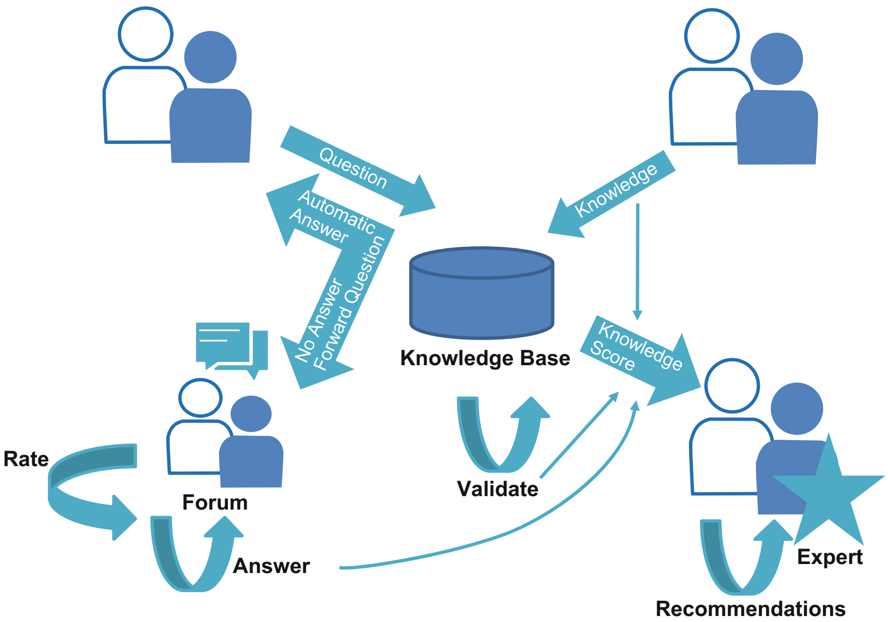


Fig. 1 Collaboration on knowledge

user-generated data. As a measurement, users can obtain "knowledge points" for adding information to the knowledge base that is marked as correct by other users in a later step. If a user collects a defined amount of knowledge points, he or she will get permission to edit more sensible data in the network. These knowledge points are also assigned to the various topics in the knowledge base, so that a user might reach the level of an expert in a specific topic (e.g., politics, science, or something even more specific for a concrete person in the knowledge base).

Knowledge points are not only used for calculating permissions but also serve as a base for the recommendation system, which recommends contents or users with similar interests and knowledge to a specific user of the platform.

As a fall-back mechanism of the question answering system described above each topic in the knowledge base is assigned a forum, where users can delegate their questions to and (hopefully) receive answers from other users of the system. Answers can be rated; these ratings are another factor for calculating knowledge points.

If users discover invalid facts in the knowledge base, they are able to mark these facts as questionable. By doing so, they automatically open a discussion around the concrete fact, where every user can vote in favor of or against the validity of this fact. An administrator or an expert of the particular topic can close the discussion, and decide to either keep or delete the discussed fact. All participants of the discussion get positive or negative knowledge points (e.g., a lower score) for the affected topic. Figure 1 gives an overview of collaboration features in ALEXANDRIA.

The platform, described above, has been open to the public since February 2012[1] as a showcase for the developed technologies, but also as a maintained platform for users interested in working on a structured knowledge database.

Technologies developed in the ALEXANDRIA use case are not only applicable for usage in a B2C platform like the actual ALEXANDRIA portal, but might also be of great benefit to an intranet scenario, where multiple employees work on specific domain knowledge.

The main research goal of ALEXANDRIA was the investigation of the interconnection of social media with semantic technologies.

Some of the main technological achievements were linguistic and semantic information extraction technologies, like a fast and precise "Named Entity Recognition And Disambiguation"-Framework, a framework for automatically tagging text documents, a framework for pattern-based information extraction, a semantically rich ontology about persons, places, organizations, events and works, a highly adaptable question answering component, and technologies for semantic analysis of media content. All technologies were developed for the German language but are adaptable to other languages and domains.

Customers using technologies that were derived from research results out of ALEXANDRIA include Siemens, Gruner + Jahr (stern.de),[2] the "Deutsche Digitale Bibliothek (DDB)" and Axel Springer Mediahouse. The ongoing research project "MIA – Cloudbasierter Markplatz für Informationen und Analysen",[3] which started within the scope of the Trusted Cloud[4] project in 2012 and ends in 2015, uses technologies derived from ALEXANDRIA and focuses on the aspects of scaling out to nearly web scale with the goal of providing flexible semantic analysis for interested partners and customers.

3 CONTENTUS: Next Generation Multimedia Libraries

The vast amounts of books, newspaper articles, pictures, videos, films and audio recordings stored in libraries, media archives and publishing houses worldwide offer an incredible wealth of information. However, due to technical, financial, or practical reasons, most of the knowledge hidden in those media is not yet readily accessible. CONTENTUS helps to unearth these treasures and make our cultural heritage as widely accessible as possible.

A Wealth of Information – Our Cultural Heritage Media archives in Germany and in the world contain incredible treasures of information: books, pictures, audio

[1] http://alexandria.neofonie.de

[2] http://stern.de

[3] http://www.mia-marktplatz.de

[4] http://www.trusted-cloud.de

Fig. 2 Example of the different types of media contained in the MIZ collection. All items in the photograph are related to the radio drama and the derived opera *The Trial of Lucullus*

recordings, films and videos that represent an important part of our cultural heritage. Much of this heritage is hidden in archives and difficult to access. In CONTENTUS we set out to develop technologies for making these treasures more easily available.

Not only does the sheer number of cultural assets pose a challenge to such technologies, but also does the diversity of media, that can be found in galleries, libraries, (media) archives and museums (GLAM institutions). Figure 2 shows an example to illustrate this. The many different types of items in the photograph are all part of the archive of the *Musikinformationszentrum des Verbandes der Komponisten und Musikwissenschaftler der DDR* (music information center of the GDR's association of composers and musicologists, MIZ).

However, with the advent of the Internet, previously unthinkable possibilities for access to and the presentation of our cultural heritage emerged: today, access to cultural and archival assets no longer needs to be restricted by time and location. Content information extracted from media allows computers to overcome the boundaries of different media types, and to handle these items in a meaningful way, even linking previously separate collections. Ideally, the media and the knowledge contained therein are readily available anywhere and at any time. But how do media collections – especially analog ones – become more accessible?

Digitization Is Only the First Step Digitization helps to preserve cultural heritage information, especially if the media involved are in danger of deterioration. However, digitization alone does not guarantee convenient or even smart access to media and information. The scan of a newspaper page from the MIZ archive, for example, is merely an image. Without additional information it is not clear how many articles are contained on the page and what these articles are about. Other than going through an entire archive, how does a user later find this particular page – or an article on this page – in a large collection?

To search for – and to find – items of interest in larger collections, metadata are needed that describe the individual media. Traditionally, these metadata are generated intellectually by experts, such as librarians or archivists. However, this is a costly and time-consuming process. With the steadily growing number of books, articles, videos, and online publications, it becomes increasingly difficult to keep up. The German National Library alone receives over 2,500 new media items each day – with a clear upward trend. Broadcast archives are faced with similar difficulties, compounded by the fact that describing and archiving audiovisual content is particularly work intensive.

The solution is to develop and use software systems that can aid librarians and archivists in their task of describing the content of media items. CONTENTUS aims to provide such support for media analysis, tailored to the requirements of the individual institutions and their media.

High quality metadata make it possible to search for media items, and distinguish among them. This, however, does not fully tap the potential of the data. What if the original is barely readable or needs special protection due to its age? What if we do not just want to find an article, but also want to understand its relationship to other media and to certain persons we might be interested in? In practice, additional processing steps are taken that go beyond digitization. These steps are the detection and removal of quality problems like scratches and dirt, often found on old filmstrips, and the enrichment of media with additional information from internal and external sources. In summary, a medium has a long way to go before becoming part of a comprehensive knowledge network. To further illustrate this process, we describe the preparation of a newspaper scan in more detail now.

From Analog Medium to Enriched Information Resource Imagine you are interested in a particular newspaper article about a music event in Berlin in 1959 (the Ninth "Biennale"), where Paul Deassau's "Guernica" was performed. A scan of the page provides the required digital representation. Quality control can detect and remove discolorations (here from a text marker) and optimize contrast and sharpness. An analysis of the page first rotates the page slightly, so that the text lines are horizontal, generates a textual representation (optical character recognition, OCR) and then identifies individual articles (page segmentation). The extracted text is particularly useful for later access. It can help users searching for certain keywords, e.g. "Biennale", to find the right article.

In addition, further analysis of the text can identify the *meaning* of certain words, in particular entities of interest, such as places (Berlin) or people (e.g. Paul Dessau). Such information about entities can then be used to semantically link the relevant references in this article to other information sources, such as authority file data (databases maintained by GLAMs, which contain information about entities, such as authors, publishers or locations) or Wikipedia articles. Before all this information

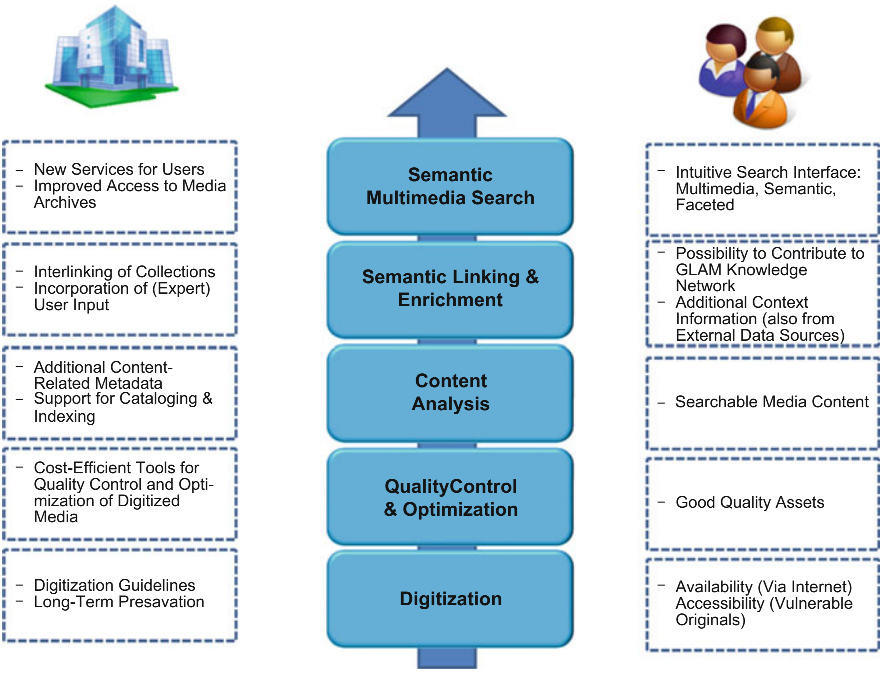


Fig. 3 Benefits of CONTENTUS technologies for GLAMs (*left*) and end users (*right*), organized by processing step

is made available to users in a semantic search environment, the generated metadata can be reviewed and extended, usually by experts.

All media types pass through a similar series of processing steps. However, solutions for digitization, quality control and content analysis are media-specific, and therefore the developed approaches differ considerably for print and audiovisual media. Consider, for instance, quality control for print and audio media: detecting and removing discolorations from a scanned book page is fundamentally different from identifying and reducing background noise in a historical audio recording. The last two steps (linking items and entities semantically and accessing them through a semantic search interface), however, are largely independent of the media types.

CONTENTUS provides solutions for all processing steps, which can be used individually or in combination. What does that actually mean for GLAMs and users? CONTENTUS technologies aim to provide better access to our cultural heritage held by GLAMs. Both the institutions and their users benefit from this, as shown in Fig. 3.

Benefits for GLAMs Digitization of old media objects can be a difficult, costly and even risky undertaking for GLAM institutions. Many collections such as the MIZ archive contain objects (for example, old magnetic tapes or film) that are in bad physical condition, due to either their age or improper handling. Every playback can lead to further damage, which may finally render the object unusable.

It is therefore vital to make sure that the digitization process runs optimally the first time. To assist GLAM institutions with this delicate task, CONTENTUS has compiled the experiences gained in the MIZ digitization process into two handbooks for digitization of audiovisual and print media (see e.g. Vogel and Effenberg 2011). Those two books describe in detail how to plan and execute digitization projects, and what pitfalls to avoid.

Media processing also includes quality control and optimization. Both are cost and time intensive steps usually performed by experts. For example, experts estimate that quality control for 1 h of video material typically takes a specialist about 4 h of work. CONTENTUS technologies provide automated support for these important processing steps, reducing the costs and time required considerably.

Analyzing the content of cultural assets and linking them to other information sources provides a richer data pool for searching. Furthermore, it facilitates a comprehensive presentation of information about every item, including relevant context information. The knowledge network comprised of media, metadata and semantically linked information sources allows for entirely new services for users to be developed and offered.

One of these services can be to open the resulting knowledge network to expert users, allowing them to add missing information to the network or correct faulty entries. By involving users in creating a comprehensive knowledge network, GLAM institutions can not only benefit from an improvement of their data quality, but can also improve customer retention.

In summary, CONTENTUS solutions help GLAMs make their valuable media archives more accessible and explore new ways of information retrieval. By developing largely automated technologies, GLAMs can benefit from this with a manageable effort, which is especially important for smaller institutions. That way, even small libraries and archives can become part of a global information infrastructure.

Benefits for Users Where GLAMs benefit from higher quality media, better metadata and semantic information, end users benefit in turn from services that use this information, as well as from new services that were not possible before.

With proper content analysis, media content can be used for searches, making previously non-searchable assets accessible. Previously unknown relationships become apparent, showing the media in their context. Semantic analyses can help provide better search results, for example by allowing users to differentiate between actual persons of interest, instead of just different names. This is especially helpful in cases where multiple people have the same name.

Additional information can be used to enrich media presentations, providing more relevant information to users. For example, links to other archives or web resources like Wikipedia can be offered.

Outlook GLAMs are faced with considerable challenges when trying to provide better access to their collections. Although technologies exist or are being developed that can help them in this undertaking, taking advantage of technological advancements is not easy. It is a challenge requiring expert knowledge in many different fields. For most GLAMs, the technical complexity is not manageable in addition to their day-to-day work. For this reason, experts from different disciplines teamed up in the CONTENTUS use case to develop concepts and solutions for turning collections of analog cultural assets into machine-comprehensible, readily accessible digital media archives. Coming from many different areas of expertise, these experts addressed the specific needs and requirements of GLAMs for each step of the process, from digitization to presentation of their cultural assets on the Internet. The development of concepts and solutions for this process in order to support GLAMs on their way into the digital domain was the primary goal of CONTENTUS.

Despite its manifold goals, CONTENTUS has managed to take research results and develop solid solutions for media archives. Many of the technologies have already been successfully employed in public and commercial projects, while others are in the evaluation and adaptation phases.

CONTENTUS print media processing technologies have been utilized in cooperation with German newspaper and magazine archives, and are part of the electronic reading room software myBIB eRoom. Similarly, CONTENTUS technologies play an integral part in several successful development projects with German broadcasters. For example, the audio mining and entity recognition technologies were employed in a cooperation with the German broadcasting company ARD in the processing of their media archive *Mediathek*. Semantic and knowledge modeling research in CONTENTUS has facilitated the provision of the German authority file data on persons, organizations and subject headings as Linked Open Data (Nandzik et al. 2010), and is an important contribution to several projects in the German library field, as well as to the German Digital Library.[5] As of September 2011, more than twenty other projects in industry and the public sector are building on CONTENTUS, strengthening Germany's economy and benefiting both the archive owners and the users. Core areas of CONTENTUS in which projects have been initiated are the processing of print media (see Konya et al. 2014), audiovisual media, and semantic linking (see Böhme et al. 2014).

These examples are only a small selection of the results of CONTENTUS. We expect, that CONTENTUS will have a great impact on the future developments of digital presentation of cultural assets in GLAM institutions. These will benefit the end user with new and better services and access to previously unavailable media and information sources.

[5] http://www.deutsche-digitale-bibliothek.de/

4 MEDICO: Towards Scalable Semantic Image Search in Medicine

The current range and impact of different imaging technologies in clinical diagnosis and care is remarkable: Medical images from various modalities such as computed tomography, magnetic resonance imaging or ultrasonography provide information about morphology, function and metabolism of the human body, and have become indispensable for detecting and differentiating pathologies, planning interventions and monitoring treatments. Medical image modalities have matured with regard to image quality and ease of use. In addition, advanced post-processing techniques for analysts of imaging data, such as using radiologists as second readers in finding and evaluating breast lesions, lung nodules or colon polyps (Bogoni et al. 2005; Doi 2007; Marten and Engelke 2007; Taylor et al. 2005) have become available. As a result, the number of image studies is growing rapidly; for instance, even small to medium-sized hospitals produce and store millions of medical images every year (Samei et al. 2004). While medical images provide a wealth of information to clinicians, the challenging issue is to establish efficient means for accessing, seamlessly integrating and searching the medical image data. Current medical image databases, called PACS (Picture Archiving and Communications System), as well as associated Radiology Information Systems (RISs), are still indexed by keywords assigned by humans or by metadata originating from the image acquisition and not by the image contents. This limitation is severely hampering clinical workflows. For instance, a clinician who wants to simply retrieve a certain image with a certain characteristic, e.g. display a large multifocal spleen lesion of a particular patient he has in mind, typically has to accomplish a cumbersome and manual process of several steps: First, he needs to query the system for the patient's name and to select an imaging study based on the acquisition date or imaging modality. He then needs to identify the most likely series, which he subsequently might load into some 2D/3D viewer for locating the spleen by browsing the images contained in the series. Only after accomplishing all those steps can he validate that he indeed found the case with the large multifocal lesions he had in mind. The advanced and efficient access to medical imaging data requires the semantic representation of medical images' content and the preprocessing of semantic image annotations for seamless integration into intelligent clinical applications. This again is an important contribution for the future Internet services[6] which will make it possible to combine and link services with one another according to the particular requirements of users.

The use case MEDICO[7] aims for advanced search and analysis technologies that exploit image semantics. The vision of the developed MEDICO system is to automatically extract meaning from medical images and to seamlessly integrate the extracted knowledge into medical processes, such as clinical decision making.

[6] http://theseus.pt-dlr.de/en/index.php

[7] http://theseus.pt-dlr.de/en/920.php

Clinician's Perspective For highlighting the practical impact of such a system, let us imagine a patient who is suffering from night sweats and exhaustion. Moreover, the lymph nodes in the neck area of the patient have enlarged over an extended period of time. Based on computer tomography images, the responsible radiologist creates a first report of the findings. In a subsequent step, the MEDICO system will perform a semantic image analysis that integrates all results and findings, i.e. the automatically identified anatomical and pathological structures; the manual descriptions by the radiologist, as well as the information from any previous findings, are placed into a common context. This becomes technically possible by storing all semantic descriptions in a database and by efficiently linking them to (a) previous examinations of the same patient, (b) patient records with similar diagnosis or treatments, and (c) external knowledge resources, such as publications that are relevant in the context of the particular symptoms of the first diagnosis. By analyzing the images from previous examinations of the patient, the radiologist can observe whether the MRI scans from 3 months earlier also have shown enlarged lymph nodes. By investigating similar cases, he can learn about the disease progress and effectiveness of treatments, and by studying relevant literature, he can review statistical reports and recommended treatments. Based on the comprehensive information about the patient history, about selected similar cases, and about the state of the art in medicine, the radiologist arrives at a diagnosis of Hodgkin's lymphoma. By linking medical knowledge with new image processing methods, knowledge-based data processing, and machine learning, MEDICO paves the way towards improved quality and efficiency of medical care, as well as the Internet of Medical Services.

Technical Perspective To implement such a scenario, low level features, segmentations and quantitative measurements, derived from image processing, are associated with concepts from medical ontologies. E.g. an automatic detector of the spleen will associate a 3D mesh with the anatomical concept "spleen" with ID 7196 in the Foundational Model of Anatomy ontology (FMA), and – if enlarged – in addition with code R16.1 ("splenomegaly, not elsewhere classified") of the International Statistical Classification of Diseases and Related Health Problems (ICD 9). A second CAD application specialized in detecting lesions in solid organs could further analyze the detected volume and label found spleen lesions as "hypodense" and "multifocal", two imaging observation characteristics defined in RadLex, the Lexicon for Uniform Indexing and Retrieval of Radiology Information Resources.

For realizing such a scenario, various challenges in terms of knowledge management needed to be tackled: The knowledge used within MEDICO is heterogeneous and stored in various distributed knowledge repositories (see Fig. 4): Medical images, radiology reports, treatment plans, expert as well as online knowledge. In accordance with the various heterogeneous data resources, different information extracting approaches are developed and applied: Besides an intelligent segmentation algorithm for analyzing the content of medical images, MEDICO relies on text mining approaches for the analyses of textual documents, such as radiology

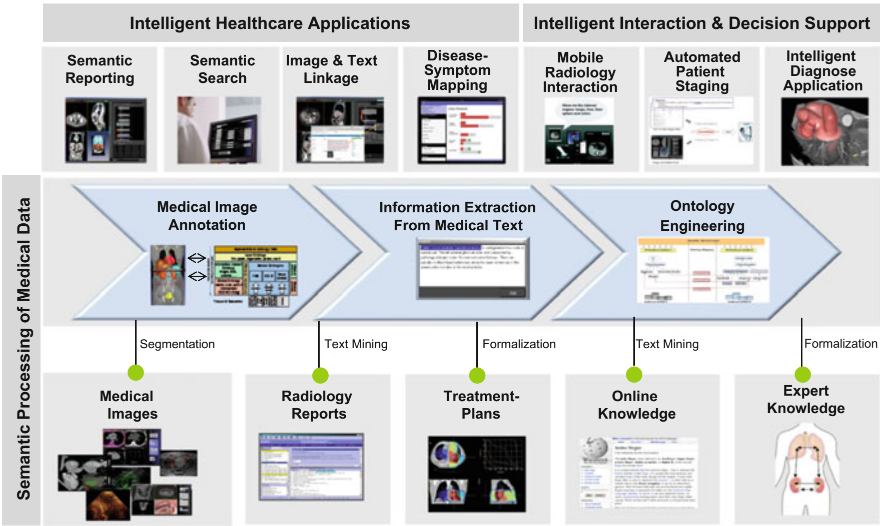


Fig. 4 Overview of the MEDICO semantic processing units, intelligent applications and workflow components

reports, or online resources, such as PubMed.[8] In addition, formalization approaches are used to capture the knowledge of medical experts and clinical best practice solutions, such as treatment plans. By providing various means for extracting the required information encompassed by the mentioned knowledge resources, MEDICO enables an integrated view over heterogeneous data resources, which is an important prerequisite for its subsequent automated semantic processing.

MEDICO Clinical Semantic Workflow The MEDICO system architecture as installed at the Imaging Science Institute, University Hospital Erlangen, is illustrated in Fig. 5. The patient's images acquired by Computer Tomography (CT) or Magnet Resonance Imaging (MRI) are stored in the hospital's image archiving system (PACS). The MEDICO system retrieves these images from the PACS, converts them into the internal image storage format and invokes the built-in image parser to segment it into disjoint regions, e.g., organs and tissues, and additionally locates discriminatory landmarks such as vessel bifurcations or bone ends. The results of the image parsing process, the semantic image annotations, are stored within the MEDICO Server and referenced in the MEDICO Annotation Ontology. The MEDICO Server is the central processing and data management unit of the MEDICO system, which links up all the system modules. It mediates the access to the knowledge layer, consisting of the MEDICO Reference Ontologies, such as FMA or RadLex, the MEDICO Annotation Ontology and the MEDICO Image Similarity

[8] http://www.ncbi.nlm.nih.gov/pubmed/

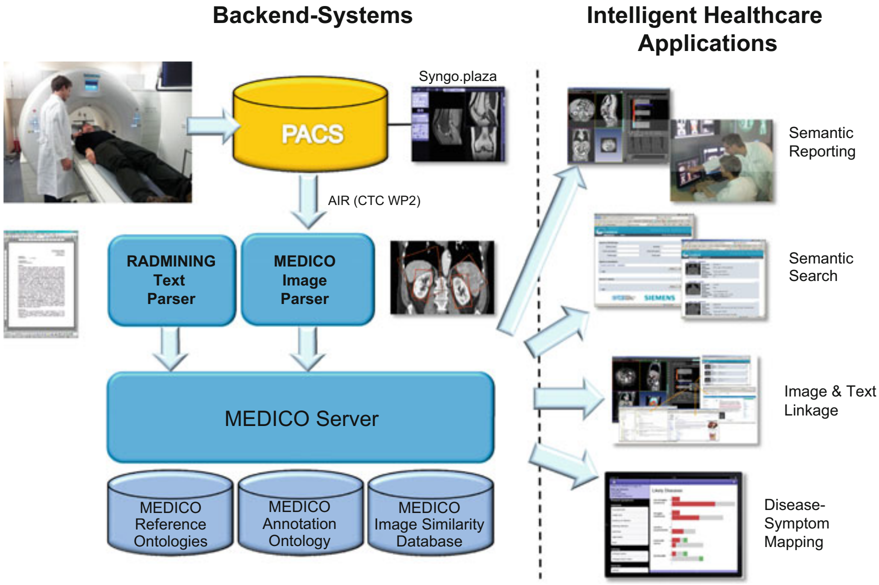


Fig. 5 The MEDICO system architecture illustrates the clinical data pre-processing steps required for the various intelligent healthcare applications

Database. The MEDICO Reference Ontologies provide the system with a consistent terminology, i.e. the vocabulary of the radiologist, and additionally anatomical and pathological knowledge. The MEDICO Image Similarity Database stores an index of image features to efficiently locate matching images in the PACS given a reference image section. This database is filled by the image parser and used by the semantic search application. As an additional modality for generating knowledge out of report texts, the MEDICO system uses a text parser developed in the associated KMU project RADMINING. The outcome of the text parser is analogous to that of the image parser stored in the MEDICO Annotation Ontology.

Overview of the MEDICO Use Case and System Figure 4 provides an overview of the MEDICO semantic processing units, the intelligent MEDICO healthcare applications, as well as the MEDICO intelligent interaction and decision support components. In Zillner et al. (2014), we describe the semantic processing of medical data that relies on intelligent algorithms for the semantic annotation of medical images, approaches for extracting information from radiology reports, and various means for the efficient management of ontological knowledge. On top of the MEDICO semantic processing unit, several intelligent healthcare applications (Seifert et al. 2014), such as semantic search, semantic reporting, the seamless integration of medical images and radiology reports as well as the efficient navigation between disease and symptom information, could be realized. In Sonntag et al. (2014) we focus on intelligent interaction approaches with the image material;

in particular, we will show what mobile radiology interaction and decision support systems of the future, which rely on semantic image annotations, may look like.

5 ORDO: Organizing Digital Information

Organizations, as well as their members, face several challenges due to the fast-paced changes in their environment. Using the abundance of information available to them efficiently – nowadays called "Big Data" – enables them to deal with these challenges. However, to actually realize the promises related to Big Data, semantic applications are needed that automatically analyze large volumes of data independently of their format, media type or location. These applications do not only need to fulfil technical requirements, but have to be implemented and operated in a financially sensible way. In particular, these applications must cope with increasing demands for data processing capability and new requirements, thus retaining investments. These kinds of technologies were developed within the THESEUS use case ORDO.

Business Challenges We are surrounded by content that is relevant to us. We can access content anywhere and anytime using smartphones, classic desktop PCs, notebooks or tablets. Furthermore, technology gets more and more intuitively usable with gestures or voice commands used in modern mobile devices.

Considering these offerings as a challenge or an opportunity depends on our ability to find the relevant content. In addition to the increasing amount of content, the available content gets more and more dispersed across different places such as folders, email accounts or social networks.

In the past, the growth of accessible content was often named negatively as "information overload". Nowadays, people seem to realize the chances that appear with the available content. This is shown by the relative distribution of the searches for "information overload" and "Big Data" in Google: From 2004 to 2006, information overload seemed to get a high search rate. Big Data seemed to be a minor topic from 2005 to 2010, with a leap in searches at the beginning of 2011.

In the authors' opinion, the reason for this mind shift is that there is an actual need to deal with Big Data. As stated in the *Extreme Information* report of Gartner (Beyer et al. 2011), it is not sufficient to just efficiently store and retrieve large volumes of content, but it is necessary to access it independently of its various formats and data sources. To provide a sound basis for business decisions, analyzing and condensing the relevant content needs to be done within a reasonable time span. Finally, available background knowledge – e.g., ontologies or open linked data – need to be considered for an analysis to create business value.

Big Data technologies enable an organization to open itself to the inside as well as to the outside. On the inside, available content from past activities is analyzed and provided within current business processes. This is needed, since the number of non-repetitive tasks increases steadily. These non-repetitive tasks demand access to available knowledge that is not known to a certain user. For example, within an investment company, analysts have to get acquainted with new technologies to make investment decisions. An example on how to support this research is the application refinder,[9] where recommendations from data sources within the organization provide information from other members. On the outside, Big Data technologies allow organizations to analyze external sources, such as customer statements or available background information. Two developments within ORDO provide an example: Attensity Analyze and Ontonaut. Attensity Analyze[10] allows access to quantitative feedback from customer statements within social networks, for example Facebook, or customer portals. Based on linguistic techniques, one can analyze the customers' statements about a product. For example, statements from customers about mobile phones – i.e. about its product features such as display or price – can be analyzed, which allows businesses to decide on the marketing strategy for the product.

Ontonaut,[11] the other example, provides named entity recognition and open linked data access as web services. These services are used by the semantic editor veeeb to detect relevant terms within an edited text. Background information and related media, such as pictures, are provided about these terms. This background information accelerates the creation of new content.

The Mission of ORDO ORDO's use case has created a highly flexible set of technologies that can be combined to support organizations as well as their members in their quest for relevant content. The main idea is to broaden the overview of a certain domain by lowering effort – or cost – related to obtaining relevant information. Currently, members of an organization obtain most of their information within the sources that are easily accessible and that are used most often. Within these sources, they "stumble upon" the content that is relevant to them. This is sufficient if the performed task is repetitive, and therefore the information needed to perform these tasks is easily available. However, as already stated, repetitive tasks are becoming less and less common. In addition, since these resources are used by other individuals as well, there is a lower chance that the information will lead to a competitive advantage. By reducing the effort required for finding and using relevant content, individuals are able to use relevant information, even if it is not located in popular sources.

This reduction of effort for the individual is the tangible effect created by semantic technologies. The individual is able to take information into account – pictorially "cherry picking" – which would not have been considered in the past due to resource constraints. The effort for the individual is directly related to precision and recall. Semantics increase precision and recall, providing individuals

[9] http://www.getrefinder.com

[10] http://www.attensity.com/products/attensity-analyze/

[11] http://www.ontonaut.net

with (a) precisely what they want, and (b) a more comprehensible overview of the results. Furthermore, they enable a user to analyze the results quantitatively.

To reduce the effort required to gain information through semantics in commercial successful applications, i.e. applications that pay off the investments made into them, the applications need to be affordable. This cost effectiveness needs to be valid for the whole lifecycle, during the initial investment, as well as during the operation. The technologies developed within ORDO enable the user to orchestrate specialized high-quality components in a highly scalable and flexible infrastructure. This facilitates the implementation of applications to realize quick returns on investments, which can be extended later, when the application has been successfully established. Such applications grow with customer demands in terms of the size of data, as well as the functionalities needed.

Another reason to employ specialized content analysis within an organization is that the statistical algorithms based on syntactical relations – like the page rank (Langville and Meyer 2006) used within a Google web search – yields suboptimal results since content is not related. For example, a presentation about a new product is not related to the marketing brochure, since they reside in different sources. Furthermore, different names might be given to the same product. For instance, the name of the development project in the presentation and the official product name mentioned in the marketing brochure. Based on this background information, relations within the content can be established that would otherwise stay unnoticed.

Results Mainly, three results of ORDO enable organizations to deal with Big Data. The industrial-strength scalable Open Source framework SMILA orchestrates the access to and analysis of data. SMILA is already in use in industrial applications as it is an established project from the Eclipse Foundation with stable interfaces.[12] Furthermore, it flexibly combines asynchronous and synchronous workflows in one framework for efficient data processing, thus providing a sound basis for commercial applications.

Neumann et al. (2014) describe how linguistics can be used to analyze unstructured textual data, and gives further examples of real-life applications. It outlines the challenges and solutions of how to “understand” textual data. Based on these linguistic technologies, organizations can draw valuable conclusions for day-to-day as well as strategic decisions such as technology development or investments.

Nick et al. (2014) give an overview of a highly scalable infrastructure to efficiently store and access data. This infrastructure – which includes SMILA – allows a real-time retrieval of up to 2,000 million documents. The section also shows the results of the scalability experiments with up to 60 processing nodes.

[12] http://www.eclipse.org/smila/

6 PROCESSUS: Management of Application, Solution and Service Knowledge in Mechanical and Plant Engineering

About 80 % of information in enterprises is available in unstructured documents. PROCESSUS and the associated small- and medium-sized enterprises (SMEs) use cases aim at extracting and unlocking this information to make it accessible in knowledge-intensive business processes. Therefore, semantic technologies to extract information are developed. The technologies range from textmining tools to NC program (numeric control) analyses to data analyses from graphical layout tools. The extracted information is stored in semantic knowledge repositories like ontologies. With suitable search technologies, the information is made accessible in business processes. This enables enterprises to perform their business processes more efficiently and strengthen their competitiveness. Three examples for the successful application of semantic technologies for supporting the management of application, solution, and service knowledge in mechanical and plant engineering are presented in this overview.

Efficient and effective business processes are key factors for successful enterprises in today's global markets. One important aspect for good business processes is to make the best possible use of existing knowledge inside the company. Within the PROCESSUS use case, embedded inside the THESEUS research program, an important step towards the development and use of semantic technologies supporting the management of knowledge in business processes has been taken. Semantic technologies are developed to ensure that users have access to the right information at the right time. This optimizes the search process by reducing unnecessary search effort and increasing the retrieval rate of needed information and therefore helps to lower costs in business processes. Three different scenarios from industry are described to present and discuss the exemplary use of semantic technologies and their particular software applications: "PROCESSUS – The semantic Enterprise", "MachInNet", and "SERAPHIM".

PROCESSUS – The Semantic Enterprise Today, companies face an increasing challenge when accessing their knowledge for company-specific processes. Growing global competition demanding new, innovative products and services, and the continuous pressure to reduce the time-to-market are only some reasons for the challenges faced by companies (Abuosba et al. 2008). Both, the reuse of existing knowledge as well as the intelligent storage of new knowledge are important factors for performing business processes efficiently (Hesse et al. 2009). For example, in the product development process the correct handling of information is one of the most important factors for successful product development (Lindemann 2009). However, the specific information required for different processes is hidden "somewhere" in the organization's content assets. Especially, the handling and management of the growing volume of unstructured data stored in different CMS, ERP and database systems is a huge challenge. These systems are mostly unable to communicate with each other, causing inconsistent and outdated information. In addition, critical information is often hidden in unstructured documents – such as texts, contracts, analyses and reports, plans, records and emails. It is typical for unstructured data that the interpretation of the meaning of the needed information is missing entirely, causing a so-called "semantic gap" in the interpretation of information. The semantic gap hinders employees' capability to access the right information at the right time (Gaag et al. 2008). This leads to unnecessary rework and work cycles that slow down the processes. Within the PROCESSUS use case, a solution was developed to overcome these current shortcomings using intelligent, innovative semantic technologies. It improves the access to existing knowledge by combining different data sources to facilitate the understanding of unstructured data. By applying this solution, knowledge-intensive business processes become more efficient.

Main Achievements As a result of PROCESSUS, a semantic business integration platform was developed and tested within different application scenarios. For example, it serves to support the automated annotation of existing documents and the following search for existing technical solutions in the automation industry. The PROCESSUS platform combines the various types of information sources found in a company's content assets. This enterprise knowledge is integrated into a specifically developed semantic knowledge model (ontology). This enables and facilitates the interpretation and deduction of relations between the data and closes the semantic gap between the specialized terminology and domain knowledge. The PROCESSUS platform facilitates the actual use of corporate information to support dynamic processes. The platform links information and documents semantically, so that their relations can be processed automatically by the computer. The location of the stored information is no longer a concern for the users. Therefore, access to the right information, at the right time, is always guaranteed.

The PROCESSUS platform incorporates tools for the semantic access of information and contextual relations in collections of unstructured data. The emphasis is placed on new components and services as well as on building and expanding existing ones. The most important standard services developed are:

- *The Facts Extraction Service* analyzes unstructured text documents. Linguistic algorithms are parameterized with instances from the ontology. Text is delivered to the service and recognized instances, i.e. facts, are returned. This service is necessary for the creation of information from unstructured documents – it reveals hidden information locked in documents with respect to the ontology.
- *The Annotation Service* extracts information from the Facts Extraction Service and enriches text documents with metadata. This service encompasses transaction handling and stores data in the ontology. The Annotation Service is responsible for tracking changes. If a document is altered, this service executes the necessary changes in the ontology. Changes can be the insertion, update or deletion of descriptive metadata in the knowledge repository.

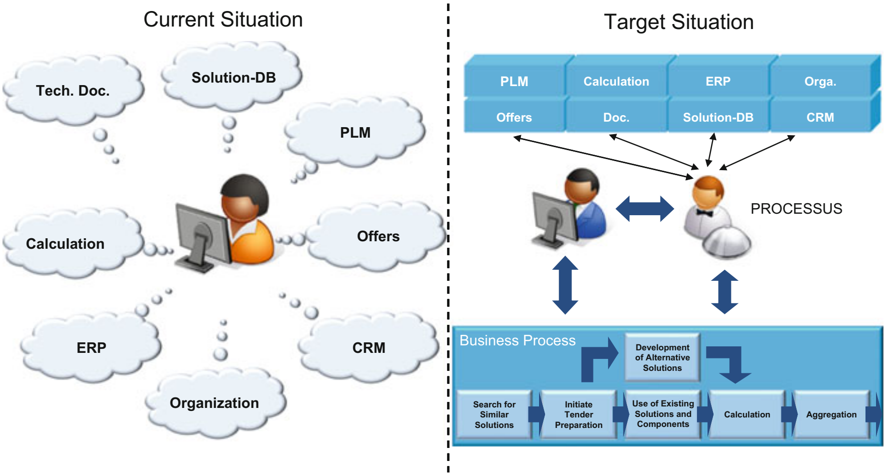


Fig. 6 The PROCESSUS Vision (According to (Franz et al. 2008))

- *The Search Service* is most essential in providing efficient access to documents. A search query is transferred to this service and is then analyzed by the Facts Extraction Service, delivering the identified instances from the ontology. With these instances, a search request is addressed at the knowledge repository, providing the best matching documents.

The PROCESSUS platform can be best understood when compared to a butler who provides the right information to the user, at the right time, as shown in Fig. 6 (Franz et al. 2008). The user is no longer lost in large amounts of different document types and databases, where relevant information is possibly hidden. The PROCESSUS platform establishes an intelligent interface to all of these knowledge sources. It provides the user with the right information – regardless of its current storage location and format – in the precise process context and process stage in which the user is, and at the exact time it is needed.

The developed PROCESSUS platform serves as a horizontal service platform allowing for different vertical applications to be built on top of it. The above-mentioned services are integrated into the PROCESSUS platform. Additionally, specific services for different domains and industries can individually be integrated as web services. The developed concepts and solutions are tested in different fields of application. For example, in the field of automation industry in mechanical engineering, the search for existing solution knowledge is addressed. Scenarios for sharing knowledge within company-internal business processes as well as scenarios for accessing knowledge from different companies are also addressed (Schmidt et al. 2011).

MachInNet – Machining Intelligence Network The technology knowledge accumulated in manufacturing companies is enormous. But many companies simply do not know what they know. Especially in smaller job shops this knowledge is “buried” in NC codes, tool layouts and other manufacturing documents. Therefore, crucial questions like

- “Have we solved a specific machining task before, and, if yes, how?”
- “What kinds of tools have been used for a certain task? Which speeds and feeds have been applied?”
- “Which tool components are needed to configure the tool assembly?”

might not be answered. This results in redundant engineering efforts and contradicts the goal of standardization. Opportunities for profitable repeat business are lost and resources are wasted by “reinventing the wheel”.

To improve this situation in the metal-working industry, CIM Aachen GmbH, EXAPT GmbH, and the Chair “I9” for Computer Science of RWTH Aachen University joined the THESEUS research program to develop a semantic, Internet-based knowledge network for SMEs. The project’s name “Machining Intelligence Network” (MachInNet) is, at the same time, its goal. MachInNet makes the large variety of technologies manageable and accelerates industrial engineering of mechanical parts.

The challenge to “unearth” this knowledge and to make it accessible for daily use was solved as part of the THESEUS research program. A new approach, using data mining algorithms and semantic search technologies, made it possible to reverse engineering data from different sources and to make it available for explicit use in technology databases, and for feature-based NCplanning. The overall goal is to create explicitly usable content for an open database. This Internet database is designed to be fed by all SMEs in the MachInNet network and offers solutions for concrete problems, using specially developed algorithms and similarity searches.

SERAPHIM Companies of the mechanical engineering and plant construction industry are increasingly reliant on the providing services in order to differentiate themselves from competitors and to bind their customers.

In the SERAPHIM project, a service platform (the SERAPHIM platform) is being developed that helps to optimize service and sales processes in the mechanical engineering and plant construction industry. On this platform, process relevant and up-to-date information specific to individual machines and plants are provided by interactive services. This enables the process participants to act quickly, based on the same, up-to-date information. The services on the platform can interact with other services on the platform as well as with external services. Services for the two scenarios “paperless repair process” and “service configurator” have been developed and have resulted in two prototypical demonstrators.

In SERAPHIM, business models were developed systematically and uniformly using a systematic approach of service engineering. By this means, gaps in the models can be closed before the market launch, which leads to higher chances of success for newly developed services. The service engineering process and service platform SERAPHIM support particularly small and medium enterprises in the professionalization of their service business.

After the end of the research project, the application partners of SERAPHIM will continue to develop their pilot systems to bring them into their productive environment. Further, similar services are planned for the future. In addition, results from the SERAPHIM project and the conceptual project work have influenced the development of a service offering for manufacturing companies.

7 TEXO: Business Webs in the Internet of Services

The TEXO use case contributed tools, methods, and concepts for the systematic use of the Internet for new ways of value creation in the services sector. This idea, the so-called Internet of Services, enables flexible business webs, i.e., the ad hoc collaboration of independent enterprises for delivering a value-adding service. Therefore, TEXO contributed to the further industrialization and automation of the service sector – the fastest growing industry sector of developed economies with increasing significance.

The service industry has become the biggest employer in Germany and tends to be a critical force for ensuring economic growth. As an example, consider the Federal Republic of Germany, where the largest part of the gross domestic product (GDP) of 2011 was generated by the service industry.

The growth in the services sector spawned many research activities and even the call for a new discipline, namely *service science* (Chesbrough and Spohrer 2006). The latter argues for a new multi-disciplinary academic approach in order to integrate academic silos and advance service innovation more rapidly. Consequently, several initiatives, institutions, and research projects have been established in recent years that revolve around the subject of service science. For example, the European Union initiated the *Future Internet Public Private Partnership* (FI-PPP[13]) aiming to advance Europe's competitiveness in Future Internet technologies and systems, and to support the emergence of Future Internet-enhanced applications of public and social relevance.

The most prominent example in Germany has been the TEXO use case of THESEUS.[14] Research in TEXO revolved around the systematic use of the Internet for enabling flexible business webs – frequently called the *Internet of Services*. Business webs are service networks where independent enterprises collaborate in order to deliver a value-adding service. The goal was to make such value-adding services accessible, discoverable, composable, deployable, and eventually tradable

[13] http://www.fi-ppp.eu/

[14] The use case was led by SAP and accompanied by industry partners such as Siemens and Attensity. Further industrial partners were Intelligent Views and ontoprise acting as solution and technology providers for semantic technologies (see also Kiss et al. 2014). In addition, there were a multitude of research centers including DFKI, Fraunhofer FOKUS, IAO, and IGD, and the FZI Karlsruhe. Finally, the Technical Universities of Dresden, Munich, Darmstadt and Karlsruhe complemented the consortium for basic research activities.

on the Internet. TEXO results allow us to generate such value-adding services, to develop innovative business models, and to ultimately establish new business webs.

As an example for a business web, consider a logistics scenario. A logistics provider might create a one-off business web to satisfy a specific shipment request of its customer. The logistics provider uses a web marketplace to flexibly compose rail and road transport services as well as container rental and customs clearance. Several real-life examples were elaborated on adjunct projects with dedicated small and medium-sized enterprises. [15]

Overall, more than 100 researchers brought forth 11 granted patents, 15 invention disclosures, and more than 200 scientific publications. Additionally, more than 100 bachelor and master theses as well as 17 PhD theses were realized in TEXO. The following sections provide an overview of the results.

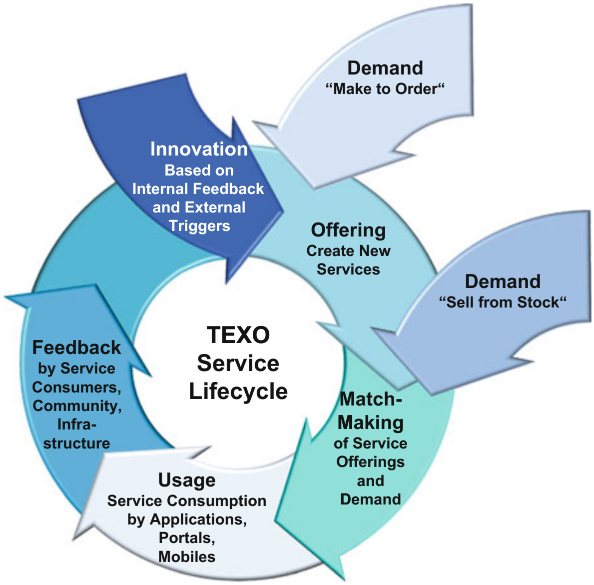
Research Results In essence, the TEXO use case has been organized along the phases of the service lifecycle (Fig. 7). TEXO provided significant scientific contributions, open-source tools and pilot applications across the whole service lifecycle.

Noteworthy results that span several lifecycle phases are (a) the *Service Governance Framework* (Winkler et al. 2009) supporting the operation of a service marketplace, and (b) *Lawful Service Engineering and Delivery* for enabling legal compliance by design in offering, matchmaking, and usage phases (Raabe et al. 2012).

Service Description Another group of results which covers all phases of the lifecycle concerns the description of services. With the growth in number and sophistication of services widely available, there is a new urgency for comprehensive service descriptions that take into account both technical and business aspects. Therefore, TEXO contributed the Unified Service Description Language (USDL) (Barros and Oberle 2012; Oberle et al. 2013), which provides a uniform conception of services across all walks of life. USDL is proposed as a normative and comprehensive master data model for commercial metadata of IT, physical, or hybrid services. More specifically, USDL allows for a unified description of business, operational and technical aspects of services. USDL aims at a holistic service description putting special focus on business aspects such as ownership and provisioning, release stages in a service network, composition and bundling, and pricing and legal aspects, among others, in addition to technical aspects. The complete USDL specification and tool chain (editor, repository and marketplace) are available as an open source solution at www.internet-of-services.com. USDL is a main outcome of TEXO and further described in Oberle (2014). In addition, TEXO brought forth several results that clarify and define the notion of a service. One major result has been the TEXO *Service Ontology* (Oberle et al. 2009), which includes *ontological foundations of service science* (Ferrario et al. 2011).

[15] In order to facilitate their applications and scenarios, TEXO developed, provided and operated a dedicated testing infrastructure – the TEXO Lab.

Fig. 7 Service lifecycle



Service Innovation The *Innovation* phase is of high importance for the success of a business web, which largely depends on offering "exciting" services and innovative business models. New and innovative ideas for future value-adding services must be created and implemented. Existing services and business models must constantly adapt to changing market requirements and customer needs.

Service Offering The innovation phase ends with a rough description for a new value-adding service. Subsequently, the *Service Offering* phase primarily comprises the activity of service engineering followed by offering the new service on a marketplace. The outcome is a fully described and implemented service archive that is deployed during the service usage phase, making the service executable, searchable and tradable.

Service Matchmaking After engineering the service and offering it on a marketplace, service consumers should be enabled to discover and select the best fitting services for their needs. This is the phase where dynamic business webs become relevant. A service consumer's need might not be realizable by a single service, but only by an ad hoc composition of services.

Service Usage The matchmaking phase ends with a contract between the service consumer and service provider. In the next step, the service consumer actually uses the service, hence the term *Service Usage* phase. In this phase, the service's performance has to be carefully monitored to ensure compliance with service level agreements (SLAs), and compensation measures may be taken. In addition, there has to be support for accessing of the service through various channels and invoking of it in various environments by service consumers. Therefore, TEXO provided a platform consisting of the newly contributed components.[16] Besides technical platform aspects, TEXO also addresses economic aspects (see Krings 2011; Statistisches Bundesamt 2012) and pricing models.

Service Feedback After the usage phase, the consumer may provide feedback of different kinds, e.g., ratings or forum discussions, which might be of interest to the service provider. In addition, service monitoring provides feedback such as logs and statistics back to the service provider for future generations of the service implementation. This feedback is input to the innovation phase and might be the starting point for future improvements or new value-adding services. Thus, the *Service Feedback* phase closes the loop of the service lifecycle in the innovation phase.

TEXO provided significant results across the whole service lifecycle. Their overall goal was to make value-adding services accessible, discoverable, composable, deployable, and eventually tradable on the Internet.

8 Conclusion

The innovative and disruptive technology solutions of the THESEUS research program which have been described in Becker et al. (2014) have already been the source for new applications. Driven by business targets, the industry partners implemented these applications to position themselves into new markets or even create new markets. While each individual partner brought in the knowledge of his relevant market, the contributions of those partners in the context of the THESEUS consortium added a new dimension. The network of individual companies assembled a variety of knowledge and experience which created new markets, established new business models, disseminated the innovations, and, as a consequence, has a significant impact on the competitiveness of the German IT industry.

References

M. Abuosba, A. Gaag, J. Franz, Semantic knowledge management – Erschließung von Lösungswissen in der Automatisierungsindustrie, in *Tagungsband vom 10. Kongress zum IT-gestützten Wissensmanagement*, ed. by M. Bentele, N. Gronau, R. Hochreiter, P. Schütt, M. Weber (CMP-WEKA-Verlag, Frankfurt am Main, 2008), pp. 351–358

A. Barros, D. Oberle (eds.), *Handbook of Service Description: USDL and Its Methods* (Springer, Berlin/Heidelberg/New York, 2012)

T. Becker, C. Burghart, K. Nazemi, P. Ndjiki, T. Riegel, R. Schäfer, T. Sporer, V. Tresp, Core Technologies for the Internet of Services, in *Towards the Internet of Services: The THESEUS*

[16]The Life-DVD at http://serviceplatform.org/wiki/SPACEflight contains a selected number of the TEXO platform software artifacts, including the USDL tool chain.

Research Program, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

M.A. Beyer, A. Lapkin, N. Gall, D. Feinberg, V.T. Sribar, “Big Data” Is Only the Beginning of Extreme Information Management, April 2011, http://www.gartner.com/id=1622715

L. Bogoni, P. Cathier, M. Dundar, A. Jerebko, S. Lakare, J. Liang, S. Periaswamy, M.E. Baker, M. Macari, Computer-aided detection (CAD) for CT colonography: a tool to address a growing need. Br. J. Radiol. **78**(1), 57–62 (2005)

C. Böhme, M. Traub, A. Bergholz, J. Hannemann, L. Svensson, Semantic linking in CONTENTUS, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

H. Chesbrough, J. Spohrer, A research manifesto for services science. Commun. ACM Serv. Sci. **49**(7), 35–40 (2006)

K. Doi, Computer-aided diagnosis in medical imaging: historical review, current status and future. Comput. Med. Imaging Graph. **37**(4), 198–211 (2007)

R. Ferrario, N. Guarino, C. Janiesch, T. Kiemes, D. Oberle, F. Probst, Towards an ontological foundation of services science: the general service model, in *10th International Conference on Wirtschaftsinformatik*, Zürich, vol. 2, Feb 2011, pp. 675–684, http://dblp.uni-trier.de/db/conf/wirtschaftsinformatik/wi2011.html#FerrarioGJKOP11

J. Franz, A. Gaag, M. Abuosb, The semantic enterprise – bringing meaning to business processes, in *14th International Conference on Concurrent Enterprising (ICE ’08)*, Lisbon, June 2008

A. Gaag, A. Kohn, J. Franz, Semantic enterprise – unleashing solution knowledge in the area of mechanical engineering, in *2nd European Semantic Technology Conference (ESTC ’08)*, Vienna, June 2008

H. Hesse, M. Abuosba, J. Lütkehellweg, M. Ott, Automatische Generierung von Unternehmensontologien – Schaffung einer Unternehmenssemantik, in *Tagungsband vom 11. Kongress zum IT-gestützten Wissensmanagement*, Bad Homburg, Sept 2009, pp. 49–56

E.M. Kiss, S. Brockmans, J. Angele, Semantic technologies for the internet of services, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

I. Konya, S. Eickeler, J. Nandzik, N. Flores-Herr, Print processing in CONTENTUS – restoration of digitized print media, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

S. Krings (ed.), *Statistisches Jahrbuch 2011 für die Bundesrepublik Deutschland mit Internationalen Übersichten*, 1st edn. (Statistisches Bundesamt, Wiesbaden, 2011)

A.N. Langville, C.D. Meyer, *Google Page Rank and Beyond* (Princeton University Press, Princeton, 2006)

U. Lindemann, *Methodische Entwicklung technischer Produkte*, 3rd edn. (Springer, Berlin/Heidelberg/New York, 2009)

K. Marten, C. Engelke, Computer-aided detection and automated CT volumetry of pulmonary nodules. Eur. Radiol. **17**(4), 888–901 (2007)

J. Nandzik, A. Heß, J. Hannemann, N. Flores-Herr, K. Bossert, CONTENTUS – towards semantic multimedia libraries, in *Proceedings of the World Library and Information Congress: 76th IFLA General Conference and Assembly – “Open Access to Knowledge – Promoting Sustainable Progress”*, Gothenburg, Aug 2010

G. Neumann, G. Paaß, D. van den Akker, Linguistics to structure unstructured information, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

M. Nick, T. Jäger, V. Nussbaum, K. Kramer, High scalability for semantic indexes, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

D. Oberle, A unified description language for the internet of services, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

D. Oberle, A. Barros, S. Heinzl, U. Kylau, A unified description language for human to automated services. Inf. Syst. **38**(1), 155–181 (2013)

D. Oberle, N. Bhatti, S. Brockmans, M. Niemann, C. Janiesch, Countering service information challenges in the internet of services. Bus. Inf. Syst. Eng. (BISE) **1**(5), 370–390 (2009), http://dblp.uni-trier.de/db/journals/bise/bise1.html#OberleBBNJ09

O. Raabe, R. Wacker, D. Oberle, C. Baumannn, C. Funk, *Recht ex machina: Formalisierung des Rechts im Internet der Dienste* (Springer, Berlin/Heidelberg/New York, 2012)

E. Samei, J. Seibert, K. Andriole, A. Badano, J. Crawford, B. Reiner, M. Flynn, P. Chang, AAPM/RSNA tutorial on equipment selection: PACS equipment overview: general guidelines for purchasing and acceptance testing of PACS equipment. Radiographics **24**(1), 313–334 (2004)

H.X. Schmidt, A. Kohn, U. Lindemann, Domain specific language in technical solution documents – discussion of two approaches to improve the semi-automated annotation, in *International Conference on Knowledge Engineering and Ontology Development (KEOD '11)*, Paris , ed. by J. Filipe, J.L.G. Dietz (SciTePress, 2011), pp. 159–166, http://dblp.uni-trier.de/db/conf/ic3k/keod2011.html#SchmidtKL11

S. Seifert, M. Hammon, M. Petri, H. Oberkampf, P. Daumke, Intelligent healthcare applications, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

D. Sonntag, S. Zillner, P. Ernst, C. Schulz, M. Sintek, P. Dankerl, Mobile radiology interaction and decision support systems of the future, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

Statistisches Bundesamt, Ausführliche Ergebnisse zur Wirtschaftsleistung im 4. Quartal 2011, Feb 2012, https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2012/02/PD12_063_811.html

P. Taylor, J. Champness, R. Given-Wilson, K. Johnston, H. Potts, Impact of computer-aided detection prompts on the sensitivity and specificity of screening mammography. Health Technol. Assess. **9**(6), 1–58 (2005)

A. Vogel, P. Effenberg (eds.), *Digitalisierungsfibel: Leitfaden für audiovisuelle Archive* (Transfer-media, Potsdam, 2011)

M. Wendt, M. Gerlach, H. Düwiger, Ask like an Egyptian – question answering in the Alexandria use case, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

M. Winkler, J. Spillner, A. Schill, SLA Management and Contract-Based Service Execution, in *Service-Oriented Computing*, ed. by L. Baresi, C.H. Chi, J. Suzuki. Volume 5900 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2009), pp. 653–655

F. Xu, H. Uszkoreit, H. Li, P. Adolphs, X. Cheng, Domain-adaptive relation extraction for the semantic web, in*Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

S. Zillner, S. Seifert, M. Erdt, P. Daumke, M. Kramer, Semantic processing of medical data, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

Further Reading on ALEXANDRIA

P. Adolphs, F. Xu, H. Li, H. Uszkoreit, Dependency graphs as a generic interface between parsers and relation extraction rule learning, in *KI 2011: Advances in Artificial Intelligence*. Volume 7006 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011), pp. 50–62

S. Ahlheid, G. Gräfe, M. Krebs, A., F.J.-P., Kuhlmann, Ansätze zur Validierung semantischer Informationstripel auf einer Web 2.0-Wissensplattform, in *Virtual Enterprises Communities & Social Networks: Workshop GeNeMe '10 – Gemeinschaften in Neuen Medien*, ed. by K. Meißner, M. Engelien, 1st edn. (TUpress, Dresden, 2010), pp. 95–106

I. Garnefeld, A. Iseke, A. Krebs, Explicit incentives in online communities: boon or bane? Int. J. Electron. Commerce **17**(1), 11–38 (2012)

W.C. Kammergruber, M. Langen, Tagging als soziales Bindeglied für Communities, in *Virtuelle Organisation und Neue Medien 2009 (GeNeMe 2009)* ed. by K. Meißner, M. Engelien (TUpress, Dresden, 2009), pp. 35–43, http://w3-mmt.inf.tu-dresden.de/geneme/

M. Langen, W.C. Kammergruber, Tagging versus ontologien? Informationsstrukturierung im Enterprise 2.0, Oct 2008, http://www.iw-live.de/wissen/2009/de/home/

M. Langen, W.C. Kammergruber, Vom Tagging zu leichtgewichtigen Ontologien – Tag-Thesaurus und soziale, semantische Suche, in *Mit Wissensmanagement Innovationen vorantreiben!*, Sept 2010, pp. 239–248

H. Li, X. Cheng, K. Adson, T. Kirshboim, F. Xu, Annotating opinions in German political news, in *8th ELRA Conference on Language Resources and Evaluation (LREC)*, Istanbul (European Language Resources Association (ELRA), May 2012)

H. Li, F. Xu, H. Uszkoreit, Minimally supervised rule learning for the extraction of biographic information from various social domains, in *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, Hissar (RANLP 2011 Organising Committee, Sept 2011), pp. 17–24, http://aclweb.org/anthology/R11-1003

C. Maaß, G. Gräfe, Sechs Thesen zum Social Bookmarking. HMD: Praxis der Wirtschaftsinformatik **46**, 100–107 (2009)

Further Reading on TEXO

M. Niemann, J. Eckert, N. Repp, R. Steinmetz, Towards a generic governance model for service oriented architectures, in *Learning from the Past and Charting the Future of the Discipline: 14th Americas Conference on Information Systems (AMCIS '08)*, Toronto, ed. by I. Benbasat, A. Montazemi (2008)

Domain-Adaptive Relation Extraction for the Semantic Web

Feiyu Xu, Hans Uszkoreit, Hong Li, Peter Adolphs, and Xiwen Cheng

Abstract In the THESEUS ALEXANDRIA use case, information extraction (IE) has been intensively applied to extract facts automatically from unstructured documents, such as Wikipedia and online news, in order to construct ontology-based knowledge databases for advanced information access. In addition, IE is also utilized for analyzing natural language queries for the ALEXANDRIA question answering system. The DARE system, a minimally supervised machine learning system for relation extraction, developed at the DFKI LT-Lab, has been adapted and extended to the IE tasks for ALEXANDRIA. DARE is domain-adaptive and has been used to learn relation extraction rules automatically for the ALEXANDRIA-relevant relations and events. Furthermore, DARE is also applied to the ALEXANDRIA opinion mining task for detecting opinion sources, targets and polarities in online news. The DARE system and its learned rules have been integrated into the ALEXANDRIA IE pipeline.

1 Introduction

Information extraction (IE) has been acknowledged as a core information technology for the constantly growing digitalized world, in particular, for the World Wide Web (WWW). The goal of IE systems is to find and link pieces of the relevant information from natural language texts and store these information pieces in a database format. As an alternative to storing the extracted information pieces in a database, these pieces could also be appropriately annotated in a markup language and thus be made available for indexing and database retrieval (Appelt and Israel 1999; Grishman and Sundheim 1996). Tim Berners-Lee defines the Semantic Web as "a web of data that can be processed directly and indirectly

F. Xu (✉) • H. Uszkoreit • H. Li • P. Adolphs • X. Cheng
DFKI GmbH, German Research Center of, Artificial Intelligence, Saarbrücken, Germany
e-mail: feiyu@dfki.de; hansu@dfki.de; lihong@dfki.de; peter.adolphs@dfki.de; xiwen.cheng@dfki.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_22,

Table 1 An example of concept entities

Concept	Extracted entities
Prize area	*Physics*
Person name	*Dr. Robert Laughlin*, *Dr. Horst Stoermer*, *Dr. Daniel Tsui*
Monetary amount	*$978,000*
Organization	*Stanford University*, *Columbia University*, *Princeton University*

by machines." However, we know that the current web is still far away from the structured setup of the Semantic Web, because the WWW was originally designed for the publication and consumption of content by humans, and not by machines. Thus, the IE technologies are important for closing the gap between the current web and the Semantic Web with two major goals:

- Allowing information providers to analyze and extract structured data from free web texts, and identify and link the data;
- Providing end users more natural and advanced search options to the web content, such as semantic search and question answering systems.

Therefore, IE is central for the ALEXANDRIA search platform, whose task is to provide intelligent semantic information access to the online information available on the Web. The central IE tasks include finding references to relevant concepts or objects such as names of people, companies and locations, as well as detecting relationships among them, e.g., the birth place of a Nobel Prize winner. Let us look at the following text (Example 1) about the Nobel Prize award event, provided by Xu (2008):

Example 1.
The *Physics prize*, also *$978,000*, will be shared by *Dr. Robert Laughlin* of *Stanford University*, *48*, *Dr. Horst Stoermer*, *49*, a German-born professor who works both at *Columbia University* in *New York* and at *Bell Laboratories* in *Murray Hill*, *N.J.*, and *Dr. Daniel Tsui*, *59*, a Chinese-born professor at *Princeton University*.

If we want to extract events of prize winning, the relevant concepts to be extracted from the above texts are entities such as *prize area*, *monetary amount*, *person name* and *organization* (see examples in Table 1). Award relevant relations include the relation between *person* and *organization* and the relation among *person*, *prize area* and *monetary amount* (Table 2).

One pillar of the ALEXANDRIA platform is its knowledge acquisition pipeline. The pipeline builds upon raw text sources. This can be either a relatively static document collection such as Wikipedia or a dynamic daily news collection, filled by a news aggregator, which gathers current issues of German newspapers and magazines on a daily basis and cleanses them for further processing. Once the content-bearing text content is identified and extracted, it is linguistically analyzed

Table 2 An example of relation instances

Relation	Extracted relation instances
person, affiliation	
	⟨ *Dr. Robert Laughlin*, *Stanford University* ⟩
	⟨ *Dr. Horst Stoermer*, *Columbia University* ⟩
	⟨ *Dr. Daniel Tsui*, *Princeton University* ⟩
person, prizeArea, monentaryAmount	
	⟨
person	{*Dr. Robert Laughlin*,
	Dr. Horst Stoermer,
	Dr. Daniel Tsui},
prize area	*physics*,
monentary amount	*$978,000*
	⟩

and everything is passed to the IE modules, which try to squeeze out every piece of information that is relevant with respect to the modeled target domains.

In ALEXANDRIA, IE has been intensively applied to enrich the ALEXANDRIA knowledge base – a rich, ontology-driven repository of world knowledge, covering famous people, organizations, locations, cultural artifacts, events, and the like. Furthermore, it is also utilized for analyzing natural language queries for the ALEXANDRIA question answering system. Moreover, one of the central topics in the IE research, namely, domain-adaptive relation extraction, has been addressed in ALEXANDRIA. The key feature of "domain-adaptive" systems is that they can be applied to new domains and applications with little effort. We have further developed the domain-adaptive relation extraction system DARE of the DFKI LT-Lab for the ALEXANDRIA IE tasks.

DARE is a minimally supervised machine learning system for relation extraction (Xu 2008; Xu et al. 2007). "Minimally supervised" means that only very little human intervention is needed to establish the system for a new domain. The DARE system can learn rules for any relations or events automatically, given linguistically annotated documents and some initial examples of the relations. Furthermore, DARE can utilize the learned rules for extracting relations or events from free text documents. The DARE system has been applied to the event detection and opinion mining tasks in ALEXANDRIA. The biographic information is a relevant part of the ALEXANDRIA knowledge database. Thus, we have conducted a systematic analysis of the extraction performance of DARE on biographic information from various social domains of Wikipedia documents, namely, politicians, business people and entertainers (Li et al. 2011). Furthermore, we have adapted DARE to be able to deal with various parsers (Adolphs et al. 2011), and have proposed a novel strategy to adapt a deep parser to the domain-specific relation extraction task (Xu et al. 2011). A deep parser delivers grammatical relations within a sentence. The DARE relation extraction machinery and its learned rules have been integrated into the ALEXANDRIA IE pipeline for the event detection and opinion mining tasks.

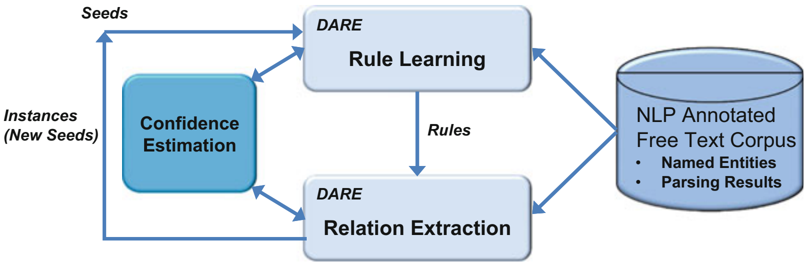


Fig. 1 DARE core architecture

2 Integration of DARE Framework in ALEXANDRIA IE Platform

DARE is a minimally supervised machine learning system for relation extraction (RE) on free texts (Xu 2008; Xu et al. 2007). It consists of two parts: (1) rule learning and (2) relation extraction (RE). Rule learning and RE feed each other in a bootstrapping framework. The bootstrapping starts from so-called “semantic seeds”, which are small sets of instances of the target relation. The rules are extracted from sentences, annotated with semantic entity types and linguistic parsing results, e.g., dependency structures, which match with the seeds. RE applies the acquired rules to texts in order to discover more relation instances, which in turn are employed as seeds for further iterations. The core system architecture of DARE is depicted in Fig. 1. The entire bootstrapping stops when no new rules or new instances can be detected. Relying entirely on semantic seeds as domain knowledge, DARE can accommodate new relation types and domains with minimal effort.

The confidence values of the newly acquired rules and instances are calculated in the spirit of the “duality principle” (Agichtein and Gravano 2000; Brin 1999; Yangarber 2001), i.e., the confidence values of the rules are dependent on the truth values of their extracted instances, and on the seed instances from which they stem. The confidence value of an extracted instance makes use of the confidence value of its ancestor seed instances.

DARE can handle target relations of varying arity through a compositional and recursive rule representation and a bottom-up rule discovery strategy. A DARE rule for an *n*-ary relation can be composed of rules for its projections, namely, rules that extract a subset of the *n* arguments. Furthermore, it defines explicitly the semantic roles of linguistic arguments for the target relation.

DARE is employed in the ALEXANDRIA IE pipeline as an IE module with deep analytic capabilities. This computational power is required by the semantic relations we want to recognize – mentioned events, together with their characterizing properties, and expressed opinions. Both targeted relation types strongly benefit from a more detailed analysis of the expressed meaning, in order to capture subtle meaning differences that cannot be covered by mainstream statistical bag-of-words approaches. Both topics will be detailed in Sects. 2.1 and 2.2, respectively.

The technical integration of DARE follows the architectural design decisions of the ALEXANDRIA platform. The IE pipeline uses Apache's open source UIMA framework[1] as a middleware for integrating the input and processing results of all NLP components. UIMA provides a general framework (i) for describing, representing and storing rich stand-off annotations of data, (ii) for implementing analytical components as well as for modelling and running pipelines on top of these components, and (iii) graphically for running components and pipelines and inspecting results. For each task, we create a specific UIMA module, which wraps the DARE system and provides the required input and output interfaces. Each module is accompanied by a formal type system, describing the kind of annotations the module is consuming and producing.

2.1 Event Identification and Extraction

In order to systematically detect common types of events, we analyzed a newspaper corpus of 50,000 documents prepared by Neofonie. Every document in this corpus belongs to one of the following news sections: *Front Page, Politics, Economy, Health, Internet, Culture, Science, Sports, Automobiles & Technology, Local, Media, Travel,* and *Miscellaneous*. Using the tf-idf weight, we determined for each section the relevant terms. From the terms with the highest tf-idf frequency in each domain, we manually identified the nine event types shown in Table 3. The relevant terms with high frequency for the corresponding event types are stored in a typed gazetteer, which serves as a resource for finding cues for mentioned events in the documents.

The recognition of event features (e.g. date, place, participants) is realized by semi-supervised learning of relation extraction rules using DARE. First, we extracted event entities with the typed gazetteer. Then we constructed start seeds for DARE with a general event name, e.g., "parliamentary elections" (DE: "Bundestagswahl"), and some possible named entities of different types (e.g. person, location, organization) to learn extraction rules in DARE. With all the seeds created this way, we used DARE to learn all sorts of rules between named entities and event names. We used further semi-automatic control methods to filter all learned rules according to their relevance and plausibility, and determined the precise roles of the (person or organization) participants in the relation. For example, from the sentence

> Die neue IWF-Chefin *Christine Lagarde* wird am Donnerstag in Brüssel an einem *Sondergipfel* zur Schuldenkrise in der Euro-Zone teilnehmen.

we can learn the following DARE rule (here shown in a shortened format):

[1] http://uima.apache.org/

Table 3 Event types and the gazetteer examples

Event type	Gazetteer size	Examples
Election	24	Direktwahl, Präsidentschaftswahl, Parlamentswahl
Conference	83	Euro-Gipfel, Bundespressekonferenz
Scandal	67	Abhörskandal, Bestechungsskandal, Abhöraffäre
Crisis	77	Schuldenkrise, EU-Krise
Social movement	8	Studentenunruhen, Aufstandsbewegung
Disaster	23	Atomkatastrophe, Atom-GAU, Taifun, Bombenexplosion
Festival	51	Filmfestival, Musikfestival
Terror	8	Bombenanschlag, Terrorattacke
Others	4	Wiedervereinigung, Party

"werden" { SB(PERSON:participant),
OC("teilnehmen") { MO ("an") {PNK(EVENT: event)} } }

where the EVENT and PERSON are NE types. With this rule, we can extract the participants of an event. The learned rules can be used to identify new event terms and their participants, which are not defined in the gazetteer. For instance, we can update the rule above by replacing the condition $type = EVENT$ with $POS = Noun$:

"werden" { SB(PERSON:participant),
OC("teilnehmen") { MO ("an") {PNK(Noun: event)} } }

2.2 Opinion Mining

An important application in ALEXANDRIA is the automatic monitoring of news for the identification of published opinions about politicians. Reported opinions are not only recognized, but the sentiment is also classified as positive, negative or neutral. Using this technology, the public image of a person, organization or brand can be tracked over time and general trends can be recognized. Opinion and trend mining are therefore crucial instruments for journalistic exploration, business intelligence, or reputation management.

We have applied DARE to learn relation extraction rules that identify the source of the opinion, the target of the opinion, and the polarity of the sentiment. The rules are learned from an annotated newspaper corpus described in Li et al. (2012). Concerning the expression level annotation, the most relevant and well-known work is the Multi-Perspective Question Answering (MPQA) corpus (Wiebe et al. 2005; Wilson 2007; Wilson and Wiebe 2005). This work presents a schema for annotating expressions of opinions, sentiments, and other private states in newspaper articles. To our knowledge, there is only one German corpus available that is annotated with opinion relevant information, which is in the domain of user-generated reviews (Schulz et al. 2010). In comparison to product reviews, opinions in newspaper articles are in general expressed with more subtle means.

Example 2(a) and (b) are the headlines of two different news articles that recount the same event, where German chancellor Merkel has obtained the Medal

Table 4 Opinion frame elements

Element	Property
Target	
Source	
Text anchor	isaIdiom, isaPhrase, isaWord, isaCompoundNoun
Polarity	positive, negative, neutral
Auxiliary	isaNegation, isaIntensifier, isaDiminisher

of Freedom by President Obama. The two authors used two different verbs: Example 2(a) with the verb *ehren* (Engl. honor) and Example 2(b) with the verb *überreicht* (Engl. hands over). The first one is strongly positive, while the second one is more or less neutral.

Example 2.

(a) US-Präsident Obama wird die Kanzlerin mit einer Auszeichnung *ehren*. (*Engl.:* The US-president Obama will honor the chancellor an award.)
(b) US-Präsident Barack Obama *überreicht* der Kanzlerin die "Medal of Freedom". (*Engl.:* US-president Barack Obama hands over the "Medal of Freedom" to the chancellor.)

Our annotation schema is inspired by MPQA and its major frame is described in the Table 4.

With the help of the annotated corpus, we are able to learn opinion extraction rules, e.g.:

Example 3.

(a) verb (kritisieren, negative): subject(source), object(target)
(*Engl.*: verb (criticize, negative): subject(source), object(target))
(b) verb (angreifen, negative): subject(source), object(target)
(*Engl.*: verb (attack, negative): subject(source), object(target))

In comparison to English texts, we are faced with the challenges of German particle verbs and complex noun compounds. Furthermore, idioms and collocations are very popular with German newspapers. We utilize the linguistic annotation delivered by the ALEXANDRIA IE pipeline. The annotation result is a valuable resource for the opinion mining research for the German language, in particular, the German political news. The distinction between context-dependent and context-independent frames is important for the estimation of the need of world and domain knowledge for a running system.

3 Conclusion

DARE is utilized in the THESEUS use case ALEXANDRIA for the information extraction task, for both automatic rule learning and extraction of facts and opinions from German news texts. The realistic tasks and data of ALEXANDRIA provided a challenge to the IE framework, which is representative of many future applications for the exploitation and transformation of unstructured web content. The experiments with various social domains give us very interesting insights, and let us better understand the challenges and opportunities behind the typical coverage of domain information in the news. The application of the DARE framework to opinion detection yielded truly promising results. The experiments with various parsers show that robust semi-deep parsers, such as the widely used Stanford Parser, are still the analysis instrument of choice if the main performance criterion for the relation extraction is recall or f-measure. Only in cases in which precision needs to be guaranteed can a deeper and less robust parser working with intellectually created grammars outperform the shallower data-driven analysis. But, the performance of deep parsers for relation extraction can be considerably improved through domain-guided re-ranking, another form of domain adaption. A side result of this work was the insight that a better parse ranking for the purpose of relation extraction does not necessarily correspond to a better parse ranking for other purposes or for generic parsing. This should not be surprising, since relation extraction in contrast to text understanding does not need the entire and correct syntactic structure for the detection of relation instances. The ease and consistency of rule extraction and rule application counts more than the linguistically correct analysis. With the evolution of parsing technology, the contribution of deep parsers may further improve.

The work on adapting the relation extraction system to the domains and tasks of ALEXANDRIA has also served as a source of ideas for future research. The insights gained by the investigation of the web-based data suggested training the system directly on the Web, without any bootstrapping in cases, where large amounts of seed data can be obtained. Indeed, for several ALEXANDRIA-relevant domains, including the domain of celebrities and other people, sufficient resources of massive seeds could be identified. After the successful conclusion of the ALEXANDRIA work, subsequent research on direct training of relation extraction by finding the patterns for rules in web content has already been started (Krause et al. 2012).

Acknowledgements The extensions of the DARE framework have been conducted through cooperations between the THESEUS ALEXANDRIA use case and the following other research projects: the German DFG Cluster of Excellence on Multimodal Computing and Interaction (M2CI) and the project TAKE (funded by the German Federal Ministry of Education and Research, contract 01IW08003). Many thanks go to Yi Zhang for his joint work on the parse re-ranking approach.

References

P. Adolphs, F. Xu, H. Li, H. Uszkoreit, Dependency graphs as a generic interface between parsers and relation extraction rule learning, in *KI 2011: Advances in Artificial Intelligence*. Volume 7006 of Lecture Notes in Computer Science (Springer, Berlin /Heidelberg/New York, 2011), pp. 50–62

E. Agichtein, L. Gravano, Snowball: extracting relations from large plain-text collections, in *Proceedings of the 5th ACM International Conference on Digital Libraries (DL'00)* (ACM, New York, 2000), pp. 85–94, citeseer.ist.psu.edu/article/agichtein00snowball.html

D.E. Appelt, D.J. Israel, Introduction to information extraction technology, in *A Tutorial Prepared for International Joint Conference on Artificial Intelligence (IJCAI '99)*, Stockholm, Aug 1999, http://www.ai.sri.com/~appelt/ie-tutorial/IJCAI99.pdf

S. Brin, Extracting patterns and relations from the world wide web, in *The World Wide Web and Databases* (Springer, Berlin/Heidelberg/New York, 1999), pp. 172–183

R. Grishman, B. Sundheim, Message understanding conference – 6: a brief history, in *Proceedings of the 16th Conference on Computational Linguistics (COLING '96)*, vol. 1 (Association for Computational Linguistics, Stroudsburg, 1996), pp. 466–471, http://acl.ldc.upenn.edu/C/C96/C96-1079.pdf

S. Krause, H. Li, H. Uszkoreit, F. Xu, Large-scale learning of relation-extraction rules with distant supervision from the web, in *Proceedings of the 11th International Semantic Web Conference (ISWC '12)* (Springer, Berlin/Heidelberg/New York, 2012), pp. 263–278

H. Li, X. Cheng, K. Adson, T. Kirshboim, F. Xu, Annotating opinions in German political news, in *8th ELRA Conference on Language Resources and Evaluation (LREC)* (European Language Resources Association (ELRA), Istanbul, May 2012)

H. Li, F. Xu, H. Uszkoreit, Minimally supervised rule learning for the extraction of biographic information from various social domains, in *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011* (RANLP 2011 Organising Committee, Hissar, Sept 2011), pp. 17–24, URL http://aclweb.org/anthology/R11-1003

J.M. Schulz, C. Womser-Hacker, T. Mandl, Multilingual corpus development for opinion mining, in *Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC '10)*, Valletta, ed. by N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias (European Language Resources Association, May 2010), http://dblp.uni-trier.de/db/conf/lrec/lrec2010.html#SchulzWM10

J. Wiebe, T. Wilson, C. Cardie, Annotating expressions of opinions and emotions in language. Lang. Resour. Eval. **39**(2–3), 165–210 (2005)

T. Wilson, *Fine-Grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of Private States*, PhD thesis, University of Pittsburgh, Pittsburgh, 2007

T. Wilson, J. Wiebe, Annotating attributions and private states, in *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky* (Association for Computational Linguistics, Ann Arbor, June 2005), pp. 53–60, http://www.aclweb.org/anthology/W/W05/W05-0308

F. Xu, *Bootstrapping Relation Extraction from Semantic Seeds*, PhD thesis, Saarland University, Saarbrücken, 2008

F. Xu, H. Li, Y. Zhang, H. Uszkoreit, S. Krause, Minimally supervised domain-adaptive parse reranking for relation extraction, in *Proceedings of International Conference on Parsing Technologies (IWPT '11)* (The Association for Computational Linguistics, Dublin, Oct 2011), pp. 118–128, http://dblp.uni-trier.de/db/conf/iwpt/iwpt2011.html#XuLZUK11

F. Xu, H. Uszkoreit, H. Li, A seed-driven bottom-up machine learning framework for extracting relations of various complexity, in *Proceedings of ACL '07*, ed. by J.A. Carroll, A. van den Bosch, A. Zaenen (The Association for Computational Linguistics, Prague, June 2007), pp. 584–591, http://dblp.uni-trier.de/db/conf/acl/acl2007.html#XuUL07

R. Yangarber, *Scenario Customization for Information Extraction*, PhD thesis, New York University, New York, Jan 2001

Ask Like an Egyptian: Question Answering in the ALEXANDRIA Use Case

Matthias Wendt, Martin Gerlach, and Holger Düwiger

Abstract The Semantic Web came with the prospect of once providing amounts of information just as immense as those now available from the Internet, ready for evaluation and analysis by machines. About a decade later, more and more data hubs (a data hub is comparable to a web site) emerged that provide information free of charge. Massive amounts of such information, also called Linked Open Data (LOD), make the vision of the Semantic Web come to life. As an example, DBpedia – by harvesting information from Wikipedia – already contains hundreds of millions of general knowledge facts. Such data can be used to conveniently make information of general interest available to the public. Above this, the structure of this information and the fact that it is present in machine readable form renders possibly more structured ways of information access. One of these technologies is Question Answering (QA) – a task that always hinged on the availability of massive amounts of information. This paper reports on our approach to implementing a QA system backed by Linked Open Data. The QA system is part of the ALEXANDRIA use case.

1 Introduction

Berners-Lee (2000) first announced his vision of the "Semantic Web" back in 1999 in analogy to the World Wide web as a web that is not only readable by humans, but "a web of data that can be processed directly and indirectly by machines". In recent years, the Semantic Web has evolved from a mere idea into a growing environment of LOD[1] sources and applications. This is in particular due to the

[1] http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

M. Wendt (✉) • M. Gerlach • H. Düwiger
Neofonie GmbH, Berlin, Germany
e-mail: wendt@neofonie.de; gerlach@neofonie.de; duewiger@neofonie.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_23,

Fig. 1 The ALEXANDRIA UI with question, first two answers and disambiguation cards for Barack Obama's place of birth. Hawaii and the USA were omitted

following current trends: The first is automatic data harvesting from unstructured or semi-structured knowledge that is freely available on the Internet, most notably the DBpedia project (Bizer et al. 2009). The second notable trend is the evolution of linked data sources that enable users to collaboratively edit the database in a similar way, as collaborative editing is allowed in a Wiki. A prominent example of such data sources is Freebase.[2] The growth of LOD gives rise to a growing demand for means of semantic data exploration. Question Answering, being the natural device of querying things and acquiring knowledge, is a straightforward way for end users to access semantic data.

In this paper, we present the QA-driven ontology design behind ALEXANDRIA,[3] a platform for exploring data of public interest comprising a system for answering questions in German. ALEXANDRIA is backed by an ontology, built from and continuously being updated with data primarily from Freebase, as well as news feeds, user-generated content and a small subset of DBpedia. Figure 1 shows the ALEXANDRIA user interface with an example question for Barack Obama's place of birth.

The system assists the user while he is entering the question by providing possible matches for persons, places, etc., then lists all matching answers that can be found in the backing knowledge base. The ontology consists of facts about persons, organizations, locations as well as works, such as books, music albums, films, and paintings. Moreover, the ALEXANDRIA ontology is designed for holding information on events relating the various resources, including temporal information and relations involving more than two participants – so called n-ary relations.

---

[2] http://www.freebase.com/

[3] http://alexandria.neofonie.de/

We will discuss what led us to the major design choices underlying our ontology – including the modeling of n-ary relations – more in depth in Sect. 3. Then we will informally describe the mapping algorithm used in our QA system in Sect. 4, and how it benefits from the ontology design. In Sect. 5 we will present an evaluation of the ALEXANDRIA Question Answering system with a bit of discussion on the results.

2 Related Work

Open domain question answering is of current research interest. There are several approaches to the subject based on linguistic analysis of natural language questions for generating queries against linked data. FREyA (Damljanovic et al. 2011) and PowerAqua (Lopez et al. 2011, 2006) are both question answering systems that are to a certain degree independent of the underlying ontology schema. Both systems work on existing Linked Open Data *as is* and can be configured to use multiple ontologies. They rely on rather shallow approaches to query mapping, in favor of portability and schema-independence. However, this also limits them to the data structures and languages used by the schemas (e.g., DBpedia does not support n-ary relations). There are also systems based on deeper, compositional mapping approaches. For example, ORAKEL (Cimiano 2004; Cimiano et al. 2007) translates syntax tree constructed by lexicalized tree adjoining grammars (LTAGs) to a representation in first-order logic which can be converted to F-Logic (Kifer et al. 1995) or SPARQL[4] queries, depending on the target knowledge base. ORAKEL also principally supports n-ary relations. Though the system is in principle very similar to the one presented in this paper, it is not proven to scale up to a large dataset. In contrast to other projects that use Linked Open Data for question answering, our approach is an attempt to combine the advantages of availability of huge LOD sources and of tailoring the T-Box to the use case of QA. While the latter facilitates the fully automated mapping of natural language questions to SPARQL queries, we trade off the possibility of useing existing labels for T-Box entities, which, combined with existing lexical resources such as WordNet,[5] GermaNet,[6] etc., boost lexical coverage. Another difference to the above-mentioned projects is that the focus of ALEXANDRIA is on answering questions in German, not English.

[4] http://www.w3.org/TR/rdf-sparql-query/

[5] http://wordnet.princeton.edu/

[6] http://www.sfs.uni-tuebingen.de/lsd/

3 Design of the ALEXANDRIA Ontology

The design of the ALEXANDRIA ontology was basically driven by practical demands of the application as well as linguistic considerations. Our approach can be seen as a unification of the "Type 4" (inference-based concept generalizations) and the "Type 3" (cross-linguistic phenomena) approaches to ontology creation presented by Hovy (2005). The knowledge base has to meet the following requirements:

- Linguistic Suitability: The data model needs to be suitable for natural language question answering, i.e. mapping natural language parse tree structures onto our data must be possible.
- LOD Compatibility: Compatibility with existing LOD sources like Freebase and DBPedia needs to be maintained in order to facilitate mass data import for practical use.
- Scalability: Large amounts of data need to be stored, maintained and updated, while keeping the time for answering a question at minimum.

One of the major aspects relating to *linguistic suitability* in ALEXANDRIA is that its target domain goes beyond what we refer to in the following as *attributive data*, i.e., data about things that are commonly known as named entities, like persons, organizations, places, etc. In addition, the domain was designed to contain what we call *eventive data*, i.e., (historic) events and relations to participants within them.

The WWW consortium (W3C) underpinned the vision of the Semantic Web with a common standard, the Resource Description Framework (RDF).[7] RDF is a standard for modeling and describing linked data using the ubiquitary notion of subject predicate object (SPO) triples, which resembles a basic grammar theory. Many RDF sources offer information in the form "*X* birth-place *Y*" and "*X* birth-date *Z*", where clearly *X* and *Y* constitute subject and object and "birth-date" and "birth-place" constitute the predicate.[8]

Natural language allows us to formulate sentences that convey more complex information. In particular, in the above-mentioned examples, facts are modeled as binary relations; however, we may arguably construct sentences with more than just a subject and an object involved. For example the sentence "Salvador Dali was born in Figueres on 11th of July 1904" exhibits the ternary pattern "*X* was born in *Y* on *Z*". This does not matter as long as singular events like birth or death are involved, since these cases can safely be handled by joining on a participant (the person) as proposed in Cimiano et al. (2007).

[7] http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/

[8] An example of this pattern is DBpedia. It is fair to say that this structure arises from the infoboxes being harvested, not a general misconception, though.

The distinction between eventive and attributive data becomes important, when relations are involved which (may) have multiple distinct instances. E.g., knowledge about people's education can only be modeled ambiguously using pure binary facts:

```
res:Angela_Merkel dom:major_in res:Physics .
res:Angela_Merkel dom:major_in res:Quantum_Chemistry.
res:Angela_Merkel
      dom:major_at res:German_Academy_of_Sciences .
res:Angela_Merkel
      dom:major_at res:Leipzig_University .
```

In this example, each single statement is correct, but a question like "What subject did Angela Merkel major in at the German Academy of Sciences?" can no longer be answered correctly by joining the binary facts, since we cannot connect the place of study with the scientific field. Similarly, relations including a time span require a different representation in order to answer questions like "Who was married to Jennifer Garner in 2006?" unambiguously.

It is possible to model such eventive n-ary facts as proposed in Pattern 1, use case 3 of the W3C Working Group Note on n-ary relations on the Semantic Web.[9] This approach is also close to the semantic model advocated in Neo-Davidsonian theories (Parsons 1990), where participants in an event are connected to the event using roles.

As for the aspect of *LOD compatibility*, it is our aim to access existing large-scale sources to populate our knowledge base. DBpedia was the first LOD source to retrieve and constantly update its data repository by crawling Wikipedia.[10] Apart from its possibilities for end users to add and update information, the majority of data contained in Freebase is obtained from Wikipedia as well. Therefore, using one (or both) of these sources is an obvious starting point for harvesting information on a broad range of popular entities, as is required by ALEXANDRIA. However, though DBpedia contains much valuable attributive data for entities of our interest, it does not offer eventive information as stated above. Also, DBpedia's T-Box does not provide a model for adding such n-ary facts, either. As opposed to DBpedia, which relies on the RDF standard, Freebase implements a proprietary format. Whereas in RDF, all information is abstractly represented by triples, Freebase abstractly represents information as links between topics. The Freebase data model incorporates n-ary relations by means of Compound Value Types,[11] also called "Mediators". A mediator links multiple topics and literals to express a single fact.

So Freebase's data model suits our requirements, but we need to use RDF to be able to use Virtuoso Open Source Edition,[12] which has proven to *scale well* for both loading and querying the amounts of data we expected.

[9] http://www.w3.org/TR/swbp-n-aryRelations/

[10] http://www.wikipedia.org/

[11] http://wiki.freebase.com/wiki/Compound_Value_Type

[12] http://virtuoso.openlinksw.com/dataspace/dav/wiki/Main/

Using the Freebase query API to pull a set of topics and links, an RDF-based knowledge base can be built according to the Neo-Davidsonian model. The API also supports querying link updates for continuously updating the knowledge base.

There are straightforward mappings of Freebase topic and mediator types onto OWL (the Web Ontology Language building on RDF)[13] classes, and of Freebase link types onto OWL properties. For example, a marriage relation is imported from Freebase and converted to RDF represented in Notation 3 (N3)[14] as follows:

```
nary:m_02t82g4
    rdf:type        dom:Marriage ;
    dom:spouse      res:Jennifer_Garner,
                    res:Ben_Affleck ;
    dom:weddingDate "2005"^^xsd:gYear .
```

The subject URI is generated from the Freebase mediator ID. The resource URIs are generated from Freebase topic names with some extra processing and stored permanently for each Freebase topic ID.

We differentiate between the following three layers of our ontology (which correspond to the three namespace prefixes alx:, dom:, and res: that appear in the examples):

- The Upper Model (namespace “alx:”) contains the abstract linguistic classes needed for a language-, domain- and task-independent organization of knowledge. The ALEXANDRIA upper model is inspired by Elhadad and Robin (1998).
- The Domain Model (namespace “dom:”) contains the concrete classes and properties for entities, events and relations of the modeled domain (e.g., Marriage, Study) as subclasses of the upper model classes and properties. It is needed to make the domain-specific distinctions which are necessary for the task of question answering.
- The A-Box consists of all “resources”, i.e. entity, event and relation instances, known to ALEXANDRIA. Entity URIs are in the namespace “res:”, n-ary event/relation URIs are in the namespace “nary:”.

Angela Merkel’s education and Jennifer Garner and Ben Affleck’s marriage would be represented in the A-box as shown in Table 1.

Syntactically, we model our domain concepts as OWL subclasses of one or more upper model concepts and our domain properties as OWL subproperties of one or more upper model properties, where the latter correspond to the Neo-Davidsonian roles mentioned earlier. The A-box facts shown in Table 1 would be represented by RDF triples as follows in N3, after deducting the upper model facts:

[13] http://www.w3.org/TR/owl2-overview/

[14] http://www.w3.org/TeamSubmission/n3/

Table 1 Upper and domain model

Upper model concept	Domain concept	U.m. role props.	Domain props.	Participant
Agentive process	Study	Agent	Student	Angela Merkel
Effective process	Study	Affected	Subject	Quantum Chemistry
Locative relation	Study	Location	Institution	German Academy of Sciences
		Located	Student	Angela Merkel
Attributive rel.	Marriage[a]	Carrier	Spouse	Ben Affleck
		Attribute	Spouse	Jennifer Garner
Temporal concept	Marriage	Start	Wedding date	2005

[a] In this example, marriage is modeled as a symmetric relation, expressing a spouse as attribute of the other spouse, i.e. carrier and attribute may be swapped

```
nary:some_id1
    rdf:type          dom:Study,
                      alx:AgentiveProcess,
                      alx:EffectiveProcess,
                      alx:LocativeRelation ;
    dom:student       res:Angela_Merkel ;
    alx:hasAgent      res:Angela_Merkel ;
    alx:hasLocated    res:Angela_Merkel ;
    dom:subject       res:Quantum_Chemistry ;
    alx:hasAffected   res:Quantum_Chemistry ;
    dom:institution   res:German_Academy_of_Sciences ;
    alx:hasLocation   res:German_Academy_of_Sciences .
nary:some_id2
    rdf:type          dom:Marriage,
                      alx:AttributiveRelation,
                      alx:TemporalConcept ;
    dom:spouse        res:Ben_Affleck,
                      res:Jennifer_Garner ;
    alx:hasCarrier    res:Ben_Affleck,
                      res:Jennifer_Garner ;
    alx:hasAttribute  res:Jennifer_Garner,
                      res:Ben_Affleck ;
    dom:weddingDate   "2005"^^xsd:gYear ;
    alx:hasStart      "2005"^^xsd:gYear .
```

Not all these triples need to be explicitly stored (materialized). Facts involving upper model concepts and properties can always be derived from domain facts and the T-box (which states, e.g., that `dom:Study` is a subclass of `alx:AgentiveProcess`, which is the (RDF) domain for property `alx:hasAgent`, which is again a super property of `dom:student`.). In ALEXANDRIA we store upper and domain `rdf:type` statements, but only upper model role facts, thus saving some space, but still saving a lot of time by avoiding too much ad hoc reasoning while executing SPARQL queries during the question answering process. We can then obtain hints to the upper model concepts and roles of interest by mapping question verbs onto domain concepts and then try to match the roles defined for the upper model classes to the respective given parts of the question.

4 Putting the Model in Action: Question Answering

As mentioned above, one of the major design goals of our ontology schema was to stay reasonably close to the phenomena and structures of natural language. Achieving this would facilitate the mapping of a natural language question to a SPARQL graph pattern that conveys the information need expressed in the question. The basic idea of the translation algorithm is to understand the problem of mapping of natural language to SPARQL as a graph mapping problem. To explain this idea, let us first give an idea of the graphs involved in the mapping.

According to a line of linguistic theory, the syntactic structure of a sentence may be represented as a graph – the dependency graph. For instance, Fig. 2 is the syntactic representation of the sentence “Mit wem ist Jennifer Garner seit 2005 verheirated?” (“Who has Jennifer Garner been married to since 2005?”). The dependency graph is constituted by a set of edges depicted as arrows in this figure. The arrow between the words “mit” and “wem”, for instance, denote that there is a certain kind of relationship between these two words: “wem” depends on “mit” – in this direction. For the sake of simplicity, we omitted the arrow between “Jennifer” and “Garner” and the one between “seit” and “2005”. The purpose of creating such a graph is basically to abstract from some linguistic variability, especially word order. To make this point clear, look at the graph on the left in Fig. 3. This graph is equivalent to the graph in Fig. 2 but it is arranged in a way that visually underlines an important property of the graph: namely, that it is a tree (each word except for the root depends on exactly one other word). This dependency graph is exactly the same as the one corresponding to the sentence “Seit 1999 ist Jennifer Garner mit wem verheiratet?”. To build a dependency graph from a string of words, we utilize a dependency parser trained on a large set of such dependency graphs. On the right-hand side in Fig. 3 the SPARQL Query graph corresponding to the sentence is given as a graph. The SPARQL Query expressed in this graph is given by:

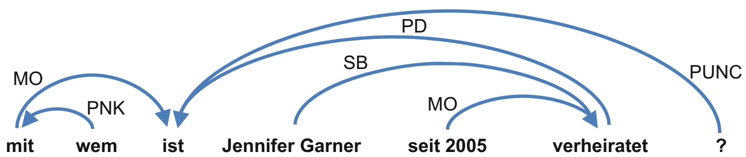


Fig. 2 Dependency parse of the sentence “Mit wem ist Jennifer Garner seit 2005 verheiratet?”

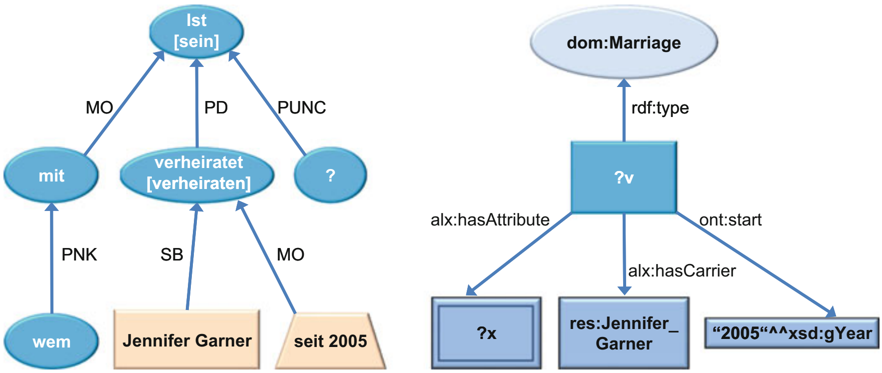


Fig. 3 *Left:* Dependency parse of the sentence "Mit wem ist Jennifer Garner seit 2005 verheiratet?" arranged in tree shape. *Right:* Graph representing the SPARQL Query corresponding to the sentence

Example 1.

```
SELECT ?x
WHERE {
    ?v a~dom:Marriage,
       alx:hasCarrier res:Jennifer_Garner,
       alx:hasAttribute ?x.
}
```

Note that the variable `?x` occurring in the `SELECT` clause is depicted in a double frame in the graph. The translation is a function of dependency graphs in the domain of SPARQL Query graphs.

Formal Notions A dependency graph is formalized like this:

Given a set L of dependency labels, a *dependency graph* for the sentence $x = w_1 \dots w_n$ is a directed graph $D = (V_D, E_D)$ with:

1. V_D a set of vertices $\{w_0, w_1, \dots, w_n\}$, and
2. $E_D \subseteq V_D \times L \times V_D$ a set of labeled edges.

The vertices are composed of the tokens in the sentence plus an artificial root node w_0. A well-formed dependency graph is a tree rooted at w_0.

Likewise, the structure of a SPARQL Select query basically consists of a graph pattern (in the "where" clause) and a projection. Given a set of variable names N_V (`?x`, `?y` ...), the set of concept names N_C, a set of role names N_P, a set of resource names N_R and a set of literals N_L defined by the ontology, we define a SPARQL Select Graph $G = (V_G, E_G, P_G)$ as:

1. $V_G \subseteq N_V \cup N_R \cup N_C \cup N_L$,
2. $E_G \subseteq N_V \cup N_R \times N_P \times V_G$,
3. The projection $P_G \subseteq N_V$.

Formally, we define the translation as a mapping $f(D)$ of a dependency graph D to a SPARQL Select Graph G.

Linguistic Processing The dependency graph is the result of the application of a *linguistic analysis* to the input sentence. An example of a resulting dependency structure may be found on the left in Fig. 3. The analysis consists in tokenization, POS-tagging[15] and dependency parsing. Dependency parsing is conducted using the MaltParser (Nivre et al. 2007), which was trained on the German Tiger corpus[16] (Brants et al. 2002). The corpus has been slightly adapted by adding a small sub-corpus of German questions and making a minor change to the set of role labels used.

To normalize surface form variation and identify morphosyntactic features, lemmatization and morphological analysis is applied to each of the tokens. This is roughly illustrated by the lemmata in square brackets at the verbal nodes (e.g. "verheiratet" has the lemma "verheiraten").

Compositional Semantics The mapping of the dependency graph to the SPARQL query is largely done in two steps: *lexicalization* and *composition*. By *lexicalization* we refer to the process of mapping tokens (lemmata) or multi-word units to corresponding ontological units.

We refer to the identification of resources (identified by resource URIs) of the A-Box as *lexical named entity identification*. For this, we make use of the title (the name of an entity) and the alternative names (consisting of synonyms and different surface forms) that are imported from Freebase into a Lucene[17] index containing the resource URI in ALEXANDRIA (e.g. `res:Jennifer_Garner`), and the OWL classes they belong to. While the user enters a question, matching entities are looked up in the index based on whole words already entered, and a disambiguation choice is continuously updated. The user can select from the found entities at any time, whereupon the respective part of the question is updated.

The second noteworthy component in lexical named entity identification is the identification of dates (and time). For these, we have adapted the open source date parser provided by the Yago project[18] to German.

All other linguistic tokens or configurations (linguistic units) corresponding to T-Box concepts are mapped using handcrafted lexica.

The complete set of mappings for the question shown in Fig. 2 is shown in Table 2. Note that the table contains only the correct entries, though there may be more than one entry in the case of ambiguities.

[15] For German tokenization and POS-Tagging we use OpenNLP with some pre-trained models. (http://incubator.apache.org/opennlp/)

[16] http://www.ims.uni-stuttgart.de/projekte/TIGER/

[17] http://lucene.apache.org/

[18] http://www.mpi-inf.mpg.de/yago-naga/javatools/

Table 2 Types of lexical mappings

T-Box Class	Custom Lexica	"wer"	`dom:Person`
		"verheiraten"	`dom:Marriage`
		"sein"	`owl:Thing`
T-Box Role	Custom Lexica	"mit"	`alx:hasAttribute`
A-Box URI	Lucene	"Jennifer Garner"	`res:Jennifer_Garner`
Literal	Date/Literal Parser	"2005"	`"2005"^^xsd:date`

Our syntax-semantics mapping is largely done by the composition of the lexical semantic entries attached to each dependency node. However, the entries seen in Table 2 still do not convey enough information about their contribution to the SPARQL query. Especially, how do we have to compose our entries to finally arrive at the SPARQL query in Example 1? How do we know that `res:Jennifer_Garner` will be object to the property `alx:hasCarrier`?

The solution is to map semantic entries into pieces of SPARQL queries, similar to templates. These templates are represented by an intermediate layer built around the notion of a *semantic description*. A semantic description represents the semantic contribution of a dependency node or (partial) dependency tree and encodes obligatory semantic complements (slots). To give some examples in an informal notation, some of the lexical entries of the example are displayed in Fig. 4. The boxes consist of the entity or variable denoted (first column), the type (class) of the entity or variable in question and (optionally) the set of properties. The prefix `?!` in a variable designation (e.g. `?!x` in the middle) means that the variable will be part of the `SELECT` clause. Initially, none of the properties has an object (we also speak of a role and a filler). A role that does not yet have a filler is called a slot and is being designated with the argument being just a `?`.

During the composition, the slots are being filled by semantic descriptions (properties), until the semantic description is satisfied, i.e., all slots have been filled.

By virtue of the lexical mapping, each linguistic unit is mapped to a set of semantic descriptions, also called *readings*.

Putting It All together The composition algorithm devices a fixed set of two-place composition operators, called actions. An action defines the mapping of two semantic descriptions related to an edge in the dependency graph to a composed semantic description, corresponding to the semantics of the subgraph of the dependency tree.

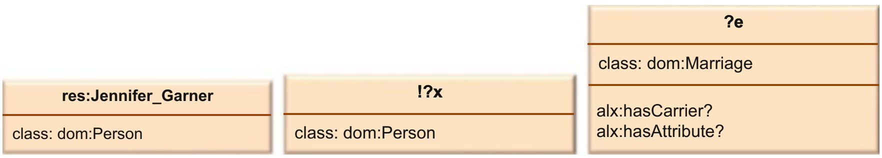


Fig. 4 From *left* to *right*: Semantic description of "Jennifer Garner", "wem" and "verheiratet"

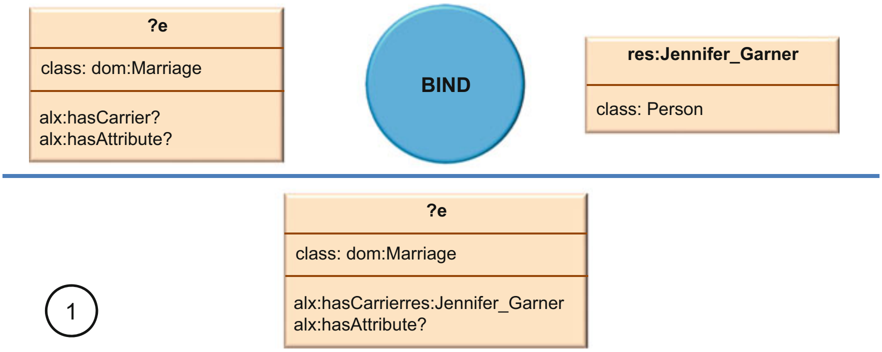


Fig. 5 Application of the operator BIND to the semantic descriptions of “verheiratet” and “Jennifer Garner”

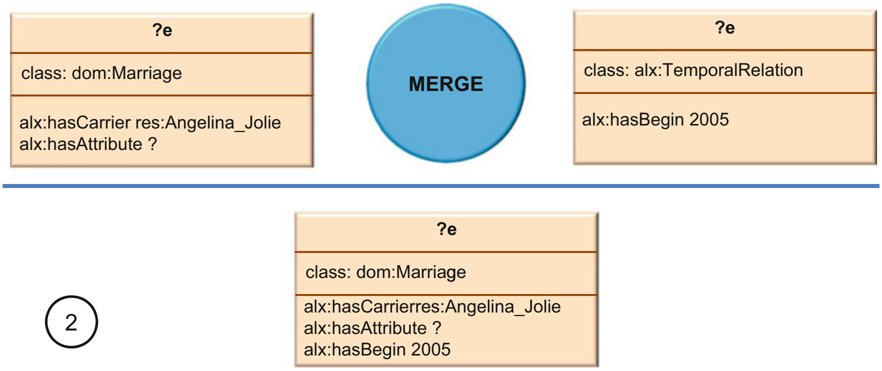


Fig. 6 Application of the operator MERGE to the semantic descriptions of “seit 2005” and the resulting description of Fig. 5

The two most important actions are BIND and MERGE. These two basic operations on the semantic descriptions involved in the composition intuitively correspond to (1) the mapping of the syntactic roles to semantic roles (otherwise called semantic role labeling), and (2) the aggregation of two nodes to one in the output graph pattern.

The application of the operator BIND to the semantic descriptions of “verheiratet” and “Jennifer Garner” is depicted in Fig. 5. The result is the semantic description of “verheiratet” with the `alx:hasCarrier` slot filled by the semantic description of “Jennifer Garner”. The operator MERGE merges the set of properties of two semantic descriptions, choosing the more specific semantic class of both. An example is given in Fig. 6, where the semantic description of “seit 2005” is merged with the semantic description derived in the previous operation.

For the composition, each of the labels in the label alphabet is assigned one of the semantic operators. For each of the edges in the dependency graph, the algorithm

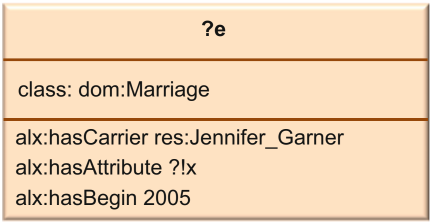


Fig. 7 Semantic description of “Mit wem war Jennifer Garner seit 2005 verheiratet?”

chooses the operator defined here to recursively compose a semantic description representing the meaning of the whole sentence. The following table shows an excerpt of this mapping:

SB	OC	PNK	MO	PD	PUNC
BIND	BIND	BIND	MERGE	MERGE	IGNORE

The action IGNORE simply skips the interpretation of the subtree. The interested reader may have a look at Wendt et al. (2012) for a detailed description of the algorithm. For brevity, it may suffice to say that when the composition algorithm is applied in a top-down manner to the dependency graph in Fig. 2 with all nodes lexicalized, the final semantic description of our running example will be the one shown in Fig. 7. The translation from this semantic description to the corresponding SPARQL query given in Example 1 is straightforward.

5 Conclusion

The n-ary modeling requires more triples for simple (binary) facts than using RDF/-OWL properties like DBpedia, because there is always an instance of a relation concept comprising `rdf:type` and participant role triples.

At the time of this writing, the ALEXANDRIA ontology contained approximately 160 million triples representing more than 7 million entities and more than 13 million relations between them (including literal value facts like amounts, dates, dimensions, etc.). We imported the triples into Virtuoso Open Source Edition, which scales as well as expected with respect to our goals.

Eighty percent of all query types understood by the algorithm (i.e. mappable onto valid SPARQL queries) take less than 20 ms in average for single threaded linguistic processing on a 64-bit Linux system running on Intel XeonE5420 cores at 2.5 GHz, and pure in-memory SPARQL processing by Virtuoso Open Source Edition on a 64-bit Linux system running on eight Intel Xeon L5520 cores at 2.3 GHz and 32 GB of RAM.

The question answering system works fast enough to be used in a multi-user web front end like http://alexandria.neofonie.de.

Table 3 Quality of results of the ALEXANDRIA question answering on the QALD-2 training set translated to German

Answer set	Avg. precision	Avg. recall	Avg. f-measure
All answers	0.25	0.27	0.25
Generated answers	0.49	0.52	0.48
Answers without data mismatch	0.59	0.57	0.57
Gen. answers without data mismatch	0.92	0.89	0.89

The time efficiency of the algorithm is in part due to the high performance of the malt parser with a liblinear model, which runs in less than 5 ms per question. By using a liblinear model, however, we trade off parsing accuracy with performance in terms of processing time per question. This sometimes becomes noticeable in cases where subject-object order variation in German leads to an erroneous parse.

The performance of the question answering system has been measured using the training set of the QALD-2 challenge.[19] As the question answering in ALEXANDRIA currently covers only German, all 100 questions were translated to German first. The results are shown in Table 3. Fifty-one of the questions had a corresponding representation in terms of the ALEXANDRIA model and a SPARQL query could be generated. The second row shows the results for these questions.

As the ALEXANDRIA knowledge base relies on data imported from Freebase to a great extent, the comparability of the results with the gold standard coming from DBpedia is limited. The comparison of both datasets results in various mismatches. For example, the comparison of questions having a set of resources as answer type is done via the indirection of using the labels. This is possible just because the labels are extracted from Wikipedia by both Freebase and DBpedia. However, some of the labels have been changed during the mapping.

Overall, we have identified the following error types:

1. Different labels for the same entities,
2. Different number of results for aggregate questions,
3. Query correct but different results,
4. Training data specifies “out of scope” where we can provide results,
5. Question out of scope for ALEXANDRIA.

Type 1 applies particularly often to the labels of movies, most of which are of the form “Minority Report (film)” in DBpedia, and “Minority Report” in ALEXANDRIA. Another source for errors (2) results when aggregate questions (involving a count) retrieve a different number of resources. The question “How many films did Hal Roach produce?” for example yields 509 results in DBpedia and 503 results in ALEXANDRIA. The third type corresponds to a difference in the dataset itself, that is, when different information is stored. For example, in

[19] http://greententacle.techfak.uni-bielefeld.de/~cunger/qald/index.php?x=challenge&q=2

ALEXANDRIA the highest mountain is "Mount Everest" whereas in DBpedia it is the "Dotsero".

The last two error types involve questions that are out of scope (4 and 5). The data model used in ALEXANDRIA differs from the model in DBpedia as a result of the considerations explained above. On the other hand, ALEXANDRIA lacks information since we concentrate on a mapped subset of Freebase. According to the evaluation, the answer "out of scope" is correct if the question cannot be answered using DBpedia.

Out of the 82 questions containing (some) erroneous results 63 belong to one of the error classes mentioned above. The last two rows of the Table 3 show the results for all questions that do not belong to any of these error classes.

References

T. Berners-Lee, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web* (HarperBusiness, New York, 2000)

C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, S. Hellmann, DBpedia – a crystallization point for the web of data. Web Semant.: Sci. Serv. Agents World Wide Web **7**(3), 154–165 (2009)

S. Brants, S. Dipper, S. Hansen, W. Lezius, G. Smith, The TIGER treebank, in *Proceedings of the Workshop on Treebanks and Linguistic Theories*, Sozopol, Sept 2002, pp. 24–41, http://scholar.google.com/scholar.bib?q=info:JVf6AeWs7nEJ:scholar.google.com/&output=citation&hl=en&as_sdt=1,5&ct=citation&cd=0

P. Cimiano, ORAKEL: a natural language interface to an F-Logic knowledge base, in *Natural Language Processing and Information Systems*, ed. by F. Meziane, E. Métais. Volume 3136 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2004), pp. 401–406, http://dblp.uni-trier.de/db/conf/nldb/nldb2004.html#Cimiano04

P. Cimiano, P. Haase, J. Heizmann, M. Mantel, ORAKEL: a portable natural language interface to knowledge base. Technical report, University of Karlsruhe, Karlsruhe, Mar 2007

D. Damljanovic, M. Agatonovic, H. Cunningham, FREyA: an interactive way of querying linked data using natural language, in *The Semantic Web: ESWC 2011 Workshops*, Heraklion, ed. by G. Antoniou, R. García-Castro, D. Fensel. Volume 7117 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011), pp. 125–138, http://link.springer.com/chapter/10.1007%2F978-3-642-25953-1_11

M. Elhadad, J. Robin, SURGE: a comprehensive plug-in syntactic realization component for text generation. Technical report, Ben Gurion University, Be'er Sheva and Universidade Federal de Pernambuco, Recife, 1998

E.H. Hovy, Methodologies for the reliable construction of ontological knowledge, in *Conceptual Structures: Common Semantics for Sharing Knowledge*. Volume 3596 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2005), pp. 91–106

M. Kifer, G. Lausen, J. Wu, Logical foundations of object-oriented and frame-based languages. J. ACM **42**(4), 741–843 (1995)

V. Lopez, M. Fernández, E. Motta, N. Stieler, PowerAqua: supporting users in querying and exploring the semantic web content. Semant. Web J. **3**, 249–265 (2011). IOS Press, http://dx.doi.org/10.3233/SW-2011-0030

V. Lopez, E. Motta, V. Uren, PowerAqua: fishing the semantic web, in *The Semantic Web: Research and Applications*, ed. by J. Domingue, Y. Sure. Volume 4011 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2006), pp. 393–410

J. Nivre, J. Hall, J. Nilsson, A. Chanev, G. Eryigit, S. Kübler, S. Marinov, E. Marsi, MaltParser: a language-independent system for data-driven dependency parsing. Nat. Lang. Eng. **13**(2), 95–135 (2007), http://dblp.uni-trier.de/db/journals/nle/nle13.html#NivreHNCEKMM07

T. Parsons, *Events in the Semantics of English: A Study in Subatomic Semantics*. Current Studies in Linguistics (MIT, Cambridge, 1990)

M. Wendt, M. Gerlach, H. Düwiger, Linguistic modeling of linked open data for question answering, in *Proceedings of the Workshop on Interacting with Linked Data (ILD '12)*, Heraklion, May 2012 (Central Europe (CEUR), 2012), pp. 75–86

Print Processing in CONTENTUS: Restoration of Digitized Print Media

Iuliu Konya, Stefan Eickeler, Jan Nandzik, and Nicolas Flores-Herr

Abstract One of the main goals of the CONTENTUS use case was to manage and improve the technical quality of large digital multimedia collections in cultural heritage organizations. Generally, there are two causes for quality impairment of digitized multimedia items: errors during the digitization process and a poor condition of the analog original. While digitization errors may be corrected by re-digitization, any deterioration of analog materials can only be counteracted by digital restoration in post-processing after digitization. This article showcases a unique technique developed in CONTENTUS to restore digitized hectograph archive documents that typically display yellowed paper and faded printing ink. The documents used in this restoration showcase belong to the archive of the Music Information Center the Association of Composers and Musicologists (MIZ) of the former German Democratic Republic (GDR), and were produced between 1960 and 1989. The hectography method was widely adopted in the GDR to copy documents at a large scale. The showcased restoration method enhances the readability of on-screen texts and, as shown by evaluation, lowers the error rate of optical character recognition. In turn, the latter improvement is expected to improve the automated extraction of semantic information entities like persons, places and organizations. The technology presented in this article is an example of how corpora consisting of visually impaired analog media can be prepared for semantic search applications based on automatic content indexing – another major goal of the use case CONTENTUS.

I. Konya (✉) • S. Eickeler
Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS), St. Augustin, Germany
e-mail: iuliu.konya@iais.fraunhofer.de; stefan.eickeler@iais.fraunhofer.de

J. Nandzik • N. Flores-Herr
Acosta Consult GmbH, Frankfurt am Main, Germany
e-mail: jn@acosta-consult.de; nf@acosta-consult.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_24,

Fig. 1 The processing pipeline for printed media within the CONTENTUS workflow

1 Introduction

The CONTENTUS (Flores-Herr et al. 2011a; Nandzik et al. 2013) use case focuses on the creation of a fully automated media processing workflow for cultural heritage organizations (Bossert et al. 2009; Flores-Herr et al. 2011b). The workflow should enable content holders to provide users with digital access to their multimedia collections. CONTENTUS techniques can be used to process and search within mixed collections of printed material (text documents, images), audio-visual content (film and video), and audio material (speech and musical recordings).

Given the enormous amount of analog content in cultural heritage organizations, an unattended processing from analog media ingest to semantic multimedia search strongly depends on automatic content analysis technologies. In case of printed media, widely used methods for information extraction are document structure analysis or Named Entity Recognition and Disambiguation (NERD), which utilize Optical Character Recognition (OCR) as a basis. However, one of the major challenges to achieving satisfactory results when applying information extraction techniques on digitized print media is a low recognition rate of OCR methods. While up-to-date OCR software produces very good results when high quality print media samples are used, the text recognition rate drops dramatically if the condition of the printed text or the paper of the original analog document is impaired, or the digitization quality is poor. While the latter can be circumvented by re-digitization, poor quality of the analog master documents has to be counteracted by digital restoration in order to increase the OCR recognition rate.

This article will give an example of how CONTENTUS restoration technologies can improve OCR results by an order of magnitude compared to traditional methods, thus enabling a meaningful semantic analysis of lower quality prints. The proposed technique treats the extraction problem of text and degradation effects as a classical blind-source separation problem. Independent Component Analysis (ICA) forms the core of our method, as it fits the showcased problem described in the article. The restoration step described in this article is part of the complete CONTENTUS processing pipeline for printed media, as shown in Fig. 1. Restoration is applied after quality assessment of the digitization step and is followed by semantic analysis steps like OCR, NERD and consequently our semantic multimedia search (SMMS).

Description of the MIZ Archive and its Importance for CONTENTUS Archives and cultural facilities nowadays contain a great variety of different analog document classes, many of which are obsolete by current standards. In order to establish a processing workflow with a proven practical relevance, the CONTENTUS use case needed to select a suitable test bed for technology development that contains both a unique and a semantically interconnected multimedia collection.

We chose the collection archived by the *Music Information Center of the Association of Composers and Musicologists* (MIZ) of the former GDR for our developments; it is owned by the *Deutsches Musikarchiv* (German Music Archive). These documents contain nearly all media types in focus of the CONTENTUS developments and are a vivid display of 40 years of the GDR music history. The German Music Archive took over the MIZ archive in 1991 and has since then continuously indexed and preserved the analog items of the collection. The German Music Archive in turn is part of the German National Library and its principal task is the central collection of sheet music and sound recordings.

A prominent document class in the context of printed archive material, such as the MIZ collection, is hectography, which was an inexpensive printing and duplication method, widely used throughout the nineteenth and twentieth centuries. The collection of MIZ documents that were produced using the hectography method consists of approximately 43,000 units, typically in DIN A4 paper with letters of roughly 12 point font size.

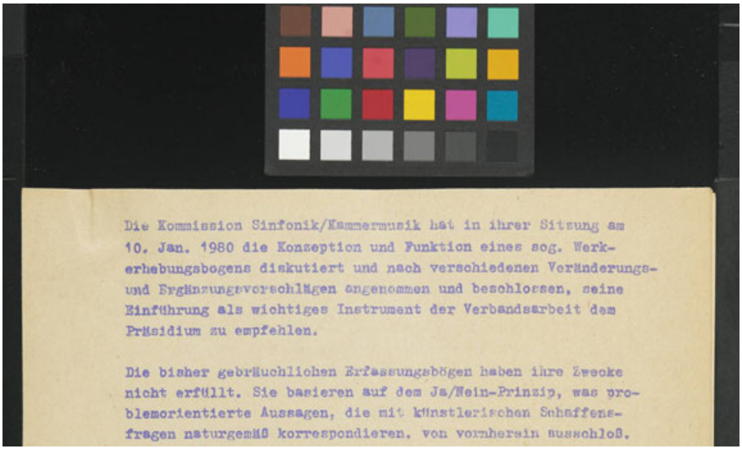
Next to hectography print media, the MIZ collection comprises 36,000 compositions in combination with program brochures, press releases, reviews and analysis and thus informs about the creativity of composers in the former GDR. A photographic archive includes pictures of composers and interpreters. This is complemented by more than 10,000 sound carriers and 1,000 music pieces in the form of sheet music by GDR musicians. In addition, a large set of various index cards, as well as a reference library of 2,700 volumes of books, completes the collection of the MIZ. This sums up to almost a dozen areas of documentation like contemporary music, cultural and educational policy, musicology, music critics, etcetera.

While the MIZ collection has already been the focus of a master thesis (Wallor 2012), its accessibility today is poor, as many media are threatened by decay. Since a noteworthy number of items – amongst others, fragile hectography prints on decayed paper – were threatened by media deterioration, the entire MIZ collection was digitized during the course of the project. As a side effect, the MIZ collection might appeal to many more historians in the future given its prospective digital searchability and availability.

The digitization of the multimedia collection of the MIZ was carried out by subcontractors converting various print media, including documents, brochures, books, sheet music, newspapers and index cards, photographs and audio content (speech and music) stored on vinyl records, shellac records, tapes as well as audio CDs.

2 Hectography: Problem Analysis and Pitfalls

Nowadays, hectography is considered to be an obsolete copying/duplication and printing method. This has not always been the case – in fact hectography was the most widespread document multiplication method in use throughout the



Die Kommission Sinfonik/Kammermusik hat in ihrer Sitzung am 10. Jan. 1980 die Konzeption und Funktion eines sog. Werkerhebungsbogens diskutiert und nach verschiedenen Veränderungs- und Ergänzungsvorschlägen angenommen und beschlossen, seine Einführung als wichtiges Instrument der Verbandsarbeit dem Präsidium zu empfehlen.

Die bisher gebräuchlichen Erfassungsbögen haben ihre Zwecke nicht erfüllt. Sie basieren auf dem Ja/Nein-Prinzip, was problemorientierte Aussagen, die mit künstlerischen Schaffensfragen naturgemäß korrespondieren, von vornherein ausschloß.

Fig. 2 Sample of one hectography scan with color reference target

nineteenth century, and until the third quarter of the twentieth century, while today, hectography documents are primarily found in archives. The process of hectography uses a mastercopy consisting of paper coated with a special mixture of gelatin. The mastercopy is written or typed with a special aniline ink. Duplicates of the mastercopy are made by means of a roller which presses the blank papers onto the gelatin. With each copy some ink is removed from the gelatin and consequently successive copies are progressively lighter. One drawback was, that even fresh duplicates obtained by hectography exhibited a low quality by current standards. Furthermore, because of the impregnation process, the paper used for hectography has the tendency to become yellow over time while the ink fades irregularly, causing a poor contrast with regard to both colors and brightness (see Fig. 2). Strong degradation effects and foldings in the used paper are common as well. These visual impairments not only affect information extraction algorithms, but also a human reader. Consequently, performing OCR on hectography duplicates without any preprocessing leads to very low recognition rates.

In general, it is not possible to solve the quality problems of hectography duplicates described above using conventional digital image enhancement methods, such as sharpening, noise reduction, contrast enhancement, etc. The use of thresholding techniques to remove unwanted background is often not effective, since the color intensities of the background can be very close to those of the foreground text. In these conditions, thresholding does either not remove the background, or eliminate essential information of relevant text (Tonazzini et al. 2004). Hence, there is a great demand for novel preprocessing methods to be developed.

However, the hectography prints of the MIZ archive are not the sole application scenario for a restoration technique: given the enormous quantity of historical texts archived by Europe's cultural heritage organizations and the growing number of

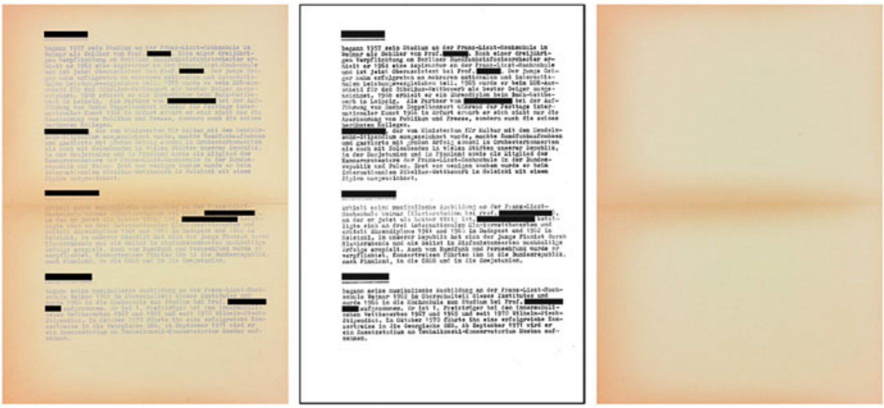


Fig. 3 Hectography separation result using the proposed method: *left* – original scan; *center* – extracted text component; *right* – extracted background and degradation component. Note: the background component was transformed back to RGB for visualization purposes and all person names have subsequently been blackened due to privacy reasons

large-scale digitization activities, there is an ever growing demand for technologies that improve the access to machine-readable text.[1]

As will be shown in this paper, the proposed preprocessing method for hectography prints improves optical character recognition results by an order of magnitude compared to traditional methods, thus forming a prerequisite for a meaningful semantic analysis of hectography prints formerly impossible.

For simplicity, in the current article we shall consider that the digitization process in itself has been perfect and did not introduce any unwanted artifacts, or all such artifacts have already been flawlessly removed. Even operating on this premise, a major problem that arises in the context of historical prints is the low quality of the documents, traceable to two main reasons:

- The outdated printing or duplication process itself, which produced only documents of low quality,
- Time-dependent degradation effects gradually decreasing the quality of the physical documents.

To summarize the observations so far: the hectography process involves using a single foreground color per document, the paper used is one having a special texture and, due to the materials employed in the copying process, the paper will tend to degrade significantly over time. Keeping these facts in mind, it seems natural to model a hectography duplicate as a weighted superposition of several independent components, namely text/drawing content, degradation effects, paper texture and noise (Fig. 3).

[1] http://www.impact-project.eu

3 Proposed Approach

The most obvious solution for the problem posed in this manner, would be the use of Principal Component Analysis (PCA) (Dunteman 1989; Jolliffe 2002). As one may recall, PCA was developed specifically for converting a mixture of correlated variables (in our case the hectography scan) into a set of uncorrelated variables (in our case the components/processes producing the image – i.e. text/drawing printing, paper degradation, paper texture and noise), dubbed principal components. As such, one may reasonably expect that the principal components will correspond to the aforementioned independent components. Indeed, our first unsuccessful attempts to solve the hectography separation problem were done using the PCA. In the following, we offer a short analysis of the causes of these failures. As it is well known, the first principal component contains the most variance of the original random variables. For most documents and hectography copies in particular, the size of the background clearly exceeds the text area, with the result that the variance of the background noise exceeds the variance of the text/line-art component by far. Therefore, the first principal component is inevitably determined by the background noise, and it is obvious that this cannot in any way lead to a cleanly separated text/line-art or paper background component. More specifically, this happens because all other principal components must be orthogonal to the first one (and to each other) (Drew and Bergner 2008; Ohta et al. 1980).

In case the aforementioned additive model holds (i.e. a hectography duplicate can be seen as a weighted superposition of several independent processes/components, namely text/drawing content, degradation effects, paper texture and noise), Independent Component Analysis (ICA) (Hyvaerinen 1999; Hyvaerinen et al. 2001) is a more powerful tool for component extraction. ICA has the great advantage of not being reliant on the orthogonality of the components and is therefore able to combine the information contained in the luminance component as well as in the color components. Since traditional image files contain three observed signals (RGB), the number of independent components to be extracted from them is limited to three as well (Hyvaerinen et al. 2001). This restriction however does not pose a major problem, since both paper-texture and degradation effects belong to related classes of phenomena, and their distinct extraction is usually not needed, i.e., it is sufficient to extract the sum of both components. In addition, the proposed approach does not consider the spatial information of the image and hence the RGB components are treated as mere random variables. The RGB components of each pixel in the image can therefore be seen as single observations.

Using the established ICA notation (Hyvaerinen et al. 2001), the problem can be stated as follows:

$$\begin{aligned} R &= \alpha_{11} \cdot Z_{\text{text}} + \alpha_{12} \cdot Z_{\text{degrad}} + \alpha_{13} \cdot Z_{\text{noise}}, \\ G &= \alpha_{21} \cdot Z_{\text{text}} + \alpha_{22} \cdot Z_{\text{degrad}} + \alpha_{23} \cdot Z_{\text{noise}}, \\ B &= \alpha_{31} \cdot Z_{\text{text}} + \alpha_{32} \cdot Z_{\text{degrad}} + \alpha_{33} \cdot Z_{\text{noise}}, \end{aligned} \tag{1}$$

where R, G, B are the observed random variables and Z_{text}, Z_{degrad}, Z_{noise} are the unknown generative components. The unknown mixing matrix is denoted by α_{ij}, $i, j \in \{1, 2, 3\}$. In conformance with the restrictions of ICA, we can assume that all components are non-Gaussian except for the noise-component Z_{noise}. Particularly, the independent component associated with the text-content Z_{text} is non-Gaussian, due to a high structure of the corresponding distribution functions. It has to be considered thoroughly how to deal with a possible noise problem. Preliminary results show, that the variance of the noise component Z_{noise} is negligible, such that this component can be omitted after whitening. This additional dimension reduction has a significant impact on the time performance of the ICA (Cichocki and Amari 2002; Hyvaerinen et al. 2001).

In the first processing step an *inverse gamma correction* is performed for all three color channels:

$$C = \begin{cases} C/12.92 & C \leq 0.04045 \\ ((C + 0.055)/1.055)^{2.4} & C > 0.04045 \end{cases}, \quad \forall C \in \{R, G, B\} .$$

This step is obligatory, since otherwise our model in (1) would not hold due to the performed gamma correction. From our experiments, however, we have observed that the actual impact of the inverse gamma correction on the final binary image is only marginal.

The next processing step comprises both the *whitening* of the random variables R, G, B, and the subsequent *dimensionality reduction*. Since we intend to reduce the dimensionality as well, we achieve this best by means of the classical PCA (Dunteman 1989; Jolliffe 2002). First, to simplify matters, it is common to subtract the means of all three random variables (Hyvaerinen et al. 2001; Jolliffe 2002). Henceforth, we will use the same names for the original and the mean-free variables.

As is well known from the theory of PCA (Dunteman 1989; Jolliffe 2002), in case the variance of one of the random variables R, G, B significantly exceeds the variances of the others the first principal component is almost completely determined by this variable. In order to avoid this phenomenon, we have to standardize the variance before applying PCA:

$$C = C/\sqrt{E\{C^2\}}, \forall C \in \{R, G, B\} ,$$

such that all random variables R, G, B have the same impact on the principal components.

Subsequently, the principal components P_1, P_2, P_3 of the random variables R, G, B must be obtained. To this end any known method can be applied. We have chosen the simple and straightforward method of eigenvalue decomposition of the corresponding covariance matrix of R, G, B. In hectography, the first two principal components P_1, P_2 declare almost always more than 90 % of the original total variance. For this reason, the last component P_3, which most likely contains only Gaussian noise, is removed.

To conclude the whitening process, we have to standardize the variances of the remaining components:

$$P_i = P_i / \sqrt{\lambda_i},\ i \in \{1, 2, 3\}\ ,$$

where $\lambda_1, \lambda_2, \lambda_3$ denote the eigenvalues of the covariance matrix of the scaled R, G, B.

In the last computation step the independent components

$$Z_{\text{text}} = \beta_{11} \cdot P_1 + \beta_{12} \cdot P_2 + \beta_{13} \cdot P_3\ ,$$
$$Z_{\text{degrad}} = \beta_{21} \cdot P_1 + \beta_{22} \cdot P_2 + \beta_{23} \cdot P_3\ ,$$
$$Z_{\text{noise}} = \beta_{31} \cdot P_1 + \beta_{32} \cdot P_2 + \beta_{33} \cdot P_3\ ,$$

are obtained using an iterative procedure described in Hyvaerinen et al. (2001) and Hyvaerinen (1999). As proposed in Hyvaerinen et al. (2001), we apply the symmetric approach and use the tanh non-linearity. Due to the dimension reduction described above, P_3 and Z_{noise} are not considered usually.

In order to decrease the number of iterations of the FastICA, an adequate starting value has to be chosen. For this purpose, we have determined the coefficients $\beta_{ij},\ i, j \in \{1, 2, 3\}$ of the demixing matrix for a large number of hectography images. We use the arithmetic average as starting value, which due to the similarity of hectography images is a better choice than random values. Additionally, the computational load may be further reduced significantly via a downsampling of the original image. A downsampling factor of $M = 2$ or $M = 3$ has performed well in our experiments.

Subsequently all preceding scalings and transformations are combined to a single transformation matrix, as described in Hyvaerinen et al. (2001) and Hyvaerinen (1999), and applied pixel-wise on the original image. Please note, that the obtained independent components $Z_{\text{text}}, Z_{\text{degrad}}, Z_{\text{noise}}$ are still kept as floating point numbers; only in the last step are they rescaled and quantized according to:

$$y = \left\lceil \frac{255}{Y_{\max} - Y_{\min}} \cdot (Y - Y_{\min}) \right\rceil, \tag{2}$$

where $Y \in \{P_1, P_2, P_3\}$.

Finally, two ambiguities specific to ICA must be dealt with: the ambiguity of the sign, and the ambiguity of the independent component ordering.

The ambiguity of the sign has the effect that in some cases some of the images representing the independent components have inverted grayscale values with respect to the original image, i.e. white background in the original image appears black in the component image. Since in text documents the amount of pixels belonging to the background usually exceeds the amount of pixels belonging to the text, this effect can be compensated rather easily by first determining the background color of the original image (after the grayscale transform), and second comparing to the background color of each component image. The previous scaling in (2) allows for using a simple static threshold of $Y_{\text{thres}} = 128$ for this purpose.

The other ambiguity of ICA is the fact that the output order of the independent components is arbitrary. As such, there is no guarantee that the first component always will in fact be the text component. This brings up the question about how to automatically identify this component. There are many possible ways of identifying the text component, some more robust than others – e.g. by using the number and properties of the connected components (Chang et al. 2004; Dillencourt et al. 1992) found in each component image.

Our approach for text component identification is a more generic one, based on two fundamental results from information theory,

1. The more "random", i.e. unpredictable and unstructured a random variable is, the larger is its (differential) entropy (Hyvaerinen et al. 2001):

$$H(X) = -\int p_X(\xi) \cdot \log p_X(\xi)\, d\xi \ ,$$

   and
2. The fact that Gaussian variables have the largest entropy among all random variables with a given covariance matrix. Both fundamental results allow us to define a measure that is zero for a Gaussian variable, and is always non-negative. This measure is known as negentropy (Hyvaerinen et al. 2001):

$$J(X) = H(X_{\text{gauss}}) - H(X) \ . \tag{3}$$

Since the brightness of letters strongly differs from the brightness of the background, text components will usually have a very structured probability density function and hence the negentropy (3) assumes a large value. On the other hand, extracted background components will have less marked inherent structure and therefore better resemble Gaussian variables. Consequently the negentropy will have a smaller value.

The main issue with negentropy is that it is computationally very difficult. In Hyvaerinen et al. (2001) several different approximations have been proposed. We employ the following one:

$$J(X) = k_1 \cdot \left(E\left\{X \cdot \exp\left(-X^2/2\right)\right\}\right)^2 + k_2 \cdot \left(E\left\{|X|\right\} - \sqrt{2/\pi}\right)^2,$$

with $k_1 = 36/\left(8\sqrt{3} - 9\right)$ and $k_2 = 24/\left(16\sqrt{3} - 27\right)$. In practice the expectation is substituted by the sample average.

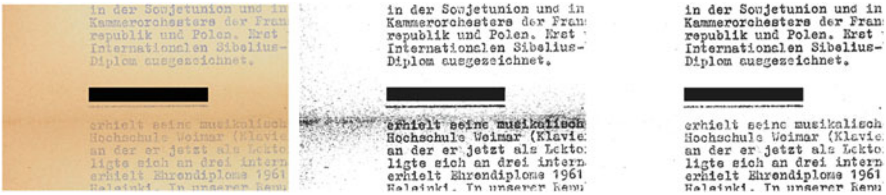


Fig. 4 *Left* – Portion of scanned hectography duplicate showing poor text quality and significant paper degradation; *center and right*: Text component after binarization, using: *center* – grayscale transformation and Otsu-binarization (Otsu 1989); *right* – proposed method. Note: all person names were blackened due to privacy reasons

4 Experimental Results

As can be seen from the zoomed-in document portion in Fig. 4, the text component is generally almost perfectly extracted. In some cases, however, the ink itself is so faint, that even this is not enough to allow us to for obtain an acceptable OCR result – this explains the relatively poor OCR results of both engines in Table 1. In such cases, classic morphological operators (Gonzalez and Woods 2001) were found to improve OCR results by an order of magnitude. Such target-specific methods do not fall into the scope of the current work, however.

In order to quantify the gains brought by the proposed method, we employ the Levenshtein distance (Levenshtein 1966) to the ground truthed page text. Our choice of the Levenshtein distance as evaluation measure was determined by the fact that it is directly proportional to the amount of human labor necessary for obtaining a perfect text result. Perfect OCR results are in general what every library or content owner would like to possess before making the digitized material available to the wider public.

Two independent OCR engines internally employing different recognition methods were used for validation, so as to ensure that the improvements are indeed generic. The OCR engines are Abbyy Finereader,[2] the leading commercial product, and Google's Tesseract,[3] arguably the best-performing free OCR engine available at present. Ground truth text templates for 22 different hectography documents were manually generated. The data set contains all together 42,825 characters and 5,833 words. In order to guarantee an unbiased comparison between the different preprocessing methods employed by the two OCR engines, the exact positions of all text lines were manually marked as well. We compare the proposed method with two classical approaches, each involving a traditional intensity-based grayscale transformation, followed by a subsequent binarization.

[2] http://finereader.abbyy.com

[3] http://code.google.com/p/tesseract-ocr

Table 1 Comparison of OCR results from two independent engines on a dataset of 22 hectography images containing 42,825 characters (5,833 words). The Levenshtein distance (Levenshtein 1966) to manually generated ground truth was used as evaluation measure

OCR engine	Otsu's method (Otsu 1989)	Sauvola et al. method (Sauvola and Pietikainen 2000)	Proposed method
Tesseract 3			
Absolute dist.	35,207	38,360	17,184
Relative dist. (%)	**82.2**	**89.6**	**40.1**
Abbyy FineReader 8			
Absolute dist.	15,193	13,754	2,182
Relative dist. (%)	**35.5**	**32.1**	**5.1**

The two binarization methods used in the comparison are the globally optimal method of Otsu (1989) and the adaptive binarization method of Sauvola and Pietikainen (2000). Both binarization methods are well known in the document analysis research community, and have consistently exhibited very good performance on document images according to recent comparisons (Gatos et al. 2011; Trier and Jain 1995). Since the hectography images are overall rather bright, a fact which would greatly disadvantage the Sauvola method, we multiply the local thresholds for each pixel by a factor of 1.2. The local window radius used was 35×35, determined to be optimal in relation to the dominant character size on the test dataset.

From the obtained results, one may see that the proposed method brings about very significant improvements in the OCR rates of both engines. The much higher error rate of Tesseract can be traced back to the fact that it uses connected components as the starting point for feature extraction. On historic documents where the ink is nearly washed out, most connected components consist of small broken parts of characters, and are as such unsuitable as a basis for the higher-level recognition task.

5 Conclusion

Within CONTENTUS, we have shown that document source-specific restoration techniques can successfully be applied to quality-impaired archive material. We proposed a novel method for foreground extraction, aimed at low-quality hectographic documents. The introduced additive generative model is not restricted solely to hectographic documents, but can robustly deal with any single-color historical documents. Our approach does not require or rely upon an a priori model for the noise component, nor on knowledge of the spatial correlation of close-by pixels. The separation of the additive components is performed by means of independent component analysis. Experimental results show that our method is able to reliably extract the foreground layer from hectography images, and provide an order of magnitude improvement in OCR performance over traditional approaches.

Our current research concentrates on the adaptation of the proposed algorithm to other historical document types, such as old typewriter documents with strong noise components. To this end we are examining the possibilities of introducing additional spatial information into our approach.

We found that – despite one of the main goals of the CONTENTUS developments being the improvement of algorithm robustness – not all source-specific impairments could be handled by one common processing path for all media sources. In these cases specifically adapted solutions for certain impairments, like low contrast samples, cannot be avoided. Future work therefore should concentrate on an automatic quality classification of the incoming media. Media having specific problems, such as the hectography material described in this article, should be automatically recognized. Once these materials are recognized, an automated workflow could be able to apply the accordant media or source-specific adaptations for processing.

Archives of cultural heritage organizations comprise media in different states of decay and deterioration. An automatic extraction of metadata through techniques that depend on the OCR recognition rate are highly likely to produce insufficient results (e.g. information extraction methods such as NER or document structure analysis) if the technical quality of the source material is below average and specific restoration algorithms, as proposed in this article, cannot be applied. Consequently, semantic search functionalities for these materials are affected as well. Upon utilizing restoration techniques as developed within the CONTENTUS use case, both the semantic analysis performed on these media assets, as well as the readability for human users, can be greatly enhanced.

References

K. Bossert, N. Flores-Herr, J. Hannemann, CONTENTUS – Technologien für digitale Bibliotheken der nächsten Generation. Dialog mit Bibl. **21**(1), 14–20 (2009)

F. Chang, C. Chen, C. Lu, A linear-time component-labeling algorithm using contour tracing technique. Comput. Vis. Image Underst. **93**(2), 206–220 (2004)

A. Cichocki, S. Amari, *Adaptive Blind Signal and Image Processing*, 1st edn. (Wiley, Hoboken, 2002)

M. Dillencourt, H. Sammet, M. Tamminen, A general approach to connected-component labeling for arbitrary image representations. J. ACM (JACM) **39**(2), 253–280 (1992)

M. Drew, S. Bergner, Spatio-chromatic decorrelation for color image compression. Image Commun. **23**(8), 599–609 (2008)

G. Dunteman, *Principal Components Analysis*. Volume 69 of Quantitative Applications in the Social Sciences, 1st edn. (SAGE Publications, Thousand Oaks, 1989)

N. Flores-Herr, S. Eickeler, J. Nandzik, S. Paal, I. Konya, H. Sack, CONTENTUS – next generation multimedia library, in *Internet der Dienste*, ed. by L. Heuser, W. Wahlster (Springer, Berlin/Heidelberg/New York, 2011a), pp. 67–88

N. Flores-Herr, H. Sack, K. Bossert, Suche in Multimediaarchiven von Kultureinrichtungen, in *Handbuch Internet-Suchmaschinen 2 – Neue Entwicklungen in der Websuche*, 1st edn., ed. by D. Lewandowski (Akademische Verlagsgesellschaft AKA GmbH, Heidelberg, 2011b), pp. 113–140

B. Gatos, K. Ntirogiannis, I. Pratikakis, DIBCO 2009: document image binarization contest. Int. J. Doc. Anal. Recognit. (IJDAR) **14**(1), 35–44 (2011)

R. Gonzalez, R. Woods, *Digital Image Processing*, 2nd edn. (Prentice Hall International, Upper Saddle River, 2001)

A. Hyvaerinen, Fast and robust fixed-point algorithms for independent component analysis. Neural Netw. **10**(3), 626–634 (1999)

A. Hyvaerinen, J. Karhunen, E. Oja, *Independent Component Analysis* (Wiley, Hoboken, 2001)

I. Jolliffe, *Principal Component Analysis*. Springer Series in Statistics, 2nd edn. (Springer, Berlin/Heidelberg/New York, 2002)

V. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals. Sov. Phys.-Dokl. **10**(8), 707–710 (1966)

J. Nandzik, B. Litz, A. Löhden, A. Heß, I. Konya, D. Baum, A. Bergholz, D. Schönfuß, C. Fey, J. Osterhoff, J. Waitelonis, H. Sack, R. Köhler, P. Ndjiki-Nya, CONTENTUS – technologies for next generation multimedia libraries. Multimed. Tools Appl. **63**(2), 287–329 (2013)

Y. Ohta, T. Kanade, T. Sakai, Color information for region segmentation. Comput. Graph. Image Process. **13**(3), 222–241 (1980)

N. Otsu, A threshold selection method from gray-level histograms. Syst. Man Cybern. **9**(1), 62–66 (1989)

J. Sauvola, M. Pietikainen, Adaptive document image binarization. Pattern Recognit. **33**(2), 225–236 (2000)

A. Tonazzini, L. Bedini, E. Salerno, Independent component analysis for document restoration. Doc. Anal. Recognit. **7**(1), 17–27 (2004), http://dblp.uni-trier.de/db/journals/ijdar/ijdar7.html#TonazziniBS04

O. Trier, A. Jain, Goal-directed evaluation of binarization methods. Pattern Anal. Mach. Intell. **17**(12), 1191–1201 (1995)

R. Wallor, Ein Ansatz zur ontologiebasierten Wissensrepräsentation. Am Beispiel des Musikinformationszentrums des Verbandes der Komponisten und Musikwissenschaftler der DDR. Master's thesis, Humboldt-Universität, Berlin, 2012

Semantic Linking in CONTENTUS

Christoph Böhme, Myriam Traub, André Bergholz, Jan Hannemann, and Lars Svensson

Abstract In this article we summarize how the use of semantic technologies can help libraries and media archives to provide better access to their collections. We show how named entity information within collections can be retrieved and used to semantically link to other collections or external data sources. Furthermore, we illustrate how users can benefit from the resulting knowledge network and we present technologies to reach these goals.

1 Introduction

Galleries, Libraries, Archives, and Museums (GLAMs) have the mission to archive and provide access to large collections of (cultural) assets. The records of the assets within these collections have been individually indexed and can be retrieved using a catalog search tool. These catalog systems have been used in libraries since the 1960s[1] and work perfectly when users look for a particular item.

However, library and archive users often do not only aim at finding a particular asset, but they look for *all* relevant items on a specific topic which are available in a collection. Furthermore, they might be interested in finding objects that are related to their research, but have not yet been identified as relevant by other scientists.

[1] http://liswiki.org/wiki/History_of_the_card_catalog

C. Böhme (✉) • M. Traub • J. Hannemann • L. Svensson
German National Library, Frankfurt am Main, Germany
e-mail: c.boehme@dnb.de; m.traub@dnb.de; j.hannemann@dnb.de; l.svensson@dnb.de

A. Bergholz
Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS), St. Augustin, Germany
e-mail: andre.bergholz@iais.fraunhofer.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_25,

Such "serendipity findings" can add an entirely new aspect to the user's research. The possibility of browsing through a collection, however, is a functionality which most catalog search systems still lack. One of the reasons is that this search method requires a database that does not contain only (bibliographic) metadata about the assets, but also content-related data that has been semantically edited.

In CONTENTUS, methods have been developed to extract relevant data from digitized objects, and to enhance them with semantic information. Using this information, a retrieval system can make use of semantic relations that exist between items within a collection, or even between items belonging to different collections.

2 Semantics in Today's Media Archives

Named Entities Semantic links can be established using named entities that are mentioned in the content of the items. A named entity can be a person, location, event, technology, sport or organization, to name but a few. While some references can be matched distinctly with the respective entity (e.g. "Eiffel Tower"), others can be highly ambiguous. For example, the name "Frankfurt" can refer either to the location "Frankfurt am Main" or to "Frankfurt (Oder)". To disambiguate the meaning of this reference, further context information is needed (e.g. *is a city in Hesse*).

With this information given, a knowledge network can be created that closely resembles the way humans use and memorize information (see Holyoak and Morrison (2005) for an introduction on theories of human thinking): by establishing associative links between objects, events and facts.

Semantic Linking After disambiguating the mentions of entities, semantic links to other relevant objects or data sources can be created. For example, a recording of John F. Kennedy's famous speech at the Rathaus Schöneberg can be linked to an entry in an encyclopedia about Berlin to provide users with additional information about the location where the speech was held.

Given such a semantically linked retrieval system, a user interested in books about a particular topic would be able to make use of content-related links to find relevant books whose bibliographic metadata do not suggest such a connection. For example, neither the title, nor the bibliographic metadata suggest that the book *Die Frau, für die ich den Computer erfand* is related to Goethe's *Faust*, yet on the first pages the author states that he was inspired to write this book by Konrad Zuse who held a speech titled "Faust, Mephisto und der Computer".

Another user might be looking for books that were written by members of the Mann family (Thomas Mann, Klaus Mann, Golo Mann, etc.). This user would exploit links that were created by enriching bibliographic metadata with further information about relations between persons (in this case, persons belonging to the same family).

These two examples show two possible types of semantic connections between two assets: They can be related by their content (similar or even same topic) or by a common entity (two books written by the same author).

3 The Role of Named Entities in Collections

Some semantic relations exist directly between objects. For instance, a commentary on Günter Grass' novel *Katz und Maus* is directly related to the novel as it discusses the book. Entities that are thematized in the commentary, e.g. in a characterization of the protagonist, can be linked directly to mentions of the same entity in the novel.

However, there is also a great number of relations which only indirectly connect objects with each other. For example, *Die Rättin* was also written by Günter Grass and is therefore related with *Katz und Maus*, even if there is no direct connection between the two. What constitutes the relation between the two books is the fact that they have been written by the same author. One way to express this would be to link those books with a "written by the same person" relation. Such a representation would pose three problems, though. First, each book should refer to every other book written by the author. If the author is a productive writer this leads to an enormous number of links required to relate all his publications. Generally, a high degree of links between items in a collection is desirable. A "written by the same person" relation, however, leads to an exponential growth of links which represent mostly redundant information. Second, whenever an author publishes a new book, the records of all his existing books need to be updated. Third, such a direct relation between the books can only express that they have been written by the same person but not who this person actually is. Luckily, these problems can be easily overcome by simply including a representation of the author along with the actual books of a collection. As a result the books can not only be related directly to each other but they can also be linked to the representation of the author. Thus, the fact that two books are written by the same author can be expressed by relating each book to the author's representation. In contrast to direct relations between the books, this solution requires only one link for each book and new books can be added to the collection without touching any existing records. Additionally, it is possible to easily retrieve all books by an author.

The exploitation of indirect relations between objects is, of course, not restricted to authors or even persons in general. It is also possible to apply the same idea to persons in other relations to an object (e.g. the directors of movies) as well as to other named entities such as organizations, locations, events or even keywords (subject headings) describing abstract concepts. By introducing entity information the degree of useful links between objects in a collection can be greatly increased. This turns a formerly loose collection of objects into a highly connected network. In this network, entities such as authors form the hubs of clusters of items which are related to them. As a result, a structure of the collection's objects emerges from these clusters.

Linking Collections Using Named Entities Named entities are not only helpful for connecting assets within a single collection, they also play a vital role when one wants to link two or more collections. The collections involved in such a task usually contain partially disjoint sets of assets. This raises the question of how to build links between the items in these collections.

Named entities can be used to bridge the divide since they possess an important property which makes them ideal candidates for linking different collections. In contrast to the assets in a collection, entities are not actually contained in a collection, but exist independently of a specific collection. Hence, different collections may contain mentions of the same (real-world or abstract) entities. This makes them ideally suited for linking items of otherwise disjoint collections.

For example, an archivist may have a large collection of movies which he wants to combine with the holdings of a library containing many biographies and newspapers. At first sight, these two collections have little in common, and the benefits of a combination are not immediately visible. Assume, though, that the archivist included the actors appearing in every movie in his records and that the librarian cataloged the biographies in her collection in similar detail. Moreover, assume that she included also entity information, which can be used to find biographies and newspaper articles about the actors appearing in the movies of the first collection. This provides the means for connecting both collections in a meaningful way.

With those links between persons, movies, books and newspapers in place the combined collections offer a much richer set of information than the individual collections could. A user searching for a movie can now be also presented with additional material related to the actors appearing in the movie. Another example is a newspaper article with an interview of an actor that can be accompanied by a list of movies in which the actor played the leading part. Note that the role of entities, when joining collections, can be understood as a form of connectors which link the assets in one collection to those in another collection. Consequently, the entirety of the entities a collection refers to can be seen as its interface for connecting with other datasets. This view of entities is illustrated in Fig. 1.

As we have seen, entities are important for enabling interlinking of collections. However, there is still a need for matching the entity references used in both collections in order to actually connect both interfaces. This process of identifying pairs of references to the same entity in both collections/interfaces can make use of context information about the entities. This information can consist only of the name of the entity or can also include additional details like geographical coordinates or biographical data. By considering additional information during the pairing process, ambiguities can be discovered and resolved, improving the accuracy of the matching. This can also minimize cataloging costs by reusing existing data.

Benefits For Users Apart from their benefits for linking collections, entity references also help to make collections more accessible to users. For instance, users searching for a book by a certain author usually think of the author as a person. Traditionally, though, they were only able to search a collection for a textual

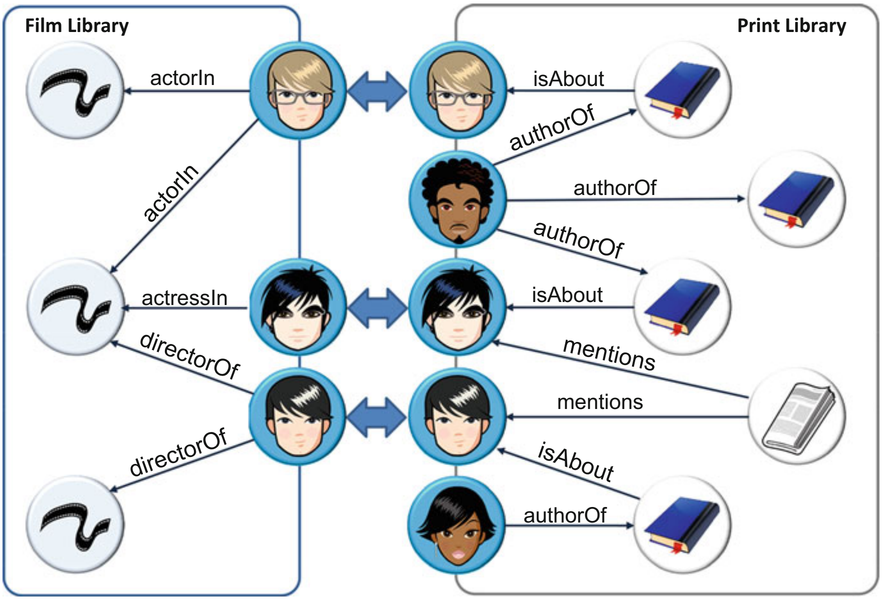


Fig. 1 Entities can be understood as the interface of a collection which can be used to link different collections together

representation of the author's name but not for the actual author they have got on their mind. Such a text-based search may produce unrelated results. For instance, a user looking for books written by the science fiction author *William Gibson* may also find books written by the New York playwright of the same name since a textual search cannot distinguish between the two authors. However, collections which contain entity references have a concept of *authors* similar to the one that users have. Thus, such a collection can be searched for all books written by the author that a user has on his mind instead of only a textual representation of her name. A smart search application can even present the user with a list of possible authors if his search request was ambiguous.

This example shows how entity references can improve accessibility of a collection by providing entry points into the collection which utilize conceptions common to humans. Consequently, the entity references in a collection are not only an interface for other collections as described above but are also part of the collection's interface to the user. In addition to supporting intelligent search, entity references can help to improve the presentation of a collection to the user. As described earlier, the objects in a collection form clusters around the entities they are referring to. Since entity references are important cornerstones for humans when structuring information (e.g., a very common approach for sorting books is by author), these clusters emerging around the entity references offer a valuable foundation for presenting the objects of a collection in a meaningful and easy to understand way in a search application to the user.

Finally, the information about the entities themselves can be beneficial for a collection's accessibility. Entity references may be annotated with detailed information describing the entities. For example, a reference to a person may not only contain the person's name but also information such as date of birth and death, a short biography or even a photograph. Moreover, an entity can also have references to other entities, allowing it to describe relations between entities such as the fact that the author Heinrich Mann is the elder brother of Thomas Mann. Such knowledge can be comfortably described using ontologies (see Sect. 5). Information about entities can be very valuable for users since it provides context information for the objects linked to the entity, thus making it easier to assess the relevance of a specific object. Furthermore, the relations between entities enable users to discover other relevant objects that they might not have thought of before, because they were not aware of the relationship between the two entities.

4 Identification and Extraction of Named Entities

As we have shown, named entities play an important role for semantically linking assets in cultural collections. They are mentioned in the metadata of the assets, as well as in the content. Metadata normally store mentions of named entities in a structured way, which makes it comparably simple to identify them. Entity mentions in the content (e.g. speech-to-text transcripts of videos) have to be extracted from continuous text, which complicates the task of identification.

Both sources therefore require specific solutions for extracting named entity references which are described in the following sections.

4.1 Utilizing Named Entities in Metadata

The metadata records of the objects in a collection are an easily available source for entity information. They usually contain references to persons, locations and organizations related to the described object. Moreover, such records are usually well structured, making it simple to identify pieces of data which refer to a certain type of entity. For instance, the "director" field of the metadata record for a movie almost certainly refers to a person, while the "studio" field most likely lists the organization responsible for producing the movie.

Even though entity mentions in metadata records tend to be easily identifiable, it is not always the case that they are already well-defined references to an entity. Often entities are only referred to by a textual representation of their name, which may or may not include a qualifier to disambiguate similar names. In order to derive entity references which are suitable for integration in a semantically linked network from such metadata, the textual references need to be disambiguated and replaced by unique identifier-based references to the entities. In order to be able to turn the textual references into identifier-based ones, a database of known entities is required. Such databases are, for instance, provided by libraries which use them to include properly disambiguated entity references in their bibliographic data. German, Austrian and Swiss libraries cooperatively manage a collection of *authority data* in the *Gemeinsame Normdatei (GND)*, which comprises about 7 million entities for persons, 1.6 million entities for organizations and 1.0 million entities of concepts and keywords including locations. The entities in these datasets are related to each other so that related persons have references to each other in their records. The development of a data model for making these data available in way suitable for semantic linking was driven by CONTENTUS and ALEXANDRIA (see Sect. 5).

With authority data being available, the disambiguation process is similar to the matching process used for connecting the interfaces of two collections. For each textual entity reference found in a metadata record the authority data is searched for matching entities in order to establish a link between the two metadata records and the entity record. This process may not only rely on the textual representation of the entity reference but may also include other information which can be derived from the metadata. For example, the fact that a book written by author X is published by publisher Y can help to disambiguate the author's name if the authority data contains information about relations between authors and publishers.

An interesting aspect of entity identification and disambiguation in metadata is that this changes the way data extracted from metadata records can be organized. Traditionally, such records are perceived as table-like data structures with a self-contained description of an object. With the introduction of entity references this view changes, though. Now, the metadata of objects no longer include all information about the objects, but only refer to it. Thus, the objects become embedded in a network containing objects and related entities. The advantages of such a networked conception of metadata are two-fold: First, it gets easier to express relations between objects as well as between objects and entities. Second, such metadata can easily be extended simply by including references to additional information, making it easier to collect, consolidate and interlink metadata sources.

4.2 Named Entity Recognition and Disambiguation in Print Media

In addition to mentions of entities in metadata, many cultural assets also contain named entities themselves. Textual and spoken-word documents are particularly rich sources of such entity references within the asset's actual content. In the context of CONTENTUS the detection is restricted to names of persons, locations, and organizations in German natural language texts and audio recordings of speech. The extraction of entity references in print media/text is a two-step process consisting of

Table 1 Example of a sentence with the tokens labeled in BIO-notation

Angela	Merkel	(	CDU	)	met	Obama	in	Berlin
B_PER	I_PER	O	B_ORG	O	O	B_PER	O	B_LOC

named entity recognition followed by named entity disambiguation. In named entity recognition, we identify for each token of the text if it belongs to a named entity of a certain type (person, location, organization). In named entity disambiguation we identify the recognized entities by matching them to entries in an ontology, in our case Wikipedia, and person data in the GND (see Sect. 4.1). In contrast to the disambiguation process used for entities found in an asset's metadata, the disambiguation process needs to consider the textual context in which the named entity was detected.

Named Entity Recognition The goal of Named Entity Recognition is to assign to every token of a text a label indicating if the token is part of a named entity. A simple approach would be to use a dictionary look-up. This approach has, however, some drawbacks. One is maintenance of the dictionary, which can be cumbersome. Another one is that in German, where all nouns are capitalized, one cannot be sure if the words "Kohl" and "Strauß" refer to persons or not. Finally, dictionary look-ups cannot generate probabilities for the correctness of the labelling. A more elaborate approach would be to use dictionary information as input to a statistical model that forms the basis of a classifier. This classifier can then classify each token of the text to generate a labelling. As an additional benefit, some classifiers automatically assign a probability or confidence to their classifications, so that results with some degree of uncertainty can be filtered out automatically.

Our method is to make use of Conditional Random Fields (CRFs), a supervised learning algorithm developed to label structured data (Lafferty et al. 2001). CRFs enable us to label sequences not by classifying each item individually, but by considering the neighborhood instead. They are an improvement over the well-known Hidden Markov Models (HMMs): Whereas HMMs consider only a fixed number of predecessors, CRFs optimize globally and allow for the inclusion of arbitrary features.

In a supervised method labeled training data is needed. We use the well-established BIO-notation to label the data (Ratinov and Roth 2009). A `B` indicates the beginning (i.e., the first token) of a named entity, an `I` indicates the continuation ("Inside") of a named entity, and an `O` indicates other tokens not part of any named entity. Table 1 gives an example sentence with its labeling.

The sentence is the unit that is treated independently in our method, i.e., no information is passed over sentence boundaries. To train the learning algorithm we model every word (token) w_i in a sentence as a vector x_i. The vector is composed of real-valued features representing the properties of the word and its neighborhood. Features can be of different types:

1. Syntactical features:
   - Capitalization of words
   - Presence of numbers and special characters, such as hyphens, etc.
   - Pre- and suffixes, such as inter-, post-, -ing, -s, etc.
2. Structural features:
   - Part-of-speech tags (indicating the role of words)
   - Parse tree information (e.g., the path to the root)
3. Dictionary features: Indicators of whether the word is listed in some dictionary of persons, locations, etc.
4. Neighborhood features: Features of the k preceding and the k following words

As listed, Conditional Random Fields also allow us to take into account properties of neighboring words, which makes it possible to represent the context of words explicitly. Thus, we represent each word w_i as a pair (x_i, y_i) where x_i is the feature vector and y_i the label (in BIO-notation, as seen in Table 1).

We train the model using the German dataset of Deutsche Presseagentur.

Named Entity Disambiguation After words have been identified to represent entities, the second goal is to identify them. For example, “Michael Jackson” is a well-known American pop star, but also a rather well-known English journalist who authored several books about beer and whisky. Named Entity Recognition only gives us the information that “Michael Jackson” is a person.

To identify the person in question, what we essentially do is relate the context of the occurrence of the person’s name to some background information we have about candidate persons. We use Wikipedia to provide us with this background information. Hence, the first step is to identify candidate persons and the second step is to select the most likely candidate. To identify candidate persons, we rely first and foremost on the complete name of the persons. Furthermore, we employ heuristics to deduce the complete name if only part of it is given in the current occurrence, i.e., we look for previous occurrences of a possibly complete name in the same paragraph or document. Last, we employ some simple rules to identify very specific people, e.g., by their function or profession. As an example think about “Joseph Ratzinger” and “Pope Benedict XVI”. To select the most likely candidate, we again perform two steps to compute a combined score. This score is a combination of a context-dependent score and an a priori context-free score. First, we determine the similarity of the current document to each of the Wikipedia entries of the candidates. To this end we compute the cosine similarity between the documents, which measures the angle between the vector representations of the respective documents. To improve the result of the computation, we perform length normalization and stop-word removal. Stop-words are very frequent words, such as articles or prepositions. Their presence would dilute the computation of document similarity. This first step gives us the context-dependent score. Second, we combine this score with the a priori score for each candidate. The a priori score indicates the context-free probability of each candidate. For example, for “Michael Jackson”, the pop star has a much higher a priori score than the English journalist. We determine the a priori score simply by counting the incoming links within Wikipedia for each candidate.

We notice that there are differences in performance depending on the type of entity. In particular, identifying locations is the most difficult, because the context information is less useful than for the other types. For example, information about a bank robbery or heavy rainfall does not help much to identify locations. On the other hand, identifying organizations is the easiest, because context information is usually useful, and the number of candidates is usually small.

4.3 Named Entity Recognition and Disambiguation in Audio-Visual Media

When analyzing texts coming from the audio portion of an audio-visual stream, we are faced with additional difficulties. First, the transformation of the stream is error-prone, so incorrect words may appear in the stream. Second, important features that help to train high-quality models are missing. Among them are punctuation and capitalization. The lack of punctuation is particularly problematic, as the sentence is our basic processing unit.

We use several heuristics to cope with these difficulties. First, we analyze pauses in the audio-visual stream to detect sentence boundaries. Second, we use dictionaries to correct transcription errors and introduce capitalization.

As can be expected, performance on audio-visual streams is somewhat lower than on pure text streams. However, given the quality of transcription our method turns out to be robust.

5 AgRelOn: An Ontology for Persons and Organizations

We argued that the most important reason for introducing semantic linking in collections of cultural assets is the resulting ability to express relations between the assets and entities. This enables the processing and presentation of cultural assets in their context. A commonly reoccurring theme in historical and sociological research, or in literary criticism, is the importance of the relations between people, as well as between people and groups. Much information on those relations is readily available, e.g. in Wikipedia, but in almost all sources, those relations are not encoded in a machine-comprehensible manner. As a consequence, it is not – or only indirectly – possible to answer questions such as “What was the relation between Charlotte von Stein and Johann Wolfgang von Goethe?”, or “What was the name of Kurt Weill’s spouse?” Another problem is that even if the information is readily available or easily extractable, we lack data models making it possible to express those relations in such a granular fashion that we can find a second cousin while searching for all relatives, whereas a search for half brothers and sisters only results in findings for those persons for whom the relation is unambiguously clarified. The solution was to develop an ontological model including a relationship hierarchy: the Agent Relationship Ontology (AgRelOn[2]). AgRelOn models relations between persons and groups in a very explicit fashion. Most use cases we considered derive from the cultural heritage domain, but the ontology is constructed to be used in all fields dealing with agents (i.e., people and groups) in one way or the other. In order to achieve this, the ontology can be extended with further relation types that are not yet defined. Within the model we can also annotate relations between agents with provenance and quality information. That way we can allow user contributed information from (potentially unknown) users in a community portal for genealogical studies on the one hand and data from research databases on the other hand to be mixed in one database, while still making it possible to retrace the origin of each piece of data and establish its reliability.

The use of AgRelOn is not confined to family relationships only, but also models more specific relations. By using the ontology we can, for example, express that one person was another person's murderer or muse, or that somebody can channel spirits. Charlotte von Stein was the muse of Goethe and Schiller, so using the relations in AgRelOn we can write [Stein *isMuseOf* Goethe] and [Stein *isMuseOf* Schiller]. Carl Gustav Jung considered his communication with spirits as constitutive for his work, so [Jung *isChannelOf* Philemon] is an important agent relation in this context. John Kennedy Toole committed suicide since he was not able to find a publisher for his novel – which later was an international success – which we can express as [Toole *hasMurderer* Toole]. At first sight, some of these relations might seem somewhat odd. They are, however, a result of what librarians record as part of their day-to-day work, which is not only to describe books, but also to record relevant facts on authors and institutions related to the publications. In Austrian, German and Swiss libraries this information is collected in the GND (see Sect. 4.1).

The first version of AgRelOn comprises almost 70 different types of relations, based on most commonly used relation types found in the GND. The relations are gender neutral and divided into private acquaintances, group associations, correspondence partners, professional contacts, family relations, spiritual contacts and relations with impact on life and death (e.g., personal physician or murderer). Those rather abstract relation types find more specific forms in concrete subrelations, such as *hasChief* – a specialization of *hasProfessionalContact* – or *hasParent* as a specific version of *hasAncestor*, which, in turn, is a more precise version of *hasRelative*.

Within CONTENTUS, AgRelOn models primarily the semantic relationships between people that are relevant in the context of the media in question, e.g., authors, or people mentioned in books, newspaper articles or TV interviews. Further, the German National Library has remodeled its authority data by integrating the formerly individual authority files for persons, corporate bodies and subject

[2] http://www.contentus-projekt.de/agrelon.html

headings into the GND. Within the GND, relations between people as well as between corporate bodies and persons are modelled in accordance with AgRelOn. The GND and thus the respective AgRelOn relations are readily available through the library's linked data service.[3]

6 Technologies for Accessing Named Entity Information

We showed how information about entities can provide a valuable addition to a collection. In CONTENTUS, technologies were developed to support providing entity information for media collections and to support matching entity descriptions from different collections. The structure of entity information usually differs from those of the assets' metadata. In particular, the increased number of references requires new methods of accessing such data. For this reason, a web service was developed in CONTENTUS, which provides ontology information about named entities. The service expects unique identifiers as input and returns either only information directly related to the requested named entity or – additionally – a set of related named entities. For example, if an entity describing a location is requested from the web service, the service will also return all locations of which the requested location is a part. The primary output format of the web service is RDF+XML (Beckett and McBride 2004); additionally, data can be returned in a custom XML format, which directly reflects the hierarchical structure of the information about an entity, as opposed to the flat RDF+XML representation, where this can only be modeled by introducing abstract entities (in RDF terminology these are called *blank nodes*). The disadvantage of introducing such entities is that the entities do not necessarily need to be described within the main entity representation, which complicates the processing of the entities. The XML structure alleviates this problem. In addition to the web service, tools were developed to support the process of matching entities from different sources. These tools provide a framework for extracting and converting metadata, so that matching algorithms can work on a normalized data format. Additionally, rules for matching location names were developed. The technologies for semantic linking developed in CONTENTUS are available as open source software as part of the Culturegraph project (Beckett and McBride 2004).

7 Conclusion

Semantic linking can improve accessibility of GLAM media collections, affecting both the ability to intelligently search collections and link different collections together. Named entities play a crucial role in this context, representing interfaces

[3] http://www.dnb.de/DE/Service/DigitaleDienste/LinkedData/linkeddata.html

between collections as well as means to formulate searches by meaning rather than words alone. There are two main sources for associating media with named entities: metadata and the media themselves. For both sources, the tools developed within CONTENTUS for extracting entity information were presented, along with an ontology for modeling the relationships between persons and organization entities. Furthermore, we showed how this information can be offered/serviced and used for matching entities. Both, the ontology and the web service are publicly available.[4][5]

Acknowledgements We would like thank our colleague Claudia Grote for her comments on an earlier draft of this article.

References

D. Beckett, B. McBride, RDF/XML syntax specification (revised). World Wide Web Consortium (2004), http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/

K.J. Holyoak, R.G. Morrison (eds.), *The Cambridge Handbook of Thinking and Reasoning* (Cambridge University Press, New York, 2005)

J. Lafferty, A. McCallum, F. Pereira, Conditional random fields: probabilistic models for segmenting and labeling sequence data, in *Proceedings of the 18th International Conference on Machine Learning (ICML '01)*, Williamstown (Morgan Kaufmann, San Francisco, 2001), pp. 282–289, citeseer.ist.psu.edu/lafferty01conditional.html

L. Ratinov, D. Roth, Design challenges and misconceptions in named entity recognition, in *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL '09)*, Boulder (Association for Computational Linguistics, Stroudsburg, 2009), pp. 147–155

[4] http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/

[5] http://www.culturegraph.org/

Semantic Processing of Medical Data

Sonja Zillner, Sascha Seifert, Marius Erdt, Philipp Daumke, and Martin Kramer

Abstract Medical images increase in quality and quantity: More and more detailed image content can be represented on the pixel level, and increasing amounts of medical images are produced in the context of clinical diagnosis. Technological solutions are needed to enhance existing clinical IT solutions helping clinicians to access and use medical images optimally. Within MEDICO, we developed methods and tools (a) to parse and describe the content of medical images, (b) to extract and annotate the related information from radiology reports, and (c) to provide and manage medical ontologies as a common language for labeling and integrating the various information entities.

1 Introduction

Today, clinicians need to switch between various applications in order to access the comprehensive, heterogeneous and distributed patient data. To improve the clinical workflow and data access, our goal is to enable the semantic integration of medical images and all their related clincial data, such as radiology reports, lab results, etc.

S. Zillner (✉)
Siemens AG, Munich, Germany
e-mail: sonja.zillner@siemens.com

S. Seifert • M. Kramer
Siemens AG, Erlangen, Germany
e-mail: sascha.seifert@siemens.com; martin.kramer@siemens.com

M. Erdt
Fraunhofer Institute for Computer Graphics Research (IGD), Munich, Germany
e-mail: marius.erdt@igd.fraunhofer.de

P. Daumke
Averbis GmbH, Freiburg, Germany
e-mail: daumke@averbis.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_26,

This helps clinicians to get better access to and understanding of a patient case without being interrupted in their workflow. Our approach in MEDICO is to make the content description of medical images and dictated radiology reports machine-accessible through annotation using medical ontologies. By representing image and text annotations as instances in the ontology, they can be serialized to a variety of formats, which again enables the interoperability among a variety of clinical systems containing patient data, such as PACS (Picture Archiving and Communication Systems) storing medical images, RIS (Radiology Information Systems) storing the current radiology report, or HIS (Hospital Information Systems) storing other clinical data such as lab values, prior reports, and administrative information about the patient.

2 Medical Image Annotation

Our goal is to make the image content information accessible for machines through annotation using medical ontologies. The creation of image annotation is accomplished semi-automatically: We are using various methods for image parsing to automatically discover and describe the medical image content. The objective of the MEDICO image parsing approaches is, to automatically annotate anatomical structures in medical images. In addition, due to the complexity of medical knowledge, we enhance the automated parsing approaches by a manual validation step that is accomplished by clinicians in order to correct, validate or extend automatically generated annotations. In the following, we will introduce the various image parsing approaches (Sects. 2.1–2.3), as well as the overall semantic image annotation workflow (Sect. 2.4).

2.1 Hierarchical Parsing of Image Content

Although whole body image scanning is common practice for the diagnosis of systemic diseases, such as cancer, and for differential diagnosis, dealing with the large amount of data of a full body scan remains a challenging issue. Therefore techniques are needed for the vast detection and segmentation of organs. There do exist various robust solutions for the segmentation of specific organs, such as the heart (Zheng et al. 2007), liver (Ling et al. 2008), pancreas (see Sect. 2.2), and lymph nodes (see Sect. 2.3). However, those solutions are organ- or disease-specific and it is costly to transfer to new organs.

For that reason, we developed a framework for hierarchically parsing whole body CT volumes and efficiently segmenting multiple organs by taking contextual information into account. The technology behind the image parsing of MEDICO is based on the Marginal Space Learning (MSL) (Zheng et al. 2007) approach. Recently, MSL has been established as a powerful machine learning-based method

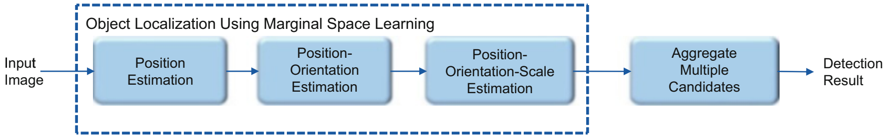


Fig. 1 The MSL method which enables MEDICO to efficiently process images, i.e., to detect 6 organs and 19 landmarks in about 20 s

for efficiently understanding the images' content. By presenting numerous examples of some anatomical structures in images, such as organs, e.g., the heart, or tissues, such as lymph nodes, the developed system learns their image appearance and recognizes them in new, unseen images. A great advantage of the developed machine learning technology is that different medical modalities, e.g., Computed Tomography (CT), Magnet Resonance Imaging (MRI) or Ultra-Sound (US), can be handled in an analogous manner; in general, retraining without touching program code is sufficient.

Figure 1 shows the implementation of MSL using a sequence of trained classifiers to estimate the pose, i.e. the position, the orientation, and the scale of an organ. The same pipeline without scale estimation can be used for detecting landmarks.

Each learned classifier is a Probabilistic Boosting Tree (PBT) (Tu 2005) with 3D-Haar-like features (Tu et al. 2006) and steerable features (Zheng et al. 2007) and trained with AdaBoost (Schapire and Singer 1999). Figuratively speaking, in the case of the position estimation, PBT superposes a set of gray-value patterns, the 3D Haar-like features, which together represent the appearance of an anatomical object in the image.

The overall image analysis process chain of the MEDICO image parsing is illustrated in Fig. 2. It consists of two main parts: the Discriminative Anatomical Network (DAN) and the database-guided segmentation module.

The purpose of the DAN is to give an estimate about the scale of the patient and the portion of the volume visible, and to detect a set of landmarks. To obtain a fast and robust system, the landmarks are connected in a graph (network). Information regarding the location of each landmark is propagated through the graph, which not only speeds up detection, but also increases detection accuracy. The database-guided segmentation module uses the output of the DAN for the detection of the position, the orientation, and the scale of the organs visible in the given volume portion. By the use of boundary classifiers, the various organs are subsequently delineated. The precision of the organ segmentation within the image parsing system was evaluated with cross-validation using a mesh-to-mesh error metric. The organ segmentation accuracy is in the range of 1.3–3.0 mm. For training, up to 457 manual organ and 591 landmark annotations were accomplished. The average detection rate is 89.3 % and it takes about 20 s to process a full body CT volume. For details about the image parsing system and evaluation results see Seifert et al. (2009).

Not all organs and content information of medical images can be detected by means of the hierarchical image parsing approach. The pancreas, for instance,

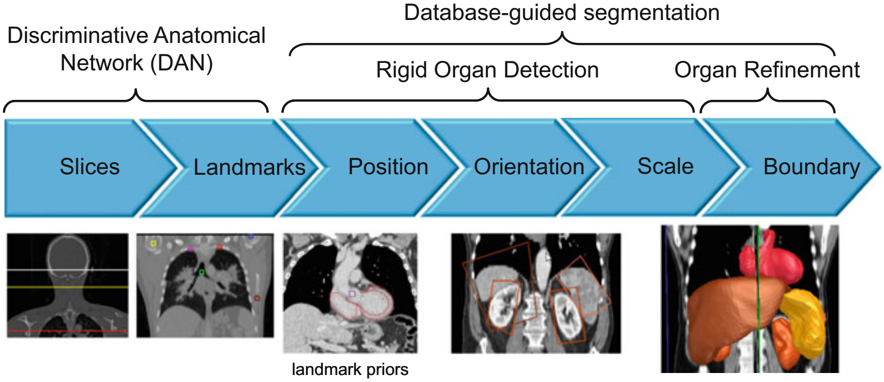


Fig. 2 The process chain of MEDICO's image parsing system

requires – due to its particular appearance that is even for human experts difficult to discover – dedicated image parsing methods (Sect. 2.2). Similar is the situation when parsing lymph nodes. Due to their irregular shape and low contrast, specific methods for the automatic detection are required (Sect. 2.3).

2.2 Pancreas Segmentation

Segmentation of pancreas tissue in CT is difficult even for a human, since the pancreas head is always directly connected to the small bowel and can in most cases not be visually distinguished. This is also the area with the highest human observer variability. Furthermore, the small bowel may also contact the pancreas at any other place. Surrounding organs like liver, stomach and spleen are also problematic though contrast agent saturation can help to differentiate these organs from the pancreas.

However, the automatic segmentation of the pancreas is a highly demanded and requested feature within clinical practice, as it paves the way towards improved clinical practice: Pancreatic cancer like ductal adenocarcinoma has a high mortality rate (5-year survival below 5 %) and is one of the most difficult cancers to treat. Patients are commonly examined using early parenchyma phase abdominal CT. Automatic delineation can support the clinician in tedious contouring work in order to cope with high-resolution data available nowadays. Furthermore, through exclusion of pancreas it can significantly help to develop automatic segmentation methods for other abdominal structures like intestine or enlarged abdominal lymph nodes.

The automatic segmentation of the pancreas is difficult to accomplish as the global appearance-based features, which are used successfully for the classification of other organs, are generally inaccurate for pancreas detection as its tissue is often

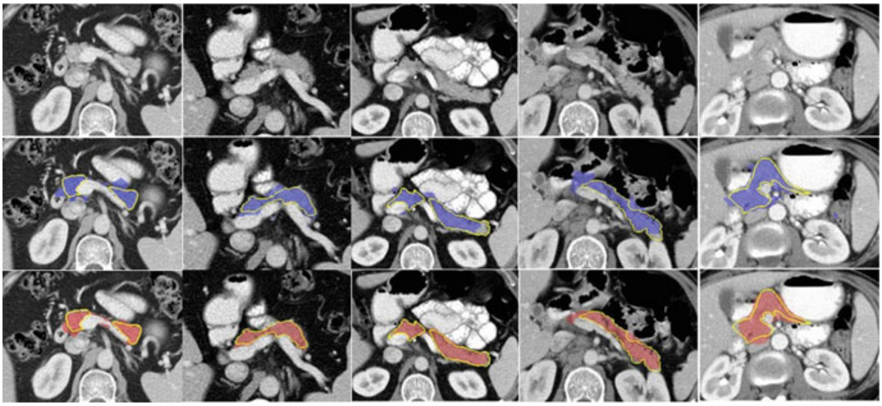


Fig. 3 Segmentation results of pancreas in exemplary datasets. Original datasets (*upper row*), detector output (*blue*, *middle row*) and final segmentation result (*red*, *bottom row*). The outline marks the ground truth

indistinguishable from the small bowel. In order to overcome this problem, our approach aims at learning local appearance features that incorporate the spatial relationship between the spleen vessel path and the pancreas position. Here, the idea is that the pancreas always lies near the spleen vessel. First, the spleen vessel is segmented in the data using vessel enhancement techniques and region growing. Secondly, a pancreas tissue classifier is built using a set of representative training datasets. This classifier incorporates features that are computed around the vessel path extracted before. These features represent local image texture. Additionally, the typical distance of the pancreas to the spleen vessel is learned by the classifier. In that way, tissue far away from the spleen vessel has a low probability of being pancreas. After applying the classifier to the image, a statistical shape model of the pancreas is initialized at the position with the highest pancreas tissue probability. The model is then adapted to the data in order to determine the final segmentation.

For evaluation, abdominal early parenchyma single phase CT data from 40 cases was acquired. Inter-slice spacing was 5 mm abdominal control standard protocol, whereas spacing within an axial slice varied between 0.6 and 0.7 mm. Manual delineation from experienced radiologists was taken as the gold standard. Threefold cross-validation was used for performance evaluation. For each fold, the statistical shape model as well as the classifiers were learned on the training data and evaluated on the test data. Figure 3 shows exemplary qualitative results separately for the detector and final segmentation. While using the detector only (middle row) already shows good accuracy compared to ground truth (bright outline), the incorporation of the statistical shape model further improves the quality (bottom row). Generally, there is very limited leakage into neighboring structures (bottom row) given the low contrast of pancreas boundaries (upper row).

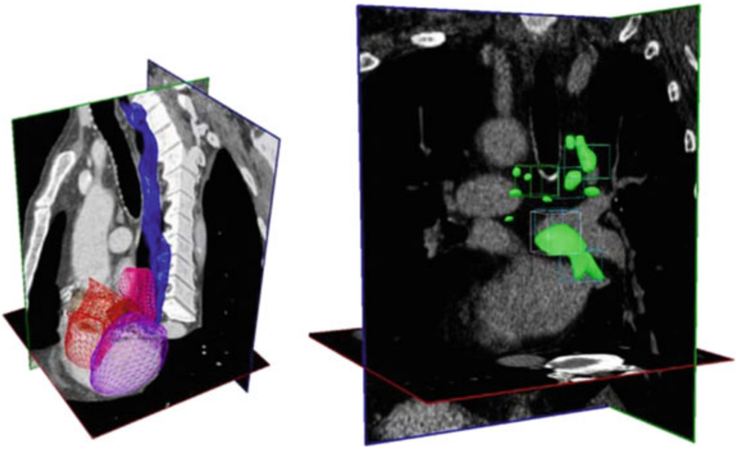


Fig. 4 The remaining chest space with removed organs is used as a basis to detect mediastinal lymph nodes (*left*). The automatically detected lymph nodes around the heart in *green* overlaid with the original CT image (*right*)

2.3 Lymph Node Detection

The automated image parsing of lymph nodes requires dedicated methods. Lymph nodes have high clinical relevance, but detection is challenging as they are hard to see due to low contrast and irregular shape. In MEDICO, a method for fully automatic mediastinal lymph node detection in 3-D computed tomography (CT) images of the chest area was developed. Discriminative learning is used to detect lymph nodes based on their appearance. Because lymph nodes can easily be confused with other structures, it is vital to incorporate as much anatomical knowledge as possible to achieve good detection rates. Here, a learned prior of the spatial distribution is proposed to model this knowledge. As atlas matching is generally inaccurate in the chest area, because of anatomical variations, this prior is not learned in the space of a single atlas, but in the space of multiple ones that are attached to anatomical structures. During testing, the priors are weighted and merged according to spatial distances. Cross-validation on 54 CT datasets showed that the prior-based detector yields a true positive rate of 52.3 % for seven false positives per volume image, which is about two times better than without a spatial prior. The knowledge in this detection approach is represented as a spatial distribution of lymph nodes that is learned with respect to anatomical structures. The algorithm makes use of the 20 salient anatomical landmarks resulting from the initial image parsing process. Examples of landmarks used are the bifurcation of the trachea, the bottom tip of the shoulder blade left and right, the topmost point of the aortic arch and the topmost point of the lung left and right. Besides the landmarks, a number of different organs such as the lungs, the heart and the trachea are used to shrink the space where the lymph nodes in the chest can normally be found (see Fig. 4).

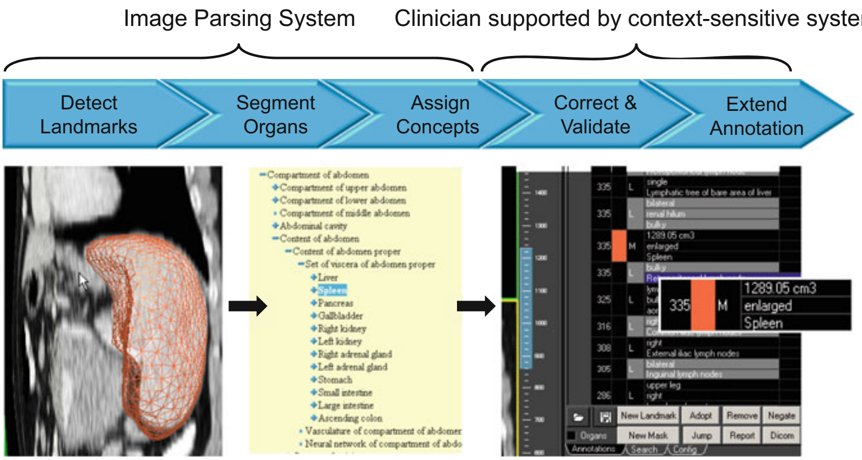


Fig. 5 The workflow for semantic image annotation

2.4 *Semantic Image Annotation Workflow*

Automated image parsing techniques are a very important instrument for capturing the content information of medical images. However, due to the vast number of anatomical structures and their pathological changes, image parsing is not yet able to fully capture the content of arbitrary medical images. The manual image annotation remains an important complement, which enables clinicians to correct, validate, or extend the automatically generated annotations. Figure 5 depicts the workflow of our semi-automated approach for the semantic annotation of medical images.

The various image parsing techniques detect the landmarks, segment organs and automatically assign ontological concepts. The mapping between anatomy and ontological concept is intrinsic by design, as the detector locates the specific anatomical structure with an apparent concept class. Subsequently, the user corrects or validates these findings and decides if he or she wants to adopt a validated landmark or organ. If the finding is validated, it will be added to the list of image annotations, i.e., a list of specified image regions, that are labeled with concepts from medical ontologies describing the particular content of the region.

3 Information Extraction from Radiology Reports

To enable the efficient integration of image and text information, the content of radiology reports needs to be annotated with concepts of medical ontologies. Radiology reports contain a wealth of information relevant for quality assurance, research, administration and teaching purposes. Radiologists need to search in radiology reports for validating a diagnosis, finding patients for a certain study, or preparing meetings and presentations. Reports are usually written as narrative text making it difficult to computationally access this information. In order to facilitate not only searching but also complex statistical evaluations, relevant information must be extracted by means of Natural Language Processing (NLP). One peculiarity of radiology reports is that they can be analyzed along three different dimensions:

- *Anatomical/spatial information*: Describing the location of a finding – e.g., the affected organ or lymph node – encompasses spatial modifiers, such as left, right or mediastinal, or location attributes of the ontological concepts, such as '*axillary lymph node*'.
- *Pathological information*: describing the pathological interpretation of the located findings in terms of size (e.g., '*enlarged lymph node*'), of density (e.g., '*lung nodules*'), or of number of occurrences (e.g., '*set of lymph nodes*').
- *Temporal information*: provides information about differences/changes of the current findings in relation to the past findings (e.g., '*… in comparison to the preceding study …*').

Another aspect of radiology reports is that they do not follow the general way of grammar. For instance, the sentences often lack verbs and punctuation, or abbreviations and negations are very common. The medical language in general is characterized by complex word forms such as '*pseudo-hypo-para-thyroid-ism*', '*gluco-corticoid-s*' or '*pancreat-itis*'. This is even more accentuated in morphologically richer languages than English, such as German. Synonymy, morphological derivation and other linguistic phenomena are used extensively: '*Pulmonary metastasis*' is used synonymously to '*lung metastasis*', '*secondary lung tumour*' or '*metastatic tumor to the lung*'. Descriptive and unassertive wordings are common, e.g. '*… subtle hypodensity near the lung dome, metastatic focus cannot completely excluded …*'. Thus, means for automatic extraction of knowledge from radiology reports have to address those textual particularities.

In RADMINING, a partner project of MEDICO, a scalable and industrial strength text mining platform[1] was applied to extract relevant information from radiology reports. Leading technologies from the field of text analysis and knowledge management are combined with knowledge of experts from various domains (radiology, IT, linguistics). Radiology reports undergo comprehensive semantic analysis, taking into account the different linguistic peculiarities:

- The first step is a syntactical and semantical preprocessing that detects sections, sentences, words, word classes, abbreviations, negations, and other entities in text. Even word parts (so-called *morphemes*) of composed words such as '*appendix', 'appendic', 'inflamed', and '-itis*' are identified by an extensive morphological analysis (Markó et al. 2005), allowing us to treat phrases like '*appendicitis*' and '*inflamed appendix*' as synonyms.

[1] http://www.averbis.de/en/averbis_extraction_platform

- The preprocessing serves as a basis for the second step, the mapping of text passages to a clinical terminology. Clinical terminologies are collections of terms which are used to represent distinct medical aspects and entities. They help in bridging the gap between clinical information expressed in natural language, and symbolic representations understood by computers. The Lexicon of Standardized Radiological Terms (RADLEX) (Langlotz 2006) intends to support the structured reporting of image findings and the indexing of teaching cases. In RADMINING, a German translation of version 2.0 (Marwede et al. 2009) has been extended and is used as primary terminology.
- In the third step, relations between concepts are identified. A frequent relation pattern is the combination of anatomical terms (e.g. '*liver*') and associated findings ('*multiple metastases*'). Another pattern is the assignment of a negation (e.g. '*no evidence of*') to a negated term (e.g. '*metastatic disease*').

As a result of semantic analysis, radiology reports are enriched with standardized metadata about their content. These semantically enriched reports serve as input for various advanced medical applications (see Seifert et al. 2014). They may be used to semi-automatically tag medical images with information from the report. They can be loaded into a data warehouse and allow comprehensive statistical analysis of large report collections. Or they are fed into a semantic search engine providing various added values for radiologists, such as the quick validation of presumptive diagnoses.

4 Ontology Engineering and Management

For the semantic processing of medical image and text annotations, we require (a) a standard and agreed terminology to label the annotations, as well as (b) a comprehensive data model to store the annotations in an efficient and integrated manner. Currently, no standard terminology for describing medical image- and text-related content exists. For covering the comprehensive domain of medical image search within the MEDICO use case, we require a set of various medical ontologies.

The interviews with clinicians and radiologists showed that medical imaging and patient data need to be considered along three different perspectives: (a) the anatomical spatial perspective that addresses body parts and their locations, (b) the radiology-specific perspective, which describes the relationships between various image modalities and anatomical regions as shown on medical images, and (c) the disease perspective that concerns the distinction between the normal and the abnormal imaging features, body parts and their locations. Upon agreement with the clinicians, from this set we decided upon the Foundational Model of Anatomy (FMA)[2] in anatomy, the Radiology Lexicon (Radlex)[3] in radiology and the ICD-9

[2]http://sig.biostr.washington.edu/projects/fm/FME/index.html

[3]http://www.rsna.org/radlex

CM[4] to represent the disease dimension. Each of the three ontologies represents a piece of knowledge that is necessary to realize the entire application. These knowledge pieces are not arbitrary, but they need to be interrelated within the context of the application. Therefore, the separate (fragments of) ontologies need to be integrated to deliver the whole picture (Sect. 4.1). In addition, we require a patient data model that provides the basis for comprehensive clincial data access by storing the image and text annotation in an integrated manner (Sect. 4.2).

4.1 Ontology Alignment

The challenging aspect for ontology aligning is the fact that the medical ontologies differ in their level of detail and coverage of particular domains. For instance, in the case of FMA and Radlex, both ontologies have strong overlaps, but both ontologies are needed for describing the required level of medical image region descriptions. This means that the alignment of these ontologies is vital for a precise semantic annotation of medical images.

The majority of methods for ontology alignment turned out to be not feasible in the medical domain, because of the large size and complex interdependencies of medical concepts. Furthermore, medical ontologies are typically rich in linguistic information. For example, the Foundational Model of Anatomy (FMA)[5] contains concept names as long as “Anastomotic branch of right anterior inferior cerebellar artery with right superior cerebellar artery”. Such long multi-word terms are usually rich in implicit semantic relations. Thus, ontology alignment approaches for the medical domain need to incorporate the described common characteristics of medical ontologies. Our proposed medical ontology alignment framework (Zillner and Sonntag 2011) has three main aspects: it suggests a combined strategy that is based on (a) the automated linguistic-based pre-processing of ontology concepts to be aligned, (b) the fine-tuning of correspondences by formulating context-specific axioms, and (c) the continuous evaluation of user feedback for composing effective and context-specific ontology alignments.

Linguistic Rules We take an information retrieval approach to discover relationships between the FMA and the Radlex Taxonomy. Therefore, we treat FMA ontology concepts as documents, which we indexed using Lucene, and match them against the search queries, which are the concepts from the Radlex taxonomy. To cope with the complex linguistic phrase structure, we enhanced the basic information retrieval approach by incorporating mapping rules that reflect linguistic features of the natural language phrases which describe a particular concept. Our assumption is that common patterns in the multi-word terms, which are typical

[4] http://www.cdc.gov/nchs/icd/icd9cm.htm

[5] http://sig.biostr.washington.edu/projects/fm/FME/index.html

for the concept labels in the medical ontologies, can be made explicit. As medical concepts are multi-term expressions, exact matches are rare. However, FMA and RadLex concepts follow a similar linguistic structure. By formally defining and using a set of linguistic-based mapping rules, we could enhance the number of alignment correspondences. For instance, by using a trigram rule that relies on a pattern of three terms, mappings of type ‘more specific’, such as (“buccinator lymph node”, “left buccinator lymph node”), could be identified, and we could enhance the number of mappings by number 6.9.

Context Information With the linguistic methods we could continously improve the recall of alignment mappings at the cost of the precision measure. For improving the precision value, we incorporated context information to reflect the particularities of the domain. For instance, we defined a function that helps to filter out alignment correspondences that contain antonym terms, i.e., mappings such as (“right buccinator lymph node”, “left buccinator lymph node”) could be avoided.

Evaluation with User-Involvement As medical/health knowledge is a very sensitive and delicate context, the evaluation of ontology alignments requires the involvement of medical experts. At the same time, as medical ontologies are large in size, the manual alignment of medical ontologies is cumbersome and impracticable. To support clinical experts in establishing ontology mappings, we combined the automatic pre-processing of alignments with continuous clinical feedback in an intelligent manner. Thus, the clinical feedback determines the composition of linguistic rules and context information to compile an ontology alignment that is most appropriate for a particular application scenario. For enabling the efficient and transparent processing of user feedback, we provided explanations of the matching results, i.e. each established ontology mapping instance incorporates the arguments of its derivation, as well as describes the type of relationship.

4.2 Annotation Ontology

The semantic annotations are stored in the *Annotation Ontology* (see Fig. 6), the design of which adheres to the Information Object Definition of the DICOM standard. The central concept of the *Annotation Ontology* scheme is the patient, represented by the OWL-class `Patient`. Every `Patient` owns a number of `Study` defined by a unique identifier and a specific time period. The MEDICO study is more than a DICOM study: it is a container for all annotations from images, texts, and clinical data within a given timeperiod. This is the cornerstone for enabling temporal queries, as well as queries considering multiple modalities.

The scheme is illustrated in Fig. 6. Its design was driven by the following requirements:

- *Linkage of text passages and image regions*: Annotations from images and texts must be stored in the same model, which should consider the fact that reports summarize annotations from multiple images.

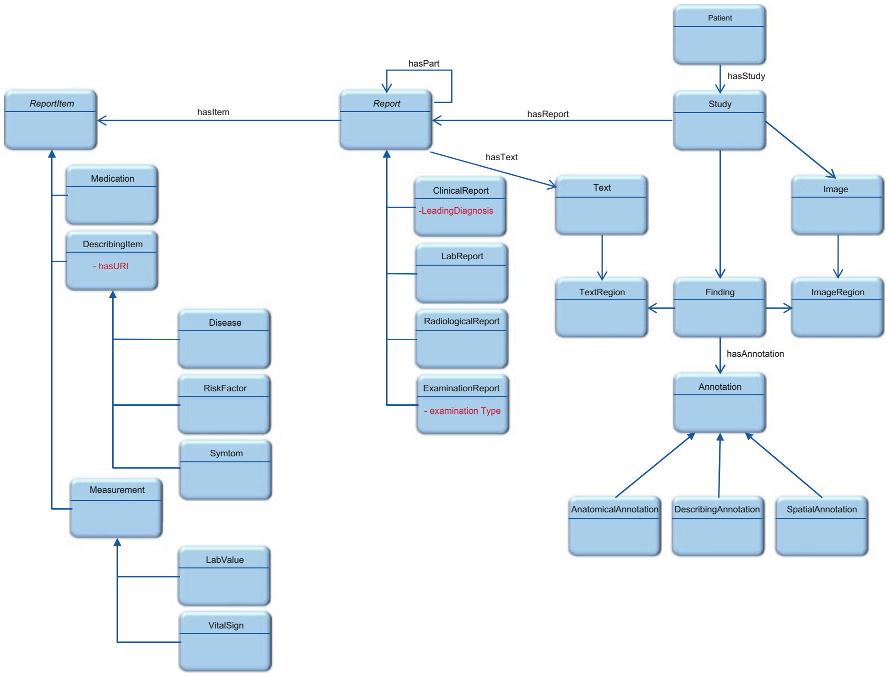


Fig. 6 Schema of the annotation ontology

- *Disease progression*: Changes to anatomy due to pathology over time should be represented. A combined examination of studies with their pre-studies needs temporal relations.
- *Multi modality*: Diagnosis often needs a synoptic view of images acquired with different modalities, e.g., CT, MRI, US. Therefore, the underlying annotation ontology should link annotations not only across time, but also across different modalities.
- *Extensibility*: In order to adopt hospitals' preferred wording, the repository of used ontologies should be extensible, e.g., some of the hospitals have already had experience with SNOMED CT. Therefore, the annotation scheme should not only incorporate RadLex and FMA, but also support further ontologies.

An image region, represented by the OWL-class `ImageRegion`, is an arbitrarily shaped spatial sub-image, which is defined as a landmark point, a triangulated mesh, or an image mask. The triangulated meshes are currently used to describe organs, detected by the image parsing system, and image masks to define scribbles.

The OWL class `Finding` relates anatomical annotations, such as liver or spleen, with the anatomy-qualifying and -describing annotations, such as enlarged, hypodense, or jagged margin. Currently, FMA and the anatomical tree of RadLex are used to define the anatomy and the imaging observation and visual modifier trees of RadLex for description. If an anatomical term of a finding is missing in the existing vocabulary, spatial annotations allow the user to paraphrase it with spatial relations, such as `nearTo` or `inBetween`, e.g., the lymph node near to renal hilus. If the finding is a specific area or volume, we can add a `Measurement` to store the values of the parameter. All other additional information can be archived by `FreeText`. To free the user from selecting the right anatomy term, we added a query expansion mechanism which recursively infers subclasses in FMA.

For non-imaging values, such as texts or lab values, the *Annotation Ontology* provides OWL-classes to handle the following types of data:

- `ClinicalReport`: supports the final report from hospital, also called doctor's letter; it is structured in several parts comprising the sections: RadiologicalReport, LabReport, RadiologicalReport and ExaminationReport.
- `LabReport`: supports 110 lab values, e.g., Alpha-2-Macroglobulin, ALT, Ammonia, Antithrombin III, aPTT, AST, base excess, basophils, bicarbonate, bilirubin, bleeding time (in-vitro).
- `RadiologicalReport`: consisting of text-based information from the radiology department, such as indication, diagnostic findings, interpretation.
- `ExaminationReport`: provides examination results of the clinician specified as a set of RadLex and SNOMED terms. It is created outside of the radiology department. The origin of the mostly text-based data is third party systems installed at the hospital.

The comprised information is modeled using `ReportItem` OWL class, which enables the system to map diseases, risk factors, symptoms (non-imaging), as well as quantitative values from lab examinations and vital signs, such as ECG, body weight and height.

In sum, the *Annotation Ontology* establishes means to efficiently store references of all kinds of medical data (e.g. images, reports) of hospital information systems in a structured way. Thus, the Annotation Ontology constitutes an important basis for the semantic integration of all related concepts, and provides the required interfaces and database functionalities for developping new client systems with respective semantic technologies.

5 Conclusion

In this article we showed in detail, how the MEDICO technology paves the way towards integrated access to clinical patient data. Based on our experiences and feedback from clinical users within the MEDICO use case, we could identify the relevant information entities from medical images and text, as well as pre-process it for improved information access in clinical practice. The semantic pre-processing of clinical data establishes the basis for various intelligent healthcare applications that are described in Seifert et al. (2014).

References

C.P. Langlotz, RadLex: a new method for indexing online educational materials. RadioGraphics **26**(6), 1595–1597 (2006)

H. Ling, S. Zhou, Y. Zheng, M. Georgescu, B. Suehling, D. Comaniciu, Hierarchical, learning-based automatic liver segmentation, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, Anchorage, June 2008 (Institute of Electrical and Electronics Engineers (IEEE), Washington DC, 2008), pp. 1–8

K. Markó, S. Schulz, U. Hahn, Morphosaurus – design and evaluation of an interlinguabased, cross-language document retrieval engine for the medical domain. Methods Inf. Med. **44**(4), 537–545 (2005)

D. Marwede, P. Daumke, K. Markó, D. Lobsien, S. Schulz, T. Kahn, Radlex – German version: a radiological lexicon for indexing image and report information. RöFo – Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren **181**(1), 38–44 (2009)

R.E. Schapire, Y. Singer, Improved boosting algorithms using confidence-rated predictions. Mach. Learn. **37**(3), 297–336 (1999). The Eleventh Annual Conference on Computational Learning Theory

S. Seifert, A. Barbu, K. Zhou, D. Liu, J. Feulner, M. Huber, M. Suehling, A. Cavallaro, D. Comaniciu, Hierarchical parsing and semantic navigation of full body CT data, in *Proceedings of SPIE Medical Imaging*, Lake Buena Vista, Mar 2009, vol. 7259. The SPIE Digital Library

S. Seifert, M. Hammon, M. Petri, H. Oberkampf, P. Daumke, Intelligent healthcare applications, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

Z. Tu, Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering, in *10th IEEE International Conference on Computer Vision (ICCV '05)*, Beijing, Oct 2005, vol. 2 (Institute of Electrical and Electronics Engineers (IEEE), Washington DC, 2005), pp. 1589–1596

Z. Tu, X. Zhou, A. Barbu, L. Bogoni, D. Comaniciu, Probabilistic 3D polyp detection in CT images: the role of sample alignment, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, New York, vol. 2 (Institute of Electrical and Electronics Engineers (IEEE), Washington DC, 2006), pp. 1544–1551

Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, D. Comaniciu, Fast automatic heart chamber segmentation from 3D CT data using marginal space learning and steerable features, in *IEEE 11th International Conference on Computer Vision (ICCV)*, Rio de Janeiro, Oct 2007 (Institute of Electrical and Electronics Engineers (IEEE), Washington DC, 2007), pp. 1–8

S. Zillner, D. Sonntag, Aligning medical ontologies by axiomatic models, corpus linguistic syntactic rules and context information, in *24th International Symposium on Computer-Based Medical Systems (CBMS)*, Bristol, June 2011. (Institute of Electrical and Electronics Engineers (IEEE), Washington DC, 2011), pp. 1–6

Intelligent Healthcare Applications

Sascha Seifert, Matthias Hammon, Marisa Petri, Heiner Oberkampf, and Philipp Daumke

Abstract Currently three client systems are provided by MEDICO: the stationary clinical radiological workplace which consists of a semantic annotation prototype to be used as a semantic reporting tool for the radiologists in-daily routine; the semantic search prototype; and a mobile system. The semantic annotation tool is designed to enable the physician to validate and complete the automatically generated semantic annotations, whereas the semantic search prototype is to be used to ensure a diagnosis by searching for similar cases in medical databases. The mobile application (Usage of mobile applications for medical purposes might be restricted in several countries.) runs on the iPad and connects to the *Semantic Server* of MEDICO, which enables the radiologist to quickly view and manipulate the semantic annotations anywhere, e.g., at the bedside, using touch screen gestures and speech.

S. Seifert (✉)
Siemens AG, Erlangen, Germany
e-mail: saschaseifert@siemens.com

M. Hammon
University Hospital Erlangen, Erlangen, Germany
e-mail: matthias.hammon@uk-erlangen.de

M. Petri
Ludwig Maximilian University Munich, Munich, Germany
e-mail: petri@dbs.ifi.lmu.de

H. Oberkampf
Siemens AG, Munich, Germany
e-mail: heiner.oberkampf.ext@siemens.com

P. Daumke
Averbis GmbH, Freiburg, Germany
e-mail: daumke@averbis.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_27,

1 Introduction

Medical image annotation is a very common task at hospitals today. Radiologists read medical images acquired at the hospital and report every peculiarity in the context of the current examination. The result is normally a natural language text, but there are also trends towards a structured way of reporting with upcoming standards such as *DICOM Structured Reports*. However, structured reporting is not well accepted by clinicians yet, due to higher efforts to deal with forms in comparison with conventional natural language reporting. MEDICO proposes a new way, called semantic reporting, which supports the clinician in navigating the image, detailing findings, and getting context-sensitive help (Seifert et al. 2010; Sonntag et al. 2014).

2 Semantic Reporting

Semantics aims to increase the efficiency of the reading workflow by these particular five means:

- Preprocess the current case by automatically generating findings, i.e. the semantic annotations, through analyzing images and texts, such that a large part of the report is already completed, and only few additions or modifications by the radiologist are needed;
- Support the radiologist during reading with anatomy-sensitive tools to efficiently browse images, create new findings or provide anatomy-related background information from the Internet;
- Link image findings to other data sources, such as texts and lab values, and enable fast switch;
- Enable semantic search for similar pathological cases in medical databases, to ensure diagnosis; and
- Infer potential diseases and propose next therapy steps from semantic annotations.

The Radiological Workflow The principle of the semantic reporting workflow proposed by MEDICO is illustrated in Fig. 1. At the beginning, images from the patient are acquired by computer tomography (CT) or magnet resonance imaging (MRI). The images are stored in the medical image database (PACS). In parallel, MEDICO starts to analyze the images for anatomy and pathology and stores the results as semantic annotations in a *Knowledge Base* accessible by the the *Semantic Server*. If prior exams are available, the images and the text reports are analyzed and stored as well. Subsequently, when the user starts the interactive semantic reporting tool, i.e., part of the radiological workstation, the images, together with the generated semantic annotations, are retrieved from the server and presented

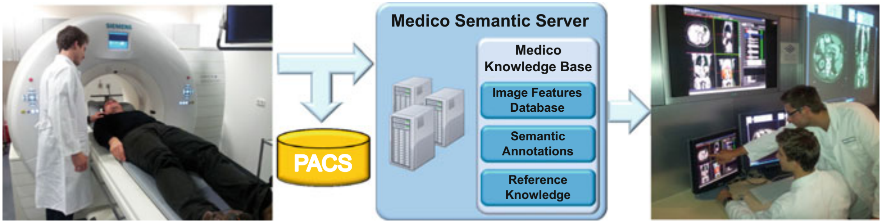


Fig. 1 The semantic reporting workflow proposed by MEDICO. The images are analyzed and the semantic results are stored in the MEDICO *Knowledge Base*, which consists of an *Image Feature Database*, dedicated to similarity search, an *Annotation Ontology* storing the semantic annotations and a database holding reference knowledge, implemented as an upper ontology on the basis of RadLex, FMA and SNOMED (Seifert et al. 2010)

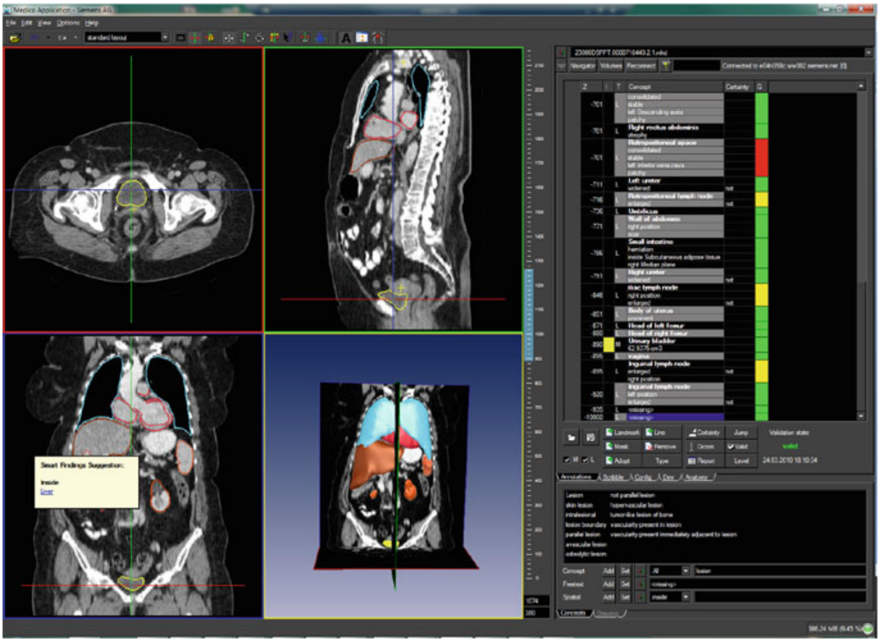


Fig. 2 The semantic annotation prototype is designed to enable the radiologist to validate and complement the semantic annotations (findings) and generate structured radiological reports

to the user (see Fig. 2). Subsequently, he can edit and extend the list of semantic annotations.

Since MEDICO parses images for the most relevant anatomical structures in oncology cases, such as solid organs and lymph node regions (see Zillner et al. 2014), the semantic reporting tool supports the user when measuring an anatomical

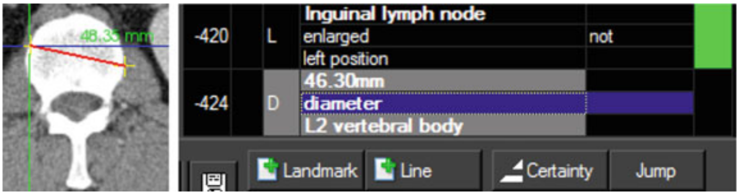


Fig. 3 Automatic labeling of measurements as a basis for disease progress diagnostics

structure in the image; e.g., for RECIST[1] by automatically locating the enclosing anatomy and storing it together with the measurement in the *Annotation Ontology*, and finally presenting it in the findings list (see Fig. 3). As an additional benefit, this allows the system to automatically associate corresponding findings (measurements) from prior exams with the current measurement by comparing the semantic information (concept and location). The measurement can then be shown in a diagram curve which lets the radiologist easily understand the progression of the disease. The progression of a measurement is one of the most important criteria for success or failure of a therapy.

In addition to the smart labeling, the radiologist will be provided with information from the Internet when creating the finding matching the considered anatomy, e.g., when examining a tumor, the TNM[2]-staging information is automatically retrieved from the Internet and presented to the user.

Alternatively, the radiologist can manually invoke the Internet information retrieval directly from the list of semantic annotations (see Fig. 4). Besides, MEDICO's semantic search (see Sect. 3) can be started, which enables the radiologist to find images matching the findings selected in the current case.

To support the radiologist, when browsing image data, the MEDICO system provides an efficient way to navigate fast within the image using an anatomy browser (see Fig. 5). This versatile tool is coupled with graphic user interface of the reporting tool and always zooms to the corresponding image region when selecting a concept in the anatomy browser.

Since MEDICO brings the semantic annotation extracted from images, texts and other data sources together in one place, intelligent applications, such as the text-to-image linking, can easily be realized. In this scenario, the parsed text will be enhanced with hyperlinks encoding the semantic information, which enable the user to directly navigate to the corresponding image location (see Fig. 4). This application is extremely beneficial for demonstration purposes at the hospitals, for

[1]Response Evaluation Criteria in Solid Tumors: standardized measuring protocol as a basis for next therapy steps.

[2]TNM Classification of Malignant Tumours (TNM) is a cancer staging system that describes the extent of cancer in a patient's body.

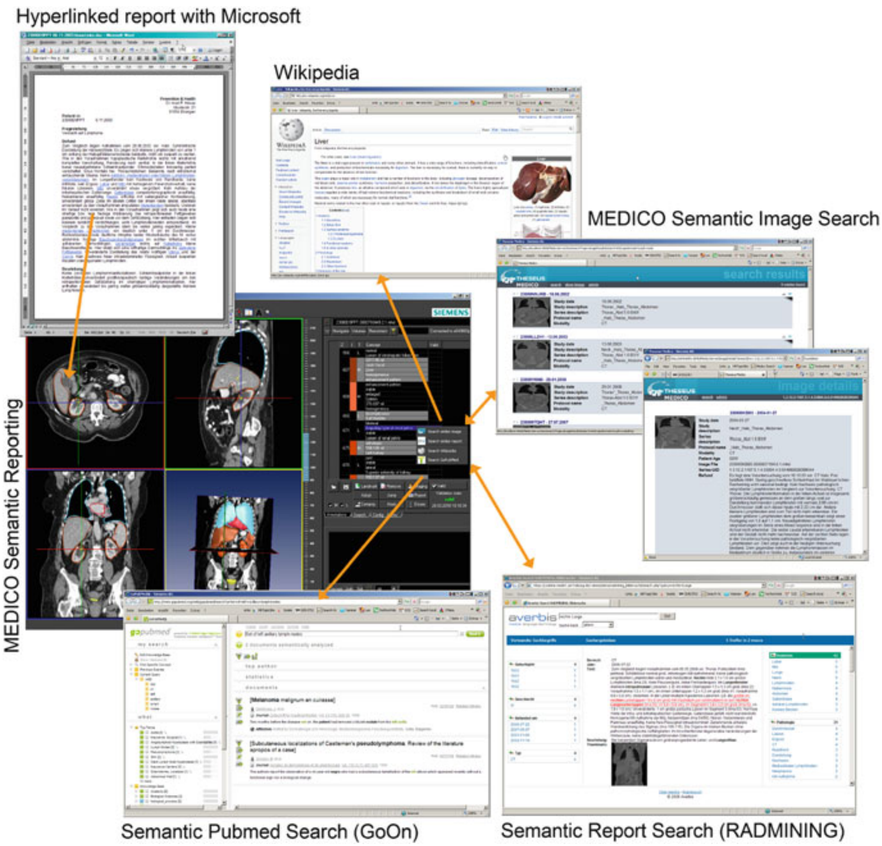


Fig. 4 The semantic reporting tool enables the radiologist to make a tailored search in an online database on the basis of the semantic annotations from the current case

the clinician, or even for the patient (Seifert 2011), to recapitulate the report received from the radiology department.

Locator Service The *Locator Service* is the technological basis for using the image content in reporting: It allows the system to query for anatomy, and to invoke the right actions given an image location, e.g., creating an anatomical label for a finding. This is done by mapping 3D locations in the image to anatomical concepts from ontology. Currently four locators are available:

- *BodyPartLocator* provides rough body region information such as head, thorax, abdomen, pelvis, extremities. This locator makes use of the landmark detection from the image parsing system to infer the body region (Seifert et al. 2009).
- *OrganLocator* uses the precise organ segmentation from image parsing to check, if a point is located inside or outside of an organ. The decision is made on the basis of a ray-intersects-mesh test.

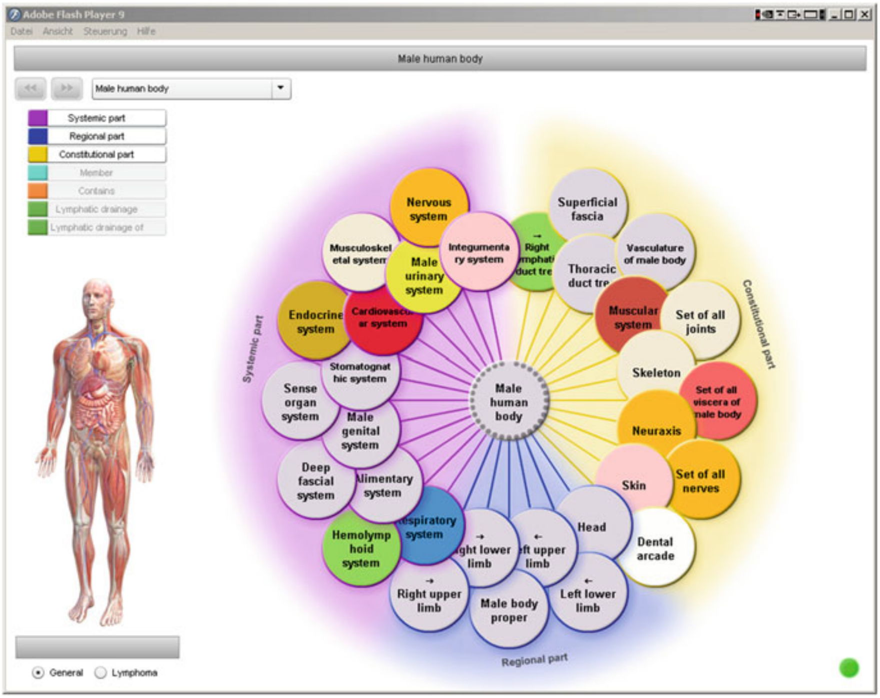


Fig. 5 Anatomy can easily be browsed using a web-based viewer; this viewer shows the concepts and relations of the FMA. The colors and icons have been carefully chosen according to physicians' custom, e.g. anatomical objects from the cardiovascular system is always visualized in *red*

- *LymphNodeRegionLocator* uses Thin-Plate-Spline registration combined with a detected landmarks approach to create an atlas for lymph node regions (Feulner et al. 2010).
- *VesselLocator* determines if the image location is inside a vessel by the use of the Euclidian distance to the center line of vessels.

A managing instance provides a unique API, controls the lifecycle of the locators, queries the individual locators and aggregates the resulting anatomical concepts. The aggregation works in several modes, such that depending on the current context, concepts for all hit anatomical structures, the most enclosing ones or the most relevant ones for examination are returned, e.g., for cardiology cases vessels are ranked higher than for oncology cases.

3 Semantic Image Search

Current medical image databases available at the hospitals today do not allow the user to search for the images' content. Therefore, the mass of images acquired daily in hospitals cannot be reused for diagnostics of other patients. However, with

Fig. 6 The browser-based user interface for the semantic search. A sample query "Search for lesions inside the spleen" is given by the user. The input field shows a list of allowed terms while typing. The resulting findings for three matching patients are listed below

MEDICO this picture changes significantly. The intelligent search application (see Fig. 6) combines a semantic search with annotations coming from the image analysis and the interactive *Semantic Reporting*, and an image similarity search comparing a specified reference image region (query-by-scribble), i.e., mostly around a lesion, with image regions in the database (Seifert et al. 2011). For the semantic search the physician does not need to give the exact keywords, as in full text search, but the system expands the given input with hyponyms, synonyms, and even with terms in other languages.

Combining these two mechanisms, the query language is tremendously extended compared to classical Content-Based Image Retrieval systems (CBIRs). The objective in MEDICO is to answer complex queries such as:

1. *Find images displaying the pelvis.*
   The query solely uses image annotations and is executed with the *Annotation Database* (query-by-concept) by the means of SPARQL.[3]

---

[3] A query language for semantic databases.

2. *Find images displaying enlarged thoracic lymph nodes.*
   It is an extension of query 1, but with an expansion of the concept thoracic lymph nodes. By the use of the reference ontologies, this results in multiple queries for:

   Thoracic lymph node → Mediastinal lymph node
   → Pretracheal lymph node
   → Esophageal lymph node, ...

3. *Find patients with similar lesions in the liver, and with thoracic lymph nodes enlarged.*
   This query searches for images containing similar regions based on the visual appearance. This is a close approach to CBIR. The main advantage of our system is that the results can be restricted by the *query-by-concept* search (here: *enlarged thoracic lymph nodes*), which limits the results to lesions within the liver. This results in an elimination of 12 % of the top ten hits which would arise without taking the semantic context into account.
4. *Find patients with thickened wall somewhere in the intestine and low hemoglobin status.*
   Since a clinical decision is rarely only image-based, the semantic search also supports lab value queries. This example also shows the query expansion capability, since intestine is here not expanded like thoracic lymph nodes into sub-types, but into different sections: *rectum*, *cecum, sigmoid*, etc.

The idea behind the similarity search is to deduce information about a query object by analyzing the most similar objects within a database of objects available in the same modality. In the case of image similarity search, these objects represent either complete images, or regions of interest defined within an image. Subsequently, the two use cases for image similarity search are detailed.

Query by Scribble Query by scribble is a classical CBIR scenario, allowing the user to quickly outline a three-dimensional Region Of Interest (ROI) within an image volume for which to retrieve similar image annotations. The quick ROI specification, called a scribble, is used to not spend too much time on defining the exact, three-dimensional border of the object of interest. An example is displayed in Fig. 7, where the user marked a lesion within the liver. The four most similar lesions and the four most dissimilar lesions in the MEDICO database are displayed to the right. The algorithm was tested on 111 liver lesions which were manually annotated with pair-wise similarity scores by a medical expert. The resulting scoring matrix could be reliably reproduced by the retrieval pipeline, which can be shown by a mean average precision of 0.78 and an average normalized discounted cumulative gain statistic (nDCG) (Järvelin and Kekäläinen 2002) of the first 10 retrieved lesions of 0.85 (Seifert et al. 2011).

ROI Retrieval Queries As a second use case of image similarity search, MEDICO provides a novel retrieval system, which enables the user to specify an ROI to be loaded from a CT scan without knowing the exact pixel coordinates. Current systems usually only support the retrieval of complete CT scans. Depending on the

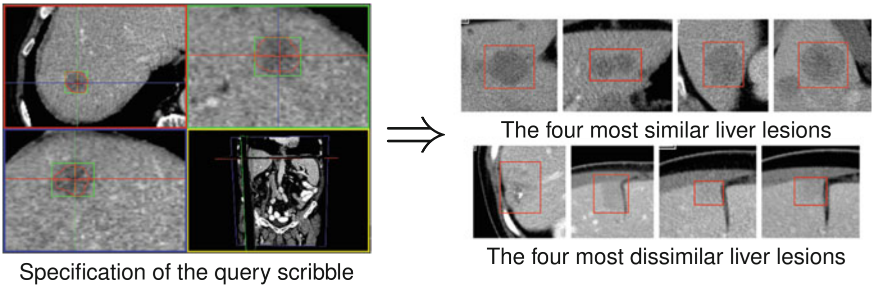


Fig. 7 The query by scribble scenario: (*left*) a scribble is selected in three 2D projections also featuring a 3D contextual view. The center slices of the four most similar liver lesions are listed to the *top right*, the four least similar liver lesions are shown at the *bottom right*

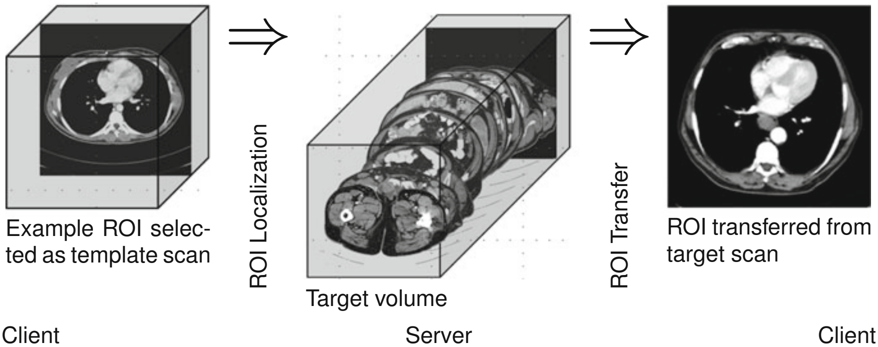


Fig. 8 ROI retrieval by example: the user specifies an ROI in an opened template scan. The system automatically locates the equivalent region within the target scan and transfers only the requested ROI

size of such a scan, the retrieval times required for loading a complete CT scan vary between a few seconds to a couple of minutes. This can be an important hindrance to the clinical routine, especially if the network is already strained from multiple parallel retrieval queries of other users. In the MEDICO framework, an ROI for a partial volume can be specified either as an anatomical concept like an organ or as an ROI in an example scan, which is already opened. The latter case is illustrated in Fig. 8 (Cavallaro et al. 2011).

When the ROI is queried by naming an anatomical concept, the framework first determines whether or not this concept has already been annotated in the semantic annotation layer for the given volume. If so, the coordinates of the ROI to be loaded are already known to the system, and they can be directly used. Else, the framework first needs to locate the ROI within the requested scan. The MEDICO system solves this problem by using an efficient atlas approach which can project any CT scan onto a standardized height scale. Using advanced methods of similarity regression, we can even map a single slice into this standardized height space, thus allowing for very flexible ways of query specification (Emrich et al. 2010; Graf et al. 2011).

4 Disease-Symptom Mapping

Clinical patient data, such as medical images and reports, establish the basis of the diagnostic process. In order to improve the access to heterogeneous and distributed clinical patient data, the MEDICO use case concentrates on extracting semantic annotations with links to concepts of medical ontologies, such as RadLex, FMA, SNOMED CT or others. These annotations capture descriptive information of clinical data as the observations made, the findings discovered and the various symptoms identified. For using this information in differential diagnosis, a relation of findings to diseases is needed. A clinician looks for findings indicating certain diseases he suspects the patient might have – for instance he looks for *cancer-indicating* findings or symptoms in the patient record. So far. this was only possible through search for specific finding as “enlarged mediastinal lymph nodes” or “enlarged spleen”. Similarly to the cognitive decision process of clinicians, who rely on their experience and expertise to interpret the findings with respect to diagnosis, the interpretation of annotations using medical background knowledge is needed. We built an initial ontology containing lymphoma-related diseases and symptoms, as well as a prototype implementation (Oberkampf et al. 2012). The created ontology and a ranking algorithm provide the basis to infer likely diseases of patients based on annotations. Further, a ranking of open symptoms, which have to be checked, and respective examinations are implemented.

Disease-Symptom Ontology Existing medical ontologies provide a comprehensive and well-structured vocabulary; however they do not have enough relations from symptoms to other relevant concepts such as diseases or examinations. In particular, existing ontologies cannot be used to infer likely diseases or plan further examinations based on annotations describing findings. We fill the missing link by proposing a disease-symptom-ontology that contains diseases, symptoms and their relations. Further the ontology is able to represent information about leading symptoms, gender- and age-specific incidence proportion of a disease, intensity of symptoms, the conditional probability of a symptom given a disease and the relative importance of different symptoms. Knowledge about the disease-symptom relations can be found in common clinical knowledge resources such as “Innere Medizin”(internal medicine) (Herold 2013), where about 300 diseases are listed and described.

Ranking and Prototype Given annotated clinical data of a patient we use the disease-symptom ontology to extract an initial set of symptoms, where an inference component ensures that also implicit symptoms are extracted. We then measure how well the actual patient’s symptoms match the typical symptoms of a given disease. The measure is based on the classical *f-measure* from information retrieval, capturing precision and recall. We extended this measure to capture also the

Fig. 9 Prototype implementation with a ranked list of likely diseases and a graphical overview of their present, open and absent (leading-) symptoms

information about leading-symptoms, gender- and age-specific incidence proportion of a disease, symptom intensity, and the relative importance of the different symptoms. A graphical visualization shows the absolute and relative amount of present symptoms and leading symptoms for each likely disease. The stacked bar graph represents the amount of present (red), open (grey) and absent (green) symptoms for each disease (Fig. 9). An interaction component allows us to change the status of open symptoms to present or absent with direct reflection in the ranking. The graphical visualization, and especially the possibility of getting an overview of open symptoms, helps to plan further examinations.

5 Semantic Search in Reports

Radiology reports are a valuable source of information, for both supporting clinical diagnosis, as well as research and education. Daily, radiologists want to validate diagnoses and look for "best practice" reports. In research and education, radiologists need to collect patient cohorts or detect representative use cases for teaching. To realize all requirements in one single search tool, a system that achieves both high correctness (precision) and high completeness (recall) of search results for a given query is needed.

At a first glance, the problem of finding relevant content in radiology reports seems to be easily solved. Entering a query term in a search engine, a user would

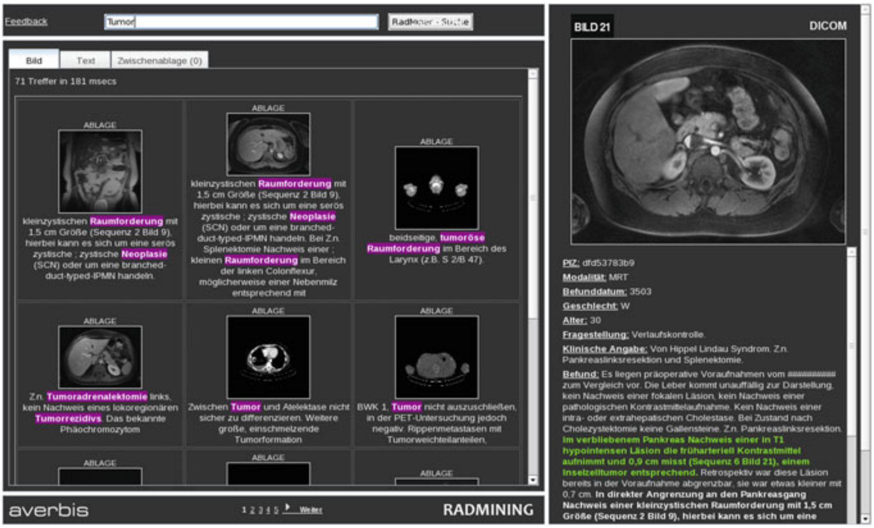


Fig. 10 Semantic search in radiology image descriptions. The search term '*Tumor*' leads to a number of results containing synonyms of the search term (e.g. '*Raumforderung*'). Text and images are linked on a sentence level via the image reference. On the *right side*, a full preview of a search result is shown including the full radiology report. The sentence containing the reference to the image is highlighted in *green*

assume to obtain a precise and complete search result. However, with this approach even the simplest queries yield poor results, as explained in the following. For instance, the query '*colon tumor*' will result in a number of false positive results containing, e.g., the phrase '… *The image shows a pancreas tumor. A colon polyp was detected as additional finding* … ' or negations such as '… *no evidence for colon tumor*'. This means that a simple free text search is not always very precise. On the other hand, reports describing the same observation ('*colon tumor*') using synonyms or quasi-synonyms, e.g. '*colorectal*' instead of '*colon*' or '*tumor*' instead of '*neoplasm*', are not being returned. In other words, a simple search rarely produces complete result sets. This is where semantics come into play: a semantic search is able to search for synonyms and child terms of a given query, e.g. search for '*tumor*' incudes search for '*lymphoma*'. Negated phrases are detected and optionally excluded from search. A semantic search is able to propose searches that are in some way similar to the entered search. These are just a few features of semantic search.

In radiology, there is another peculiarity. Reports are always associated to images they refer to. Often, radiologists mention a "key image" out of a series of images that can best show a given diagnosis, e.g. '… *shows a known tumor* (IMA 18) … '. This information is particularly useful for linking text passages to single images and thus for creating a semantic link between reports and images.

In THESEUS, a semantic radiology search engine called RADMINER was developed, realizing various features of semantic search (see Fig. 10). The system carries out a comprehensive analysis of both queries and text (see Zillner et al. 2014). A search for '*colon tumor*' includes searches for '*colorectal tumor*', '*colon cancer*' or '*neoplasm in the large intestine*'. In the search result, both the relevant text passage and the corresponding key image are shown. When clicking on a result, a more comprehensive preview of image and report is shown on the right side of the screen. With one click, the user can open a DICOM viewer to see the whole DICOM image. One has access to previous and follow-up examinations in order to compare the progress of a finding. A cache allows us to store characteristic findings for later reuse.

The system proved to achieve higher recall and precision with respect to the top n results compared to ordinary search implementations (Daumke et al. 2010). It supports radiologists with a fast and sure diagnosis at the workplace. With RadMiner, presumptive diagnoses can be validated easily and effectively by means of comparison with similar cases, saving valuable time and resulting in improved medical care of patients. RadMiner not only provides for the creation of patient collectives, but is also an important prerequisite for comprehensive data mining methods with which new medical findings can be made.

6 Conclusion

Intelligent healthcare applications powered by MEDICO's semantic technology canl change the way radiologists are working today. Several examples were presented which make clear how semantics may improve efficiency in reporting, enable more precise diagnosis by comparing current cases with similar cases found by semantic image search, and help with therapy decisions by disease-symptom reasoning. Finally, the relevance of semantics in searching reports was demonstrated. As a consequence of using semantics for the next generation of healthcare applications, not only the physician's work will change dramatically, but also the interoperability of hospital systems will change when in the future systems are exchanging knowledge based on Semantic Web standard, instead of just data as today.

References

A. Cavallaro, H.P. Kriegel, M. Schubert, M. Thoma, Region of interest queries in CT scans, in *Advances in Spatial and Temporal Databases*, ed. by G. Goos, J. Hartmanis, J. van Leeuwen. Volume 6849 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011), pp. 56–73

P. Daumke, S. Holste, S. Ambroz, M. Poprat, K. Simon, D. Marwede, E. Kotter, Informationsextraktion aus radiologischen Befundberichten. Technical report, Universitätsklinikum Freiburg, Universitätsklinikum Leipzig, Averbis, 2010

T. Emrich, F. Graf, H. Kriegel, M. Schubert, M. Thoma, A. Cavallaro, CT slice localization via instance-based regression, in *Proceedings – SPIE the International Society for Optical Engineering*, vol. 7623 (SPIE, Bellingham, San Diego, 2010)

J. Feulner, S. Zhou, M. Huber, J. Hornegger, D. Comaniciu, A. Cavallaro, Lymph node detection in 3-D chest CT using a spatial prior probability, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '10)*, San Francisco, June 2010 (Institute of Electrical and Electronics Engineers (IEEE), Washington DC, 2010), pp. 2926–2932

F. Graf, H. Kriegel, M. Schubert, S. Pölsterl, A. Cavallaro, 2D image registration in CT images using radial image descriptors, in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, Toronto. Volume 6892 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011), pp. 607–614

G. Herold, Innere Medizin, 2013. http://www.herold-innere-medizin.de/index.htm?information.htm&0

K. Järvelin, J. Kekäläinen, Cumulated gain-based evaluation of IR techniques. ACM Trans. Inf. Syst. **20**(4), 422–446 (2002)

H. Oberkampf, S. Zillner, B. Bauer, Interpreting patient data using medical background knowledge, in *Proceedings of the International Conference on Biomedical Ontologies (ICBO '12)*, ed. by R. Cornet, R. Stevens. Volume 897 of CEUR Workshop Proceedings. CEUR-WS.org, 2012. http://dblp.uni-trier.de/db/conf/icbo/icbo2012.html#OberkampfZB12

S. Seifert, Gesund sein und bleiben - Der intelligente Arztbericht und das elektronische Rezept, in *Zukunftsbilder der digitalen Welt. Nutzerperspektiven im internationalen Vergleich*, 4th edn. (BMWI, Munich, 2011), pp. 186–197

S. Seifert, A. Barbu, K. Zhou, D. Liu, J. Feulner, M. Huber, M. Suehling, A. Cavallaro, D. Comaniciu, Hierarchical parsing and semantic navigation of full body CT data, in *Proceedings of SPIE Medical Imaging*, Lake Buena Vista, vol. 7259, Mar 2009. The SPIE Digital Library

S. Seifert, M. Kelm, M. Möller, S. Mukherjee, A. Cavallaro, M. Huber, D. Comaniciu, Semantic annotation of medical images, in *SPIE Medical Imaging: Advanced PACS-Based Imaging Informatics and Therapeutic Applications*, San Diego, vol. 7628, Mar 2010. The SPIE Digital Library

S. Seifert, M. Thoma, F. Stegmaier, M. Hammon, M. Kramer, M. Huber, H.P. Krieger, A. Cavallaro, D. Comaniciu, Combined semantic and similarity search in medical image databases, in *SPIE Medical Imaging: Advanced PACS-Based Imaging Informatics and Therapeutic*, Lake Buena Vista, Mar 2011. The SPIE Digital Library

D. Sonntag, S. Zillner, P. Ernst, C. Schulz, M. Sintek, P. Dankerl, Mobile radiology interaction and decision support systems of the future, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (eds.), *Towards the Internet of Services: The THESEUS Research Program* (Springer, Berlin/Heidelberg/New York, 2014)

S. Zillner, S. Seifert, M. Erdt, P. Daumke, M. Kramer, Semantic processing of medical data, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

Mobile Radiology Interaction and Decision Support Systems of the Future

Daniel Sonntag, Sonja Zillner, Patrick Ernst, Christian Schulz, Michael Sintek, and Peter Dankerl

Abstract Clinical care and research increasingly rely on digitized patient information. There is a growing need to store and organize all patient data, including health records, laboratory reports, and medical images. Medical images have become indispensable for detecting and differentiating pathologies, planning interventions, and monitoring treatments. The effective retrieval of images builds on the semantic annotation of image contents and intelligent interaction with the image material. The semantic annotation of image contents has an automatic and a manual component. In our work, we heavily rely on automatic organ, tissue, and disease detection, which represents one of the main technical research questions in MEDICO. In this article, however, we will focus on intelligent interaction with the image material, i.e., what mobile radiology interaction and decision support systems of the future, based on automatic detectors, may look like.

D. Sonntag (✉) • C. Schulz
DFKI GmbH, German Research Center of Artificial Intelligence, Saarbrücken, Germany
e-mail: Daniel.Sonntag@dfki.de; Christian.Schulz@dfki.de

S. Zillner
Siemens AG, Munich, Germany
e-mail: sonja.zillner@siemens.com

P. Ernst
Max Planck Institute for Computer Science, Saarbrücken, Germany
e-mail: pernst@mpi-inf.mpg.de

M. Sintek
DFKI GmbH, German Research Center of Artificial Intelligence, Berlin, Germany
e-mail: Michael.Sintek@dfki.de

P. Dankerl
University Hospital Erlangen, Erlangen, Germany
e-mail: Peter.Dankerl@uk-erlangen.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_28,

1 Introduction

With traditional user interfaces (see MITK, Cerner Millennium, Philips XIRIS, or Siemens *syngo*.via), users may browse or explore visualized patient data, but little to no help is given when it comes to the interpretation of what is being displayed. Semantic annotations should provide the necessary image information, and semantic human computer interaction (HCI) systems should be used to ask questions about the image annotations while engaging the clinician in a natural speech dialog. In addition, the semantic background model should be expressive enough to assist the clinician in the patient finding process by taking an active role in his or her decision process.

Radiology workstations of the future, which are based on a semantic patient search as implemented in MEDICO, should therefore rely on HCI systems that

- Reduce turnover times and annotation errors; and
- Facilitate structured reporting.

The problem with current HCI technology is that a clinician cannot directly create a structured report while scanning the images: in this eyes-busy setting, he or she can only dictate the finding to a tape recorder. After the reading process, he or she can replay the dictation to manually fill out a patient's finding form or delegate other personnel. But since the radiologist has to check the form again, task delegation does not save time. In addition, automatic detector results have to be verified or corrected in the most convenient way for which no HCI concepts exist in the research literature. In addition, the implementation of advanced medical applications requires means for the automated post-processing of medical image annotations. But it often remains unclear which steps of the automatic procedures should be presented to the user, and which decisions should be supervised by the clinician in the form of dialog feedback.

Mobile radiology interaction and decision support systems of the future should therefore address these problems by, first, extending the (only) desktop-based reporting practice, and second, incorporation of automatic inference services in order to infer previously unknown and relevant medical knowledge about the contents of the patient images.

We address these issues by making possible dialog-based radiology image reporting (Sect. 2) and image metadata reasoning processes for improved clinical decision support by automatic patient staging. The stage of a cancer is a description (usually numbers I–IV, with IV having more progression) of the extent to which the cancer has spread (Sect. 3). A further ontology modeling approach about the "Semantic Heart" is also presented as one example of future directions in knowledge representation towards our goals (Sect. 4). Section 5 provides a conclusion.

2 RadSpeech

RadSpeech[1] is the design and implementation of a multimodal dialog system for structured radiology reports. Towards this goal, several incremental developments have been made. In order to allow a radiologist to annotate special image regions in a database format and to search for similar cases, we developed a desktop-based manual annotation tool called RadSem (Müller et al. 2009) in the initial phase of the MEDICO use case. Anatomical structures and diseases can be annotated while using the auto-completion combo-boxes with a search-as-you-type functionality. The resulting annotation was accurate but very time-consuming. In addition, RadSem did not fulfill the special requirement that clinicians have: to have access to a coherent view of image data within their particular diagnosis or treatment context in the radiology department while they are skimming many image series and thousands of pictures in a minute's time. Although it is widely reductive to put it this way, a senior radiologist has three main goals:

1. Access the images and image (region) annotations,
2. Complete them (by, e.g., tissue characteristics), and
3. Refine existing annotations.

The radiologist's daily task with patient image finding stations is also illustrated in Fig. 1. After evaluating RadSem in the clinical environment, we argued that these tasks can best be fulfilled while using a multimodal dialog system, and, according to that vision, in a first step, we experimented with a large touchscreen installation (Sonntag and Möller 2010). Since the results with speech interaction were very promising, we tried to opt for the mobile context in a second step. Our mobile system[2] is expected to provide the radiologist with the ability to review images when outside the laboratory, and to make a diagnosis without having to be back at the workstation.[3]

With our technology developed over the last 5 years (Sonntag 2010; Sonntag et al. 2010), we implemented the first mobile dialog system on the iPad and iPhone, which is tuned for the radiology domain and makes the complete system unique. Our solution not only provides more robustness compared to speech-to-text systems (we use a rather small, dedicated, and context-based speech grammar which is also very robust to background noise), it also fits very well into new radiology reporting processes which will be established in Germany and the U.S. over the next several years: in structured reporting you directly have to create database entries instead of text.

Two aspects are implemented and shown in the YouTube demo. First, the inspection of and navigation through the patient's data, and second, the annotation of

[1] http://www.dfki.de/RadSpeech/

[2] Usage of mobile applications for medical purposes might be restricted in several countries.

[3] http://www.youtube.com/watch?v=uBiN119_wvg

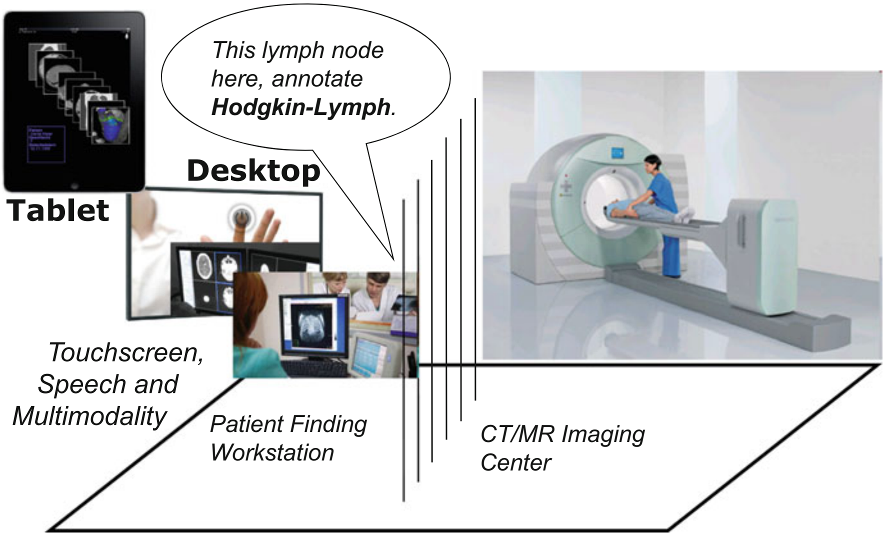


Fig. 1 Complete interaction scenario including wireless connection between the MEDICO server and the mobile RadSpeech interface

radiology images by the use of speech and gestures. We implemented a middleware that invokes the MEDICO back end server directly. This middleware is implemented by means of the ontology-based dialog platform (ODP, see Fig. 2): the central ODP is responsible for routing messages between the dialog system's connected components. Such components include a speech recognizer (ASR) and a speech synthesis (TTS) module. On the client (e.g., an iPad), only a slim application is needed to encode the speech and gesture input and receive the speech synthesis to be played. New clients can be added easily by downloading an iPad app.

Consider the following example dialog: A radiologist treats a lymphoma patient. The patient visits the doctor after chemotherapy for a follow-up CT examination. The focus of the speech-based interactions is the following sub-dialog.

1. U: "Show me the CTs, last examination, patient XY."
2. S: Shows corresponding patient CT studies as DICOM picture series and MRI images and MRI videos."
3. U: "Annotate this picture with 'Large Cell Lymphoma' (+ pointing gesture) and 'shrunken' (+ pointing gesture)."
4. S: Shows the new annotations on the image and confirms a database update.

With the help of such dialog interaction scenarios, RadSpeech strives for more efficiency during the medical finding process. By using ontological semantic knowledge of anatomy, diseases, and disease characteristics, the second goal of more structured finding reports including semantic image annotations can be pursued. The semantic dialog system should be used to annotate images and ask questions

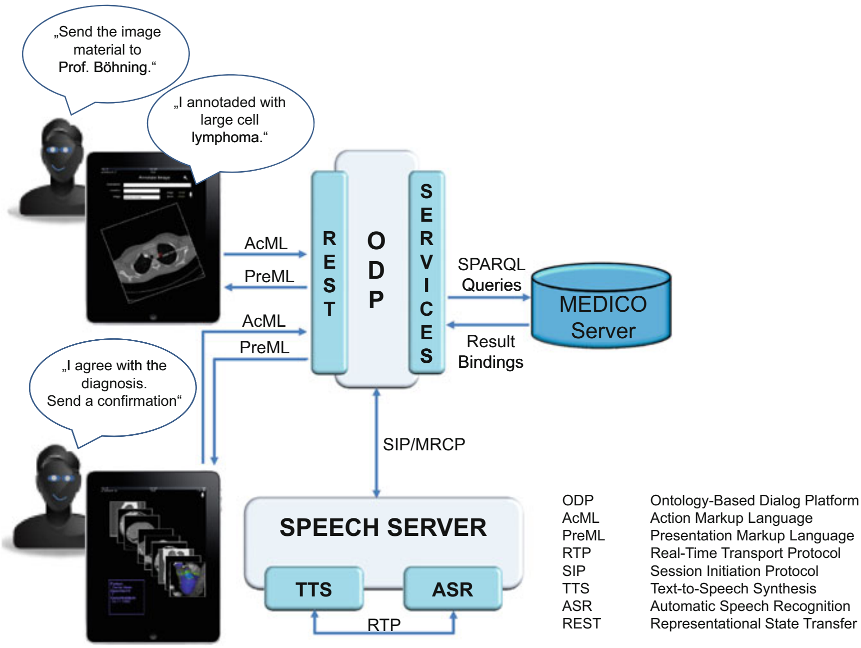


Fig. 2 RadSpeech's technical architecture including the MEDICO server access

about the image annotations while engaging the clinician in a natural speech dialog. Different semantic views of the same medical images (such as structural, functional, and disease aspects) can be explicitly stated, integrated, and asked for. This is the essential part of the knowledge acquisition process the clinical users are involved in. We extended this idea to diagnostic quality control and user interaction.

3 Diagnostic Quality Control and Multimodal User Interaction

The goal of diagnostic quality control, as envisioned in MEDICO, is to use medical image content information for the automated staging of lymphoma patients. We concern ourselves with the inference process of how explicit descriptive knowledge about a particular image finding (e.g., the knowledge about the number of enlarged lymph nodes) relates to the patient degree and can be inferred automatically.

The staging information is paramount when clinicians assess an individual patient's progress and decide on subsequent treatment steps. The automatic staging of lymphoma patients relies on automatically detected image metadata (Seifert et al. 2010), and combines it with additional clinical knowledge. We use the formalized Ann Arbor staging system (Wittekind et al. 2005), which recognizes four different lymphoma stages. Stage I indicates that the cancer is located in a single region, and Stage II that it is located in two separated regions confined to one side of the diaphragm.[4] Stage III denotes that the cancer has spread to both sides of the diaphragm, and Stage IV shows diffuse or disseminated involvement of one or more extra lymphatic organs.

The new automatic staging use case we describe (published in Zillner and Sonntag 2012) also introduces a new medical application for the automated classification of lymphoma patients in well-defined categories. Each category, i.e., the stage of a lymphoma patient, is determined by the number, location, and distribution of lymphatic occurrences. This information is captured within image metadata which are automatically extracted by the MEDICO software and used as input to our automatic staging process. Technically, the image metadata are extracted from the DICOM headers and additional information about organs and landmarks is extracted from the image regions automatically (preprocessing by MEDICO detectors). In addition, a spatial learning algorithm can be used to generate an atlas for plausible anatomical configuration, which serves as input to an OWL-based reasoner to derive the correct patient stage (Zillner and Sonntag 2012).

Our goal is to use medical image content information for the automatic staging of cancer patients. In particular, we developed a formal and explicit representation of the lymphoma staging system that allows us to automatically classify lymphoma patients by means of existing reasoning procedures. The automatic staging results can be presented to the user, and can be confirmed by a subsequent interactive user dialog that paves the way towards improved clinical diagnosis. The dialogs presented in this section are extensions of the RadSpeech dialog system for image annotation in the context of automatic lymphoma staging.

3.1 Automated Staging of Lymphoma Patients

To derive the stage of lymphoma patients automatically, we carried out the following steps within the MEDICO use case:

- We established a staging ontology in OWL DL following the rationale of the Ann Arbor Staging System (Zillner 2010), which allows to determine the patient data classification within the reasoning process. The Ann Arbor ontology consists of a set of defined classes (OWL classes described by necessary and sufficient constraints) capturing the information that leads to a patient's lymphoma stage (i.e., the number, type(s), and the distribution of a patient's indicated lymphatic occurrences). According to the semantics of each patient class, it will be classified within the reasoner.

[4] A sheet-form-like internal skeletal muscle that extends across the bottom of the rib cage. The diaphragm separates the thoracic cavity (heart, lungs, and ribs) from the abdominal cavity.

- We transformed the patient data of the annotation procedure in Seifert et al. (2009) into an OWL representation by using the OWL API.[5]
- We identified the relevant ontology fragments and established the required ontology alignments (Sonntag et al. 2009).
- We integrated all manually created and aligned standard ontologies, i.e., the aligned medical ontology fragments, the patient ontology, and the Ann Arbor ontology, and executed the reasoning process on top of the integrated ontological model for staging. While the reasoning process enables us to classify patients by integrating knowledge captured by the external medical ontologies, the resulting ontological model captures the inferred stages of patients in an explicit form.
- The knowledge captured in the inferred model, in particular the deduced staging information, can be queried by SPARQL.[6]
- The result of the staging retrieval step can be presented to the user by using RadSpeech's dialog system; inconsistencies can be resolved.

The resulting diagnostic process behind these steps, the automatic classification and staging of patients, helps to significantly improve the quality and efficiency of clinical care: its quality can be improved by the automatic detection of noticeable, and possibly contradicting, findings that indicate which patients require a more detailed analysis by the clinicians. Efficiency can be improved by using the knowledge about patients' stages and contexts to optimize the clinical workflow, and subsequently, by automatically initiating diagnostic sessions with dedicated experts through a system-initiative dialog interaction.

3.2 Improved Quality Control by Interactive System User Dialog

Back to our CT examination dialog of a real-world scenario, where we now consider an advanced stage lymphoma patient who has already been treated with three chemotherapies using the so-called CHOP protocol. As the accomplished treatments did not help to improve the patient's health condition of our example case, he or she has been referred to a specialist hospital. In the specialist hospital, the clinician takes part in the following human-machine speech dialog:

5. U: "Show me the complete patient record of patient x."
6. S: Shows and summarizes a discharge letter that indicates lymphoma stage IV and corresponding CT images.
7. U: "Which stage's indicated?"
8. S: "Lymphoma stage IV."

---

[5] http://owlapi.sourceforge.net

[6] http://www.w3.org/TR/rdf-sparql-query/

The patient's discharge letter covers details about the last medical diagnoses as well as findings, assessments, progressions, and recent therapies. Our example patient's discharge letter indicates an Ann Arbor IV-Stage. This information can also be extracted and made available to the dialog system by automatic information extraction (IE) techniques as explained in Seifert et al. (2014) or other natural language processing (NLP) of the discharge letter.

In addition, in the specialist hospital, the patient has been screened using CT. The image material can now be used to run our automatic staging process. As a result, the two opinions can be compared and presented to the user as a system initiative dialog fragment (in case of a contradiction).

9. S: "Automated staging of patient x indicates lymphoma stage III. Please note that contradicting diagnoses have been found. Do you want to see the results again to solve the conflict now?"
10. U: "No. Suggest an expert for a second opinion."

These results (and the resulting dialog) stem from authentic clinical cases, such as the following: By automatically analyzing and annotating the medical images, 16 enlarged and pathological lymph nodes on both sides of the diaphragm have been identified. However, no indication of the involvement of extra-lymphatic organs has been noted. Relying on the formal Ann Arbor classification criteria, the ontology-based staging approach classifies patient x as Ann Arbor III-Stage. The clinician's diagnosis contradicts the results of the automated reasoning application. As the staging grade strongly influences clinicians in their sequential treatment decisions, this issue is of considerable importance for achieving effective diagnostics and treatments (and counts as a contradictory result). By highlighting such contradictory results, special clinical cases can be spotted and potential medical treatment errors can be reduced (Fig. 3). It is important to note that contradictions do not necessarily have to be resolved directly. But in many medical cases, they provide reasons to either ask for a second opinion to be considered, or indicate that a more detailed analysis and examination is required.

In principal, two different causes can lead to such a staging contradiction. Accordingly, these options can also be asked for by the dialog system to resolve the "conflict" immediately.

11. S: "Do you want to interpret the results to solve the conflict now?"
12. U: "Yes."
13. S: "Interpretation 1: The patient's health condition has improved significantly lately. The more recent findings represent the current health condition of the patient."
14. S: "Interpretation 2: The first and second results show different conclusions."

The progress of health conditions is a very important observation in clinical diagnoses. Often the changes are not yet documented explicitly in the findings. Such incidents are not exceptional, as clinicians typically avoid making definite statements about the patient's health condition; rather, they indicate the documentation of related indications.

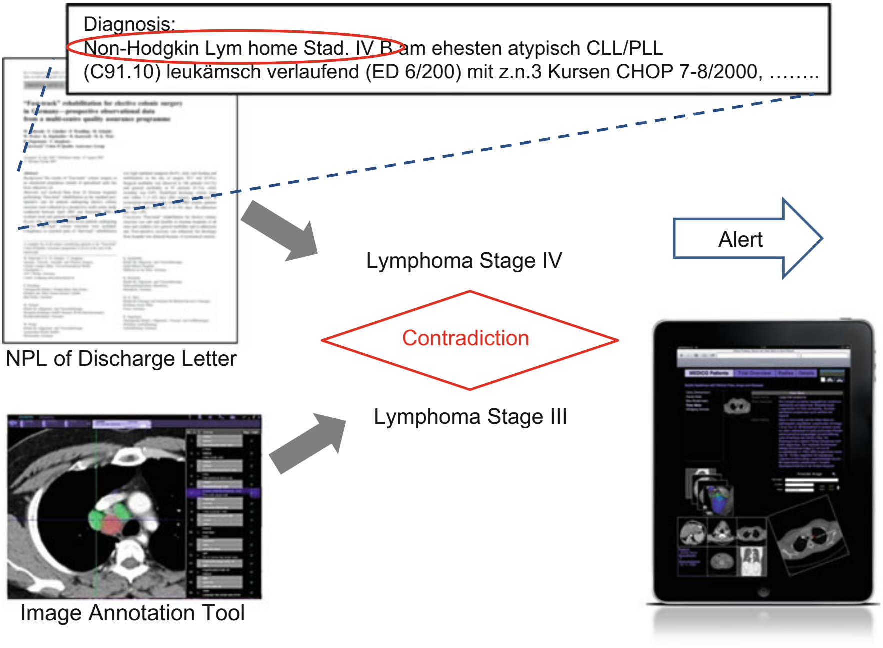


Fig. 3 Automated staging enhances the quality of diagnosis by uncovering conflicts and providing feedback

15. U: "I select interpretation 1."
16. S: Repeats staging process, displays results according to interpretation 1 and confirms a database update of the staging process.

4 Semantic Heart and Reasoning Processes

The MEDICO extension "Semantic Heart" aims to help the decision making process by creating a systematic representation and usage of knowledge about the human heart and its diseases. The project focusses on the modeling of a decision support for Acute Myocardial Infarction (AMI). AMI is defined as the death of myocardial cells. According to WHO, 65,228 deaths were caused by AMI in 2004 in Germany alone. Actually, cardiovascular diseases are the leading causes of death in the developed world.

The main goals of Semantic Heart are to develop an ontology model that explains and organizes the semantic dependencies and interactions in the context of AMI, and to develop an interactive decision support application that allows for better management of AMI, including disease prevention and earlier diagnosis. Semantic

Heart's relevant medical domains are anatomy, pathology, and pharmacology. The related knowledge model is based on four different external ontologies and terminologies:

- *Foundational Model of Anatomy*: a domain ontology representing knowledge about human anatomy. The FMA is one of the largest computer-based knowledge sources in the biomedical sciences. It contains 75,000 classes and over 120,000 terms.
- *Disease Ontology*: A well-structured ontology defining diseases, their synonyms and links to various other sources.
- *National Drug File*: A terminology describing drugs and their ingredients. It contains 19,883 clinical drugs, 7,023 ingredients, and 578 drug classes.
- *Clinical Measurement Ontology*: A structure describing around 530 different clinical measurements.

The reasoning processes, which also require user interaction, include reasoning over lab values, pharmacology, and diagnosis.

Reasoning Over Lab Values: We expressed a wide spectrum of clinical measurements in the Semantic Heart knowledge model by creating and integrating a clinical measurement ontology. The new representation had to cover all clinical measurements required during the diagnosis and treatment of myocardial infarction. Moreover, for each measurement we needed the crucial information for its application, i.e., the normal range, the consequences if a value is too high or too low, to which group of patients a specific measurement applies, etc.

Pharmacology Reasoning: We used the pharmacology knowledge defined in the National Drug File, in conjunction with information about the patient's lab values and diseases. Leveraging these connections we were able to automatically determine drug-drug interactions, drug-drug actions, and drug contra-indications. Drug-drug interaction information can be used to alert the clinician about any unsafe interactions that exist between two drugs that the patient is currently prescribed. Drug-drug action information can be used to raise alerts if two drugs have the same or similar properties, such as mechanism of action or physiological effect, that can get compounded in an unsafe manner. Drug contraindication information can generate alerts when it is suspected that the patient has a specific disease for which the given drug is an unsafe option. The prospective dialogical integration may look like this:

17. U: "This patient will be treated with drug x."
18. S: "He is already treated with drug y. There are known drug interactions for x, classification: major."
19. U: "Other drugs with similar effects as x which do not interact with drug y?"
20. S: "Drug z has similar effects as drug x with minor interaction with drug y: minimally clinically significant. However, due to the age of patient x, treatment with z should not be prescribed for more than 10 days."

Diagnostic Reasoning: Combining the former two use cases with a rule system which assigns symptoms to diseases, the smart assistant as a combination of the new alerts of diagnostic reasoning and RadSpeech's dialog system extensions should be able to determine the most probable diseases for a given set of inputs which constitute lab values, drugs, and symptoms. As a result, the back end also computes the relevancy of a specific disease, and what symptoms contributed to a particular diagnosis.

5 Conclusion

We have presented a mobile semantic speech dialog system for the radiologist. Our new prototypical dialog system will provide the radiologist with the ability to review images when outside the laboratory, to annotate important image regions while using speech and gestures, and to make a structured diagnosis without having to be back at the workstation. There is also a growing interest in the automatic processing of medical image content and the semantic integration of explicitly expressed content into clinical applications. We introduced an application for the automated staging of lymphoma patients using image metadata information. The focus of the presented work lies in the combination of intelligent user interaction with intelligent reasoning procedures, based on the image semantics. In that regard, we provided a first example of an extended human-machine speech dialog with the following three key benefits of our solution within MEDICO: (1) improve patient safety and reduce errors by having access to semantic image information using speech dialog; (2) increase operational performance with automatic reasoning for patient staging to create more efficient and effective workflows; (3) improve patient safety and reduce errors by having access to semantic image region information and a speech-based dialog system which is able to communicate conflict by system initiative and solve diagnostic conflicts in the course of the system-user dialog. We also developed an ontology model that explains and organizes the semantic dependencies and interactions in the context of AMI, which allows for a better management of AMI, including disease prevention and earlier diagnosis.

In our future work, to improve mobile radiology interaction and decision support systems of the future, we aim to formalize and integrate related staging systems, such as the TNM classification, into advanced medical dialog applications. The next integration steps for the heart ontology model are also straightforward: the reasoning about lab values, pharmacology, and diagnosis allows a dialog engineer to formulate adequate dialog-based interaction about drug-drug interaction information for example, which can be used to alert the clinician about any unsafe interactions and wait for appropriate user decisions. Our experts predict that mobile dialog technology will dominate mobile healthcare in just a few short years. The speech-based dialog system RadSpeech is currently a part of a larger clinical study about the acquisition of medical image semantics at Siemens Healthcare, the University Hospital in Erlangen, and the Imaging Science Institute (ISI).

References

M. Müller, S. Regel, M. Sintek, RadSem: semantic annotation and retrieval for medical images, in *The Semantic Web: Research and Applications*. Volume 5554 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2009), pp. 21–35

S. Seifert, A. Barbu, K. Zhou, D. Liu, J. Feulner, M. Huber, M. Suehling, A. Cavallaro, D. Comaniciu, Hierarchical parsing and semantic navigation of full body CT data, in *Proceedings of SPIE Medical Imaging*, Lake Buena Vista, vol. 7259, Mar 2009. The SPIE Digital Library

S. Seifert, M. Hammon, M. Petri, H. Oberkampf, P. Daumke, Intelligent healthcare applications, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

S. Seifert, M. Kelm, M. Möller, S. Mukherjee, A. Cavallaro, M. Huber, D. Comaniciu, Semantic annotation of medical images, in *SPIE Medical Imaging: Advanced PACS-Based Imaging Informatics and Therapeutic Applications*, San Diego, Mar 2010, vol. 7628. The SPIE Digital Library

D. Sonntag, *Ontologies and Adaptivity in Dialogue for Question Answering*. Volume 4 of Studies on the Semantic Web (AKA/IOS Press, Heidelberg, 2010)

D. Sonntag, M. Möller, Prototyping semantic dialogue systems for radiologists, in *Proceedings of the 2010 Sixth International Conference on Intelligent Environments (IE '10)*, Kuala Lumpur (Institute of Electrical and Electronics Engineers (IEEE), Washington DC, 2010), pp. 84–89

D. Sonntag, N. Reithinger, G. Herzog, T. Becker, A discourse and dialogue infrastructure for industrial dissemination, in *Spoken Dialogue Systems for Ambient Environments*. Volume 6392 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2010), pp. 132–143

D. Sonntag, P. Wennerberg, P. Buitelaar, S. Zillner, Pillars of ontology treatment in the medical domain. J. Cases Inf. Technol. **11**(4), 47–73 (2009)

C. Wittekind, H. Meyer, F. Bootz (eds.), *TNM Klassifikation Maligner Tumoren*. UICC International Union Against Cancer, 6th edn. (Springer, Berlin/Heidelberg/New York, 2005)

S. Zillner, Reasoning-based patient classification for enhanced medical image annotation, in *The Semantic Web: Research and Applications*, ed. by L. Aroyo, G. Antoniou, E. Hyvönen, A. Teije, H. Stuckenschmidt, L. Cabral, T. Tudorache. Volume 6088 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2010), pp. 243–257

S. Zillner, D. Sonntag, Image metadata reasoning for improved clinical decision support. Netw. Model. Anal. Health Inform. Bioinform. **1**(1), 37–46 (2012)

Linguistics to Structure Unstructured Information

Günter Neumann, Gerhard Paaß, and David van den Akker

Abstract The extraction of semantics of unstructured documents requires the recognition and classification of textual patterns, their variability, and their interrelationships, i.e., the analysis of the linguistic structure of documents. Being the integral part of a larger real-life application, this linguistic analysis process must be robust, fast and adaptable. This creates a big challenge for the development of the necessary linguistic base components. In this drill-down, we present several dimensions of this challenge and show how they have been successfully tackled in ORDO.

1 Introduction

According to Thibodeau (2010) "data is expected to grow by 800 % over the next 5 years, and 80 % of it will be unstructured data." Methods for extracting useful information from such unstructured sources are still dominated by simple word-level strategies, based either on simple linguistic operations, or on no linguistics at all. They mostly just pick up the isolated words, and consequently apply some word-based statistics to generate useful information. For coarse-grained applications, like

G. Neumann (✉)
DFKI GmbH, German Research Center of Artificial Intelligence, Saarbrücken, Germany
e-mail: Guenter.Neumann@dfki.de

G. Paaß
Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS), St. Augustin, Germany
e-mail: gerhard.paass@iais.fraunhofer.de

D. van den Akker
Attensity Europe GmbH, Saarbrücken, Germany
e-mail: david.van.den.akker@attensity.com

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_29,

word-based indexing, word-based spelling correction, and word-cloud generation, such techniques have been proven to be satisfactory.

It has been observed, however, that the higher the demands are for the extraction of meaningful structure from the original text, the less satisfying any word-level analysis becomes. For example, in the case of the extraction of named entities or relations, it is not just the individual words that are important, but also the syntactic structures in which they are configured, as well as the functional roles which these structures have in their clause, and last but not least the semantic context in which these roles and structures appear. Consider as an exemplification an application that is designed to update a company's database with the latest revenue numbers of the competition as they get reported on the public portals that these companies typically maintain. For example, "According to IDC's Worldwide Quarterly Server Tracker, IBM led in factory revenue gains to achieve 30.5 % market share in the second quarter of 2011, compared with HP's 29.8 % share."

Actually, this integrated word-level and syntax-level approach of Natural Language Processing (NLP) has been a driving power followed in almost all application scenarios and realized in ORDO, as we will make clear in the following sections.

2 From Symbol to Sense: Applied Linguistics

Natural language produces a staggering amount of ambiguity at the levels of word sense, phrase structure and syntactic roles, but most dramatically in the transformation of structure into the meaning that the sentence conveys. Mastering the high degree of potential ambiguity is the main challenge for achieving efficient, and hence usable NLP technology.

Ambiguities may often be resolved within the sentence scope. For instance, the choice between "lie" for "tell untrue things" and for "take a horizontal position" might be resolved locally with the help of contextual clues. Sometimes though, such ambiguity types cannot be resolved within the sentence scope. Worse, other ambiguity types can never be resolved at that level, such as most pronoun ("him", "it") and common noun ("the company", "the defendant") references. Even if resolution may be achieved elsewhere in the same text, NLP is still in charge: in recent years, the extraction of cross-sentence syntactic and semantic relations has come within reach. Some encouraging results will be presented later in this article.

However, even optimal extraction of cross-sentence semantic relations will leave a residue of unresolved ambiguities. Every text contains references that cannot be understood just by reading that text. One text will assume that you know that "Arab Spring" denotes a political uprising, not a climatological season. Another will trust that you grasp the strictly metaphorical use of the term "Titanic". A third presupposes the knowledge that one can see through glasses, but not through wood. Knowledge of these pragmatic relations is required for understanding the full meaning of a text, yet they are for the large part unattainable by software today.

Broad-domain ontologies can help in bridging this gap, but much of our common knowledge happens to be stored only in our minds.

Clearly, the road from symbol to sense is long and winding. NLP is steadily moving along, though. It is able to grasp more sense today than was even considered theoretically possible 10 years ago. Three independent factors have enabled this development:

1. A revolutionary increase in affordable computational power, thanks to faster processors and the arrival of cloud computing,
2. An equally revolutionary increase in the availability of multilingual data, thanks notably to the emergence of social media on the Web,
3. Inspired by the former factors, an unprecedented focus on NLP semantics.

In the following sections, we will specify the results of this focus, starting with the underlying classical components, moving further to recent developments in automatically trainable multilingual syntactic analyzers and finally to their integration and exploitation for the recognition and extraction of domain-specific entities and relations among them. All these technologies have been exploited and used in the different application scenarios of ORDO, for example in the area of innovation monitoring or semantic analysis of web content.

3 What Linguistics Can Do for You

Automatic morphosyntactic analysis has found its way into a host of text-oriented applications. The one component not generally considered a commodity today is the analysis of phrases and syntactic functions. We will address this in some detail at the end of this section, after a brief description of the classical components.

Widely used classical components are *language identification* (determination of the source language of the input text), *tokenization* (segmentation of a document input into a stream of linguistic units – tokens, words, numbers, punctuation, etc.) and *POS-tagging* (selection of the correct part-of-speech using statistical models and/or heuristic rules). *Stemming* is the process of the identification of the stem of a given word form with a morphological analyzer and a lexicon lookup. A stem carries the core syntactic and semantic properties of all its word form variations (e.g. "buys", "bought", "buying" have the same core semantics as the stem "to buy"). *Decompounding* segments a word into a sequence of stems; e.g., the German noun "Hybridelektrokraftfahrzeug" ("hybrid electric vehicle") can be decomposed into the following stems: "Hybrid#elektro#kraft#fahrzeug". Compounding is a very productive process and cannot simply be handled by lexical lookup. Furthermore, compounds are notoriously ambiguous, e.g. "Verdichtung#verhältnis" and "Ver#dichtung#verhältnis" are two possible segmentations of the German noun "Verdichtungsverhältnis" ("compression ratio"), but only the first one has an acceptable meaning. Heuristically specified semantic rules are usually needed for identifying the best decomposition. Almost all of this standard functionality is widely available either as Open Source packages, or as professional software development kits. The software has already had a big impact in boosting the on-market development in a wide and diverse range of areas of the semantic analysis of unstructured documents like text analytics, information extraction, semantic search, deep-question answering and ontology learning.

4 Syntactic Analysis

As stated in the introduction, the analysis of phrase structure and syntactic functions is a dynamic field in NLP, which is in contrast to the components listed above. Syntactic analysis has only relatively recently started to find its way into commercial NLP application, for the simple reason that it is a computationally heavy operation, precluding acceptable performance of NLP systems until even a few years ago. However, with the fast-growing demand for semantically driven applications, as those exploited in ORDO, the status of syntactic annotation changed from unaffordable to indispensable almost overnight. Practically all intelligent semantic analysis requires some level of syntactic annotation for satisfactory accuracy.

A large and widely varying number of theories and formalisms is available for syntactic parsing and function assignment of natural language text. Most of them, however, aim at full coverage, and are consequently quite elaborate, even by today's standards of processing speed. For TME, a more goal-oriented formalism has been chosen. Its syntactic analysis just provides the coverage and level of detail that subsequent semantic components need and no more than that. Currently, TME offers such annotation classes for five languages: English, German, French, Spanish, and Simplified Chinese.

These syntactic annotation classes have been developed in a proprietary formalism, called SALSA (Simple and Adaptive Language for Semantic Annotation). It was originally a regular expression language, allowing reference to any and all linguistic annotations already produced. For the purpose of syntactic function assignment and coreference resolution, SALSA has been extended with a global cache mechanism. Details follow further down.

5 Trainable Fast Multilingual Dependency Parsing

In the previous section, we have learned that goal-oriented formalisms for syntactic analysis, like the one used in TME, are promising approaches for reaching the high demands of semantically driven applications. In this context, rule-based approaches are often used where the rules are defined manually. Although it is known that rule-based approaches can achieve high levels of accuracy, they usually lack wide coverage, i.e. they might not be able to identify enough syntactic variations. Consequently, weak robustness is a critical issue. As an alternative, in recent years

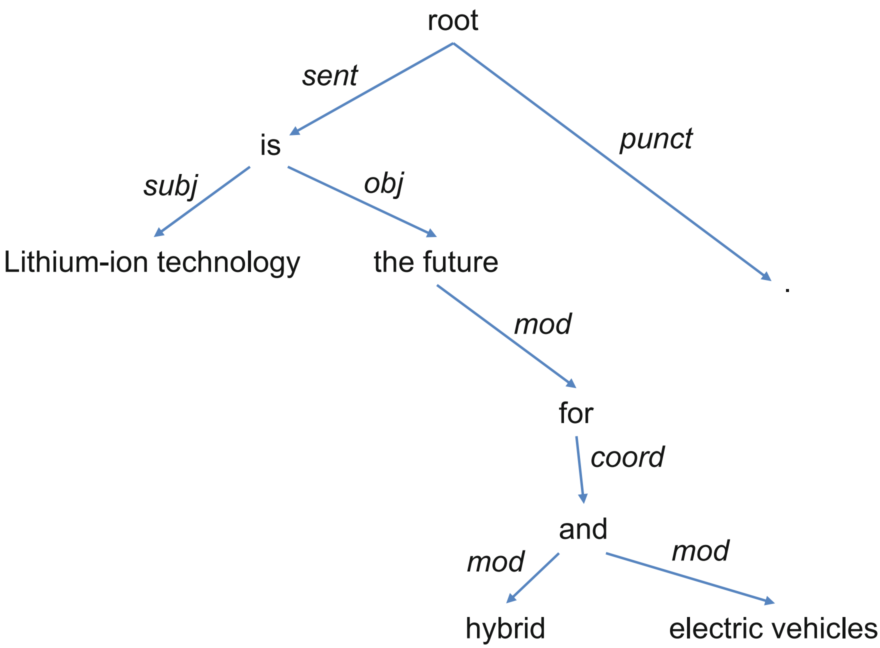


Fig. 1 Simplified dependency tree for sentence "Lithium-ion technology is the future for hybrid and electric vehicles." Edges are decorated with labels for grammatical functions

a lot of research energy has been brought into the development of robust trainable syntactic parsers that are able to compute the complete syntactic relationships of arbitrary sentences quickly, robustly and accurately.

Such trainable parsers receive as input a (usually very large) set of sentences (also called a "treebank"), each of which is already (manually or semi-automatically) annotated with the correct syntactic tree. A parsing engine together with a statistical-based learning algorithm is applied to automatically learn a model of all possible syntactic decisions that can be induced from the treebank. This acquired model is then used to determine the syntactic structure of any new sentence. Since the annotation schema is basically the only language-specific parameter, such a trainable parsing system is inherently multilingual, because it can process treebanks of any language.

The syntactic annotation schema of a treebank usually follows a linguistic theory or formalism. However, dependency theory in particular has been shown recently to be very suitable for achieving the necessary degree of robustness and efficiency in such a learning environment (Kübler et al. 2009). The dependency structure of a sentence is a rooted tree (more precisely, a rooted acyclic graph), where the nodes are labelled with the words of the sentence, and the directed edges between the nodes are labelled with the grammatical relations that hold between pairs of words. Figure 1 is an example of this. Dependency structures are appealing because they already represent a "shallow" semantic relationship. They are a suited data structure for many semantic applications, e.g. semantic search and relation extraction.

As part of ORDO, DFKI has developed the MDParser, which is a very fast multilingual trainable dependency parser, which also exists as a SMILA component (Volokh and Neumann 2011). MDParser has been adapted to a new highly efficient linear multiclass classifier to obtain high speeds. It is now able to process up to 50,000 tokens (∼2,000 sentences) per second. For the traditional English WSJ test data, MDParser scores 89.7 % UAS (Unlabelled Accuracy Score) and 87.7 % LAS (Labelled Accuracy Score). This performance is comparable with that of other state-of-the-art parsers; it is about three to five times faster than the widely used and previously fastest known dependency parser MaltParser (Hall et al. 2006).

The MDParser has been exploited successfully in a number of different semantic applications, such as recognizing textual entailment and trend analysis (Hong et al. 2011). As an example for the latter case, we developed a software demonstrator called "TechWatchTool" using SMILA. It aids companies in detecting emergent technologies in a particular field, and in identifying associated key players and their cooperative networks. It currently supports three scenarios:

1. Retrieval of patents and publications,
2. Ontological presentation of a knowledge domain, and
3. Identification of new trends in relevant documents.

The system has been developed in collaboration with ThyssenKrupp Steel AG and combines various methods used in bibliometrics, information extraction, and knowledge technologies. It integrates advanced NLP for detecting and extracting relevant facts from unstructured documents. Especially for relation extraction, the MDParser is used as a basis for the extraction of relational features. TechWatchTool provides personal and group-level access via a browser application and an interactive graphical interface, as shown in Fig. 2.

6 The Extraction of Named Entities and Semantic Relations

A central annotation task is the extraction of meaningful semantic information from text. Important subtasks are the extraction of named entities (names of persons, locations and organizations) as well as the extraction of semantic relations between these entities. For example, in the sentence "Metro-Chef Eckhard Cordes hat von der Konzerntochter Real im März 2008 gefordert, eine Umsatzrendite von drei Prozent zu erwirtschaften," we have the person "Eckhard Cordes", the companies "Metro" and "Real", the date "März 2008" and the percentage "drei Prozent". As nouns in German are in general capitalized the detection of names is much more difficult than in English. In addition, names often contain words which also have other meanings. Therefore, we require a context-sensitive detection algorithm. We used conditional random fields (McCallum and Sutton 2007) to detect the name phrases of a sentence by taking into account the roles and attributes of consecutive words in a sentence.

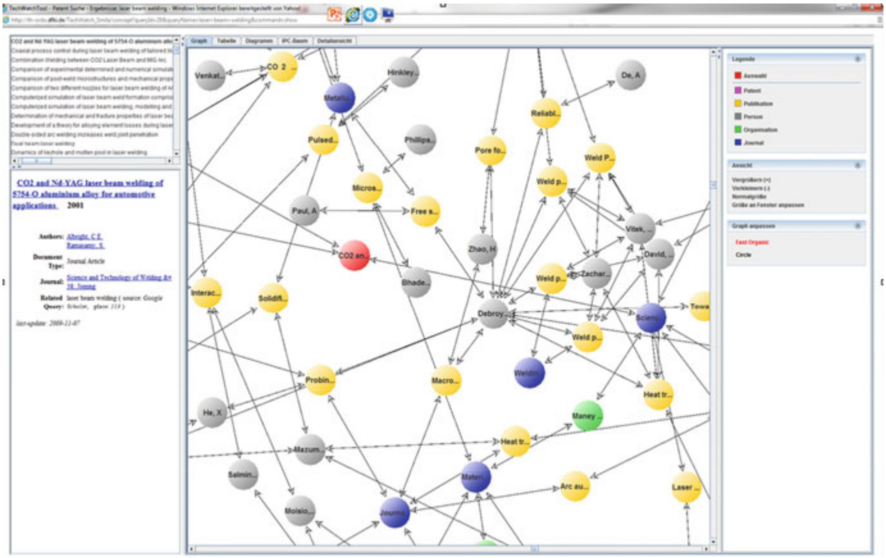


Fig. 2 Screenshot of TechWatchTool – retrieval and navigation in patents and publications

A major challenge for their application is the extraction of relevant input features for the German language.

Often, extracted entities have several possible meanings. In Wikipedia, the term “Metro”, for instance, can be a rapid transport system, an airport, a city, an administration, a newspaper, a retailer, a theatre, a film, etc. To be able to interpret an entity, we have to disambiguate a mention of an entity using its context. We have developed an advanced disambiguation system (Pilz and Paaß 2011) based on latent representations, which is able to find the correct meaning of an entity with high reliability and assign it to its corresponding article in Wikipedia. The system also detects if an entity is not covered in Wikipedia. The approach is language-independent, as it works for the English, German and French Wikipedia.

Relation detection requires that two named entities are selected as arguments. In the above example, for instance, we have the relations “Eckhard Cordes” is-leader-of “Metro” and “Real” is-subsidiary-of “Metro”. To detect such relations it is no longer sufficient to exploit the sequence of words but we need the structure of a sentence as described by a parser. In THESEUS, we evaluated dependency parsers and phrase grammar parsers and found that a combination of both structures gives the best results. Using annotated training examples, we trained a classifier using these structures which predicts whether the target relation holds between these arguments (Reichartz et al. 2010). For the member-of relation, for instance, we achieved an F-value of more than 80 %.

As an application within ORDO, Fraunhofer IAIS has developed a tool for Commerzbank, the second largest bank in Germany, which has its headquarters in Frankfurt. Similar to many other large banks, Commerzbank has only a few

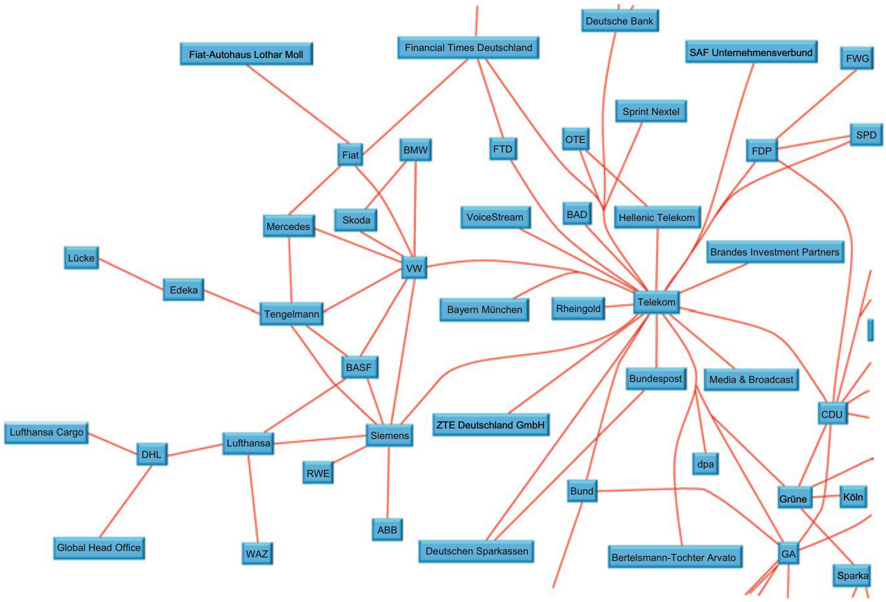


Fig. 3 Company relations extracted from the *General-Anzeiger* in Bonn

administrative centers that handle the credit business with medium and small enterprises. Although this structure is efficient, the centers often lack sufficient knowledge about the local markets and enterprises within small regions. The tool extracts information about the economic situation of markets and enterprises from local sources.

An important indicator of the economic situation of a company is the economic success of its business partners or its competitors. This posed the task of extracting and monitoring the relationships between companies from local news, especially online pages of local newspapers. To allow screening of large text repositories and the regular monitoring of relations, machine learning approaches from text mining were used. Figure 3 shows a network of relations between businesses extracted for the region of Bonn from the *General-Anzeiger* (the regional newspaper).

After downloading the web pages and cleaning the HTML pages, two text mining tools were employed. By using named entity recognition models, the names of companies were extracted from the text (2,209 companies for Bonn). Subsequently a relation extractor was trained and applied to extract relations between companies, e.g., company A does business with company B, A is a competitor of B, or A acquires B. As an input for the relation extractor, the text of the sentences is enhanced by many features that are indicators of a relation. Different types of information, relevant for the bankers could be extracted by using the detected relations:

- Slump in sales or net loss for the year for an enterprise. This also indicates possible problems for business partners and competitors.
- Cooperation between companies, acquisition of companies.
- Increase in turnover, growth of staff.
- Good order situation in the building sector.

If critical developments are detected, this information is given to the representatives of the bank who take appropriate measures. Currently, the extension of the system to include more relationships, as well as its application to a larger number of news sources, is being prepared.

7 Cross-Sentence Analysis

As mentioned in the introduction, one of the more difficult tasks in text mining is the resolution of ambiguity outside the sentence scope. A typical example is pronominal reference, which usually crosses sentence boundaries, as in: "George1 sent Martha2 flowers. He1 wanted to make up with her2." It is obvious, that the level of understanding of a text is enhanced by the proper resolution of pronouns like "He" and "her" in the example above.

First and foremost, this applies to the resolution of person names. Standard named entity extraction operating at the sentence level is able to identify only proper names as person referents, not pronouns. It is not able either to link a variant of a person's name (called alias) to its full (or canonical) referent. Consequently, standard methods fail to generalize mentions such as "George Burton", "Mr Burton", "George", "he" and "him" by linking them to the same canonical referent, "George Burton". Issues like these can be addressed by the introduction of a cache mechanism that stores likely candidates with their attribute sets for any type of disambiguation resolution. Such a cache has been developed for TME. Feature matching and stem similarity are the predominant heuristics used by the cache. In addition to named entity extraction, TME deploys the SALSA cache for certain syntactic function annotations, and for semantic applications such as sentiment analysis and resume annotation.

A functionality called "Dossier Creation" is currently under development; it heavily relies on the cache mechanism. Its general idea is to create automatically qualified structured information about a certain coherent class of objects from unstructured or semi-structured data sources. Dossier Creation promises a large variety of applications, from targeted data mining for research (intelligence, finance, patents) to normalization of large chaotic – i.e. unstructured – data repositories (HR, medical, technical, legal).

8 Conclusion

This article has outlined the needs and challenges for extracting semantics from unstructured documents from a linguistic perspective. All of the described language technology has been exploited and realized in ORDO, where some of the most important results and their embedding into innovative applications have been described in more detail. These results allow us to draw valuable conclusions from otherwise unstructured texts. Linguistics provides the basis for dealing with the Big Data challenge on the content level. From an application perspective, future work will further the creation of dossiers. Here, it might also be interesting to take into account cross-lingual phenomena since several of the above-mentioned linguistic components can already process several natural languages. Another promising line of R&D is the new field of "textual inference", a sort of applied textual semantics. A sub-area of this new field, "recognizing textual entailment", has already been explored successfully in ORDO with promising results. Beyond the mere research, ORDO has demonstrated that text analysis with linguistics is a robust technology. Its integration in day-to-day business processes creates added value, and facilitates the automation of tasks, which often still require human intervention.

References

J. Hall, J. Nilsson, J. Nivre, MaltParser – a data-driven dependency parser, in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC '06)*, Genoa (European Language Resources Association (ELRA), May 2006), pp. 2216–2219

L. Hong, F. Xu, H. Uszkoreit, TechWatchTool: innovation and trend monitoring, in *Proceedings of the International Conference on Recent Advances in Natural Language Processing 2011 (RANLP '11)*, Hissar, ed. by G. Angelova, K. Bontcheva, R. Mitkov, N. Nicolov (RANLP 2011 Organising Committee, 2011)

S. Kübler, R. McDonald, J. Nivre, *Dependency Parsing*. Synthesis Lectures on Human Language Technologies (Morgan & Claypool, New York, 2009)

A. McCallum, C. Sutton, An introduction to conditional random fields for relational learning, in *Introduction to Statistical Relational Learning*, ed. by L. Getoor, B. Taskar (MIT, Cambridge, 2007), http://www.cs.umass.edu/~mccallum/papers/crf-tutorial.pdf

A. Pilz, G. Paaß, From names to entities using thematic context distance, in *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM '11)*, Glasgow, ed. by C. Macdonald, I. Ounis, I. Ruthven (Association for Computing Machinery (ACM), 2011), pp. 857–866, http://dblp.uni-trier.de/db/conf/cikm/cikm2011.html#PilzP11

F. Reichartz, H. Korte, G. Paaß, Semantic relation extraction with Kernels over typed dependency trees, in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York (Association for Computing Machinery (ACM), 2010), pp. 773–782, http://portal.acm.org/citation.cfm?doid=1835804.1835902

P. Thibodeau, Gartner: the Bell tolls for Windows XP, in *Computerworld*, Oct 2010, http://www.computerworld.com/s/article/9191663/Gartner_The_bell_tolls_for_Windows_XP

A. Volokh, G. Neumann, Automatic detection and correction of errors in dependency treebanks, in *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT '11)*, Stroudsburg, vol. 2 (Association for Computational Linguistics, 2011), pp. 346–350

High Scalability for Semantic Indexes

Markus Nick, Thorsten Jäger, Volker Nussbaum, and Kai Kramer

Abstract ORDO has developed the infrastructure to deal with the Big Data challenge. It allows an organization to semantically enrich and access large amounts of unstructured data. This infrastructure can be run on regular server hardware, with additional nodes added when the amount of data increases. However, adding additional nodes also increases the likelihood of malfunctioning nodes. Therefore, the infrastructure implements failover concepts. Data distribution within the infrastructure takes place without employing a central – and maybe failing – communication point.

1 Introduction

The most prominent goal for high scalability is the processing and handling of Big Data (Beyer et al. 2011). As a requirement for the THESEUS activities within ORDO, the magical number of one billion records was stipulated at the start of the project. This huge number of records should be processed on a cluster of standard workstations (hereafter called node) without the need for special infrastructure components such as InfiniBand. As a consequence, reasonable scalability must be provided by any system fulfilling the aforementioned goal. Thinking of a cluster of nodes of size N that is capable of processing 100 million records within a given performance constraint, and then doubling the volume, i.e., adding another 100 million records, will typically lead to performance degradation. With an ideal linear scalability, doubling the number of cluster nodes will resolve this degradation, i.e., a cluster of $2N$ nodes will be able to process 200 million records with the same performance as before. Besides the primary goals of scalability and processing high

M. Nick (✉) • T. Jäger • V. Nussbaum • K. Kramer
Empolis Information Management GmbH, Bielefeld, Germany
e-mail: markus.nick@empolis.com; thorsten.jaeger@empolis.com; volker.nussbaum@empolis.com; kai.kramer@empolis.com

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_30,

data volume, the following secondary challenges and goals have to be addressed as well:

- *Maintainability*: since the entire system runs in a cluster of nodes, the system administrator needs support for basic operations, such as software deployment, system start/stop, cluster monitoring and process monitoring. The latter should also support drill-downs for effective failure analysis and – where needed – manual troubleshooting. In addition, cluster reconfiguration, i.e., increasing/decreasing the number of cluster nodes, should be supported. This is of particular importance in cloud scenarios in order to adapt the cluster to rising or declining amounts of data or requests.
- *Reliability*: operating a cluster of nodes, one has to accept that "failure is normal". First, the probability of a node failure increases with the number of nodes involved. Second, network errors may also appear more likely in a cluster with a high number of nodes, and thus a high number of cables, switches and network devices in general. Third, the probability of software failures (crashes of individual processes) is higher due to the sheer amount of input data that has to be processed, and that is typically not the outcome of a well-defined authoring process.
- *Efficiency*: in general, resources should be used whenever available, but resources should also not be wasted.
- *Performance*: every operation within the system should be performant, which basically holds for two types of operations: asynchronous operations should process a large amount of data within guaranteed time frames. As an example, the data ingest process should be throughput maximized, but should also offer quick visibility of each individual record. Synchronous operations should in general be sufficient for interactive usage. As an example, retrieval must not take longer than 500 ms, even on very large data sets.

In addition to the above-mentioned technical aspects, the *functionality* itself has to be evaluated carefully and then implemented. Certain algorithms, such as "associative search", based on standard cosine measures within the well-known vector space Incorporating Term Relevancies (IDF), bring up additional challenges: the IDF is no longer computable on a single node because of the volume of data.

2 General Solutions

From an indexer's perspective, the basic solution for the main challenge, volume and scalability, is *data partitioning* (i.e. *sharding*). In general, a logical index comprising a large number of records will be partitioned into several physical index parts (i.e. *shards*) that reside on the individual cluster nodes. Thus, every node has to deal with a (small) fraction of the entire index and may be involved in retrieval requests. The index distribution and partitioning will also allow for fast and low-latency *retrieval*.

On the other hand, throughput-optimized *index creation* (i.e. *data ingest*) has to be performed in an asynchronous bulk-oriented manner. Actual data processing is distributed among all cluster nodes. With this, an efficient processing of high volume data input is guaranteed. For details refer to the SMILA section.

For both types of operations, a highly scalable and fail-safe persistence layer is required. Within ORDO, the Distributed Object Store (*DOS*) has been designed and implemented to optimally serve both types of operations. However, a fail-safe persistence layer is not sufficient for solving the reliability requirements. In addition, fail-safe "behavior" (*fail-over*) is part of every individual component in a way that it is able to cope with unavailable services of any kind.

Furthermore, both *scale-up* and *scale-out* are very important concepts in designing a truly scalable and efficient system. The more obvious requirement is scaling out, i.e., scaling over different physical nodes. In addition, the scale-up within a single node should provide reasonable "parallel processing" on typical modern-day workstations that offer two to four CPUs each with 2–8 cores, which results in a basic computing power of 4–32 cores for a single node. For efficiency and performance, this computing power is fully utilized. Last but not least, a specific service offers all kinds of administrative operations and thus allows for easy system maintenance even in a cluster comprising a large number of nodes.

Before looking deeper into some of the aforementioned aspects, we have to stretch a very important theorem called CAP (Gilbert and Lynch 2002). Basically, the CAP theorem states that in every distributed system one can ensure only two of the following three aspects: *C*onsistency, *A*vailability, *P*artitioning. With data partitioning being the key to scalability, as a simplified consequence of the CAP theorem you have to choose between consistency and availability. The most prominent large-scale systems such as Google or Amazon sacrifice consistency and favor availability. The well-known database pattern ACID, demanding strong consistency, has been replaced in such systems by an alternative called BASE (*B*asically *A*vailable, *S*oft state, *E*ventually consistent) (Pritchett 2008; Vogels 2009). Since consistency has not been formulated as an ORDO requirement, we also favor availability over strong consistency.

3 Distributed Object Store (DOS)

Designing a persistence layer for a scalable distributed system, one has to be very clear about the actual requirements. The following are a subset of the requirements within the ORDO scope:

- Serve as persistence layer for ALL large objects that have to be stored within the system independently of their nature (e.g. record bulks, index shards) or type (e.g. temporary data, permanent data).
- Guarantee fail-safety within configurable constraints (e.g. number of nodes that may be down) for every single object that is stored in the persistence layer.

- Support fail-safety for small objects (e.g. single records) entering the system without compromising throughput performance.
- Support cluster reconfiguration with automated data rebalancing.
- Offer garbage collecting functionality for deletion of temporary or inconsistent data.
- Allow for low-latency retrieval on index shards stored in the persistence layer.

Currently, none of the available market or Open Source solutions fulfils these requirements. For example, Hadoop with HBase and HDFS offers distributed computing based on the map-reduced framework that was introduced by Google. Within Hadoop, the last two requirements are currently not fulfilled. In general, Hadoop fulfils a subset of requirements related to distributed and fail-safe data ingestion. However, there is no support for low-latency access to index shards to fulfil retrieval performance requirements.

The DOS has been designed to fulfil all fail-safety requirements as stated above. In addition, it offers low-latency access to index shards via memory mapping (mmap), a crucial feature for low-latency retrieval. As a conclusion, the DOS offers an "intelligent" way of handling data and does not act as a passive storage service. The DOS is based on the following design principles:

- "All nodes are equal". This way, maximum fail-safety can be achieved since one or more distinguished *master nodes* do not exist. The state of the persistence layer is stored in the persistence layer itself.
- The DOS utilizes the file system as an underlying physical persistence layer. Thus, there is no need for separate database installation and maintenance.
- Objects stored in the persistence layer are replicated over the cluster nodes of the persistence layer. The degree of replication is freely configurable for every logical store. Note that object replication is the only way to offer read/write access in case of node failures.
- Object distribution among cluster nodes happens according to a consistent hashing function.[1] Again, the distribution is not the task of a master node. Instead, the distribution can be calculated at any arbitrary cluster node based on a hash ring.
- From a client perspective, the DOS offers transparent fail-over and recovery mechanisms. For example, in case a read access to an object fails because of a node or network failure, the object is read from a replica and transparently delivered to the client.

As mentioned above, the physical distribution of data objects is determined by a hash ring function which is based on consistent hashing. Simplified, every cluster node has a fixed position on a hash ring, based on the hashing function. In addition to that, a position for every data object on the same hash ring can be calculated. Turning clockwise from that position, the cluster nodes can be determined that will

[1] http://en.wikipedia.org/w/index.php?title=Consistent_hashing&oldid=482752550

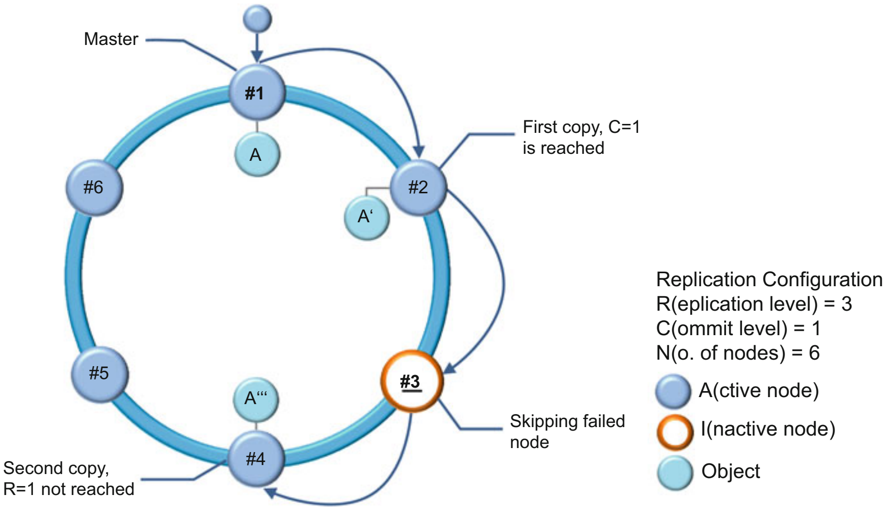


Fig. 1 Replication with one-node failure during import

serve as physical storage points for that object and its replicas. Figure 1 shows an example.

Assume the cluster consists of six nodes labelled #1 to #6, and the nodes are distributed equally on the hash ring based on these six nodes shown in the above picture. Further assume that a client wants to store object A in the DOS. The storage command is a simple PUT call with object A as the parameter via the DOS API. In order to determine the physical storage position, the DOS calculates the hash value for object A, i.e., its position on the given hash ring. The first cluster node in the clockwise direction is the first physical storage point ("master") for object A. Without replication, object A would be stored on node #1, and the operation is finished. With the shown replication level of $R = 3$, the DOS tries to write three more replicas of object A. The physical storage position is again determined by following the hash ring clockwise. In the example, nodes #2, #3, and #4 would store the three replicas of object A. In the case of a node or process failure, the failed node will be skipped. The failure of node #3 hinders storage of a replica on that particular node. In this case, the storage continues with the next node. Please note that there are always $R + 1$ write requests, with R being the replication level. This logic is independent from the number of failed nodes, i.e., the physical storage positions are independent from the node failures, and can be calculated at any time by any cluster node.

A drawback of this approach is that the replication level for an object might not be reached although enough cluster nodes are available. In the above example with such a "dynamic relocation", a third replica could have been stored on node #5. The number of copies that must be written successfully is configured by the commit level $C = 1$. If it is not possible to write at least $C + 1$ copies, the entire PUT operation fails, which is immediately returned to the client. If at least C + 1 copies are written, the client will receive an OK as a result of the PUT request. Hereafter, potentially missing replicas are written asynchronously in the background to reach the full replication degree. Besides the functionality of a fail-safe persistence layer for large data objects, the DOS offers an API for the storage of small data objects (records, key-value objects) as part of a large data object.

For this the DOS provides efficient lease management to ensure that the changes to data are atomic and the storage of the data takes effect instantaneously after the operation.

When recovering from a failure, every DOS node synchronizes itself with all other nodes. Since the hash ring is static, i.e., the position of every data object is stable over time, every node can calculate the positions for every data object based on the hash ring, the object's hash (i.e. its position on the hash ring), and the configured replication level. Thus, missing objects on a recovered node can be determined without master nodes, keeping track of all physical storage points and failed operations. This recovering process runs asynchronously in the background and does not block any other read/write requests to the DOS.

The DOS also supports entire cluster reconfiguration, i.e., adding nodes to an existing cluster or removing nodes from a cluster. An advantage of the utilized consistent hashing is that cluster reconfiguration requires minimal data transfer. Think of increasing the cluster size by one node from N to N + 1. With consistent hashing, only $1/N$ of the data objects will be relocated while the majority of all data objects including the replicas will remain on the same cluster node as before. Again, the physical storage points of all objects can be calculated on every node without the need for additional control tables or master nodes. The DOS also offers powerful Garbage Collection (GC) mechanisms. Traditionally, temporary data, e.g., intermediate results of asynchronous workflows, are physically deleted by such a GC. As a consequence of sacrificing consistency, (temporarily) inconsistent data may reside within the DOS. This is handled by intelligent GC algorithms that resolve inconsistencies and guarantee consistent states with small delays that are commonly accepted when implementing BASE solutions.

4 High Scalable Semantic Indexer (HSSI)

Besides a distributed fail-safe persistence layer, the actual indexer, hereafter referred to as HSSI, is probably the most challenging part in designing a distributed scalable retrieval system. As opposed to a single-node single-partition index, a logical index will now be physically partitioned. The index partitions (i.e. index shards) are stored in the persistence layer, whereby every index partition is a data object in the DOS. Since all data objects are distributed according to the mentioned hash ring logic, all physical index partitions for a single logical index are distributed among all cluster nodes as well. It follows that retrieval is not restricted to a single cluster node, but

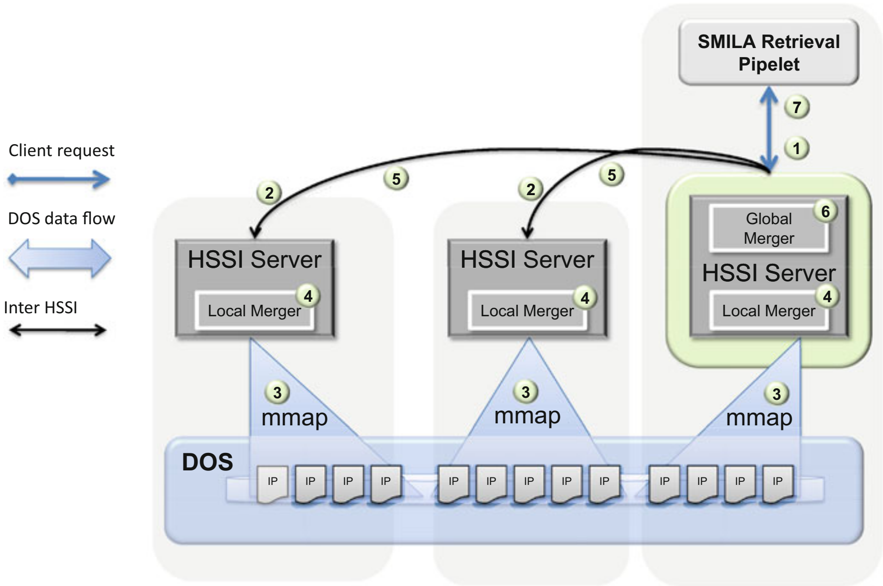


Fig. 2 Architecture of ORDO information access technology

rather spreads over all nodes that have at least one index partition in their local DOS instance.

In Fig. 1, a typical retrieval request is depicted. The light grey boxes depict three physical nodes of a cluster. Every cluster node runs an instance of the HSSI server. This HSSI server instance performs the retrieval on its local Index Partitions (IPs), i.e. those index partitions that are stored on this node according to the DOS' hash ring. When a retrieval request is sent from a client to an arbitrary HSSI instance (1), the request is distributed to all other HSSI instances (2). The results (3) on the local IPs are merged in the respective instance (4) and sent back to the caller (5), which acts as a "master" for this retrieval request. The master collects all retrieval results from all instances and performs the final result merging (6) to obtain the result list that is delivered to the client (7). Since the entire process is synchronous from a client's perspective, speed is very important, which is the reason for the DOS offering the low-latency access via memory map to certain kinds of data objects such as index partitions. With enough main memory per cluster node, the index can be kept in the node's RAM, which is the only way to guarantee low-latency retrieval.

It is important to notice that there is no notion of a single "HSSI master instance". The HSSI instance receiving a client request acts as a master for exactly this request. For the next request, a different HSSI instance may act as the master. As with the DOS, we have strictly avoided a single point of failure in the system design.

Again, fail-safety is very crucial for success. In combination with the persistence layer's fail-safety, HSSI also offers full fail-safety up to a configurable degree of node/process failures. Assume the middle node in Fig. 2 is down due to a power failure. All other nodes in the cluster will immediately be notified upon detection of a failed communication to this node. From this moment on, retrieval requests will no longer be issued to this node. Since retrieval results in HSSI are guaranteed to be complete, the index partitions originally evaluated in the middle node must now be evaluated by the HSSI instances on the remaining two nodes. With a replication level $R = 1$, the DOS stores one replica of each index partition that will now be utilized during retrieval. According to the hash ring distribution, the replicas are almost equally spread over the remaining cluster nodes, i.e. the additional retrieval load for the five index partitions is also spread over the remaining nodes. This is a huge advantage compared to other approaches where entire nodes are replicated. Here, one of the two remaining nodes would "take over" for the failed node, resulting in much higher performance degradation, because the node "taking over" automatically becomes the bottleneck for all subsequent retrieval requests. With equal distribution of the additional load based on the hash ring distribution, degradation is only minimal, especially in a large cluster.

In addition, HSSI also offers fully automatic recovery. Assume the failed middle node comes up again. Since all data objects are asynchronously recovered by the DOS, and HSSI – like DOS – does not rely on a control or master node, the recovered HSSI instance is capable of fully recovering its state. As soon as this recovery is finished, the instance is back into play and will receive retrieval requests from other instances again. For optimal retrieval performance, HSSI also supports scale-up, i.e. local index partitions can be evaluated in parallel. Thus, retrieval performance automatically benefits from multi-core architecture. Ideally, the number of local index partitions corresponds to the number of cores available.

The real world is a world of continuously changing data and for that a search engine should offer the user a simple mechanism to adapt the index to these changes. Consequently, HSSI has to deal with a situation of continuous modifications like the insertion or deletion of data, which of course enforces the change of the index shards. As for databases, resharding becomes necessary whenever the capacity of logical shards deviates too much from a configured target capacity. Together, DOS and HSSI offer two strategies for resharding of index shards to guarantee the optimal shard size. The first strategy offers a continuous optimization process to merge index shards to an optimal final index shard. The second strategy offers a compact process to remove marked records for deletion from the index shard. Thereafter, the shrunken index shard is again a part of the optimization process. The intelligent GC algorithm of the DOS removes the old data in an efficient manner. Resharding is typically a very costly operation (George 2011). Resharding is performed as a background operation and will guarantee homogenous index partitions with respect to query performance throughout the entire cluster. The design of the resharding process ensures the correct visibility of the index data across the index shards which are part of the optimization process.

5 Evaluating High Scalability of Semantic Technology

Big Data is one of the major challenges for the coming years. To process today's volumes of data and Big Data, scalability is the enabling "feature". As mentioned before, scalability has two dimensions: "scale-up" and "scale-out". Scale-up refers to scaling over the number of cores within nodes/computers in a cluster. Scale-out refers to scaling by adding more nodes. Scale-up is limited by the constraints of the hardware, i.e. the number of CPUs and cores per computer, IO bandwidth, etc. Thus scale-up is not sufficient to address Big Data; scale-out is the enabler for processing Big Data.

To scale-out without limits, the system needs to scale linearly. Linear scalability means, for example, that for doubling the volume of records with the same time for the ingest, the number of nodes needs to be doubled; doubling the number of users with the same retrieval performance requires doubling the number of nodes; doubling the performance with the same number of users and records requires doubling the number of nodes. Linear scalability has technical and theoretical limits. For example, a technical limit is the TCP/IP communication overhead. Theoretical limits are set by Amdahl's law (Rodgers 1985). Amdahl's law is a model for the relationship between the expected speed-up of parallelized implementations of an algorithm relative to the serial algorithm, under the assumption that the problem size remains the same when parallelized.

In a controlled experiment, the ensemble of the following ORDO technologies was evaluated with respect to linear scalability: SMILA, text mining engine, highly scalable semantic indexer, distributed object store, administration/operations functionality. The scalability tests were conducted by Fraunhofer ITWM in their Hercules cluster. The tests have positively evaluated the scalability for both types of processing: (1) throughput-oriented bulk processing – such as for indexing – and (2) low-latency access.

The architectural principle is a shared-nothing architecture. In this kind of architecture, each node has its own resources for processing its tasks. Such resources are CPU, main memory (RAM), hard disks and network connection. These nodes are usually connected with a Gigabit LAN.

The hypothesis is that the ensemble of the ORDO technologies scales out almost linearly. To test the scale-out, the type of nodes/computers and the number of documents is kept stable. Only the number of nodes is changed in the experiment. The scale-out is evaluated for both types of processing in SMILA: (1) throughput-oriented bulk processing – e.g. for indexing – and (2) low-latency access.

Low-latency access is built on in-memory computing technology. Thus, testing requires having sufficient main memory on the nodes in the cluster to hold the index in the memory. As the nodes of the cluster had only 8 GB of main memory, the smallest number of nodes and the amount of data was limited. With these constraints, 20 nodes and 71 million documents were the lower end for the tests. The test has evaluated the scalability for 20, 40 and 60 nodes with positive results.

In the following, we give some details about the hardware, the test data and the test procedure.

Hardware: Fraunhofer ITWM conducted the evaluation on their Hercules cluster. The primary purpose of the Hercules cluster is the computation of simulations and visualizations. The hardware is equipped respectively: this cluster consists of 260 computers/nodes that are equipped with 2.33 GHZ Intel Xeon Quadcore CPUs. Each node has 8 GB of main memory (RAM) and is connected to a parallel file system based on GPFS. This file system provides data transfer rates of up to 1,200 MB/s. In addition, each node has 30 GB of local hard disk storage. The network connection is 10 GBit/s InfiniBand as well as a redundant Gigabit Ethernet connection.

In-memory computing and information access had different constraints, which required the adaptation of the test setup to the available hardware: the RAM size of 8 GB was a limit for testing the low-latency access, which is based on in-memory computing technology. Furthermore, the local disk space of 30 GB was not sufficient to store all index data. Thus, the GPFS had to be used for the tests, which is regarded as being sufficiently similar to local disk access with respect to latency for this evaluation.

Test data: The test data was derived from the Wikipedia articles in German, English, Spanish, Portuguese, French, Italian, Dutch, Polish and Rumanian. All Wikipedia articles for these languages comprise approximately 10 million documents of 4 KB on average. Earlier tests with the TREC GOV2 dataset (25 million documents) as well as experiences of other datasets have shown, that huge document sets have approximately one new term per document. Thus, to make the volume realistic with respect to the term volume, each copy of a Wikipedia document was extended with a unique new term. Sizing calculations with the RAM and disk size per node resulted in approximately 70 million documents on 20 nodes as the maximum size of the test dataset. Thus, the mentioned Wikipedia articles were copied six times and the generated terms were added to each copy. This led to a test dataset of 71,085,195 documents.

Test procedure: The high scalability tests were conducted with 20, 40 and 60 nodes. Each of these tests consisted of the following steps:

1. Initialization of the nodes,
2. Automatic deployment of the software on all nodes,
3. Import/indexing of the test data while monitoring the cluster,
4. Evaluation of the import regarding duration and number of partitions,
5. Checks for availability of the cluster after the import,
6. Information access tests, and
7. Random checks of the data distribution in the distributed object store.

Furthermore, the system was configured with a replication degree of 2. This means that the distributed object store copies the data twice and therefore the loss of two nodes is possible without losing data. This is the same setup as with a pilot customer. The scale-out was evaluated for both types of processing in SMILA:

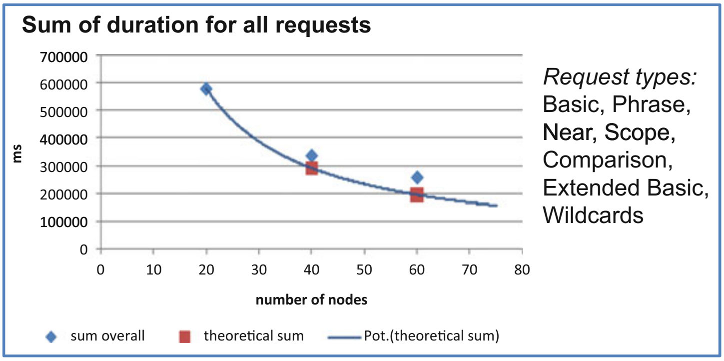


Fig. 3 Measurements of request time again show a linear scalability

1. Throughput-oriented bulk processing, and
2. Low-latency access.

Throughput-oriented bulk processing was evaluated with the indexing/ingest scenario. For this purpose, the indexing time was measured for 20, 40 and 60 nodes. With linear scalability, the import time should be cut in half when the number of nodes is doubled. The actual measurements show some small variations, which are within the normal range of technical variability. Thus, the results demonstrate the linear scalability for the scale-out over the number of nodes for the throughput-oriented bulk processing.

Low-latency access has been evaluated with the retrieval scenario. For the test of the retrieval scenario, a set of test queries was defined. These test queries covered typical requests and requests that cause a high system load. Since response time depends on the complexity of the query, the queries were categorized into request types according to their complexity class.

The sum of the response time of all test queries evaluates the general scalability of the retrieval scenario. The test queries were executed for 20, 40 and 60 nodes (Fig. 3). Using the result for 20 nodes as a baseline, the theoretical response time sum for 40 and 60 nodes has been calculated by cutting the response time in half when doubling the number of nodes in the cluster. This is shown as a trend line in Fig. 3. The figure shows that the results are almost linear for the retrieval scenario.

The analysis of the scalability by request type gives further insights. In general, the response time is shorter with more nodes. However, the scalability effect becomes less than linear with more complex queries. This is particularly demonstrated with the phrase queries. The reason is that with complex queries, the calculation of the intermediate results on the nodes has a higher probability of taking longer on some nodes. This then slows down the calculation of the overall query result.

6 Conclusion

The results of the experiment have shown that the technologies scale in a linear manner for indexing and almost linear for low-latency retrieval. This has been demonstrated for 20, 40 and 60 nodes with a test dataset of 71 million documents. A pilot customer has been using ORDO technologies in a cluster with 32 nodes in 24/7 operations without downtime for more than 1 year. In this cluster, each node is equipped with a dual quad core CPU and 96 GB of main memory. Furthermore, the ORDO technologies were used for real-time import of Twitter feeds with more than 400 million documents.

Combining these results for typical in-memory computing hardware leads to the following volumes: 100 GB main memory holds 25 million documents of 4 KB raw text on each node. This results in 1,000 million documents on 40 nodes and 1,500 million documents on 60 nodes for the low-latency access. Due to further improvements in the index structure evaluated in a test with 250 million documents, ORDO technologies are now able to deal with up to 2,000 million documents.

Since batch-oriented processing can easily scale beyond this amount of documents, the technologies developed within ORDO provide a reliable and extensible foundation for dealing with the Big Data challenge.

References

M.A. Beyer, A. Lapkin, N. Gall, D. Feinberg, V.T. Sribar, "Big Data" is only the beginning of extreme information management, Apr 2011, http://www.gartner.com/id=1622715

L. George, *HBase: The Definitive Guide*, 1st edn. (O'Reilly Media, Köln, 2011)

S. Gilbert, N. Lynch, Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services. ACM SIGACT News **33**(2), 51–59 (2002)

D. Pritchett, BASE: an acid alternative. Queue Object Relational Mapp. **6**(3), 48–55 (2008)

D.P. Rodgers, Improvements in multiprocessor system design, in *Proceedings of the 12th Annual International Symposium on Computer Architecture (ISCA '85)*, Los Alamitos (IEEE Computer Society Press, June 1985), pp. 225–231

W. Vogels, Eventually consistent. Commun. ACM Rural Eng. Dev. **52**(1), 40–44 (2009)

Integration of Semantic Technologies for Business Process Support in the Automation Industry

Hans-Josef Hesse, Andreas Kohn, Jörn Lehmann, Helena Hashemi Farzaneh, and Ditmar Ihlenburg

Abstract The reuse of existing information is an important factor for efficient business workflows in mechanical engineering. However, a considerable amount of this information is currently stored in unstructured text documents and therefore not easily accessible. Currently, no existing software system gives access to this hidden information. Companies will benefit from using linguistic algorithms in combination with semantic technologies to extract information from unstructured documents. This article exemplifies the successful use of semantic technologies, combined with textmining tools to extract facts from unstructured data. An example from the field of automation industry is used to describe and evaluate the presented approach. A function-based knowledge model for the classification and description of technical solutions is developed. An innovative software system supports the annotation of unstructured documents and the subsequent search for technical solutions. Previously inaccessible and hidden knowledge is unlocked and can be utilized more efficiently in business processes.

H.-J. Hesse (✉)
Empolis Information Management GmbH, Bielefeld, Germany
e-mail: hans-josef.hesse@empolis.com

A. Kohn • H.H. Farzaneh
Technische Universität Munich – Product Development, Garching, Germany
e-mail: andreas.kohn@pe.mw.tum.de; hashemi@pe.mw.tum.de

J. Lehmann
VDMA Verlag GmbH, Frankfurt am Main, Germany
e-mail: joern.lehmann@vdma.org

D. Ihlenburg
Festo AG & Co. KG, Esslingen, Germany
e-mail: dihlenburg@web.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_31,

1 Introduction

The PROCESSUS use case focusses on users in enterprises who face special challenges in complex and communication-centric business processes, such as purchasing or product development. For these individuals, typically called "knowledge workers", it is essential to obtain access to existing solutions and project information. The retrieval of correct and relevant information within the context of the business process is one of the key issues in enterprises today. The amount of digital information is growing by a factor of 8 from 2011 to 2016 (Gantz and Reinsel 2011). Knowledge workers have to bridge the gap between the general availability of information and the correct and quick access to it. The vision of PROCESSUS is to develop a semantic business integration platform to support the knowledge worker by using Web 3.0 technologies in the form of semantic knowledge models and innovative software algorithms. The developed algorithms are intended to support business processes within the enterprise between different divisions, as well as business processes between cooperating enterprises, such as supplier and customer or cooperation partners. The knowledge model delivers the vocabulary and the relations between business objects to improve communication and the flow of information. PROCESSUS focuses on the field of automation industry as an exemplary application for the developed semantic business integration platform. Particularly, the important and challenging interaction processes between suppliers and customers concerning solution knowledge are supported.

2 Customer and Supplier Interaction and Handling of Solution Knowledge in the Automation Industry

Customer requirements for industrial solutions in the German automation industry are characterized by increasingly high complexity and extremely specialized demands. Accordingly, the amount of necessary communication and coordination between suppliers and customers during product development grows significantly resulting in more and more cross-company activities (Lüthje 2000; Reichwald and Piller 2009). Thus, real value generation increasingly occurs within networks. Traditional marketing activities shift from transactional behavior to relational behavior. Communication, interaction, and the dialog between suppliers and customers become increasingly important (Bonnemeier et al. 2007; Tuli et al. 2007). Modern and next generation innovation management requires full integration of suppliers and customers in the product development process. Especially, the interaction process for the exchange of information between the involved parties is currently insufficiently supported (Ihlenburg 2011; Reichwald and Wegener 2008). From the suppliers' point of view, the interaction process between suppliers and customers can be best described by an acquisition processes. In contrast, for the customers, this process phase is a requirements-oriented search for possible solutions that fulfil the specified demands, similarly to an innovation process (Ihlenburg 2011). Customers search for solution knowledge in their specific context and prefer to readily find what they are looking for.

A survey on the future of Internet platforms in the automation industry reveals interesting facts about searches for existing solutions and usage of information channels (Danzinger et al. 2008). In this study, 371 participants from different companies were asked about their demands for Internet platforms. As a major result, it was identified that the available information channels and present search mechanisms are time-consuming, and the result of a search process is often not satisfying. For example, experts with specialized knowledge within the company who can help in a certain situation are often unknown, due to decentralized organizational structures (e.g. national sales offices). Furthermore, the search for appropriate solutions, based on personal communication, is a challenging task. In most cases, knowledge concerning possible solutions is presented in unstructured documents such as success stories, marketing information, email communication or project folders. Due to the use of different classification systems or thesauri and various linguistic descriptions of facts, a full text search often fails. Existing databases, as well as search engines, do not offer efficient access to knowledge documents outside the company or solutions currently under development. This results in a “semantic gap” between technical problems and their technical descriptions.

Bridging the semantic gap between the technical problem and its technical description is the overall challenge in supporting the search for solution knowledge (Gaag et al. 2009). In general, the author of a technical description (technical editor, salesperson, engineer, etc.) creates the documents using company or domain-specific linguistic patterns, combined with taxonomies, dictionaries, thesauri and other classification systems. Companies differ in the use of vocabulary for equivalent or similar solutions (Schmidt et al. 2011). Therefore, the common search mechanisms of keyword-based or faceted searches in databases clearly reveal disadvantages concerning the number and quality of retrieved search results, as well as their time-consuming character. Static structures, such as classification systems, thesauri or dictionaries, do not offer much support to the knowledge worker who wishes to search for unknown solutions based on a model-independent description.

3 Knowledge Model and Software System for Solution Handling in the Field of Automation Engineering

A knowledge model in the form of an ontology for the classification and description of engineering solutions in the automation industry was developed (Gaag et al. 2009) to support processes for the handling of solution knowledge as described above. It is applied within a software system called “Solution Assistant” that enables an improved handling of technical solutions (e.g. reference projects). The Solution Assistant is based on the developed PROCESSUS semantic business integration platform and supports the automated annotation of existing documents and the following search for these documents. It uses the three standard services of the PROCESSUS platform: Facts Extraction Service, Annotation Service and Search Service. Using the ontology, the Solution Assistant is able to describe and classify technical solutions from unstructured text documents. A document describing a technical solution can be annotated with respect to the ontology using textmining technology. Annotation, in this case, means the identification of entities in the document which match the concepts and definitions of the ontology and the creation of the appropriate metadata for the document. Thus, the technical documents are integrated into the ontology and can be accessed more easily. The Solution Assistant also supports the intelligent access to information and provides a semantic search for similar solutions. Here, the ontology is used for the computation of similarities between technical solutions. Measurements like graph distance and vector similarities are used to compare different technical solutions. In terms of generalization and specialization, it is possible to change the context within the ontology. This results in an increase or decrease in the number of displayed solutions. This innovative procedure in searching for similar solutions enables the transfer of existing solutions from one industrial sector to another. Thus, the Solution Assistant supports the process of innovation by suggesting unexpected solutions. This procedure facilitates the reuse of existing solutions with all its benefits; reducing the time for product development, accompanied by reduced costs and increased quality by referencing existing and functional solutions. In contrast to systems such as product configurators, it is a knowledge model and a software system that supports the selection process of technical solutions, not by creating new systems, but with a computer- aided selection process, based on abstract engineering principles stored in the ontology.

The classes and relations of the ontology are described in the next section. Then, the identification of instances in unstructured text documents by linguistic algorithms is presented in detail. Subsequently, the main functionalities of the Solution Assistant (search and annotation) are illustrated.

4 Classes and Relations of the Ontology for Handling Solution Knowledge

The ontology for the classification and description of engineering solutions is displayed in Fig. 1, showing the most important concepts and relations. The central concept is the *solution* represented and described in unstructured text documents. The solution is related to the concept *function*, *function owner*, *company*, *industrial sector* and *properties*. The solution realizes required functions. A function is the combination of an *operation* and an *object*. Figure 1 shows the function "transfer bottle" as an example. Here, the function is composed of the operation "transfer" and the object "bottle". The function, as a combination of operation and object,

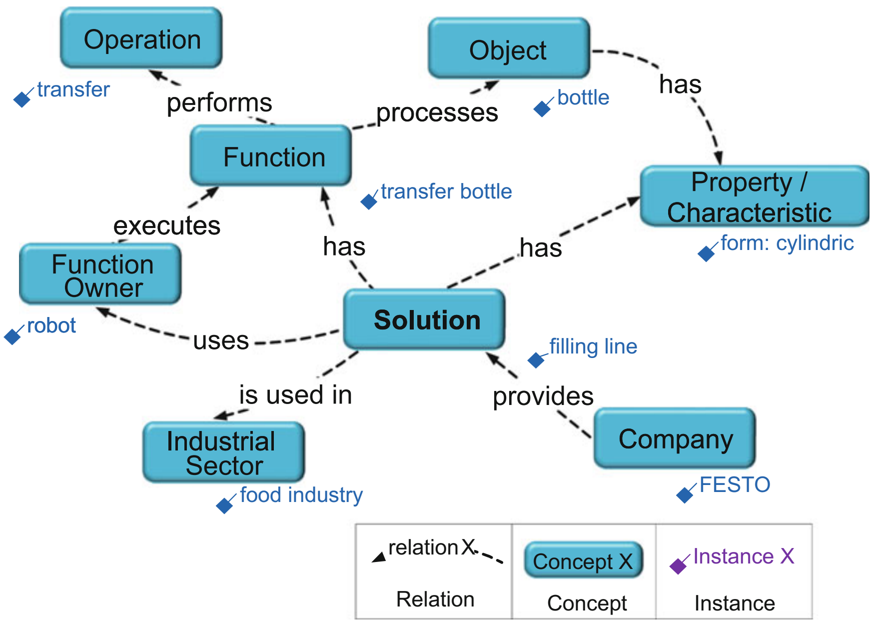


Fig. 1 Core structure of the ontology with instance examples

has a direct impact on the state of the object and it is used to modify one or more properties of an object. E.g., the function "transfer bottle" uses the operation "transfer" to change the property "position" of the object "bottle". The concept function owner represents the technical realization (e.g., robot, cylinder, . . .) of the solution which executes the required function. In addition, technical solutions have domain-specific properties and characteristics to fulfil the requirements. Industrial sector and company serve as additional meta-information to characterize the solution.

The concepts and relations modeled in the ontology are used by the software system for computing the similarity between technical solutions. The number of mutual instances describing solution documents is one measure for the similarity. For example, two solutions are similar if they both realize the function "transfer bottle". Additional concepts, such as the goal of performed operations (Gaag et al. 2009), and the properties of handled objects – e.g., the shape – provide further abstract measures for this computation. The abstract description of operations regarding their operation goal and the properties of the objects provides an appropriate abstraction layer for categorizing the technical solutions within the mechanical engineering sector in order to detect similarities.

5 Linguistic Algorithms for Facts Extraction Using Textmining Capabilities

The Solution Assistant uses the Facts Extraction Service of the PROCESSUS platform for analyzing unstructured text documents. The ontology with its concepts, properties, relations and instances provides the basis for the semantic classification of technical solution documents. In order to make the information in unstructured technical solution documents available, the relevant entities inside unstructured solution documents have to get related to the ontology. The resulting network of information represents the persistent knowledge about technical solutions. This knowledge can be used by knowledge workers to improve and support their daily work.

Linguistic algorithms, developed in cooperation with the THESEUS Use Case ORDO, are used to parse textual input from solution documents to extract the instances from the ontology. These algorithms are parameterized by concepts and instances in the ontology. Analysis results are stored in the ontology as data structures called "semantic annotations". The identification of instances, as exemplarily shown in Fig. 2, is also called "entity tagging". The exemplary sentence "The KR 150 is taking four crates from the conveyor belt", taken from a solution document, describes the handling of crates with bottles in the beverage industry. Here, "KR 150" is the name of a certain industrial robot. The linguistic algorithm analyzes the textual description and extracts detailed information about the content of each sentence such as tokens, word stems or text patterns. This builds the basis for matching the text with the ontology. Semantic annotations can be applied to contiguous or discontinuous text patterns. A pattern may be anything, logical or illogical, as long as it can be matched within the scope of semantic tagging.

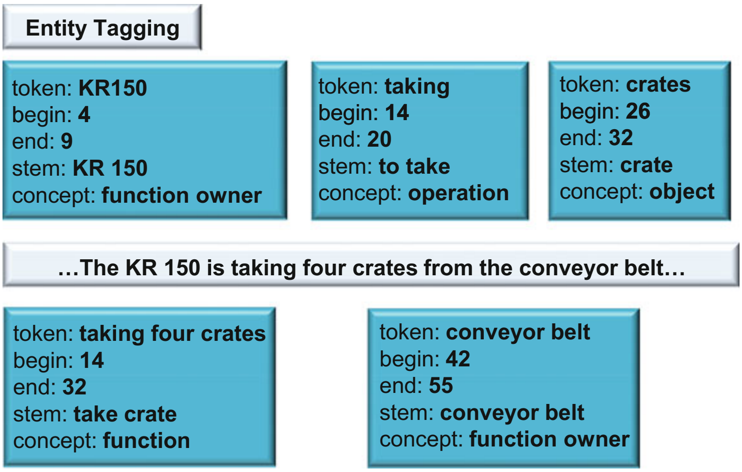


Fig. 2 Example for entity tagging using textmining

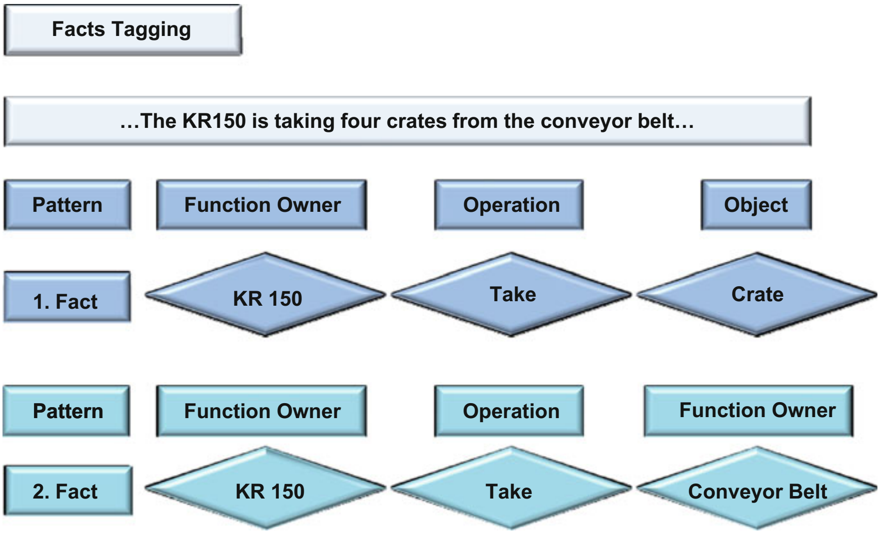


Fig. 3 Example for facts tagging using text mining

Currently, the scope of semantic tagging is one sentence. Typically, contiguous patterns constitute named entities, such as the product specification ("KR 150" – the two tokens "KR" and "150" are direct neighbors in the sentence) or the function owner ("conveyor belt"). But equally valid patterns would be word stems ending with "ation", verbal infinitives, or noun phrases. For example, the combination of a verbal infinitive for the operation and a noun phrase for the object suggests the existence of a function (in Fig. 2: "take crate" – the token "take" is an operation and the token "crate" is an object).

Discontiguous text patterns usually constitute facts. The occurrence of instances belonging to different concepts within sentence boundaries are supposed to be facts as shown in Fig. 3. By using the relations modeled in the ontology, senseless facts are identified and can be dismissed. For example, the analysis of the sentence "KR 150 is taking four crates from the conveyor belt." reveals two possible facts. The first fact obeys the pattern that a function owner (KR 150) performs an operation (take) on an object (crate). This relation is modeled within the ontology and can be proved by the ontology. The second fact states that a function owner (KR 150) executes a function which has another function owner as object (conveyor belt). A relationship of this type is not modeled in the ontology and the fact is therefore senseless. This example shows that the derived facts can be verified with respect to the ontology by using textmining algorithms.

The combination of semantic technologies and textmining capabilities is useful for analyzing knowledge hidden in unstructured data, e.g. in text documents, and for making that knowledge explicit. In the following sections, the application of the textmining capabilities within the Solution Assistant's search and annotation functions are presented.

6 Semantic Search Using the Solution Assistant

The Solution Assistant provides the possibility of searching for existing information stored in technical solution documents by using the Search Service of the PROCESSUS platform. This use of semantic technologies allows intuitive access to solution knowledge in the automation industry. A solution is described by its properties stored in the ontology. The knowledge worker searches for solutions based on more or less defined query strings. As shown in Fig. 4, the user asks for a solution to "handle bottles with the KR 150 in the food and luxury industry". This text string is passed to the system and it is analyzed. It is important to give the user feedback about the recognized concepts and instances from the ontology. Thus, the user is able to reconsider the search request and can modify it, if necessary. The result list displays the title of the solution document, the dominant function, with respect to the search string, and a short summary of the solution. The similarity feature, computed on the basis of the ontology, reveals a ranking which lists the best solution first. Using the relevance setting in the upper right corner, the user is able to set a threshold. Documents with a similarity lower than the threshold are disregarded.

In Fig. 4, the semantic filter bar is displayed at the right. This filter lists all instances related to the documents in the result list, ranked by their weight with respect to the ontology. This allows the user to browse through the content of the ontology by extending the filter list. The solution range can be effectively reduced by choosing the appropriate filter settings. If the user starts with a general query, the system offers a broad search result list. The user may inspect the displayed instances of the different concepts and can successively decrease the offered solution range. By applying various filter strategies, the user is able to extract several possible solutions which fulfill the same functionalities but utilize different technologies, e.g. the function transferring bottles can be executed by a robot or a conveyor belt. So, the system supports the process of product development by making existing products, technologies or services better available.

7 Semantic Annotation Using the Solution Assistant

In addition to semantic search, the Solution Assistant provides the possibility of relating process-relevant metadata with the solution documents. This is done by using the Annotation Service of the PROCESSUS platform. It is based on the Facts Extraction service and enriches solution documents with the appropriate metadata. Figure 5 shows a screenshot taken during the annotation process, when a document is uploaded to the platform.

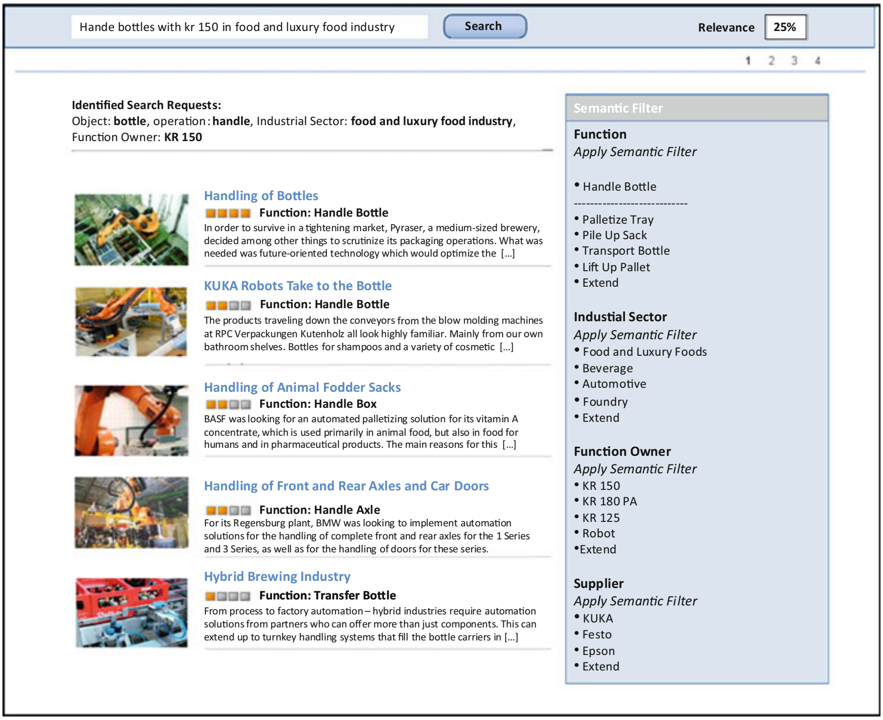


Fig. 4 Semantic search GUI in the PROCESSUS prototype

The system allows for document upload with a related picture describing the solution, as well as the option to view the document's content. Besides the "normal" metadata, such as title and description, the system must effectively support the creation of metadata with respect to the ontology. Therefore, during upload, the document is analyzed by means of textmining by the Facts Extraction Service and the Solution Assistant suggests appropriate recommendations for functions, industrial sector, function owner and supplier, as well as properties. The user must choose the correct annotation and the metadata is stored in the platform. By using the Annotation Service, the existing textual information about technical solution is enriched with the appropriate metadata in the ontology. The subsequent search then uses this metadata which, enables a more structured and detailed search.

8 Conclusion

The PROCESSUS use case has shown manifold possibilities for the use of semantic technologies for handling knowledge in a technical context. It has proven the successful application of the PROCESSUS platform for supporting the

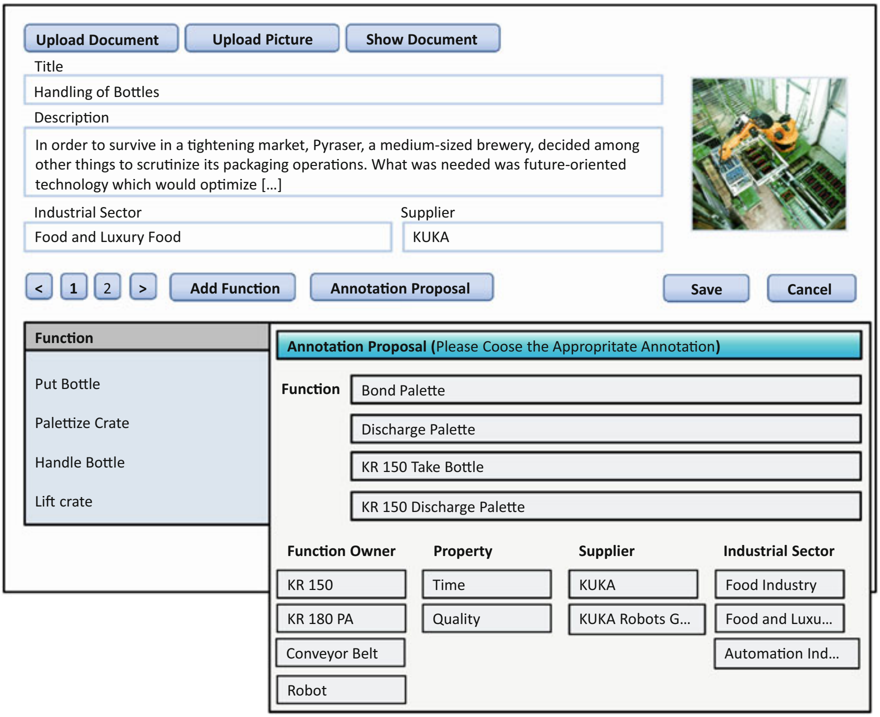


Fig. 5 Semantic annotation GUI in the PROCESSUS prototype

knowledge-intensive business processes in the automation industry. The application of the Facts Extraction Service, the Annotation Service, and the Search Service enables improved access to existing knowledge about technical solutions. In the course of the development of the PROCESSUS knowledge model, the significance of the involvement of the future user became particularly apparent. In order to understand the underlying knowledge base and make the best possible use of the hidden information, the user needs to understand the structure of the knowledge model. In addition, the use of language is a factor that has to be considered carefully. Language analyses performed on solution documents for the PROCESSUS knowledge model provided interesting results. In solution documents from different companies within the same industrial sector, significant differences in the use of language and terminology can be observed (Schmidt et al. 2011). Consequently, the software based on the knowledge model must be capable of supporting translation and mapping of different terms used for the same instances (normalization). This is relevant for both core functionalities of such software, semi-automated annotation and search.

These observations highlight a number of starting points for future research work. To expand the PROCESSUS perspective to the handling of knowledge throughout the entire product development process, a closer look can be taken at different models used during the process (Kohn et al. 2012). This research can provide recommendations for the choice of models which facilitate knowledge transfer to customers and knowledge reuse for future development processes. An additional focus of further work is the enhancement of knowledge models and model transfer usage to a cross-discipline application (Kaiser et al. 2012). Transferring results from disciplines, such as biology, to technical products supports the implementation of new technologies. In this case, differences in model use and their specific structure, and the use of language play an even more important role. The use of knowledge models is challenging, but provides new perspectives for better collaboration and cross-discipline implementation of technologies.

Beyond the described scenario in the field of automation industry, experiences made in PROCESSUS and the developed services of the semantic business integration platform can be transferred to many other sectors and business applications. The use of semantic technologies allows for exploitation of implicit knowledge in business processes by making it explicit and creating added value. For example, requirements in customer care and customer service applications can be supported by PROCESSUS technologies (Beinhauer and Kayser 2012). Semantic technologies provide a significant value in identifying process-relevant information. The topics and facts contained in a customer request can be identified with respect to a service ontology. Browsing through the knowledge repository in order to search for similar problems or customer requests is possible. A significant service benefit can be achieved for product lines with shorter cycle times and increased numbers of similar product variants. An integration with human resource management is well worth discussion, especially when shared services are considered. Due to organizational distribution to different systems, an ontology with its stored knowledge is the perfect common denominator needed to drive processes related to employee management.

References

M. Beinhauer, S. Kayser, Wissensmanagement im HR-shared service center, in *HR-Servicemanagement – Produktion von HR-Dienstleistungen*, ed. by W. Appel, W. Felisiak (Oldenbourg Wissenschaftsverlag, Munich, 2012), pp. 180–184

S. Bonnemeier, C. Ihl, R. Reichwald, Wertschaffung und Wertaneignung bei hybriden Produkten – Eine prozessorientierte Betrachtung. Technical report 3, Technische Universität München (TUM), Munich (2007), http://www.hyprico.de/publikationen/AP_Prozesse.pdf

F. Danzinger, C. Ihl, K. Meisel, R. Reichwald, *Die Zukunft von Internetplattformen im Maschinen- und Anlagenbau*, 1st edn. (VDMA Verlag, Frankfurt am Main, 2008)

A. Gaag, A. Kohn, U. Lindemann, Function-based solution retrieval and semantic search in mechanical engineering, in *Proceedings of the 17th International Conference on Engineering Design (ICED '09)*, ed. by M. Norell Bergendahl, M. Grimheden, L. Leifer, P. Skogstad, U. Lindemann, vol. 8 (The Design Society, Palo Alto, 2009), pp. 147–158

J. Gantz, D. Reinsel, Extracting value from Chaos (2011), http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf

D. Ihlenburg, *Interaktionsplattformen und Kundenintegration in Industriegütermärkten – Akzeptanzfaktoren, Wettbewerbsvorteile und Kundennutzen am Beispiel des Maschinen- und Anlagenbaus*, 1st edn. (Gabler, Wiesbaden, 2011)

M.K. Kaiser, H. Hashemi Farzaneh, U. Lindemann, An approach to support searching for biomimetic solutions based on system characteristics and its environmental interactions, in *Proceedings of the 12th International Design Conference (DESIGN '12)*, Dubrovnik (The Design Society, 2012), pp. 969–978

A. Kohn, M. Lutter-Günther, M. Hagg, M. Maurer, Handling product information – towards an improved use of product models in engineering design, in *Proceedings of the 12th International Design Conference (DESIGN '12)*, ed. by D. Marjanovic, M. Storga, N. Pavkovic, N. Bojcetic, Dubrovnik (The Design Society, 2012), pp. 1719–1730

C. Lüthje, *Kundenorientierung im Innovationsprozess: Eine Untersuchung der Kunden-Hersteller-Interaktion in Konsumgütermärkten*, 1st edn. (Deutscher Universitäts-Verlag, Wiesbaden, 2000)

R. Reichwald, F. Piller, *Interaktive Wertschöpfung: Open Innovation, Individualisierung und neue Formen der Arbeitsteilung*, 2nd edn. (Gabler, Wiesbaden, 2009)

R. Reichwald, K. Wegener, Controlling im Kontext Hybrider Produkte Controlling **20**(8/9), 467–472 (2008)

H.X. Schmidt, A. Kohn, U. Lindemann, Domain specific language in technical solution documents – discussion of two approaches to improve the semi-automated annotation, in *International Conference on Knowledge Engineering and Ontology Development (KEOD '11)*, ed. by J. Filipe, J.L.G. Dietz (SciTePress, Paris, 2011), pp. 159–166. http://dblp.uni-trier.de/db/conf/ic3k/keod2011.html#SchmidtKL11

K.R. Tuli, A.K. Kohli, S.G. Bharadwaj, Rethinking customer solutions: from product bundles to relation processes. J. Mark. **71**(3), 1–17 (2007)

Machining Intelligence Network: Data Mining and Semantic Search in Manufacturing Industry

Ines Färber, Sergej Fries, Götz Marczinski, Thomas Seidl, and Nils-Per Steinmann

Abstract The "Machining Intelligence Network" (MachInNet) project tackles the challenges to "unearthing" manufacturing knowledge from NC codes (numerical control codes), tool layouts and other manufacturing documents, and to making it accessible for daily use, e.g., for feature-based NC planning. A new approach using data mining algorithms and semantic search technologies makes it possible to reverse engineer data from different sources and make it available for explicit use with the help of a semantic, Internet-based knowledge network. The business rationale of MachInNet is to help SMEs (small and medium-sized enterprises) to manage a large variety of technologies, to avoid redundant engineering efforts, and to accelerate industrial engineering of mechanical parts.

1 Introduction

Industrial enterprises are well aware of the "treasure of machining intelligence" that their organizations hold. No wonder substantial progress has been made in recent years to bring to bear this knowledge the day-to-day business. Summarized with the buzzword "Digital Factory", advanced engineering systems, like simulation software to safeguard manufacturing processes or feature-based process planning to automate industrial engineering, have been created. These systems, based on respective databases, enable us to automatically create operation plans and NC programs. The question is: How are the respective databases built up? The manufacturing

I. Färber (✉) • S. Fries • T. Seidl
RWTH Aachen University, Aachen, Germany
e-mail: faerber@informatik.rwth-aachen.de; fires@informatik.rwth-aachen.de; seidl@informatik.rwth-aachen.de

G. Marczinski • N.-P. Steinmann
CIM Aachen GmbH, Aachen, Germany
e-mail: mc@cimsource.com; ste@cimsource.com

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_32,

knowledge needed for these databases is available, but only in the shape of NC programs.

Product Lifecycle Management (PLM) systems have been developed to serve as the database for the Digital Factory. Following the concept of "product-process-resource" they are supposed to provide all information for the industrialization of a product in a cross-linked manner. The flip side is that PLM Systems are very demanding with respect to data content because in order to really serve as the database for the Digital Factory, processes and resources need to be represented in the system. Why not ask the suppliers to provide the respective data? The information should be in their databases.

The challenges of feature-based planning and product lifecycle management have two things in common:

1. In each case large data pools are available.
2. The context of "manufacturing" predefines a semantic framework for the data.

The key question for the MachInNet consortium was: How could the fact that a sufficient supply of data is already available be used to build up a searchable database? In contrast to the "traditional" approach of predefining the necessary data structure (reference model) to create a database and then collecting the required data, the starting point of database design should be the data itself. Prospective solutions for that new approach may be found in the domain of data mining. Data mining is defined as the systematic application of methods to a data pool with the aim of recognizing new patterns (Han and Kamber 2006). The starting point is not a predefined reference model, but the data itself.

Especially for SMEs a third challenge needed to be tackled. SMEs do use high-tech NC machines, but they do not necessarily operate costly engineering software like the aforementioned PLM or feature-based planning systems. The build-up of a knowledge database is not enough in this context. What is needed is access to the manufacturing knowledge, a service rather than a software. This requirement led the MachInNet consortium to the idea of the machining intelligence network.

2 The New Approach of Data Mining

To support manufacturing engineering the concept of "product-process-resource" to cross-reference all data relevant for the industrialization of mechanical parts needs to be put in practice. We talk about product data (work piece including the specification of the material), process data (of machining process) and resources (tools and production equipment). The good news is: Data of all three domains is available. Every supplier of tools for machining operations has drawings of nearly each tool. Any manufacturing operation has 500–2,500 NC programs available, in many cases even much more. Further, a lot of tool layouts and tool lists related to the NC programs can be found. The bad news: NC programs are no more than a sequence of path and switch commands. Sometimes there might exist a line of comment or "readable" subroutines, but otherwise the manufacturing know-how exists only implicitly in a "paint-by-numbers" fashion.

Which Data Mining Algorithms are Appropriate? The goal of data mining is to recognize patterns in large data pools that possibly no engineer might have thought of before. The next step is to assign a meaning (semantic) to these patterns. This is done either by "human-machine" interaction (an expert assigns a meaning to each pattern), or it is achieved automatically, as the data mining algorithm conducts a matchmaking of the recognized patterns with well-established patterns. Therefore you need a structured pool of as many data as you can get as a reference. In the case of tool data this was the CIMSOURCE tool database with more than 350,000 tool items. In the case of NC programs MachInNet used the extensive databases of selected EXAPT users.

Simply stated, the challenge is to transform inhomogeneous data pools into machine-readable formats. To meet this challenge MachInNet's research work had to distinguish graphics and NC programs as two fundamentally different data sources.

Data Mining in Tool-Graphics: A substantial data source from the supplier's side are technical drawings, ideally as CAD files (mainly DXF), in some cases digital or printed catalogues and complementary Excel charts. To generate explicitly useful data from these sources, a two-step approach proved to be effective. The first step is necessary for classifying the graphics. The reference to a known tool category makes it much easier in the second step to search for the data to be extracted. Because of the classification you "know" what to look for.

To classify tool drawings, methods for the similarity search in graphical structures are applicable. As part of the MachInNet project three different methods

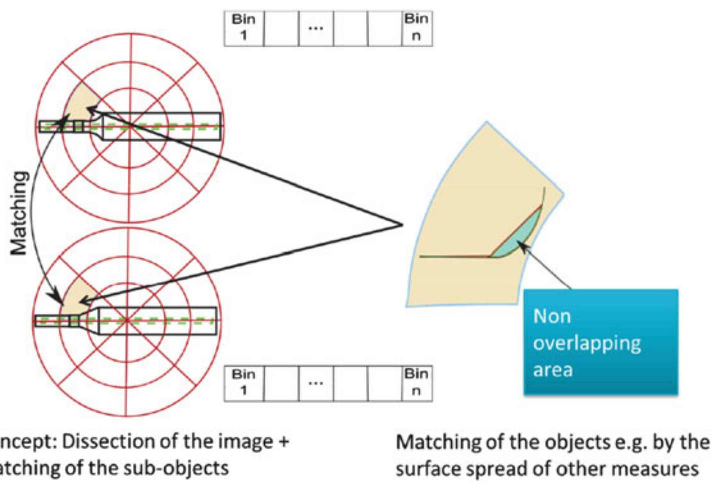


Fig. 1 Concept of the FormHist-approach (Marczinski and Steinmann 2012)

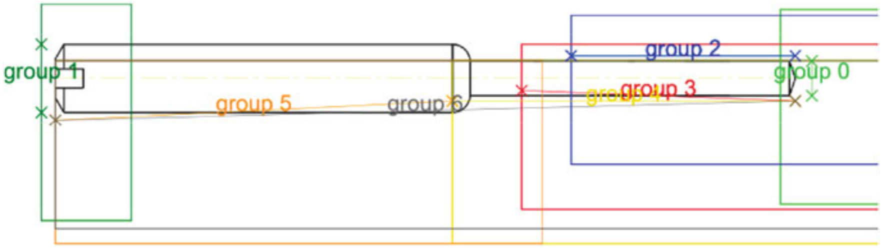


Fig. 2 Data extraction based on clustering

have been investigated. Two promising approaches (*EMDPol*, *VisWords*) rely on the dissection of a tool into a number of subcomponents. Another approach (FormHist) relies on the dissection of the graphics, which are then subdivided in smaller sections by the help of a grid. Thereby the comparison of two tools is reduced to a number of more focused and localized comparisons, leading to a reduced runtime of the data mining algorithms (Fig. 1). The FormHist approach proved to be the most appropriate for classifying the data pools of tool drawings.

For the automatic extraction of parameters from the categorized drawings a clustering method was applied to the dimensioning sections of the drawings. The clue of clustering is an overlay of already categorized drawings of the CIMSOURCE database with similar drawings according to the FormHist classification. Dimensions with similar position and orientation make up a cluster and may be referenced to the target structure (Fig. 2).

Knowledge-Extraction from NC programs: The analysis of NC programs is based on the widespread concept of features. A feature represents a machining object, which is created with a single tool. Simple features such as a threaded hole should be distinguished from complex features, such as a plane milling operation. The recognition of simple features is fairly easy, because there is a one-to-one relation of NC path and feature. Rule-based methods are used to identify this kind of feature.

The automatic recognition of complex features relies on the fact that regardless of the kind of feature, each feature is created by narrowly adjacent machining paths (Fig. 3).

To search interrelated regions in an NC program, we use the density-based clustering algorithm DBSCAN (Ester et al. 1996) in MachInNet. DBSCAN delivers clusters, whose objects are interrelated via a density criterion. Through the iterative application of the algorithm all relevant NC paths are collected and thereby the feature is identified. With the supportive information of the tool diameter and under the assumption that the work piece has a plane surface, the machined object may be extracted.

In addition to the recognized features, machining parameters like speed, feed and work piece materials are extracted. The inherent technological knowledge for solving a specific machining task is represented by exactly this combination of

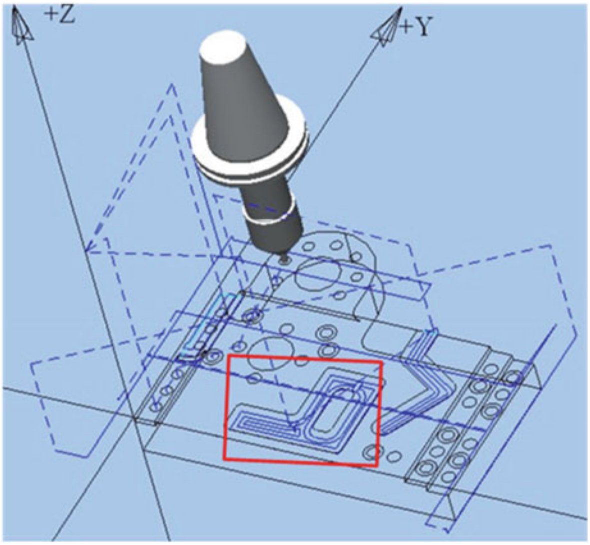


Fig. 3 Marked in *red* is a complex feature, which is generated by a sequence of machining paths (Marczinski and Steinmann 2012)

parameters describing the tool and cutting grade, the work piece material, the generated geometry and the cutting conditions (speed, feed, ...).

3 Business Cases

The results of MachInNet provide manufacturing operations the methods to solve day-to-day problems fairly pragmatically. This includes:

- Transfer of “NC encoded” manufacturing technology to new machines to shorten the ramp-up;
- Search for existing solutions to actual machining tasks to avoid “reinventing the wheel” and to promote standard processes;
- Consolidation of the technology know-how from different digital sources into a single source as foundation for the Digital Factory;
- Build-up of a technology database (feature library) to speed up industrial engineering and raise the quality level of machining processes;
- Standardization of machining processes and deproliferation of the range of cutting tools in use.

The pilot applications of MachInNet at four manufacturing companies ranging from aerospace to capital equipment manufacturing as well as at two manufacturers of cutting tools proved the value of the research results.

Proof of Concept with Tool Data The “legacy” data pools of two tool manufacturers were used as a test bed for the methods and algorithms developed by MachInNet.

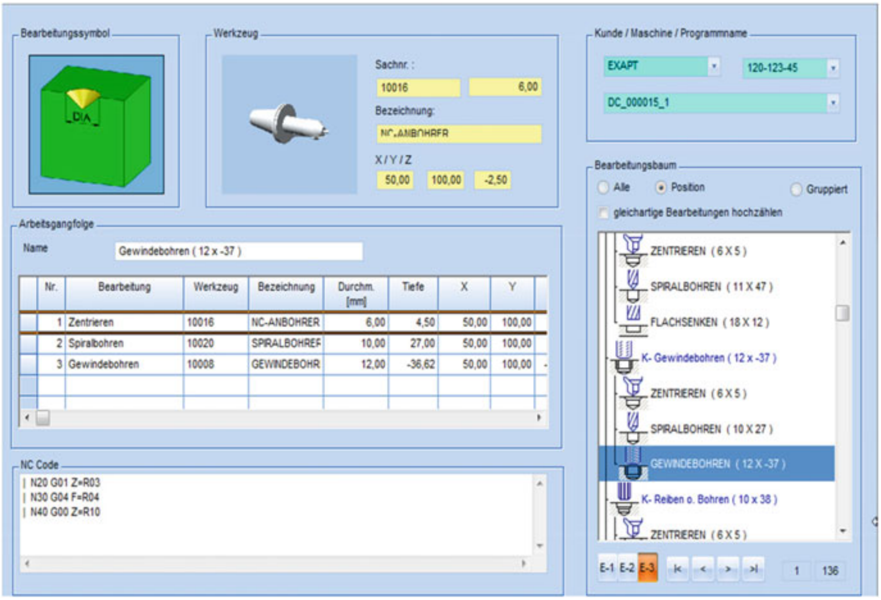


Fig. 4 The NC program analysis yields explicitly usable manufacturing know-how (Marczinski and Steinmann 2012)

The test case showed that the manual effort of data preparation could be reduced to 20 % of the former level. And the ultimate check for data quality with the help of the ToolsUnited Internet platform was also positive. The quality of the extracted data is sufficient to populate generic CAD models, so 3D CAD models could be generated automatically (Marczinski and Steinmann 2012).

Reverse Engineering of NC Programs in Practice To evaluate the algorithms and methods developed by MachInNet, four companies from different branches provided extensive sets of NC programs. In addition to that, the related tool information was provided in the form of tool lists (Excel, formatted lists, . . .), and presetting data was provided from the shop floor.

Tool information is important for the evaluation of NC processes because they provide additional information. Therefore it is possible to make use of the interplay of the various data mining methods. Comparable to the approach of dragnet investigation, the information extraction might be narrowed down step by step. If, for example, the data mining algorithm during the NC program analysis "assumes" a milling operation, the tool analysis "cross-checks" indications about whether the respective tool number relates to a milling tool. The results of the test case have been imported to the Exapt-FDO database, which facilitates the display of a list of single operations and of the frequency of occurrence of specific operation sequences. It is possible to compare cutting parameters for similar machining features and thereby help to standardize machining processes (Fig. 4).

The problem that NC programs for nearly identical machining processes may turn out very different depending on the product to be machined, the machine tools applied and the cutting tools in uses will become a matter of the past.

4 Conclusion

As part of PROCESSUS the MachInNet project focuses on industrial engineering for mechanical parts, a business process crucial for the competitiveness of manufacturing companies. MachInNet uses the basic concepts of PROCESSUS and adds specialized data extraction mechanisms for the manufacturing domain.

The data driven approach of MachInNet eases the setup of technology databases substantially. Already existing data pools may be used to generate value-adding know-how.

With the help of various data mining methods it is possible to unearth knowledge and to feed advanced engineering systems with machine-readable information. This way the “treasure of experience” is surfaced and may be secured and extended independently of the current IT systems for engineering and programming and beyond the current generation of machine tools and engineering experts.

Tangible results of the first pilot applications have been achieved in the following areas:

- Stabilization, quality increase and certification of machining processes (e.g., safety-related parts, such as in the aerospace industry),
- Cost reduction through the deproliferation of machining processes,
- Reduced effort to qualify new employees.

In one respect the MachInNet consortium needed to cut back from the original goals. The pilot projects and the demonstrators that have been developed were all tested offline. So the next step will be to overcome the reluctance of manufacturing companies to use web-based services to solve engineering problems. The fear that competitors might benefit from their own knowledge or even infringe intellectual property rights held industrial businesses back from using the Machining Intelligence Network online.

As part of PROCESSUS the MachInNet project focuses on industrial engineering for mechanical parts, a business process crucial for the competitiveness of manufacturing companies. MachInNet uses the basic concepts of PROCESSUS and adds specialized data extraction mechanisms for the manufacturing domain.

References

M. Ester, H.P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD’96)*, Menlo Park, ed. by E. Simoudis, J. Han, U. Fayyad (AAAI, 1996), pp. 226–231

J. Han, M. Kamber, *Data Mining – Concepts and Techniques*, 2 edn. (Morgan Kaufmann, San Francisco, 2006)
G. Marczinski, N.P. Steinmann, Von Data Mining und semantischer Suche – Neue Wege zum Aufbau von Feature-Bibliotheken und Technologiedatenbanken. Zeitschrift für wirtschaftlichen Fabrikbetrieb (ZwF) **107**(1–2), 82–86 (2012)

Further Reading

C. Beecks, I. Assent, T. Seidl, Content-based multimedia retrieval in the presence of unknown user preferences, in *Advances in Multimedia Modeling*. Volume 6523 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2011a), pp. 140–150
C. Beecks, J. Lokoc, T. Seidl, T. Skopal, Indexing the signature quadratic form distance for efficient content-based multimedia retrieval, in *Proceedings of the 1st ACM International Conference on Multimedia Retrieval (ICMR'11)*, Trento (ACM, Berlin/Heidelberg/New York, 2011b)
C. Beecks, S. Wiedenfeld, T. Seidl, Improving the efficiency of content-based multimedia exploration, in *20th International Conference on Pattern Recognition (ICPR'10)*, Istanbul (IEEE, Washington, DC, 2010), pp. 3163–3166. http://dblp.uni-trier.de/db/conf/icpr/icpr2010.html#BeecksWS10
S. Günnemann, I. Färber, E. Müller, I. Assent, T. Seidl, External evaluation measures for subspace clustering, in *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM'11)*, Glasgow (ACM, New York, 2011), pp. 1363–1372
P. Kranen, S. Günnemann, S. Fries, T. Seidl, MC-Tree: improving bayesian anytime classification, in *Scientific and Statistical Database Management*. Volume 6187 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2010), pp. 252–269
Y. Rubner, C. Tomasi, L. Guibas, A metric for distributions with applications to image databases, in *6th International Conference on Computer Vision (ICCV'98)*, Bombay (IEEE, Washington, DC, 1998), pp. 59–66
J. Sivic, A. Zisserman, Video Google: a text retrieval approach to object matching in videos, in *Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV'03)*, Nice, vol. 2 (IEEE, Washington, DC, 2003), pp. 1470–1477

Integrated Service Engineering (ISE)

Holger Kett, Matthias Winkler, and Kay Kadner

Abstract Service Engineering comprises many different aspects of defining and creating a service. This especially applies to the Internet of Services (IoS) domain as it is understood by TEXO, where services are not limited to a certain web service technology, abstraction level of functionality or application domain. In this context, service providers are increasingly forced to cooperate to strengthen their core competencies and to create partner networks to offer new services. In order to facilitate the collaboration process, business ecosystems are built which allow service providers to offer their own, as well as combined, electronic services with other providers over the Internet. In TEXO, the Integrated Service Engineering methodology (ISE) has been developed which provides a comprehensive, interdisciplinary, and model-driven approach to electronic service development in business ecosystems. One major aim of ISE is to integrate business as well as technical stakeholders in the electronic service development process and support them with appropriate models, methods, and tools.

1 Introduction

Software development as an engineering discipline relies on formally defined models and methodologies. Models serve as development artifacts which can explicitly be measured, validated and adapted. Methodologies, on the other hand, control the software development process (Weisbecker 2002). Balzert (2000), for

H. Kett (✉)
Fraunhofer Institute for Industrial Engineering (IAO), Stuttgart, Germany
e-mail: holger.kett@iao.fraunhofer.de

M. Winkler • K. Kadner
SAP AG, SAP Research, Dresden, Germany
e-mail: matthias.winkler@sap.com; kay.kadner@sap.com

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_33,

example, introduces methods and models which strongly focus on the development of software products. However, software is increasingly offered as a service (König et al. 2006; Lorenz et al. 2007; Weiner et al. 2010), so new models and methodologies have to be developed to cope with the changed requirements.

The TEXO use case created an integrated approach for model-driven service development, offering and trading services via a web-based service marketplace, and executing them on a hosted service platform (see Barros and Dumas 2006; Janiesch et al. 2008; Winkler et al. 2009).

Prior to designing the service processes and the service concept, a service provider or a consortium of service providers needs to agree on the services which are offered over the Internet and the delivery process. Those aspects are designed within a service concept (see Bullinger et al. 2003), or a business model (see Kett et al. 2009; Osterwalder 2004), and further refined by a business process model and an IT-oriented service concept. The Integrated Service Engineering ISE methodology has been developed within TEXO to support the engineering process of services in the context of the Internet of Services and reaches from a strategic business model perspective to a concrete IT-oriented perspective (Kett et al. 2009, 2008).

2 Challenges of Bridging the Business and Technical Worlds of Electronic Service Development

Information Representation Nowadays, organizations generate a wealth of information describing strategies, goals, business processes, standards, IT infrastructures, etc. This information, which can be classified as unstructured, semi-structured, and structured, constitutes a precious source for developing software as a service, a so-called electronic service. The set of all the information produced by an organization is often termed as organizational memory (de Vasconcelos et al. 2003). Unstructured information usually characterizes documents such as Word files, spreadsheets, and presentations. Semi-structured information is generally associated with data instances which do not conform to a formal data model but still contain a clear structure, such as XML. Finally, structured information can typify relational databases and formal models.

Depending on their background and skills, people tend to use different structures to define information. For example, business stakeholders tend to use unstructured data to describe future strategies, financial data, SWOT analysis (Strength, Weaknesses, Opportunities, Threats), Porter's Five Forces (a framework for business strategy development), and balanced scorecards. On the other hand, IT professionals usually rely on semi-structured or structured information to describe formal models such as UML activity diagrams, MOF (Meta-Object Facility) and BPMN processes (see Fig. 1).

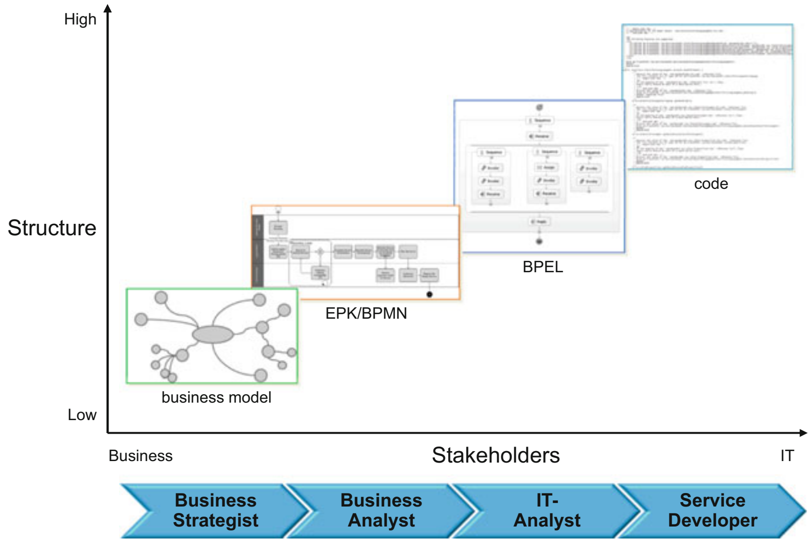


Fig. 1 Levels of structured information (Adapted from Kett et al. 2008)

Unstructured information drives numerous business processes but often organizations cannot leverage this information efficiently, leading to inconsistent communication between stakeholders, duplication of effort, poor decision-making, and higher costs (Herschel and Jones 2005). The differences between the various structure levels that are used to embody information create a representational gap that needs to be closed in order to unify and bridge the business and IT perspectives on electronic services. If this gap is not closed, unstructured information generated by business stakeholders will not be available to the IT professionals for electronic service development.

Distinct Contexts and Views Business and technical professionals have invariably different backgrounds, skills and mindsets. This fact makes the alignment of the business and IT perspectives on electronic services a challenging task. On the one hand, the concerns of business and IT perspectives are distinct. Business professionals are concerned with aspects related to financial assets, marketing missions, return-on-investment, competition, value chains, and SWOT analyses. On the other hand, technical people are concerned with functions, data, components, integration, compatibility, programming, scalability, API, concepts and their relationships, formal models, UI, and rules. As a result, it is common that different people use the same word to describe distinct entities and distinct words to refer to the same entity (Fensel et al. 2002).

A suitable methodology for electronic service development can use the notion of ontology to aggregate distinct models (e.g. SWOT analysis, Porter's Five Forces, balanced scorecards, UML activity diagrams, BPMN, etc.) from distinct communities to represent and manage both organizational information containers and contents. This option allows the representation of information in a way that it facilitates knowledge sharing and reuse between the stakeholders involved in the electronic service development process.

Model Integration and Model-Driven Architecture IoS-based services require not only a specification of their internal functioning, the data being handled, and their interfaces, but also specific information for trading them on a service marketplace, such as pricing and legal aspects. Another important point is that the development of a service is a complex process starting with an idea and a business model. During different development phases additional information about a service is captured, refined and transformed, i.e., starting from a business problem to be solved an abstract business process is created and later transformed into a concrete process implementation. Finally the development process results in a concrete software product. The complex nature of IoS-based services requires stakeholders from different disciplines (e.g., business, legal, technical) to work jointly in the development process.

In order to manage this complexity, a model-based development approach was proposed in TEXO, where different models capture distinct information about a service and thus help to break down its complexity. A model may refine the information captured in another model during the transformation from business idea to a concrete business process implementation. It may also represent different aspects of a service, i.e., its internal process and pricing or legal information. Each model is developed by specific experts and is designed to fit their specific needs.

Information in different models can be overlapping and changes in one model may require changes in one or more other models. Therefore, a mechanism for synchronizing information in different models was needed. In order to achieve that, the different models are integrated via model transformations, which ease the development process.

Tool Support One of the challenges in the area of electronic service development is to make suitable tool support available to engineers. In order to provide a common approach, one common set of tools should be used by the different types of stakeholders. Since these tools are used by various stakeholders with different perspectives (i.e., business and technical perspectives), they should provide a global view on the models that are used to describe electronic services. In order to support collaborative development of an electronic service a common repository is needed. This includes shared access to multiple models in the repository, i.e., every participant involved in service engineering should see the same version of a service's models. This enables an instant synchronization (and visualization) of changes.

Such sets of tools should not only provide shared model access, but should also simplify the development process by providing assistance and guidance where possible. Some tools might be incompatible to others, preventing the use of unified resources. Therefore, the construction of roadmaps is another challenge. Roadmaps inform users about the tasks that need to be carried out, when, and how. This way, they provide guidance on the sequence of models that should be modeled, and which suitable tools to use. In this context, the most challenging task is creating an association between the development, innovation, design, and implementation phases, and the integration of different representation formalisms such as unstructured information (e.g. natural language) and structured information (e.g. formal MOF models).

3 Integrated Service Engineering (ISE) Framework

The ISE framework in TEXO is based on the approach of the Zachman Framework (Zachman 1987) and supports the four phases of business model development, requirements analysis, service design, and service implementation. These phases are assigned to different abstraction layers (perspectives), i.e., strategic, conceptual, logical, and technical (see Fig. 2). Therefore, the ISE framework provides selected data models and methods for the specific players of each layer. The models can be replaced by other adequate models if they better suit the requirements of an organization. In order to reduce the complexity, each layer is based not only on one complex model for the whole layer, but is also broken down into five dimensions and corresponding models: service description, business processes, human resources, data assets, and business rules.

A dimension is a part of an abstraction layer and consists of its own models (artifacts). The artifacts of each dimension can be transformed from one abstraction layer to another so that information created on a higher level of abstraction can be transformed to a lower level of abstraction. Nevertheless, automatic transformation reduces the required effort by users, who otherwise would have to do the transformation manually. The ISE framework is built in line with the following rules (Kett et al. 2008):

- The dimensions do not need to be analyzed and modeled in a specific order.
- All dimensions of one layer are combined into a complete model for the layer-specific stakeholders.
- Each abstraction layer features a simple, basic and unique model.
- The layers are separated and clearly differentiated (no overlap).
- In order to achieve an integrated approach, information of the artifacts is transformed from one layer model to another and back (iterative approach).

Figure 2 illustrates that the degree of business orientation in the innovation, strategic, and conceptual perspectives is high and decreases in the next two perspectives when the electronic services are technically designed and implemented. In the runtime perspective, the electronic service is launched and offered on the market and, thus, business and technical aspects increase in importance. The four abstraction layers of the ISE methodology are described below. Many, but not all models could be evaluated within TEXO. The models which are mentioned are considered as recommendations and can be exchanged with models which better suit certain organizational contexts.

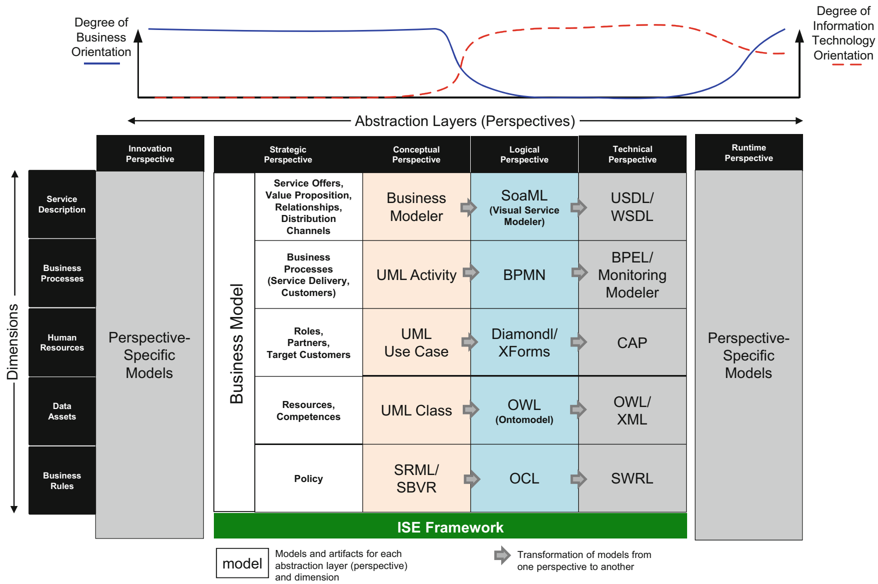


Fig. 2 ISE Framework (Adapted from Kett et al. 2008)

Strategic Perspective ISE starts with the development of the strategic perspective of an electronic service. The main aim of this perspective is the identification of market and business requirements for an electronic service, which are derived from the market, market players, competitors, environment, company's strategy, organization, and capabilities. The market and business requirements are identified with established business methods which are well-known among business strategists, e.g., Porter's five forces model, resource-based view on strategy, SWOT, cost-profit analysis, BCG matrix (market growth versus market share), Porter's competitive strategies, the Porter analysis of the value chain, the four Cs – customer, competition, cost, and capabilities, and the four Ps – prize, product, place, and promotion. The semi-formal results of these methods are then transferred into a formal business model (see Kett 2011; Kett and Franczyk 2010; Kett et al. 2009; Osterwalder 2004). The main elements of the business model are: the service (e.g., value proposition), the customer interface (e.g., target customer, distribution channel, and customer relationship), the infrastructure management (e.g., processes, resources, competences, partners and business rules), and the financial aspects (e.g., cost structure and revenue model). The latter, is considered in the above mentioned references and is added as an additional row in the ISE framework.

Prior to developing the strategic perspective, innovative service ideas are identified and evaluated with the target of filtering service ideas with a high business potential. In the strategic perspective, only the remaining service ideas are refined.

We propose the following models for the formation of the strategic layer:

- *Service description* – the electronic service itself (value proposition), the way in which the electronic service is marketed and sold (distribution channels), the relationships to customer groups, and the revenue model are parts of the business model ontology that are the focus of this dimension.
- *Business process* – this dimension stresses the infrastructure management of a developing electronic service, and, thus, the issues of how an electronic service is delivered (processes, partners and business rules) as well as what know-how and resources are needed for the delivery (resources and capabilities).
- *Data assets* – a list of relevant data that is required for the delivery of electronic services.
- *People* – the people model lists the relevant stakeholders who are involved in the delivery of the electronic service and its consumption (target customers, actors, and partners).
- *Business rules* – this dimension lists the main policies that need to be considered and refined during the development of an electronic service and delivery process.

In this abstraction layer, a strategic business model is developed including a coarse-grained service concept. This perspective includes a strategic decision-making process which is usually conducted by business strategists, business developers, and the management of an enterprise.

Conceptual Perspective The conceptual perspective follows the strategic perspective and refines its results. Whereas the business model describes the exchange of value between business partners, the process model shows operational characteristics of how the value exchange is conducted (Dorn et al. 2007).

Thus, during the conceptual perspective, business analysts have the task of operationalizing and implementing the strategic artifacts that were developed during the strategic perspective. Business analysts have knowledge about the market the electronic service will be offered in, including important processes, organizational structures, crucial assets, and domain constraints. They have the responsibility of analyzing and expressing requirements for the electronic service from their perspective. Additionally, business analysts know about models, modeling, architecture, and transformation. Their responsibility is to transform the domain experts' perceptual requirements into diagrams. These diagrams serve as a basis for communication so that domain experts can agree or disagree with the electronic service design analysis (Zachman 1987).

The models for the conceptual layer must comply with the following requirements:

- *Service description* – the Service Business Modeler tool uses a Domain-Specific Language (DSL) (Gronback 2009) in order to document service properties. The properties can include the illustration of electronic services (Baida et al. 2004) in terms of functionality (Oaks et al. 2003), and monetary, quality, legal, and security aspects (O'Sullivan 2006). On this basis, potential service consumers search, rank, compare, select, and substitute electronic services. An electronic service can also be described by the service property ontology introduced by O'Sullivan (2006) and classification approaches such as UNSPSC[1] or eCl@ss.[2]

- *Business process* – the aim of the workflow dimension is to show the electronic service behavior. It presents tasks in sequence or in parallel that need to be carried out in order to fulfill electronic service functionality. Tasks represent a company's capability. Additionally, workflow models make use of data, people, and rule concepts (Bhattacharya et al. 2007; zur Muehlen et al. 2007). Appropriate model notations for the workflow dimension include event-driven process chains (EPC) (Scheer et al. 2000), and the UML activity diagram.[3]
- *Data assets* – data conceptualizes information, which is necessary to perform the different tasks described in the workflow model. Information comes from within organizations (intangible resources) and from outside organizations. Data can be described using the UML class diagram and entity-relationship diagrams (Chen 1977).
- *People* – the people dimension embodies organizations (actors), organizations' hierarchies, and roles within organizations. People own capabilities and carry out or are responsible for tasks within a workflow. Model notations for the people dimension are the UML use case diagram and organizational charts.
- *Business rules* – the aim of rule descriptions is to implement rules and policies with which an electronic service must comply. Model notations for the rule dimension include Semantic Rule Model Notation (SRML) and Semantic Business Vocabulary (SBVR) (zur Muehlen et al. 2007).

Logical Perspective The logical perspective refines the conceptual perspective. This perspective acts as the bridge between business requirements and technical specifications. IT architects are responsible for translating conceptual models into formal models without any platform-specific dependency. This allows the generation of service implementations for different technological platforms. In the following, requirements and appropriate models for the logical perspective are presented:

- *Service description* – the logical service description consists of an abstract interface describing the services' different capabilities, interactions, business and technical properties, and provider information. SoaML (Amsden 2006) is a UML profile that offers a graphical notation in order to model service-oriented architectures along with services and their interfaces, actors, and dependencies. It served as a base for the Visual Service Modeler.
- *Business process* – workflows or processes describe the services' behavior. Various workflow and process modeling notations exist. Business process modeling

[1] http://www.unspsc.org

[2] http://www.eclass.de

[3] http://www.omg.org

notation BPMN[4] offers graph-based notation for the modeling of processes, actors, activities, as well as the sequence logic.

- *Data assets* – for formalizing services' data, the Ontomodel is applied. The Ontomodel uses OWL ontologies[5] to describe the data. Alternatively, the Core Component Technical Specification (UN/CEFACT 2009) may be applied. The specification offers semantics and a methodology for designing data assets in a technology-agnostic fashion and links well with UML class diagrams and XML schema definitions.
- *People* – for modeling, the people dimension for the logical perspective, the ISE framework, may include the XForms specification (Boyer 2009). It offers a means to describe form data that is intended for usage as services' input and output. These forms can then be used for HTML websites or other technical target platforms. Alternative models may be models which focus on the description of user interfaces, such as Diamodl (Trætteberg 2007).
- *Business rules* – business rules, also known as behavioral knowledge, are represented as constraints over data entities. The ISE framework makes use of the object constraint language (OCL[6]).

Technical Perspective The technical perspective concludes the ISE framework. This perspective is always dedicated to one specific technology platform. Consequently, IT developers must choose realization technologies for the platform accordingly. The ISE framework is valid for every platform. However, in this text, web service technology is the chosen target platform. IT developers use logical models as input and create IT artifacts in order to implement services. The following technologies may be used for doing so:

- *Service description* – the web service description language (WSDL) (Chinnici et al. 2007) comprises information about operations, inputs and outputs as well as interfaces of a service. However, in order to trade services via a service marketplace, further information such as pricing model and payment options, legal information, available service bundles, and quality of service is needed. To support that, the Unified Service Description Language USDL was developed within TEXO[7] and integrated into ISE.
- *Business process* – services' behavior may be implemented with different technologies, including programming languages such as Java or .NET, and process execution languages such as BPEL (Jordan and Evdemon 2007) and XPDL (van der Aalst 2004). These languages are executed by application servers and workflow engines. In order to monitor the business processes which are executed, the Monitoring Modeler was developed within TEXO which supports the creation of event-based monitoring models. The Monitoring Modeler is based on the

[4] http://www.bpmn.org/

[5] http://www.w3.org/TR/owl2-overview/

[6] http://www.omg.org/spec/OCL/2.0/

[7] http://www.w3.org/2005/Incubator

BPEL-process and applies event-based rules which concern the monitoring of key performance indicators (KPIs) of the categories quantity, quality, feedback and finance (Vidačković et al. 2010).

- *Data assets* – for data, the ISE framework relies on XML Schema Definitions (XSD) or OWL ontologies. It links in well with the core components' technical specification that is used in the logical perspective as well as with the WSDL for the service description.
- *People* – to define the user interfaces and how humans interact with services, the ISE framework makes use of the Canonical Abstract Prototypes (CAP) notation (Constantine 2003). Alternatives are HTML forms that are generated from XForm specifications. These forms allow users to provide input for service execution and to display services' results.
- *Business rules* – the rules can be implemented in the Semantic Web Rule Language (SWRL), which combines OWL and RuleML.[8]

4 Implementation: The ISE Workbench

In order to apply the ISE framework in practice, a tool was developed that incorporates all necessary models as well as transformations between them (Scheithauer et al. 2011). The ISE workbench provides one platform for four distinct roles: business strategist, business architect, IT analyst, and IT developer. Furthermore, the workbench offers tools to model five distinct service aspects: service description, business processes, data assets, people, and business rules. The service design utilizes business process modeling and its notations such as BPMN and UML, whereas the service implementation relies on web service standards including BPEL and WSDL as well as USDL. ISE workbench builds on Eclipse's rich client platform (RCP), which allows the integration of existing tools as well as offers a platform for novel tool development. The tool embodies a total number of 20 editors in order to model the five service aspects for each of the four roles. OMG's query view transformation (QVT) specification is the basis for model transformation implementation, e.g. BPMN to BPEL.

5 Conclusion

Since the concept of electronic services that are offered and executed over the Internet is still in a ramp-up phase, no customized methodologies exist for the engineering of those services. As a result, it is inevitable that front-runners will carry out electronic service planning and development in a manual, ad hoc,

[8] http://www.w3.org/Submission/SWRL/

subjective, time-consuming, and error-prone fashion. This will lead to electronic service solutions that are disorganized, behind schedule, over budget, or cancelled. Therefore, the development of the ISE framework provides key design practices and artifacts to the planning and development of electronic services. ISE brings a set of benefits to the creation of electronic services:

- *Active engagement*: All stakeholders (e.g., business strategists, business analysts, IT analysts, and IT developers) who can influence the engineering of electronic services participate in the development to guarantee that business objectives are accomplished. ISE supports those stakeholders by providing specific perspectives on the electronic service development.
- *Model-based approach*: The ISE framework supports the engineering of electronic services by a set of models that describe or specify service structure, functionality, and behavior. Each model is assigned to specific stakeholders and to specific dimensions. The different models are synchronized via model transformations. Thus, an evolutionary change in one model triggers changes in adjacent models.
- *Compliance with regulations*: Since ISE strongly relies on formal models to design electronic services, it provides a paradigm that may guarantee that architected electronic services comply with relevant laws, policies, and regulations. For example, ISE provides information on billing, payment, security, and confidentiality. Therefore, financial information and security policies can be easily checked.
- *Technology standardization*: Electronic services must conform to existing standards in order to be compatible with data, applications, services, communication, integration, and security. Since the models devised within ISE and the main output of ISE are a set of technical standards, formal models of those services are independent of specific technological choices and thus of platforms.
- *Clear semantics*: ISE relies on two constructs to provide semantics for electronic services: using models and using a business ontology. The set of models that is made available by ISE provides strong semantics to stakeholders due to their high level of familiarity and standardization. On the other hand, the business ontology allows the construction of a common vocabulary of concepts (data) consistently throughout the organization, which is shared across perspectives and dimensions and is accessible for all models to perform their functions.

References

J. Amsden, UML profile and metamodel for services RFP UPMS services metamodel. Technical report, Object Management Group (OMG), Sept 2006, http://www.omg.org/news/meetings/tc/special-events-dc/soa-pdf/Amsden-SOA_Profile_RFP.pdf

Z. Baida, J. Gordijn, B. Omelayenko, A shared service terminology for online service provisioning, in *Proceedings of the 6th International Conference on Electronic Commerce (ICEC'04)*, Delft (ACM, New York, 2004), pp. 1–10

H. Balzert, *Lehrbuch der Software-Technik*, 3 edn. (Spektrum Akademischer Verlag, Heidelberg, 2000)

A.P. Barros, M. Dumas, The rise of web service ecosystems. IT Prof. **8**(5), 31–37 (2006)

K. Bhattacharya, C.E. Gerede, R. Hull, R. Liu, J. Su, Towards formal analysis of artifact-centric business process models, in G. Alonso, P. Dadam, *Business Process Management*. Volume 4714 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2007), pp. 288–304

J.M. Boyer, XForms 1.1. W3C proposed recommendation, 2009, http://www.w3.org/TR/2009/PR-xforms11-20090818/

H. Bullinger, K. Fähnrich, T. Meiren, Service engineering methodical development of new service products. Int. J. Prod. Econ. **85**(3), 275–287 (2003)

P.P. Chen, The entity-relationship model: a basis for the enterprise view of data, in *Proceedings of the National Computer Conference (AFIPS'77)*, Dallas (ACM, New York, 1977), pp. 77–84

R. Chinnici, J. Moreau, A. Ryman, S. Weerawarana, Specification: web services description language (WSDL) version 2.0 part 1: Core Language, June 2007, http://www.w3.org/TR/2007/REC-wsdl20-20070626/

L.L. Constantine, Canonical abstract prototypes for abstract visual and interaction, in *Interactive Systems: Design, Specification, and Verification*. Volume 2844 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2003), pp. 1–15

J. de Vasconcelos, C. Kimble, A. Rocha, Organisational memory information systems: an example of a group memory system for the management of group competencies. J. Univers. Comput. Sci. **9**(12), 1410–1427 (2003)

J. Dorn, C. Grun, H. Werthner, M. Zapletal, A survey of B2B methodologies and technologies: from business models towards deployment artifacts, in *Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS'07)* (IEEE, Washington, DC, Waikoloa, Hawaii, 2007), pp. 1–10

D. Fensel, C. Bussler, Y. Ding, V. Kartseva, M. Klein, M. Korotkiy, B. Omelayenko, R. Siebes, Semantic web application areas, in *Proceedings of the 7th International Workshop on Applications of Natural Language to Information Systems (NLDB'02)*, Stockholm, June 2002, pp. 27–28

R.C. Gronback, *Eclipse Modeling Project: A Domain-Specific Language Toolkit*, 1 edn. (Addison-Wesley Longman, Amsterdam, 2009)

R.T. Herschel, N.E. Jones, Knowledge management and business intelligence: the importance of integration. J. Knowl. Manage. **9**(4), 45–55 (2005)

C. Janiesch, R. Ruggaber, Y. Sure, Eine Infrastruktur für das Internet der Dienste, in ed. by J. Hofmann, A. Meier *Webbasierte Geschäftsmodelle* (dpunkt Verlag, Heidelberg, 2008), pp. 71–79

D. Jordan, J. Evdemon, Web services business process execution language version 2.0, Apr 2007, http://docs.oasis-open.org/wsbpel/2.0/wsbpel-v2.0.html

H. Kett, A Business model approach for service engineering in the internet of services. Int. J. Serv. Sci. Manage. Eng. Technol. **2**(4), 1–8 (2011)

H. Kett, B. Franczyk, A business model approach for service engineering in the internet of services, in *Informatik 2010: Service Science: neue Perspektiven für die Informatik*, ed. by K. Fähnrich (Köllen Verlag, Bonn, 2010), pp. 539–544

H. Kett, G. Scheithauer, N. Weiner, A. Weisbecker, Integrated service engineering (ISE) for service ecosystems: an interdisciplinary methodology for the internet of services, in *eChallenges Conference*, Istanbul, ed. by P. Cunningham, M. Cunningham, Oct 2009, pp. 1–8

H. Kett, K. Voigt, G. Scheithauer, J. Cardoso, Service engineering in business ecosystems, in W. Ganz, F. Kicherer, A. Schletz, *New Horizons for the Role and Production of Services (RESER'08)*, Stuttgart (Fraunhofer Verlag, 2008)

M. König, B. Guptill, B. McNee, J. Cassell, SaaS 2.0: Software-as-a-Service as next-gen business platform, April 2006. http://saugatucktechnology.com/research/latest-research/1971-239ssr-saas-2-0-software-as-a-service-as-next-gen-business-platform.html

O. Lorenz, G. Martzahn, T. Rüdiger, T. Renner, A. Pols, eBusiness-Barometer 2007/2008, in *eBusiness 2007/2008: Jahrbuch der deutschen Wirtschaft*, ed. by O. Lorenz (Wegweiser, Berlin, 2007), pp. 8–61

P. Oaks, A.H. ter Hofstede, D. Edmond, Capabilities: describing what services can do, in *Service-Oriented Computing – ICSOC 2003*, Trento, M.E. Orlowska, S. Weerawarana, M.P. Papazoglou, J. Yang. Volume 2910 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2003), pp. 1–16

A. Osterwalder, The business model ontology – a proposition in a design science approach. PhD thesis, Université de Lausanne, Lausanne, 2004. http://www.hec.unil.ch/aosterwa/phd/osterwalder_phd_bm_ontology.pdf

J. O'Sullivan, Towards a precise understanding of service properties. PhD thesis, Queensland University of Technology, 2006

A. Scheer, M. Nuettgens, W.M.P. van der Aalst, J. Desel, A. Oberweis, ARIS architecture and referencemodels for business process management, in *Business Process Management*, W. van der Aalst, J. Desel, A. Oberweis. Volume 1806 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2000), pp. 376–389

G. Scheithauer, K. Voigt, M. Winkler, V. Bicer, A. Strunk, Integrated service engineering workbench: wervice engineering for digital ecosystems. Int. J. Electron. Bus. **9**(5/6), 392–413 (2011)

H. Trætteberg, Hybrid tool for user interface modelling and prototyping, in *Computer-Aided Design of User Interfaces V*, ed. by G. Calvary, C. Pribeanu, G. Santucci, J. Vanderdonckt (Springer, Berlin/Heidelberg/New York, 2007), pp. 215–230

UN/CEFACT, Core components technical specification – version 3.0. Technical report, United Nations – Centre for Trade Facilitation and Electronic Business (UN/CEFACT), Sept 2009. http://www.unece.org/fileadmin/DAM/cefact/codesfortrade/CCTS/CCTS-Version3.pdf

W.M.P. van der Aalst, Business process management demystified: a tutorial on models, systems and standards for workflow management, in *Lectures on Concurrency and Petri Nets*, ed. by J. Desel, W. Reisig, G. Rozenberg. Volume 3098 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2004), pp. 1–65

K. Vidačković, I. Kellner, J. Donald, Business-oriented development methodology for complex event processing – demonstration of an integrated approach for process monitoring, in *Proceedings of the Fourth ACM International Conference on Distributed Event-Based Systems (DEBS'10)*, Cambridge (ACM, New York, 2010), pp. 111–112

N. Weiner, T. Renner, H. Kett, *Geschäftsmodelle im Internet der Dienste: Trends und Entwicklungen auf dem deutschen IT-Markt* (Fraunhofer Verlag, Stuttgart, 2010)

A. Weisbecker, Software-Management für komponentenbasierte Software-Entwicklung. PhD thesis, Universität Stuttgart, Stuttgart, 2002

M. Winkler, J. Spillner, A. Schill, SLA management and contract-based service execution, in *Service-Oriented Computing*, ed. by L. Baresi, C.H. Chi, J. Suzuki. Volume 5900 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg/New York, 2009), pp. 653–655

J.A. Zachman, A framework for information systems architecture. IBM Syst. J. **26**(3), 276–292 (1987)

M. zur Muehlen, M. Indulska, G. Kamp, Business process and business rule modeling: a representational analysis, in *11th International IEEE EDOC Conference Workshop (EDOC'07)*, Annapolis (IEEE, Washington, DC, 2007), pp. 189–196

A Unified Description Language for the Internet of Services

Daniel Oberle

Abstract In order to enable communication and trade between providers and consumers of services, the Internet of Services requires a standard that creates a "commercial envelope" around a service. Therefore, research in TEXO centered around a description language for arbitrary services, such as human or platform services. Consequently, the *Unified Service Description Language* (USDL) is a major outcome of TEXO and allows a normative and balanced unification of service information. The unified description established by USDL is machine-processable, and considers technical and business aspects of a service as well as functional and non-functional attributes. USDL has been built according to the design science approach, whose major research questions are answered in this article.

1 Introduction

According to the design science approach (Hevner et al. 2004), each designed artifact requires a relevant need. In the case of USDL, there is both a *socio-economic* and an *IT-related* need. With respect to the *socio-economic* need, the services sector is an economic growth driver in most developed economies. As an example, consider the Federal Republic of Germany, where the largest part of the macroeconomic value of 2009 was generated by the service industry (Krings 2010). This trend coincides with the ongoing industrialization of the services sector in developed economies. Similar to the area of physical products, *Standardization* (David and Steinmueller 1994) is the basis and prerequisite of every further development of a domain. Therefore, standards will play a significant role also in the services industry. Standards are expected to drive the professionalization

D. Oberle (✉)
SAP AG, SAP Research, Karlsruhe, Germany
e-mail: d.oberle@sap.com

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_34,

and industrialization of the service industry, to increase transparency, to lead to higher value services, and eventually to contribute to the overall development of the service economy. The need for standardization also comprises a normative service description as outlined in für Normung (DIN).

The *IT-related* need considers specialized roles in service networks that need to disclose and exchange, as well as comprehend, business information about services. In addition to consumers requesting services, emerging third-party intermediaries who can deliver and further monetize services in the form of brokers, channel partners and cloud providers also raise questions about how the high-level details of services can be understood without full reliance on service providers (Barros and Dumas 2006). As the forms of services being exposed become more sophisticated, the existing efforts for describing and accessing services pose limitations. These efforts leave open the way in which consumers understand details of business operations, pricing, legal and other implications of using services.

As a response to both needs, the *Unified Service Description Language* (USDL) (Barros and Oberle 2012; Oberle et al. 2013) has been developed and represents a major outcome of the TEXO use case. USDL allows us to capture business, operational and technical aspects of human to automated services. The remainder of this article addresses USDL's remaining research questions, namely its structure, construction, evaluation, and whether it is a progression compared to state-of-the-art service description languages.

2 Example

Consider a German medium-sized company which is a world market leader in manufacturing of gas turbines. The company relies heavily on exports, usually to Europe and the US. However, the company has just received an order by a Russian company so it has to ship to Siberia for the first time. Therefore, the manufacturing company has to use a new (in an ideal case all-inclusive) service bundle for shipping, forwarding, clearance, etc. Consequently, the manufacturing company consults a logistics service marketplace for discovering a suitable service bundle.

The logistics marketplace features a range of 2PL[1] and 3PL services,[2] accompanying services such as customs clearance, insurances, and container rental, as well as 4PL services,[3] typically bundling the aforementioned services. In

---

[1] 2PL – A second-party logistics provider is an asset-based carrier, which owns the means of transportation, e.g., shipping lines, airlines, truck, or rail companies.

[2] 3PL – A third-party logistics provider is a firm that provides a one-stop shop service to its customers of outsourced logistics services for part or all of their supply chain management functions.

[3] 4PL – A fourth-party logistics provider or lead logistics provider is a consulting firm specializing in logistics, transportation, and supply chain management.

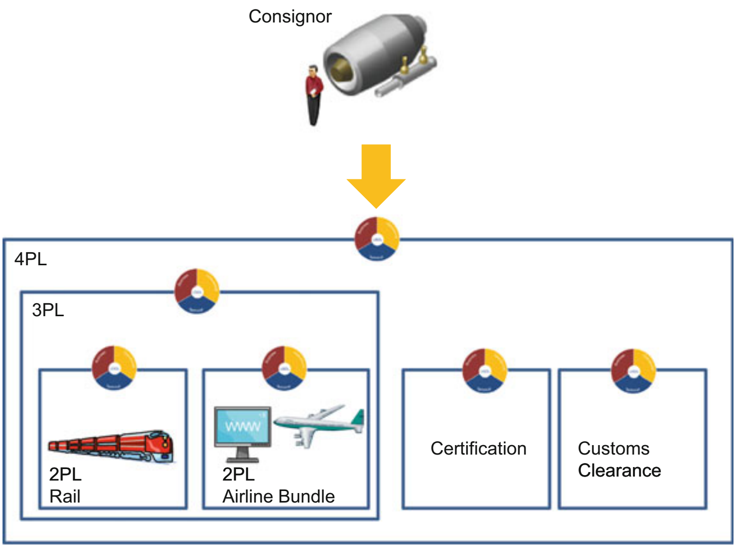


Fig. 1 Logistics scenario as running example (Barros et al. 2012)

particular, the manufacturer requires a Gost-R certification[4] for exports to Russia as well as a customs clearance.

As depicted in Fig. 1, the service offered by 4PL is actually made up of other service bundles and compositions. The business, operational, and technical aspects of the bundles, compositions, and individual services – as well as their dependencies – can be captured by USDL. The corresponding USDL description of the 4PL benefits the provider since the bundled and composed services can be easily recombined with other packages. The USDL description also benefits the marketplace as a prerequisite for automated processing with established tools. Finally, the consumer benefits by better comparability through the normative description.

3 Structure

The Unified Service Description Language (USDL) is proposed as a normative and comprehensive master data model for commercial meta data of IT, physical, or hybrid services. More specifically, USDL allows for a unified description of business, operational, and technical aspects of services (see Fig. 2). USDL aims at a holistic service description, putting special focus on business aspects, such as ownership and provisioning, release stages in a service network, composition and

[4]http://www.gost-r.info/

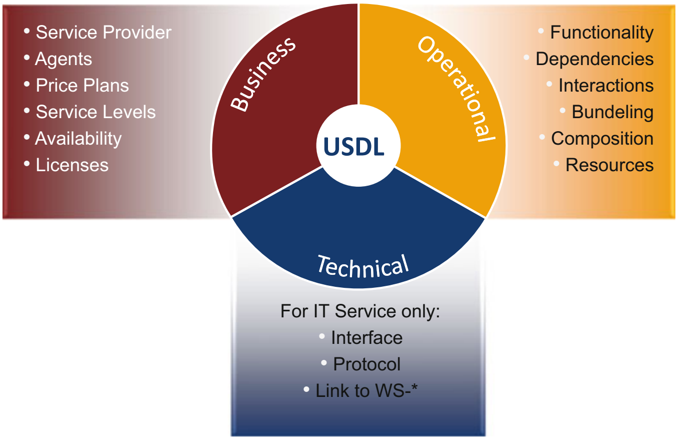


Fig. 2 USDL unifies business, operational, and technical master data of a service

bundling, pricing, and legal aspects, among others, in addition to technical aspects. It proposes a consolidated foundation for service-based systems, enabling different roles for participating in diverse aspects of provisioning in service networks.

In this section, we have a closer look at the structure of USDL. According to the design science approach, the structure is one of the research questions that have to be addressed in every scientific presentation of an artifact. Given the complexity of the service domain, USDL has been designed with the requirements of conceptualization and modularity in mind. With respect to the former, UML class models are used for capturing USDL. With respect to modularity, USDL is split into several packages (according to UML terminology). Each package represents one USDL 'module' and contains one class model. The resulting split in modules follows from prominent non-functional aspects, such as pricing and legal constraints, how services are interfaced with for delivery and service-level agreements, which partners have responsibility for the service and details about service functionality[5]:

- *Service Module*: The Service Module focuses on the essential structure of a service, i.e., the building blocks of a service, because of the complexity entailed. This arises from the conceptual diversity of service types, across the human-to-automation continuum, which brings services of different "shapes and sizes" into view. Also related to the structure of services are the orthogonal and adjacent issues of service dependency, related to the resources/services supporting a service's delivery and composition, concerning the different parts put together for a service's functionality. In our example, the Service Module's modeling

[5]The following description is taken from Oberle et al. (2013). Note that the Technical and Foundation modules are not discussed due to space limitations.

constructs allow us to capture the fact that *4PL* consists of other bundles, compositions, and individual services.

- *Participants Module*: The Participants Module captures information about the organizational actors that are important for the provisioning, delivery and consumption of a service. The Participants Module distinguishes between service owners (e.g., *2PL Rail* actually owns the transportation means), service providers (e.g., *3PL* has the service delivery responsibility for *2PL Rail*), stakeholders (having regulatory, commercial or other designated interests in the service), intermediaries (having specialist provisioning, such as a broker or cloud provider, beyond the original provisioning), and end consumers.
- *Functional Module*: The Functional Module allows the capture of service functionality at an abstract level, regardless of the proximity of the service on the human-to-automation continuum, and free from technical implementation details. USDL supports the capture of service functionality in different layers, for different levels of concern. Accordingly, the Functional Module combines well-established concepts of capability modeling and functional decomposition. In our example, we could capture that *2PL Rail* offers *point-to-point transport on rail* with inputs such as *shipping instructions* and *cargo*, or preconditions, such as *maximum weight* and *size* of the cargo.
- *Interaction Module*: The Interaction Module captures the behavioral aspect of services and complements the structural focus of the Functional and Technical modules. The behavioral aspects of services concern how involved participants interact with the service. For example, *2PL Rail* requires a complex interaction that runs through the phases *order entry*, *booking and pre-carriage*, and *main carriage*. Each phase consists of several interactions, such as *place order* and *receive order confirmation* in the *order entry* phase.
- *Pricing Module*: The Pricing Module concerns the charging of services as mutually understood by those who own or deliver services and those who consume them. USDL has a hierarchical structure for service pricing starting with one or more price plans. The service *2PL Airline Bundle*, for example, offers a price plan that consists of multiple dependent price components. The main parts are the *Base Component*, a flat fee of 350€ for each order, and the *Cargo Component*, which prices per weight and distance. For shipments below 1 ton, the shipper has to pay 35€ per 25 kg and 100 km flown. At above a one ton cargo weight the price falls to 25€. In order to guard itself against volatility of currency exchange rates, a *Currency Adjustment Factor* of 10 %, and a minimum 30€, is applied to the Cargo Component. Finally, the local sales tax of 5 % on the sum of the three components is added.
- *Legal Module*: The Legal Module addresses the need for legal certainty and compliance in service networks and in trading services on marketplaces. Participants need to know about the terms of usage of a particular service, for example, liability, privacy, or copyright. A simple example of a software license concerns the service *Shipment Manager* of the *2PL Airline Bundle*, which grants licensees a non-exclusive right to reproduce the service. This reproduction process takes the form of accessing a web UI through an Internet browser or installing a

software client on PCs owned by the licensee. Even though the service and software are owned by the provider, German law requires a natural person to be identified as the author, which in this example is the developer of the software.

- *Service Level Module*: The Service Level Module is kept generic and does not specify how concrete service levels on concrete aspects should be specified. Instead, its main purpose is twofold. First, it provides the glue between abstractly specified service level issues in other USDL concepts. Second, it allows for incorporation of arbitrary attribute and expression languages for specifying service-level constraints. As an example, consider the expected delivery duration, i.e., a service level of *2PL Airline Bundle*.

4 Construction

Besides the structure of USDL, design science also requires the discussion of how an artifact was built. Therefore, this section emphasizes the construction of USDL which basically began as early as in 2007 during the start of the German publicly funded TEXO use case led by SAP Research. September 2008 saw the first version of a *Universal* Service Description Language. This first version was directly modeled in a simple XML Schema which was structured according to the segments in the circular USDL logo (see Fig. 2). Accordingly, the XML Schema was composed of business, operational, and technical elements (Cardoso et al. 2009). A version 2.0 followed in December 2008 which however only added some additional elements to the XML Schema required for internal use cases (Cardoso et al. 2010).

A significant change happened throughout 2009 in the lead up to the Future Internet Assembly (Tselentis et al. 2009) when USDL was prominently disclosed to the public. It was decided to use class models instead of the XML Schema. Reasons for that were mainly limitations in expressive power of version 2.0, which required considerable extensions to the model, thus motivating the switch to a proper tool chain and modular design. This led to the first milestone (M1) of USDL version 3.0, which was documented by detailed specifications available at the community site www.internet-of-services.com, and released in September 2009. Different milestones (M1–M5) were released, each increasing the scope and capabilities of USDL.

Version 3 has been built and evaluated in a collaborative and interdisciplinary way by SAP Research. That means about a dozen researchers at SAP Research have contributed to USDL by bringing in their expertise from different backgrounds (computer scientists, including security and SLA experts, business economists, legal scientists, etc.); they are spread over different locations around the world. The modeling has been managed by a central governance body that coordinated the different contributions. All this has been carried out in the context of several publicly funded research projects, most prominently TEXO, where services from various domains including cloud computing, service marketplaces and business networks, have been investigated for access, repurposing, and trading in large settings. Other projects contributed as well, including German Federal Ministry of Education and Research projects (e.g., Premium Services), EU DG INFSO projects (e.g., FAST, SLA@SOI, or SOA4ALL), or the Australian Smart Services CRC.

Subsequent iterations of USDL also include the contributions and evaluation feedbacks of project partners. On the one hand, specific academic partners were given mandates to incorporate or refine particular aspects in USDL. On the other hand, industrial partners of the research projects, e.g., Siemens, evaluated USDL by case studies in their setting. Finally, the scope of input was broadened even wider by approaching a standardization body. As a first step to standardization, a W3C Incubator group was founded including additional players of academia and industry.

5 Evaluation

USDL has been designed to address the needs described in Sect. 1. Consequently, the design science approach requires a rigorous demonstration of the artifact's utility, quality, and efficacy to address the needs. This task is usually called an evaluation (see Barros and Oberle 2012).

First, several case studies were carried out: services in the energy domain, services for mobile users, manual services for insurances, and services for B2B integration (Schäffler et al. 2012). Each case study originated from a different company and provided its own conclusion. Case studies are the first step to a comprehensive validation of USDL, because they can create feedback regarding usability, feasibility, completeness, etc. of certain aspects of USDL. The case studies allowed for a coverage analysis to judge whether USDL meets their corresponding modeling needs and whether USDL offers unnecessary modeling constructs. In essence, most constructs were required at least once, but also some deficits are apparent which were input to future versions of USDL.

Second, requirements for a service description language from potential USDL users were identified in Matzner and Becker (2012). Through an anonymous, written multi-stage survey process, the study elaborates a set of requirements. The requirements can be used to ex-post test if the features of the USDL actually address the users' needs and to recheck the underlying assumptions of the USDL design and development process. The observations above suggest a very broad support of the respondents with the provided list of requirements.

Finally, Birkmeier et al. (2012) focus on evaluating the expressive power of the language to specify *software* services from a theoretical point of view. In particular, it discusses which information should be provided in order to support the discovery and combination of software services. This evaluation showed that USDL provides the most detailed approach to date to comprehensively describe software services, which nevertheless should be harmonized in some aspects.

6 Progression

After highlighting the need, relevance, structure, construction, and evaluation, the question turns to how USDL positions itself against other service description languages. The design science approach requires that USDL represent a progression compared to the state of the art. Indeed, there exists a plethora of service description efforts that can be grouped into different strands summarized below (see Barros and Oberle 2012, for a detailed discussion). Each of the strands has its own motivation and representation needs for capturing service information.

Unlike any of the other approaches, USDL serves a reference purpose, its class models can be used to facilitate model-driven software engineering, and its XSD can be used for information exchange. However, it is mainly the content and normative character of the USDL modules that sets USDL apart from many of the related approaches discussed below. For example, UDDI, WSMO, or OWL-S only prescribe tiny schemata and leave the modeling of service description concepts (such as a generic schema for defining a prize model or licenses) to the user. Both W3C SAWSDL and W3C SA-REST are designed to be agnostic of any service description schema. This also holds for W3C SML. USDL progresses the state of the art since USDL builds on and acknowledges the efforts of the SOA strand, yet adds business and operational information.

The first strand of service description efforts is the field of *Service-oriented Architectures* (SOA). Because of their IT focus, most approaches limit their attention to the field of software architecture. Originally, several standards bodies specified several dozens of different aspects which are collectively known as WS-* (including WSDL, WS-Policy, WS-Security, etc.). Since one of the key components of a SOA is a service registry, the OASIS standards body introduced the concept of Universal Description, Discovery and Integration (UDDI), i.e., a specification for a platform-independent registry. UDDI services shall be discovered via white pages (address, contact, and known identifiers) or yellow pages (industrial categorizations based on standard taxonomies), as well as green pages. However, UDDI hardly prescribes any schema for such information. As the concept of SOA matured, calls for support in software and service engineering increased. Hence, the OMG standards body dedicated its focus to software engineering for SOA, and, subsequently defined the Service-oriented architecture Modeling Language (SoaML) (Object Management Group (OMG) 2008). Finally, the multitude of description efforts and different definitions of SOA led to a Reference Model.[6]

A second strand consists mainly of ontologies in the field of *Semantic Web Services*. As presented in the seminal paper, viz., (McIlraith et al. 2001), the main goal of semantic web services approaches is automation of discovery, composition, and invocation of services in an SOA by ontology reasoners and planning algorithms. The most prominent efforts are OWL-S and WSMO (Roman et al. 2005).

[6] http://www.opengroup.org/soa/source-book/ontology/

Considering the need to attach semantic descriptions to existing WS-* descriptions and the multitude of ontologies available, the W3C came up with a recommendation called Semantic Annotations for WSDL (SAWSDL). Similarly there is a W3C Member Submission called Semantic Annotations for REST (SA-REST).

There are overarching efforts that concentrate on the bigger picture of *service systems* or service science also taking into account socio-economic aspects. Alter (2008) was one of the first to realize that the concept of a service system is not well articulated in the service literature. Therefore, he contributes three informal frameworks as a first attempt to define the fundamentals of service systems. Current research in this strand is represented by the work of Ferrario et al. (2011), which can be seen as a continuation and formalization of Alter's approach. Although differing in its main notions, they present a reference ontology for ontological foundations of service science, which is founded on the basic principles of ontological analysis. In turn, this reference ontology forms the core part of the TEXO Service Ontology, which extends it by ontology modules for pricing, legal, innovation, or rating information (Oberle et al. 2009).

The final strand is driven by schools of business administration and business informatics, and focuses on capturing the purely economic aspects of services regardless of their nature (with less or no focus on IT services and software architectures). Led by the Fraunhofer IAO research institute, the German standard DIN PAS 1018 (für Normung DIN) essentially prescribes a form for the description of services for tendering. The structure is specified in a non-machine-readable way by introducing mandatory and optional, non-functional attributes specified in natural language. The PhD thesis of Emmrich (2005) has a similar motivation, only that this work focuses on product-related services, such as maintenance, and is specified in UML. The PhD thesis of O'Sullivan (2006) adopts a wider scope and contributes a domain-independent taxonomy that is capable of representing the non-functional properties of conventional, electronic and web services. Ioan (2010) presents a syntactic translation of O'Sullivan's work in the proprietary WSML language. The goal is to extend the aforementioned WSMO by non-functional properties for automation of discovery, composition, invocation, and, in particular, ranking of services in an SOA. de Kinderen and Gordijn (2008) introduce the e3Service ontology to model services from the perspective of the user needs. This offers constructs for service marketing, but in a computational way, such that automated reasoning support can be developed to match consumer needs with IT services. The main focus of this work is to generate service bundles under the consideration of customer needs.

7 Conclusion

The article summarized one of the central outcomes of the TEXO use case, viz., USDL. Since USDL has been built according to the design science approach, the article discussed its need and relevance, its structure, its construction, its evaluation, and its progression compared to the state of the art. In addition, TEXO developed a comprehensive USDL tool chain, including editors, stores and marketplaces, allowing flexible deployment scenarios.

References

S. Alter, Service system fundamentals: work system, value chain, and life cycle. IBM Syst. J. **47**(1), 71–85 (2008)

A. Barros, D. Oberle, *Handbook of Service Description: USDL and Its Methods* (Springer, Berlin/Heidelberg/New York, 2012)

A. Barros, D. Oberle, U. Kylau, S. Heinzl, Design overview of USDL, in *Handbook of Service Description: USDL and Its Methods*, ed. by A. Barros, D. Oberle (Springer, Berlin/Heidelberg/New York, 2012), pp. 187–225

A.P. Barros, M. Dumas, The rise of web service ecosystems. IT Prof. **8**(5), 31–37 (2006)

D.Q. Birkmeier, S. Overhage, S. Schlauderer, K. Turowski, How complete is the USDL? in *Handbook of Service Description: USDL and Its Methods*, ed. by A. Barros, D. Oberle (Springer, Berlin/Heidelberg/New York, 2012), pp. 521–538

J. Cardoso, A. Barros, N. May, U. Kylau, Towards a unified service description language for the internet of services: requirements and first developments, in *Proceedings of the 2010 IEEE International Conference on Services Computing (SCC'10)* (IEEE, Washington, DC, Miami, Florida, 2010), pp. 602–609

J. Cardoso, M. Winkler, K. Voigt, A service description language for the internet of services, in *Proceedings 1st International Symposium on Services Science (ISSS'09)*, R. Alt, K. Fähnrich, B. Franczyk (Logos, Berlin, Leipzig, Germany, 2009), pp. 229–240

P. David, W.E. Steinmueller, Economics of compatibility standards and competition in telecommunication networks. Inf. Econ. Policy **6**(3/4), 217–241 (1994)

S. de Kinderen, J. Gordijn, e3service: an ontological approach for deriving multi-supplier IT-service bundles from consumer needs, in *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS-41)*, Waikoloa (IEEE Computer Society, 2008), p. 318

A. Emmrich, Ein Beitrag zur systematischen Entwicklung produktorientierter Dienstleistungen. PhD thesis, University of Paderborn, 2005

R. Ferrario, N. Guarino, C. Janiesch, T. Kiemes, D. Oberle, F. Probst, Towards an ontological foundation of services science: the general service model, in *10th International Conference on Wirtschaftsinformatik*, Zürich, vol. 2, Feb 2011, pp. 675–684. http://dblp.uni-trier.de/db/conf/wirtschaftsinformatik/wi2011.html#FerrarioGJKOP11

D.I. für Normung (DIN) (ed.), *Grundstruktur für die Beschreibung von Dienstleistungen in der Ausschreibungsphase (PAS 1018)* (Beuth, Berlin, 2002)

A.R. Hevner, S.T. March, J. Park, S. Ram, Design science in information systems research. MIS Q. **28**(1), 75–105 (2004)

T. Ioan, Modeling and ranking semantic web services based on non-functional properties in web services. PhD thesis, University of Innsbruck, 2010

S. Krings (ed.), *Statistisches Jahrbuch 2010 für die Bundesrepublik Deutschland mit Internationalen Übersichten* (Statistisches Bundesamt, Wiesbaden, 2010)

M. Matzner, J. Becker, Requirements for a service description language – findings from a Delphi study, in *Handbook of Service Description: USDL and Its Methods*, ed. by A. Barros, D. Oberle (Springer, Berlin/Heidelberg/New York, 2012), pp. 503–519

S. McIlraith, T. Son, H. Zeng, Semantic web services. IEEE Intell. Syst. **16**(2), 46–53 (2001)

D. Oberle, A. Barros, S. Heinzl, U. Kylau, A unified description language for human to automated services. Inf. Syst. **38**(1), 155–181 (2013)

D. Oberle, N. Bhatti, S. Brockmans, M. Niemann, C. Janiesch, Countering service information challenges in the internet of services. Bus. Inf. Syst. Eng. (BISE) **1**(5), 370–390 (2009). http://dblp.uni-trier.de/db/journals/bise/bise1.html#OberleBBNJ09

Object Management Group (OMG), Service oriented architecture Modeling Language (SoaML) – Specification for the UML Profile and Metamodel for Services (UPMS). Technical report, Adaptive, Capgemini, EDS, Fujitsu, Fundacion European Software Institute, Hewlett-Packard, International Business Machines, MEGA International, Model Driven Solutions, Rhysome, Softeam, 2008. http://webcache.googleusercontent.com/search?q=cache:XOtyNHTMUcEJ:www.omgwiki.org/SoaML/lib/exe/fetch.php%3Fid%3Dspecification%26cache%3Dcache%26media%3Dsoaml_specification_2008-11-01.pdf+&cd=1&hl=de&ct=clnk&gl=de&client=firefox-a

J. O'Sullivan, Towards a precise understanding of service properties. PhD thesis, Queensland University of Technology, 2006

D. Roman, H. Lausen, U. Keller, J. de Bruijn, C. Bussler, J. Domingue, D. Fensel, M. Kifer, J. Kopecky, R. Lara, E. Oren, A. Polleres, M. Stollberg, D2v1.1. Web Service modeling ontology (WSMO), Feb 2005. http://www.wsmo.org/TR/d2/v1.1/

M. Schäffler, A. Thede, B. Leferink, K. Kadner, A. Horch, M. Kintz, M. Weidmann, M. Weiten, Case studies, in *Handbook of Service Description: USDL and Its Methods*, ed. by A. Barros, D. Oberle (Springer, Berlin/Heidelberg/New York, 2012), pp. 465–486

G. Tselentis, J. Domingue, A. Galis, A. Gavras, D. Hausheer, S. Krco, V. Lotz, T. Zahariadis (eds.), *Towards the Future Internet – A European Research Perspective* (IOS, Amsterdam, 2009)

Semantic Technologies for the Internet of Services

Eva Maria Kiss, Saartje Brockmans, and Jürgen Angele

Abstract The TEXO use case in the THESEUS research program offers technology and solutions for service marketplaces addressing in particular small and medium-sized enterprises (SMEs), which profit most from a standardized and open platform with out-of-the-box functionality. ontoprise,[1] a leading company in semantic technologies since 1999, contributed the TEXO semantic infrastructure as a collaborative development and runtime environment for the Service Ontology. By providing the TEXO partners with a continually improved semantic infrastructure, we learned about the chances and challenges of using semantic technologies in the Internet of Services. As an illustrative example for the advantages of these semantic technologies in a B2B context, we present a proof of concept for an ontology-based recommendation system, which relies on the semantic infrastructure, and includes enhanced data integration support.

1 Introduction

Recent events, such as Facebook going public,[2] confirm the continuing progress of the Internet of Services and the closely related semantic technologies, while other, less noted work (Davis et al. 2013; Kiss 2013) reports that these novel

[1]As of June 28, 2012, SemEO Services GmbH, Darmstadt, Germany has acquired all intellectual property rights (IPRs) from ontoprise, in particular for OntoBroker, OntoStudio and other products and software mentioned in this article. The business activities of ontoprise will be continued by semafora systems GmbH, Darmstadt, as licensed by SemEO.

[2]United States Securities and Exchange Commission, Facebook Registration Statement Form S-1, http://www.sec.gov/Archives/edgar/data/1326801/000119312512034517/d287954ds1.htm.

E.M. Kiss (✉) • S. Brockmans • J. Angele
Ontoprise GmbH, Karlsruhe, Germany
e-mail: kiss@ontoprise.de; saartje@muzemix.be; juergenangele@googlemail.com

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_35,

paradigms are only slowly adopted and give cautious estimates for possible positive developments. The truth is somewhere in between. The new technologies are slowly becoming established, some parts being commercially more successful than others. Top-down models of these technologies are still a matter of intensive research. Due to programs such as THESEUS, significant progress has been made in realizing suitable infrastructure and valid applications.

Business Scenarios from a User's Perspective We live in times of complex technological and social developments. Part of the Year 2K promises (Autere and Korhonen 2004) of Internet and mobile communication have come true, part of them are still a matter of research, heated discussions and sustained efforts. Nevertheless, the building of a global world without walls and the ubiquitous presence of the Internet has changed our daily lives. A description of some typical scenarios reminds us how service society, Internet and mobile have created the Internet of Services. For example, a business person and an assistant will fly on short notice to a 3-day workshop abroad, which will require detailed travel planning. The business person will use a number of online services and travel planners offered by flight, rail, and other specialized travel companies, and the comparison of prices and timetables will take some time. Next, they will prepare some last-minute slides, at the latest when they are in the train or at the airport. Preparation of the slides requires gathering information from different sources, most of them online or in the head of people moving around with other business. They will use Wi-Fi to access required information in the company's portal and on the Web. At the same time they will try to stay in touch with family and friends, calling or texting them, and making appointments. At the destination, they need to find the way to the hotel, and can use an online pedestrian navigation system on the smart phone, which guides them in the right direction. At the hotel, they use an online service to find a suitable restaurant nearby and another one to check in online for the return flight. This description of typical scenarios in the life of an average business person today reminds us how service society, Internet, and mobile communication have created the Internet of Services, which has made life much easier already.

Scientific Progress and Market Dynamics How did this happen? What will be the next step? To answer the first question, we know that credit is due to the big newcomers, such as Amazon, eBay, Google, and Facebook, who recognized the potential and focused on growth and commercial exploitation. Credit is also due to the big IT players that took up the development and supported its consolidation, and credit is due to the large numbers of researchers at universities and within companies who paved the way since the very beginning of information science. It started with first ideas on Artificial Intelligence – to create robots – and semantic technologies – to teach robots think like humans, by performing logical reasoning. Semantic technologies evolved to provide support for modeling the world, support for knowledge management, for search and data integration.

To answer the second question, we can make some more or less accurate predictions based on the knowledge of today's still open problems and the knowledge of the potential of the technology we have. One of the problems is that today the Internet of Services still leaves too much work to the paying customer. Our travelling business person now has to manage quantities of online services with his or her respective rules and accounts, and has to be technically expert at doing this in the best possible way. Instead of their being free to follow their own business and being assisted in doing so, the technology is controlling the lives of people in a rather pushy way. Yes, we want the freedom to choose and combine services, but no, we do not want to spend 80 % of our time in making the optimal decisions. It becomes even more complicated when the unexpected happens, be it strike, accident or bad weather conditions. The agenda has to be rescheduled and all relevant parties have to be informed, a situation where IT support is still insufficient.

More than 10 years later, the same business persons might simply tell their mobile phones that they want to attend a workshop on short notice and need to handle such and such last minutes duties, business-related and personal. The system will assemble the required information and return recommendations for alternative travel routes and hotels, according to the registered profile and preferences of the business person. The system will process inferred knowledge from available resources on the Web and proprietary information from the company's network, respecting the privacy settings of person and company. The Internet of Services has already made a big step towards this future, in the form of research work, pioneer applications and early adopters. In order to fully attain the envisioned features, we will need more of a stable semantic infrastructure and a network of services that actually make life easier, such as trustworthy recommendations systems to direct us to the needed help in each situation.

Semantic Technologies Semantic technologies are technologies that aim to encode meanings separately from data and content files, and separately from application code. They provide an abstraction layer above existing IT technologies that enables bridging and interconnection of data, content, and processes. Semantic technologies offer solutions to a range of problems, such as modeling, decision support, prediction, and data integration. High-level modeling provides a formalism that is comprehensible for business experts and can be shared in social networks, encoding of business logic in rules that can be modified, combined and reused in a production environment, and the ad hoc integration of heterogeneous data sources.

The advantages offered by semantic technologies today are clear to experts in the field. Yet for a non-technical person, it may be difficult to grasp their value, since they often are used on top of other technologies and due to their relative novelty, require the knowledge of new concepts ranging from basics to hype. We remind the reader of some basic concepts from the domain of semantic technologies, first of all, the central concept of ontology. In the context of semantic technologies, an ontology is a shared data model used for knowledge modeling, which provides a formal description of a given domain of knowledge (Gruber 2009). An ontology consists of a data model (classes and the relationships between them, rules and queries) and data (class instances). The most commonly used ontology language today is the W3C recommendation OWL (Calvanese et al. 2009), a family of Description

Logic languages, which is used in combination with SPARQL as query language and SWRL as rule language. Besides OWL, we have the family of frame-based ontology languages F-Logic/ObjectLogic (Angele et al. 2009; Kifer et al. 1995), which use the same language for ontology, rules and queries. A reasoner is a middleware that performs reasoning on ontologies, i.e. the deduction of explicit knowledge from implicit knowledge.

2 Semantic Infrastructure for the Internet of Services

The Internet of Services as we know it has been enabled by joint research work within national and international projects. One of these projects is the THESEUS TEXO use case, which bundled for several years the efforts of leading German industry and research partners with the goal of providing a stable environment for the Internet of Services. One of the main results of the TEXO use case is the Service Ontology (Oberle et al. 2009), an ontology that models the stack of semantically enriched services. The Service Ontology exploits the concept of separating user roles in a services lifecycle: technical experts will provide service descriptions using Unified Service Description Language USDL,[3] legal experts will contribute the general terms on the usage, marketing experts will contribute pricing details, etc. The Service Ontology is an interlinked layer of ontology modules grounded in a Core Service Description module, enhanced with domain modules for different service governance aspects (technical, legal, business model), and offers add-on industry modules. The separation of roles and the modular approach imply that domain experts carry responsibility only for their respective module and require support for collaborative ontology engineering. In consequence, the Service Ontology has been built collaboratively through the joint effort of TEXO use case partners, using a common semantic infrastructure.

Providing a Semantic Infrastructure An application relying on ontologies to store the information model for the system's domain knowledge is built and run in the environment offered by a semantic infrastructure. In order to implement the ontology modules of the application, the system architect first decides what the ontology needs in term of expressiveness and reasoning performance. In the next step, the architect chooses a modeling environment to support the knowledge engineering process. The modeling of ontologies needs the expertise, knowledge, and know-how of many different types of persons. Service engineers, ontology experts, and domain experts are required to create sound and usable results. Dedicated workshops might provide a common ground and understanding; however the more detailed modeling will often take place in a spatial and temporal distributed manner, thus requiring support for collaborative modeling. Then, a runtime environment is

[3] http://www.w3.org/2005/Incubator/usdl/

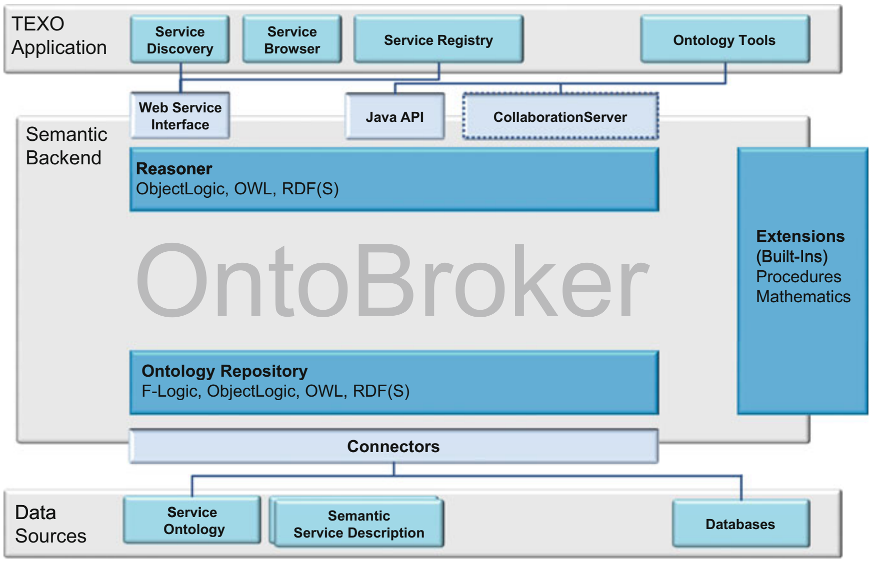


Fig. 1 Semantic infrastructure for TEXO

needed for accessing the information stored in the ontologies in online and offline scenarios. In addition, the architect will explore the possibilities of embedding the knowledge-based system in existing legacy systems and of exposing interfaces. Finally, options for maintenance and upgrades are considered and the cost of the system is estimated, possibly in the form of a Return Of Investment (ROI) analysis.

The TEXO use case architects designed a platform according to the above principles. Here we focus on the knowledge model of the overall platform, the Service Ontology. The main design decisions for the Service Ontology were to build it (1) based on a foundational ontology, (2) in a modular way, (3) using ontology design patterns and (4) using the W3C recommendation OWL as ontology language. Within TEXO, the infrastructure of choice was built on top of the OntoBroker product suite[4] (Angele 2012), which at the beginning of the project offered state-of-the-art required functionality and was subsequently developed to match the requirements fully.

TEXO*'s Semantic Infrastructure* (Fig. 1) comprises two main parts: a modeling environment (OntoBroker CollaborationServer with OntoStudio clients) and a runtime environment (Semantic Back End on top of OntoBroker). The modeling environment provides modeling functionality of new service descriptions for the service provider, as well as browsing and discovering functionality for the service consumer. The semantic back end operates as inference engine for all TEXO components and applications connected via web service interfaces, such as the

[4]http://www.semafora-systems.com/en/products/ontobroker/

TEXO Service Browser, the TEXO Semantic Service Discovery and the context-aware Semantic Service Discovery. We briefly describe the main components.

OntoBroker. The OntoBroker product suite comprising the reasoning engine OntoBroker and the ontology engineering workbench OntoStudio is an environment for the development and management of ontologies as well as a runtime for running ontology-based systems. OntoBroker provides native support for the ontology language F-Logic/ObjectLogic and for all major W3C Semantic Web recommendations including RDF(s), OWL and SPARQL. This makes OntoBroker applicable to a wide range of scenarios. It is possible to use OntoBroker as a high-performance RDF triple store, for typical OWL reasoning tasks (e.g., consistency checks) for conjunctive queries against an OWL ontology or for high performance rule-based reasoning.

OntoStudio is an Eclipse-based ontology engineering workbench with an integrated OntoBroker. Functionality includes import and export of all W3C ontology languages, the creation, validation and maintenance of rule-based models, a mapping editor for semantic data integration, a graphic rule editor, a query editor and an integrated test environment. The ontologies are organized in ontology projects, each project containing one or more modules. Modules are access- and security-protected. OntoBroker can be run in remote mode as CollaborationServer, in which case the ontology repository is accessed remotely by ontology engineers using their respective OntoStudio clients. Within TEXO, we have extended the basic implementation for F-logic to enable efficient modeling of OWL ontologies and added change management functionality.

OntoModel is a UML-based editor for OWL ontologies. The UML extension mechanism allows for the development and maintenance of OWL ontologies with Model Driven Architecture (MDA) technologies. When working with large ontologies it is possible to use several diagrams for a single ontology. A module extraction mechanism allows the separation of a part of an existing ontology into an independent module. The final version of OntoModel has been published under EPL License on http://sourceforge.net/projects/ontomodel/.

The development of the Service Ontology was carried out by an interdisciplinary team of domain experts in different areas such as legal, ratings, and business models, each expert being responsible for the module representing their domain of expertise. The foundational and core modules were modeled by an expert team formed by software vendors (SAP, ontoprise). The ontology modelers were equipped with OntoStudio clients connected remotely with the OntoBroker Collaboration Server and the ontology repository hosted on a dedicated virtual machine. Those familiar with UML used OntoModel as preferred editor to model their ontologies.

Lessons Learned The ontoprise TEXO team maintained the infrastructure for the collaborative development of the Service Ontology over a time span of 3 years. During this time, the modeling environment itself has been changed from originally supporting F-Logic to supporting OWL and ObjectLogic. OWL was introduced to satisfy requirements on interchangeability. However, a W3C recommendation such as OWL comes in different dialects and is not 100 % standard, so our first experience was that considerable amount of work had to be done to ensure data exchange between different systems. At the end of the project, our infrastructure has an OWL implementation that conforms to the recommendation, but when importing and exporting OWL, manual steps are commonly required.

We learned that the collaborative development of an ontology over several months or years will lead to slightly differing models. This stems from the different levels of expertise and different contexts of domain experts, and from the fact that a user-friendly tool such as OntoStudio allows easy modeling even for non-experts. Modularization of the ontology partly addressed this problem. We found that a combination of a priori use of best practices and a posteriori use of semantic data integration works best. Adopting methodologies that support collaborative modeling helps to create better and formally correct models. On the other hand, agreeing and applying methodologies and teaching domain experts requires time. In practice, it is not possible to do it always 100 % according to the textbook. A certain degree of inaccuracy and mismatching will occur within the ontologies and between the different models. This mismatching can be addressed afterwards by using semantic data integration. We used this combined approach successfully for several subsequent applications.

3 Ontology-Based Recommendation System

During the exploitation phase, we implemented a proof of concept for an ontology-based recommendation system relying on the TEXO Semantic Infrastructure.

Recommendations are well known to anyone who has ever bought an item on Amazon. Did you buy books on some topic of your interest in the past? Did you rate the book, the service and the vendor? Next time you search for a book, you will receive recommendations for similar books or for items related to the topic of the book. Amazon explains, that "These recommendations are based on items you own and more". The underlying process is roughly the same for all recommendation systems. Customers are asked to rate the items they bought or services they used. Based on their ratings and on the products bought by customers with similar profiles, they will receive recommendations for new products that might interest them. In a nutshell, a recommendation system recommends products to customers based on the customer's rating and the rating of similar customers.

Things become challenging because B2C Marketplaces in a global market potentially have customers and products of the order of millions (Linden et al. 2003). Thus, significant research effort has been sponsored to develop sophisticated algorithms that solve performance issues when dealing with huge numbers of customers and products. Recommendation algorithms are classified in collaborative filtering ("recommend products used and rated by similar customers") and content-based ("recommend products based on product characteristics or customer's history"). These methods are combined into hybrid recommendations or extended to context-aware recommendations, which take account of the customer's context.

Requirements and Design Decisions Software as a Service (SaaS) is an increasingly popular software distribution model in which applications are hosted by a vendor and accessed by customers over the Internet. From the perspective of an SaaS customer, the most important requirement is to find the right service for their specific demand. According to a Forrester Research Study (Forrester Research 2009), one main barrier for buyers in adopting SaaS besides cost and security is that they cannot find the specific application they need. An important business goal of B2B Portal providers is to ensure that prospective and signed customers find the applications they need. A straightforward way of doing this is anticipating the customer's needs and give personalized product recommendations, which leads to *Requirement (R1): The system shall be able to provide personalized product recommendations that can be modified on the fly in a flexible way*.

A B2B Marketplace differs from a B2C portal in several aspects. Firstly, it is characterized by complex buying decisions and a well-defined set of buying situations: scoping, routine task, occasional purchase. A typical scenario is the use of recommendations during the scoping process, when prospect customers try to find the configuration of services that supports their business best. These different buying situations require that recommendations be context-aware, modular and easy to change. For example, during the first critical period, new customers will receive recommendations for products based on purchases of similar customers. Later, when they are signed, they will receive recommendations for product add-ons based on their history, domain of interest and current context. Second, a B2B marketplace will have long-term customer relationships; the number of customers will be comparatively small and less prone to fluctuations than in B2C relations. Customers are companies with an established organizational structure, a complex business agenda, and a business goal of their own. To support their business, they will purchase a configuration of services rather than a single service, resulting in *Requirement (R2): The system shall be able to model hierarchies of customers, products, and product configurations in a flexible way*. Since the buying decisions concern larger budgets and have larger implications for the buyer than in B2C marketplaces, it is more important that the recommendations given are exact matches that can be explained, which is our *Requirement (R3): The system shall provide explanations*.

Requirements *(R1)–(R3)* resulted in the design decision to exploit semantic technologies by modeling the domain of knowledge as an ontology, realizing the recommendations as rules with explanations and integrating heterogeneous data sources via semantic data integration. We further decided to use ObjectLogic as ontology language, because it offers the possibility of modeling the domain of knowledge and business rules within the same framework, and has an integrated explanation feature. According to the identified business scenarios, a number of recommendation strategies have been analyzed in depth and represented as business rules.

Recommendation Strategies We have analyzed recommendation strategies described in Blanco-Fernández et al. (2008) and adapted them to fit the requirements of the B2B marketplace, giving special attention to the fact that (1) products are bundled in configurations, and (2) products are recommended to users who have a function within an organization and use the portal in different scenarios. We use different types and combinations of recommendation strategies depending on knowledge of the company's business and user context. Intuitively, a sales manager of a regular customer from a well-represented industry, such as IT Services, will receive recommendations based on his or her history and on the products used by similar customers with similar configurations, while a small company (SME) with a rather unique type of business will receive recommendations based on inferred knowledge from its own history, and on information from external data sources. The basic recommendation strategies used by our system are *(RS1)–(RS3)* as follows.

(RS1) *Recommend products from similar configurations.* This recommendation strategy recommends products from configurations similar to the customer's purchased product configuration. The similarity of configurations is determined by a similarity measure.

(RS2) *Recommend products of similar customers.* This recommendation strategy recommends products used by customers similar to the reference customer. The similarity of customers is given by metrics such as industry sector, number of employees, and budget. While the similarity of customers is usually defined by some measure depending on the transaction history, we define it only based on the customer's company information, as described in Table 1. This information is available for new customers without transaction history as well as for signed customers with transaction history.

(RS3) *Recommend products from the customer's domain of interest.* This recommendation strategy recommends products from product groups where the customer has already purchased products, which is interpreted as the domain of interest of the customer.

To return better recommendations, strategies *(RS1)–(RS3)* can be combined into mixed strategies.

(RS4) *Recommend products from similar configurations of similar customers.* This recommendation strategy combines two elementary recommendations and will return a better fit of recommendations.

(RS5) *Recommend product add-ons for products from the customer's domain of interest and used by similar customers.* Strategy *(RS5)* takes the results (i.e., intersection of answer sets) of recommendations strategies *(RS2)* and *(RS3)* and recommends product add-ons that match the recommended product. Under this strategy, sales employees of an automotive supplier will receive recommendations for a book on "Activity Management", because the product "Activity Management" is in their domain of interest, and because the product "Activity Management" is used by similar customers.

Architecture and Implementation The architecture of the recommendation system relies on the semantic infrastructure depicted in Fig. 1. The B2B portal as client application queries OntoBroker by sending the identifier (ID) of a customer

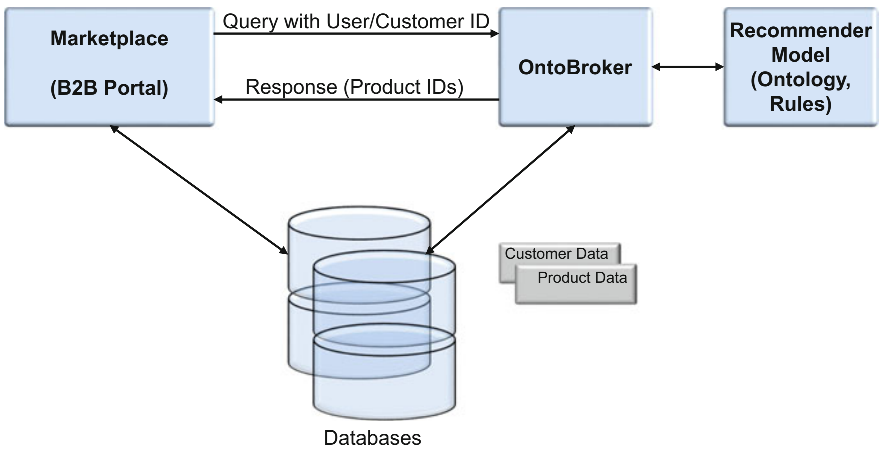


Fig. 2 Architecture of the ontology-based recommendation system

or user and receives as answer set the identifiers (IDs) of the products to be recommended (Fig. 2). The domain of knowledge is modeled as business ontology ("Recommender Ontology"). Application data, such as customer and product data, is stored in databases and matched to the ontology using rule-based semantic data integration. The Recommender Ontology itself does not contain any data and is used as the unifying common data model.

The main design decisions for the Recommender Ontology were to build (1) a unifying data model reusing existing CRM and product models, (2) in a modular way, (3) using rules to model the recommendations, and thus (4) using ObjectLogic as ontology language. Ontology modeling was carried out collaboratively by a business professional expert in CRM and recommendation systems, and an ontology engineer with expertise in rule modeling.

The domain of knowledge in the case of a B2B portal is defined by the data model of a standard CRM System such as SAP CRM. The core model is shown in Fig. 3. The main concept groups are customers, products/services, and ratings.

A customer of a B2B marketplace uses a configuration of products on terms that are specified in an order or contract. Customers are modeled as hierarchies of organizations and products are modeled in a hierarchy of products. Organizations belong to an industry sector and have business units located in different locations. Organizations have employees, i.e., persons who belong to a business unit and have a specific function in the organization. Products are offered by a vendor and belong to a product type. Products are bundled in configurations.

The recommendation strategies are represented as ObjectLogic rules. In the first step, the metrics for determining similarity of customers and product configurations are modeled as rules defining *similarTo* relations. We define in Table 1 that a customer *?Customer1* is similar to a customer *?Customer2* and write

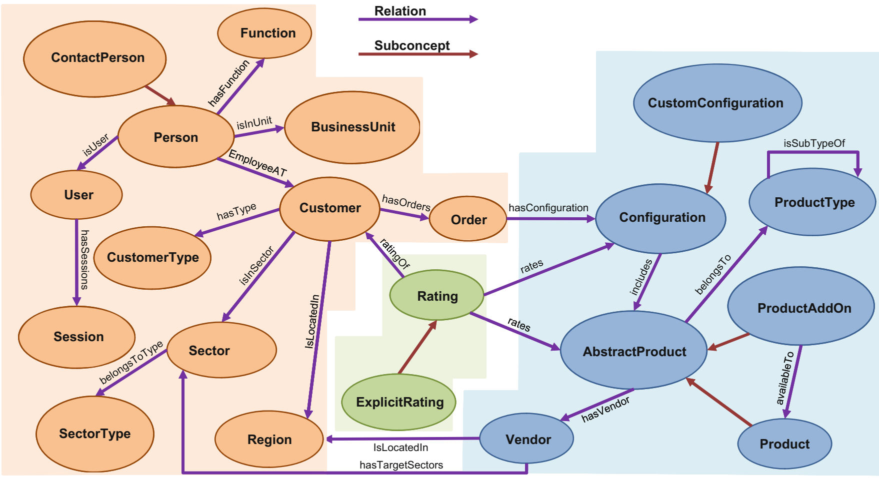


Fig. 3 Data model of the ontology-based recommendation system

?Customer1[similarTo->?Customer2] if they have the same sector, similar number of employees, and similar business volume.

In the next step, the recommendation rules are modeled using the previously defined metrics. As an example, the rule named *recommendProductOfSimilarCustomers*, which is represented in Table 2 in natural language and in ObjectLogic syntax, models the recommendation strategy (RS2). The ObjectLogic syntax conveys an intuitive idea of the language while a modeler might prefer the UML-like graphical rule editor to design rules.

The Recommender Ontology provides a generic data model for customer and product relations, and is not restricted to specific industry sectors or product types. In particular, it is possible to consider USDL services as a subtype of products. A typical usage of the Service Ontology in this context is to describe services in USDL, enrich them with legal and pricing information from the corresponding modules of the Service Ontology, import them into an ObjectLogic ontology using the OntoStudio OWL import functionality, and map this ObjectLogic Service Ontology with the help of a MappingAssistant (Angele and Gesmann 2006; Noessner et al. 2011) to the Recommender Ontology.

The rules that represent the recommendation strategies have a modular structure, allowing for complex rules to be assembled from elementary rules using conjunction. Business experts can thus modify the recommendations on the fly, without

Table 1 ObjectLogic rule defining the similarity of customers

Rule (ObjectLogic syntax)

`?Customer1[`**`similarTo`**`->?Customer2] :- ?Customer1:Customer AND`
`?Customer1[sameSectorType->?Customer2] AND`
`?Customer1[hasSimilarNofEmployees->?Customer2] AND`
`?Customer1[hasSimilarBusinessVolume->?Customer2].`

Table 2 ObjectLogic rule defining recommendation strategy *(RS2)*

Rule (Natural language)
If
?Customer1 is a Customer and has Configuration ?Config1 and
?Customer2 is a Customer and has Configuration ?Config2 and
?Customer1 is similar to ?Customer2 and
?Product is in ?Config2 and not in ?Config1
Then
recommend to customer ?Customer1 the product ?Product.
Rule (ObjectLogic syntax)
?Customer1[**recommendProductOfSimilarCustomer**s->?Product] :-
?Customer1:Customer[hasConfig->?Config1] AND
?Customer2:Customer[hasConfig->?Config2] AND
?Customer1[similarTo->?Customer2] AND
?Config2:Configuration[hasSetDifference(?Config1)->?Product].

having to know details about the modeling of the elementary rules, or how to program, or having to redeploy the application. For example, in the rule *recommend ProductOfSimilarCustomers* in Table 2, it is possible to modify the recommendation strategy by replacing the metric *similarTo* with another one, *moreSimilarTo*.

4 Conclusion

We believe that the Internet of Services with semantic technologies as enabler will further establish itself within the next few years, sustaining B2C and B2B marketplaces and related applications with a unified architecture. As result of the THESEUS research program, the semantic infrastructure has become more stable and mature, and has better support for data integration and interoperability, thus increasing acceptance among more conservative buyers. The semantic technologies community has started to offer solutions beyond the hype, oriented to the needs of business and market, such as recommendation systems or semantic service discovery, while keeping alive its youthful impetus and innovation factor.

References

J. Angele, OntoBroker – mature and approved semantic middleware, in *Semantic Web – Interoperability, Usability, Applicability*, **5**(3) 2014, pp. 221–235, IOS Press

J. Angele, M. Gesmann, Data integration using semantic technology: a use case, in *2nd International Conference on Rules and Rule Markup Languages for the Semantic Web*, Athens, Nov 2006, pp. 58–66

J. Angele, M. Kifer, G. Lausen, Ontologies in F-logic, in *Handbook on Ontologies*, ed. by S. Staab, R. Studer. International Handbooks on Information Systems (Springer, Berlin/Heidelberg/New York, 2009), pp. 45–70

J. Autere, T.O. Korhonen, Innovations in the internet and mobile era: the real dot.com revolution web, in *Handbook of Product and Service Development in Communication and Information Technology*, ed. by T.O. Korhonen, A. Ainamo (Springer, Berlin/Heidelberg/New York, 2004), pp. 171–192

Y. Blanco-Fernández, J.J. Pazos-Arias, A. Gil-Solla, M. Ramos-Cabrer, M. López-Nores, J. García-Duque, A. Fernández-Vilas, R.P. Díaz-Redondo, Exploiting synergies between semantic reasoning and personalization strategies in intelligent recommender systems: a case study. J. Syst. Softw. **81**(12), 2371–2385 (2008)

D. Calvanese, J. Carroll, G. De Giacomo, I. Herman, B. Parsia, P.F. Patel-Schneider, A. Ruttenberg, U. Sattler, OWL 2 web ontology language – profiles, 2009

M. Davis, M. Radoff, A. Panasenko, R. Riestra, R. Lockwood, M.E. Beltran, D3.2 final demand driven mapping report. Technical report, The VALUE-IT Consortium, 2013. http://www.value-it.eu/c/document_library/get_file?p_l_id=12732&folderId=12741&name=DLFE-1014.pdf

Forrester Research, Enterprise and SMB software survey, North America and Europe, Q4 2008. Technical report, Forrester Research, 2009

T. Gruber, Ontology, in *Encyclopedia of Database Systems*, ed. by L. Liu, M.T. Özsu (Springer, Berlin/Heidelberg/New York, 2009), pp. 1963–1965

M. Kifer, G. Lausen, J. Wu, Logical foundations of object-oriented and frame-based languages. J. ACM **42**(4), 741–843 (1995)

E.M. Kiss, ONTORULE D8.4 market intelligence report, 2013. http://ontorule-project.eu/outcomes?func=fileinfo\&id=51

G. Linden, B. Smith, J. York, Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Comput. **7**(1), 76–80 (2003)

J. Noessner, F. Fallahi, E.M. Kiss, H. Stuckenschmidt, Interactive data integration with MappingAssistant, in *The 10th International Semantic Web Conference (ISWC'11)*, Bonn, Oct 2011, pp. 1–4. http://iswc2011.semanticweb.org/fileadmin/iswc/Papers/PostersDemos/iswc11pd_submission_01.pdf

D. Oberle, N. Bhatti, S. Brockmans, M. Niemann, C. Janiesch, Countering service information challenges in the internet of services. Bus. Inf. Syst. Eng. (BISE) **1**(5), 370–390 (2009). http://dblp.uni-trier.de/db/journals/bise/bise1.html#OberleBBNJ09

Part IV
Program Organization and Dissemination

Fostering Innovation with the THESEUS Research Program

Hendrik Speck, Nico Weiner, Ralph Traphöner, Thomas Renner, Herbert Weber, and Walter Mattauch

Abstract The THESEUS research program, launched by the German Federal Ministry for Economic Affairs and Energy (BMWi), focuses on the development of intelligent information management technologies. THESEUS is envisioned to facilitate access to information, and to merge and connect information to establish the basis for the development of new Internet services.

1 Introduction

The goal of the THESEUS research program[1] is the development of promising technologies, products and services which can be transformed from core technologies and prototypes into innovative tools, products and services. In addition to government subsidies of 100 million €, the participating partners (industry, research and universities) provide the same amount, so that a total of

[1] http://theseus.pt-dlr.de/en/index.php.

H. Speck (✉)
FH Kaiserslautern, Zweibrücken, Germany
e-mail: contact@hendrikspeck.com

N. Weiner • T. Renner
Fraunhofer Institute for Industrial Engineering (IAO), Stuttgart, Germany
e-mail: Nico.Weiner@iao.fraunhofer.de; thomas.renner@iao.fraunhofer.de

R. Traphöner
Empolis Information Management GmbH, Bielefeld, Germany
e-mail: ralph.traphoener@empolis.com

H. Weber • W. Mattauch
Fraunhofer Institute for Open Communication Systems (Fokus), Berlin, Germany
e-mail: herbert.weber@fokus.fraunhofer.de; walter.mattauch@fokus.fraunhofer.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_36,

200 million € could be invested in cutting-edge technologies for the Internet of Things and the German research landscape. Through the incorporated companies the risk could be distributed and a stronger direct engagement of the program partners could be achieved.

In THESEUS 60 industrial and scientific partners worked together on six different application scenarios (*Use Cases*, see Kuhlmann et al. 2014) and in the CORE TECHNOLOGY CLUSTER (CTC, see Becker et al. 2014). The application scenarios were explicitly focused on technologies, services and the corresponding industrial utilization. The core technologies were determined through an analysis of the technological requirements at the beginning of the program. In turn, in the CORE TECHNOLOGY CLUSTER technologies and services were developed that were requested in several application scenarios and that benefitted from a centralized processing and consequent networking of services and capabilities.

The proposals were processed and the program was coordinated in a central program office. Independent consultants accompanied the work during the term of the program, special requirements regarding usage, publication and networking supported the goal of the program, and an independent research group evaluated and documented the entire process.

Especially, the continued support and the conscious concept of the program structure and innovation program could guarantee a direct usage of the research findings. The separation of application scenarios and core technologies led to synergy effects. The sustainability was ensured through the promotion of innovation at different levels, in which large enterprises, research institutions, SMEs and individual developers could work together on different assignments based on common technologies.

2 Organization, Coordination and Controlling

The German Aerospace Center, Deutsches Zentrum für Luft- und Raumfahrt (DLR) took on the role as project executing organization for the research program (PT-DLR). In this position DLR was responsible for financial controlling and program management, analyzing technological trends and controlling of assessment and evaluation processes.

Within the scope of the THESEUS research program Empolis took on the central and far-reaching role of program coordinator. Following the formation of the consortium and notification of the program, the Program Office acted as an interface between the representatives of the THESEUS Steering Committee,[2] the PT-DLR and the BMWi.

[2]The steering committee had the overall responsibility for organizing the execution and implementation of the research program carried out by the consortium partners. It made program-related decisions for the entire program and coordinated monitoring of the implementation of the various projects.

The THESEUS Program Office (TPO) coordinated the work of the consortium and provided comprehensive supporting services such as public relations, program management (work organization and milestone planning) as well as internal and external networking of all partners. TPO also coordinated the implementation of developed strategies and recommendations in collaboration with the steering committee. Furthermore, TPO was the consortium representative to BMWi, as well as PT-DLR. In addition to this, TPO also negotiated contacts and supported dialog with parties interested in the research program. TPO maintained the program networking with the THESEUS Joint Research and coordinated the cooperation of all parties among themselves.

3 THESEUS SME

In the second funding level of the THESEUS research program, the THESEUS SME program, small and medium-sized companies from Germany were invited to access the technologies developed within THESEUS, to develop new products and services and to create new service and business models.

THESEUS SME intended to establish competency networks between individual SMEs and other program partners, shorten development times, increase the competitiveness of the partners and enable synergies. The required and early cooperation and knowledge transfer led directly to stronger cooperation between the individual program partners; 12 additional application scenarios and 30 SMEs enriched the research program.

The direct integration into the THESEUS research program allowed start-ups and founders to avoid the usual discrimination of such enterprises, interested partners could apply for federal subsidies as or for venture capital, and there were no limitations regarding size, scale and structure of the projects.

The THESEUS SME program was relatively open; required was only the thematic relationship to the research program. Partners were selected based only on the quality of the project idea, innovation and value, as well as on the connection and thematic proximity to the results of the CORE TECHNOLOGY CLUSTER and the addition or expansion of the research scenarios through new applications and technologies. THESEUS SME succeeded in promoting innovation across several institutions. Innovation, growth and employment was strengthened not only in major research institutions, but also in SMEs and the affiliated companies.

In the assessment phase, proposals were selected to be considered in the subsequent consultation and application phase. A final decision about scope and amount of funding was made by the project sponsor and donor. Generally binding rules were established in the agreements of the consortium; special use agreements for specific application scenarios were made in individual agreements.

4 THESEUS Talent

The third level of the THESEUS research program, the THESEUS Talent Competition, was targeted towards the uncovered developer scene in Germany. The innovative potential from pupils, students, researchers, scientists and programmers, even from outside of the THESEUS research program, should be made available to the program. In addition to the awards, the program offered marketing and promotion services in press and media, mentoring and integration opportunities, as well as internships, diplomas and master projects.

DLR carried out a preliminary assessment of the submissions and created a list of potential candidates, which was then evaluated by an independent jury from science and industry. The winners of the THESEUS Talent Competition were honored at a ceremony in the summer of 2008.

The assignments of the THESEUS Talent Competition focused on concrete and relevant problems and research scenarios from individual application scenarios. In addition, an open challenge from the context of the application scenario, with more leeway for creative ideas and approaches, was offered. Assignments included, for instance:

- How can quality management programs provide an automatic analysis of user generated submissions and content?
- What could the semantic visualization of music look like based on the archived data of the music as well as on data inferred from the individual files and pieces, including moods, harmonies, tact and dynamics?
- How can interfaces between individual visualization components of the Semantic Web be standardized?

5 Transversal Groups

The transversal groups have a cross-sectional position in the THESEUS research program. Both groups, the business model and the system architecture group, support use cases, as for example, TEXO or PROCESSUS with basic information and services in their specific topic.

Based on theoretical and empirical research, the business model group provided a place for exchange of expertise regarding the design of business models for the duration of the THESEUS research program. The main goal of the group was to neutrally catalyze discussions within the sub-projects about business model variants for the solutions created. Therefore, typical activities of this group during the last years did include hosting of workshops with partners from all use cases, and hosting of larger events like "Stuttgarter E-Business Tage" to create awareness for the business model topic and to create guidelines and publications about the design of business models. For the whole duration of the project of 4 years, the group was led by Fraunhofer Institute for Industrial Engineering (IAO) in Stuttgart, Germany.

The task of the transversal group on system architectures was to gather information on the different architectural approaches throughout the THESEUS research program. It identified best practices and gave advice with respect to the state of the art in service oriented architectures to foster standardization and the use of standards.

The main achievements of the business model group were methods and models for the structured business model design. The resulting framework was called "[moby] – methodology for business agility" and supports the creation of business models in a step-by-step manner. Additionally, the group introduced the web-based software for visual business model design at the end of 2011, called the "[moby]-Business Model Designer".

The key result of the architecture group's work is a reference model for the description of applications in the Internet of Services. It distinguishes four types of services, ranging from data access to applications, to provide an easy-to-understand perspective on service-based systems.

6 Monitoring Collaborative Research in THESEUS

With the growing dependence of economic growth and prosperity on technological innovations, successful research has become a critical success factor in many countries and societies. Research is meant to represent the first step in the innovation chain towards the deployment of new products and services into marketplaces.

Significant budgets have been assigned now to research in many countries to succeed in the innovation race. In many cases the size of research projects and the volume of resources allocated to projects has grown significantly with the growth of the research budgets to the extent that their management has become very complex and risky. THESEUS, e.g., did use a total budget of more than 200 million € and the "Future Internet Public-Private Partnership" Project of the European Commission just starting is expected to consume 600 million € over the next 5 years. As a consequence, the monitoring and management of these programs and projects require new techniques for their successful completion. This led THESEUS to the separation of the "monitoring" of the THESEUS work on one hand and its day-to-day program coordination and management on the other hand.

The monitoring task "THESEUS Joint Research" was commissioned by the BMWi and assigned to an outside organization that was charged to act on a "middle ground" between the different THESEUS projects and their program management and the public sponsor of the program. It compared the technologies and applications developed in THESEUS with similar European and international initiatives and linked the THESEUS consortium with international partners and initiatives. The agent for the joint research was the Fraunhofer Institute for Software and Systems Engineering (ISST).

In this position it was meant to be of service to both of these sides. THESEUS Joint Research has been providing information and counseling likewise to the ministry as their primary client and to the collaborators of the THESEUS use cases and CORE TECHNOLOGY CLUSTER.

The profiling of THESEUS results led to the implementation of a THESEUS asset store, a repository with some 1,800 entries representing the THESEUS research results. A slim search application on the Web[3] allows for accessing THESEUS results through a full text query, but also by filtering according to the type of result, the organizations involved, and in other ways.

In the future, THESEUS Joint research will continue to work on these kinds of research management tools and intends to develop an ambitious instrument for the evaluation and the dissemination of research results.

7 Conclusion

Especially through the structural design, but also through the multi-layered approach to promote innovation, the THESEUS research program succeeded in connecting potential stakeholders and interest groups, which would not be possible with a traditional tender process.

During the term of the THESEUS research program, more and more program partners experienced the possibilities and the standard-enabling competitive advantages of the open source movement, and used them within their application scenarios and business models. Through participation in standardization agencies an early awareness for and development of standards could be made possible, resulting in a stronger reflection of the requirements of the private sector and also of social values and norms.

THESEUS SME and the THESEUS Talent Competition could integrate different developer and innovation potential; potential employees could be recruited. It remains to be desired to provide such programs with a long term perspective, which will allow synergies independent from theme and project.

The structural development, including a further development of such methods and content to promote innovation, the strengthening of the usage and networking aspects, increased reliance on open source and participation in the development of standards will be commonly used in the near future. Even the European Commission has recognized this – the FPP programs of the European Union are oriented along the structure of the THESEUS research program.

If future programs to promote innovation are designed accordingly, then THESEUS, fulfilling the Greek myth, in which Theseus must enter the labyrinth of the Daidalos and kill the Minotaur with a sword, will do more than justice to the historical reference. The Ariadne thread used by THESEUS, the prototype of a semantic navigation system, will then be understood as a true measure for the promotion of innovation.

[3] http://www.theseus.joint-research.org/theseus-ergebnisprisma.

References

T. Becker, C. Burghart, K. Nazemi, P. Ndjiki, T. Riegel, R. Schäfer, T. Sporer, V. Tresp, Core technologies for the internet of services, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

F. Kuhlmann, J. Hannemann, M. Traub, C. Böhme, S. Zillner, A. Cavallaro, S. Seifert, B. Decker, R. Traphöner, S. Kayser, U. Lindemann, S. Prasse, G. Marczinski, R. Grützner, A. Fasse, D. Oberle, The THESEUS use cases, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

From Idea to Market: The THESEUS Innovation Center for the Internet of Services

Gudrun Quandel

Abstract How can a novel abstract theme like the Internet of Services be made readily understandable to a broad target audience? And what kind of support can be given to the transitional phase of developments made in the THESEUS – Technologies for the Internet of Services – research program? It was the quest for answers to these two basic questions that led to the decision to found the THESEUS Innovation Center for the Internet of Services in 2009.

1 Introduction

Since mid-2010 the THESEUS Innovation Center has served as a platform for disseminating the aims, objectives and concrete results of this ambitious program – but also as a marketing platform for the R&D products and solutions of the Center's some 60 partners. The core of the Center consists of some 40 demonstrators. Clustered in the five theme areas – Medicine, Knowledge, Services, Business Processes and Multimedia – they give interested parties and, above all, potential users the opportunity to discuss, explore and try out the applications they themselves envisage. Regular events and training programs on selected themes with bearing on the Internet of Services are also held. And the Innovation Center will continue its existence after the official close of the THESEUS research program, with a new strategic focus on Data and Services for the Internet. For more information on the THESEUS program see Speck et al. (2014).

Located in Berlin, the THESEUS Innovation Center for the Internet of Services is a place where people can gain direct hands-on experience with the Internet of Services and explore its myriad possibilities. Some 500 m^2 of space, attractively

G. Quandel (✉)
Fraunhofer Institute for Telecommunications (HHI), Berlin, Germany
e-mail: gudrun.quandel@hhi.fraunhofer.de

W. Wahlster et al. (eds.), *Towards the Internet of Services: The THESEUS Research Program*, Cognitive Technologies, DOI 10.1007/978-3-319-06755-1_37,

situated in the heart of Berlin, showcases a broad range of prototypes from results of the THESEUS research program, and an array of interactive applications with ample space for dialog between researchers and users – all with the overarching aim of establishing the Internet of Services securely in the small and medium enterprise sector, in industry and in public administration. Yet it's not just the demonstrators alone, but the shape and feel of the Center itself, which is designed to aid in understanding what the Internet of Services stands for. Conceptual planning, design and organization of the THESEUS Innovation Center are by the Fraunhofer Institute for Telecommunications (HHI).

The more highly abstract a theme is – and for many people the Internet of Services is even more abstract than the Internet of Things or the Internet itself – the more important it becomes to communicate and present it in an analog way. This basic idea was paramount in shaping the conception of the Center and its space and enfolding environment. Above all, the idea of visualizing something which by its very nature cannot be seen – like data, services or the Internet – was designed to promote understanding – and especially the kind of understanding that enables every visitor to recognize how the issue impacts on their own situations and businesses – and to take action based on this understanding.

"Visualization" was the first major challenge – visualization in pictures and in words. The space itself and its layout should aid in promoting an understanding of what the THESEUS research program and the Internet of Services is all about. It was such kinds of considerations that led to the development of a space that tells a story as each visitor enters the world of the Internet. The architecture of the space and its uniform optical and graphic design create the impression that what you've entered is a place without corners and hard edges, a space of near-zero gravity. A place that mirrors an understanding of the Internet itself as an environment without top or bottom, without right angles and so on. And this is the impression that many visitors do indeed get. Obviously, the laws of physics cannot be abolished but they can be temporarily pushed into the background.

This impression is corroborated by the second part of the story: a broad array of data are in motion in this virtual space, symbolized here through partly animated polygons which intelligent instruments developed in the THESEUS research program trawl, find, identify and bring together, creating new Digital Assets or new values for new knowledge, new products and new services. The room itself, with its innovative design and story lines, first transports the visitor emotionally into a virtual world which is the starting point for all subsequent discussions and engagement with the Internet of Services.

2 From the Idea to the Marketplace

The Innovation Center for the Internet of Services is designed to substantially shorten the route that leads from the original research idea to the marketable product for the THESEUS research program. Through its special concept, the THESEUS

Fig. 1 The *Demonstration/Presentation Area* of the THESEUS Innovation Center

Innovation Center gives tangible direct experience of the whole process from research to marketplace.

The *R&D Lab* is the "heart" of the Innovation Center. It is the starting point of the journey from idea to market-ready product with the right kind of business model. It is where experts and project groups meet, where system chains are forged, and prototypes from various providers tested. And where interested visitors have the opportunity to take a "behind the scenes" look at a research program or process.

The results of the R&D work in the THESEUS research program – mainly demonstrators and prototypes – are showcased in the *Demonstration/Presentation Area* where they can be discussed and explored by interested parties and potential users (Fig. 1). Developments and scenarios of the Center's partners are clustered in the five application areas of Medicine, Knowledge, Services, Business Processes and Multimedia. Such a practically oriented presentation makes it easy for visitors to the Center to get a clear general idea as the groups of various demonstrators relate to one specific theme. Basic technologies and application scenarios in each theme area are linked together so that visitors can follow their own particular interests in terms of technologies or applications without neglecting the other side. The generously proportioned space of the *Communication Area* is available for functions and events for groups both large and small. This is an area for the in-depth discussion of topics and issues that serves especially to promote dialog between users, mount training programs and build networks. The working and *Innovation Area* is designed for in-depth workshops or incubators for the development of new ideas and strategies.

The THESEUS Innovation Center implements the innovation process in the language of architecture. The number of manifold ways of use for such an innovation environment means that the THESEUS Innovation Center for the Internet of Services is at one and the same time an open research and development laboratory, a showroom for the latest research findings, an Information Center for the Internet of Services, a communication platform for research, industry and politics, a try-out platform for applications, an incubator for new ideas covering all aspects of the Internet of Services, and a transfer and marketing instrument.

3 Live and Hands-On and Tangible: Demonstrators and Events

Behind the market-ready product stands the prototype, the demonstrator. Once the intensive run-in period is complete, the demonstrators stand ready to exemplify the Core Technologies (Becker et al. 2014) and typical use cases (Kuhlmann et al. 2014) for the Internet of Services. Clustered in five application areas – Medicine, Knowledge, Services, Business Processes and Multimedia – the demonstrators show the present status of development while also acting as the springboard for in-depth discussions and feedback.

Presentation of the demonstrators – and by the end of 2011 the THESEUS Innovation Center had 38 of them – is generally the starting point or highlight of every event. The excitement of the live experience and the practical opportunity to apply what is being demonstrated to the visitor's own situation play a large part in creating the sustainable long-term effect the Center aspires to: what are the benefits for me as a small company, a medium-sized company, a utility operator, a media enterprise, etc.? And what exactly must I do to harness this solution either as the answer to a specific problem or as the basis for a new Internet-driven company?

The demonstrators in the THESEUS Innovation Center are also used for training programs on special themes of particular relevance to the third part of the program. Executive Training programs give decision-makers an introduction to future technologies and their impact on old and new industry sectors, while application training enables those taking part to try out R&D solutions in their own particular contexts. A selection of demonstrators showing the breadth of the development spectrum is given in Table 1. One prominent feature of the THESEUS Innovation Center for the Internet of Services is its "theme days" or "theme weeks" which offer events, covering all aspects of the Internet of Services, such as *Business Models for Cloud Computing*, *Efficient Use of Digital Media*, *SMILA – The Framework for Semantic Applications*, *Digital Imaging in Medicine*, *Social Media Mining* or SME Conventions and Business Start-up Days. The THESEUS Innovation Center is also ideal as a platform for new formats true to its goal of establishing new contacts and developing new possibilities of practical import such as "Business Speed Dating", dedicated to building contacts between the partners in THESEUS and prospective users, or "Science Slam", a kind of short competitive presentation leading to new ideas and new contacts. In an 18 month period, a total of 100 events were held attracting just under 3,000 participants from industry, government and academia.

4 The Innovation Center as a R&D Marketing Instrument: Lessons Learned

At the very beginning the THESEUS Innovation Center for the Internet of Services was an experiment embarked on out of the need to give visibility to and to support the marketing of R&D results achieved in the THESEUS research program as well

Table 1 Selection of demonstrators

Theme Area Medicine
- MEDICO, a knowledge platform enabling physicians to compare patient illnesses in a patient archive

Theme Area Knowledge
- ALEXANDRIA, a collaborative knowledge platform that enables the simple use of semantic technologies
- SemaVis, enables visualization of information and its contexts and correlations and thus offers insight into the way information will be presented in the future

Theme Area Services
- USDL Editor, enables the rapid and easy description of services
- Automobile damage claims settlement and eco-calculator use transparent web-based services to safeguard the interests of claimants

Theme Area Business Processes
- PROCESSUS helps identify ontological concepts
- TechWatchTool – The TechWatchTool aids companies in identifying new technologies and key players together with the networks they are embedded in
- SMILA – An Open Source Framework for semantic access to heterogeneous data sources

Theme Area Multimedia
- Text 2.0 uses an eye tracker to display information at the point when the reader wants it
- Semantic Video Retrieval
- "Simple" picture analysis results enable answers to be given to content-based enquiries in a database
- CONTENTUS enables retrieval of texts with semantically enriched information

as to give the Internet of Services a secure and sustainable positioning, particularly in the small and medium-sized enterprise sector. Even though the Innovation Center was only established in the third and final part of the THESEUS research program, its aims and objectives have thus far been implemented with notable success: Results achieved by THESEUS effectively find their way to target groups including creative IT companies, branch players and SMEs; selected formats enable target group aligned placement of themes and issues with bearing on THESEUS; correlations between individual developments could be made; and THESEUS partners use the platform created by the Innovation Center for marketing their products. On top of this, a number of lessons were learned which in general have led to a further refinement of this R&D marketing instrument, as e.g. the integration of an Innovation Center in a project from the very beginning, the allocation of funding for an Innovation Center as well as the definition of benchmarks for marketing success. Further experiences were gained in the early involvement of potential users in development (feedback), the clear target group orientation (customer focus) in terms of events, and in training THESEUS program partners in R&D marketing. In general it can be said that an Innovation Center always proves its worth if a project

- Is exceptionally multilayered and/or treats complex abstract issues,
- Involves a large number of partners each working highly independently on R&D solutions,
- Addresses a broad array of different target groups,

- Deals with a theme that has, or should have, strong relevance for social and economic development,
- Deals with a theme of intense topical interest.

5 Conclusion

The THESEUS Innovation Center for the Internet of Services has now created a comprehensive infrastructure for the marketing of R&D products and services. Based on the work done thus far in the THESEUS context, the THESEUS Innovation Center is poised to develop into a Competence Center for Data and Services in the Internet – its present working title. Big Data is now the subject of intensive debates and discussions throughout the world in view of the urgent need for the joint processing and evaluation of ever larger stocks of unstructured data from different sources in different formats. Furthermore, such data needs real-time, not retrospective, processing if it is to be fit for purpose for present and forward-looking decision-making. The new strategic focus on Data Analytics and Visual Analytics will bring together and synergize expertise from the THESEUS context with the activities of further partners from industry and science. Cloud Computing is another theme of high relevance to the Big Data context. And the – distributed – allocation of resources for the new challenges Data Analytics brings with it makes the development of Internet-based services a further key focal point.

The Competence Center for Data and Services in the Internet is open to all programs and projects of industry and science in Germany and Europe. Its partnership program will secure its financing and give partners access to its wide range of offers. The THESEUS Innovation Center for the Internet of Services is an experiment in the sustainable implementation of R&D results.

References

T. Becker, C. Burghart, K. Nazemi, P. Ndjiki, T. Riegel, R. Schäfer, T. Sporer, V. Tresp, Core technologies for the internet of services, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

F. Kuhlmann, J. Hannemann, M. Traub, C. Böhme, S. Zillner, A. Cavallaro, S. Seifert, B. Decker, R. Traphöner, S. Kayser, U. Lindemann, S. Prasse, G. Marczinski, R. Grützner, A. Fasse, D. Oberle, The THESEUS use cases, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)

H. Speck, N. Weiner, R. Traphöner, T. Renner, H. Weber, W. Mattauch, Fostering innovation with the THESEUS research program, in *Towards the Internet of Services: The THESEUS Research Program*, ed. by W. Wahlster, H.J. Grallert, S. Wess, H. Friedrich, T. Widenka (Springer, Berlin/Heidelberg/New York, 2014)