Knowledge-Based Semantification of Business Communications in ERP Environments

Marios Meimaris and Michalis Vafopoulos

School of Electrical & Computer Engineering, National Technical University of Athens
{m.meimaris,vaf}@medialab.ntua.gr

Abstract. The complexity of information in business environments is increasing in rates that are difficult for companies to handle using traditional solutions. The diversity of information and communications that stem from remote and heterogeneous sources creates the need for information management in multi-dimensional contexts and platforms, such as web sources, mobile devices and databases. As web technologies evolve, the volume, temporality and heterogeneity of available data exhibits a degree of dynamicity that is beyond manual control. Without these limitations, the aggregation of this information can provide new value-driving layers in business environments. Moreover, when this data is semantically structured, knowledge sharing and rule-based management allow for new ways of controlling the flow of information, both internally and externally. Focusing on the extraction of *information about communications* and its exploitation in platform-independent manners, we propose, design and implement a set of semantic components extending ERP systems in order to assist semantic web interoperability, to provide a basis for intelligent knowledge management and to unify communication level platforms under shared sets of principles.

Keywords: Semantic Web, Linked Open Data, Business Intelligence, Knowledge Management, Ontologies.

1 Introduction

Modern technologies and design approaches make the web an exceptionally dynamic environment, where data and information is created, consumed, edited and deleted within very small periods of time. This information can be in the form of communication between people and organizations, as in the example of social networking, or in the form of electronic publishing of facts and otherwise static data with varied time scopes. The dynamicity of these types of information makes it necessary for people and systems to be capable of being constantly up to date and ready to respond to challenges, problems and difficulties that result from change.

In the web business world, when it comes to responsiveness such aspects are crucial. Within business contexts, these types of information form a level that will be called the *Business Communications Level*. They can stem from both internal (e.g. databases, employee mobile phones etc.) and external (e.g. social media, web

A. Haller et al. (Eds.): WISE 2011 and 2012 Combined Workshops, LNCS 7652, pp. 159–172, 2013.

repositories etc.) sources. In the traditional sense, such data is usually neither reusable nor easily interchangeable between heterogeneous and remote sources. It is even more difficult to take advantage of such data sources in cases where response times have to be minimal, as is the case in the business world.

Aggregating these types of data and information and combining them through a unified schematic structure will become essential for business entities in order to survive and adapt to the required responsiveness. Moreover, the use of data that stems from mobile devices can provide and quantify contextual and operational information that was not taken into consideration before. For this reason, it is within our belief that providing platforms with the ability to support interoperability through interchanging diverse communicational data is a step toward an evolved business environment. We have designed and implemented a set of components that follow knowledge-based, semantic web and open data directives, as well as facilitate the extraction and reuse of mobile phone data under a shared ontological substrate, as an extension to traditional Enterprise Resource Planning (ERP) systems, in order to propose an ontology for communications and demonstrate how semantic extensions thrive in environments that are rich in diverse information.

This paper is organized as follows: section (2) will provide the background, section (3) will describe the conceptual modelling, section (4) will describe the system implementation and sections (5) and (6) will provide an evaluation as well as present the results and a brief discussion of this work's contribution.

2 Background

2.1 Knowledge Management and Business Intelligence

Generally, Knowledge Management (KM) can be considered as a loosely defined set of methodologies and techniques that aim at externalizing what is considered as internal, or tacit knowledge. A particular subset of Knowledge Management, as is suggested in [1] is *Business Intelligence* (BI), which is considered to be a knowledge level that provides analytical tools to support decision-making. With the incorporation of new technologies, classical BI functionality is extended to include context identification and socially-aware components, this way taking advantage of the richness of information that stems from social web as well as mobile environments.

Within corporate contexts, information can be found both internally and externally. Internal sources include databases, company knowledge, mobile communications and any other types of information that are directly controlled by the company. External sources include web 2.0 (social media, blogosphere etc.), semantic web services, open data repositories, the linked open data (LOD) cloud and so on.

2.2 Internal Sources

Databases and ERP Systems

It is usual for companies to hold their data in local databases. More often than not, these are traditional relational databases that use schemata designed and created in company-specific manners. In many cases, they are incorporated in Enterprise Resource Planning (ERP) or similar systems. ERP systems provide a framework for managing business-related resources and assets (both human and non-human) with respect to projects, processes and specific tasks that need to be accomplished. This way, interdepartmental cooperation and coordination can be achieved. Successful deployment of ERP systems reduces overall operating costs and improves enterprise data management and interdepartmental information exchange [2].

Smaizys & Vasilecas [3] argue that the absence of widely accepted standards and formalisms on business rules is what leads ERP system designers to time-consuming solutions which can often be both misleading and discouraging. It is therefore necessary for this data to be structured in ways that allow for sharing and combining with external data under shared and loosely coupled principles. Furthermore, deploying semantic technologies on ERP systems is essential for the development of web service oriented architectures in the semantic web [4].

Mobile Devices

Diverse smartphone functionalities have led companies to incorporate them into their operation management models. Graf and Tellian [5] talk about the benefits of using smartphones as trackers in Supply Chain Management, through the use of data connections and GPS. Furthermore, the app market has led companies to develop applications for managing human resources. Modern smartphones permit context-aware application development. For instance, they include metadata such as geospatial and chronological usage data. Furthermore, in the case of open operating systems, data extraction is feasible, making usage contexts and communication patterns identifiable and tractable.

Little work has been done to assess the use of mobile communication data, and it is within our hypothesis that its aggregation leads to a level of information, called the *Communication Level*, which makes possible the derivation of more complex information regarding the user and the developing communicational contexts they participate in. We build *communication profiles* of users and assess the derived networks.

2.3 External Sources

External data can generally be considered as dynamic or static. Dynamic data originates from social media and otherwise dynamic sources, such as Google+, Twitter and Facebook, web 2.0 platforms in general and so on, whereas static data includes unchanging (or slowly changing) information, such as business registries and other repositories. Both are powerful sources of information, but must be handled differently.

Social Media

We consider social data to be of high business value, because, if interpreted correctly, they hold key information and knowledge about public sentiment. Social media provide added value in many ways. Two main value drivers associated with them are *social media marketing* and *sentiment mining*. Sentiment analysis can provide valuable insight on the public opinion, as well as provide useful results in e-commerce [6].

(Linked) Open Data
Open Data refers to the notion of openness (i.e. free public access) in certain kinds of data which can range from crime reports to government decisions and from bus schedules to registry information about corporate entities. Open datasets are often designed to be machine-understandable and for this reason it is common for them to be shared through open document formats such as JSON, RDF and CSV. A particular case of data openness is that of *Linked Open Data*, which refers to a set of methodologies and guidelines for interlinking data on the web. The advantages of using open datasets such as OpenCorporates.com are that there is no need to store data locally, or tend to the needed updates, since such larger data catalogues publish and curate updated data.

The shift to semantic web technologies and (linked) open data paradigms in the business world is guided by the provision of unique descriptions of corporate entities as reference, publishing of information relevant to company-related affairs, such as accounting facts and figures, operating decisions and board changes and so on (e.g. *OpenCorporates.com*), as well as publishing offers for business transactions that include products and services (e.g. GoodRelations). For a discussion of the implications of an integration of linked open data in business environments, see [7].

3 Conceptual Design

3.1 Test Scenario

The domain of sales was considered as a test domain for the system. The graph-like nature of sales is based on people (actors in general) and their communications, thus making it a suitable test bed for our implementation. It is expandable, ranging from the presales to the after-sales service levels and many different tasks can be identified within it. The advantage of using semantic technologies in this approach is that a common substrate of meaning is employed in a context and platform independent way, for both incoming and outgoing data.

Communication instances stem from diverse sources, such as mobile phones, computers, landlines etc. Especially in large enterprise networks, this results in a highly interconnected graph of dynamic nature. Furthermore, the created network is not isolated from the external world and sometimes the boundaries are not distinct. The scenario is defined as follows:

> *The implemented system is used by a sales company. The combination of external and internal information triggers sales threats regarding prices and offers of particular products/service. The system notifies with alerts when threats are triggered. These might be concerning the same goods offered by competitors, or similar goods with matching functionality (e.g. a competitor company issues a better offer about a particular product, thus threatening our company's existing clientele). The implemented system handles the alerts in two ways:*

a. Automatic suggestions for handling a threat. These include identification of a subset of employees that is suitable to handle the threat, assignment of particular clients/client sets to these employees, suggestion of suitable communication strategies and so on.

b. Provision of related information and statistics for the particular threat for educated manual handling.

Past information drawn from the system is shared among salespeople within the company. The system automatically uses a set of classifications on the market and clients in order to outline the conditions, the results of which are made available to the salesperson. Finally, the salesperson responds to the threat by adapting her operational plan according to these. The results of her attempts are quantified and measured before they are made part of the shared knowledge, thus rebooting the cycle.

The purpose of the system is to demonstrate how corporate decision making can be assisted. The system's input is made up of the sources that were described in section II. The company's salespeople handle both new and existing clients. The system builds *communication profiles* for both clients and employees, based on past data. Therefore, when an existing client needs to be approached, the salesperson can consult the client's profile in order to select a strategy of approach. In the case of new clients, the market profile shows what communication strategy has been successful product-wise in the past.

3.2 Ontologies and Vocabularies

The *local* or *communications* ontology was created mainly to represent the *Communications* domain, which forms a large part of our approach. This has been created from scratch, always having in mind the principle of *minimal ontological commitment* [8], which simply states that in ontology design, the representational claims should be exactly as many as needed for the description of the domain of interest, in order to avoid knowledge redundancy. Classes can be seen in figure 1, with simplified names to assist readability. However, reusing existing ontologies is generally preferred over creating new ones. In our approach, parts of existing ontologies have been employed in order to describe high-level concepts and relationships such as people, organizations and companies, products and services and so on. The reuse of existing ontologies is a default recommendation, as it is quicker, less expensive and drives information sharing between diverse sources.

In our implementation, several external ontologies have been reused, such as the *GoodRelations Ontology* [9] and the *Friend-of-a-Friend* (FOAF) vocabulary. *GoodRelations* provides a conceptual framework for representing e-commerce offers of goods in a way that can be processed by web services. It is adopted by large corporations such as Google, Yahoo! and BestBuy. *FOAF* [10] is an ontology for the description of individuals and organizations, and the relationships between them. It is widely accepted as a reference model for the description of people and groups. Class equivalence between ontologies ensures the interlinking of these concepts in the structural level (e.g. lo cal:Agent is equivalent to foaf:Agent).

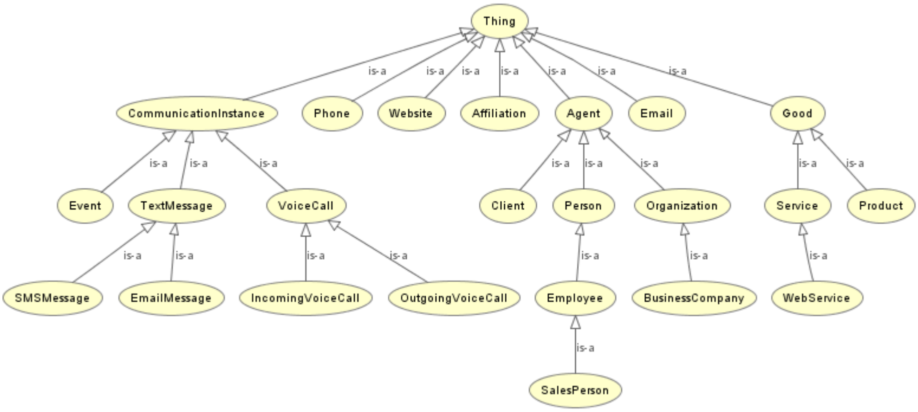

Fig. 1. Local and Communications Ontology

4 System Implementation

4.1 Technological Choices

The implemented system and all associated methods and classes are written in Java. Semantic web and ontology integration are done with the use of the Jena Semantic Web Framework [11]. Demonstration is implemented in the form of JSP (JavaServer Pages) on top of simple HTML, resulting in a web environment capable of providing basic demonstrative functionality.

We have chosen to expand on OpenERP [12], which has rounded capabilities and small size. Thomas Herzog [13] shows that OpenERP (cited as TinyERP at the time of writing) covers all necessary functionalities with a smaller set of database tables than the competing systems. For the mobile communications part, we use Android as the scenario's mobile environment. Android's programming framework is built on top of Java classes, making the incorporation of external Java libraries natural.

Data is converted and stored in the RDF format. Storing and linking data with semantic (ontological) structure is particularly easy with the use of RDF, a w3c recommendation. We have selected *OpenLink Virtuoso* to store RDF triples. The choice was based on the fact that it has rounded capabilities and good querying performance, as suggested by standardized benchmarking tests [14]. Furthermore, Virtuoso contains java libraries for connecting with the Jena Semantic Web Framework, making it less costly for the implemented system to perform RDF data transactions between the server and the various external platforms.

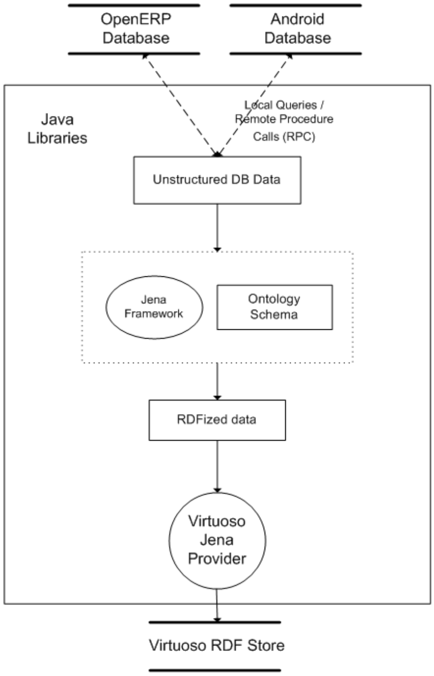

Fig. 2. Data extraction, RDF conversion and storing

4.2 Information Retrieval and Semantification

Various data sources are used for information retrieval in this scenario. Each of them is queried differently, according to the technologies it is subject to. They are converted to RDF using the Jena framework, the system's ontologies and a set of conversion rules (figure 2). After the conversion, they are stored in the quad store. RDF data is then offered depending on the context, and the use of negotiation rules ensures that the availability of the data is limited and controlled accordingly. In practice, this would mean that different subsets of the data will be made available to different parties at different contexts.

4.3 Information Aggregation and Utilization

Content Aggregation

As has already been established, ERP systems form the centre of the scenario, in the sense that the ERP is the main context driver. For this reason our implementation stands as an extension to its functionality, providing several layers, both structural and functional, around the domain of interest, mainly focusing on the benefits of the communicational approach. The structural layer is a consequence of the provision of semantics, while the functional layers are the sum of all extra functionalities that are implemented on top of the system's current ones. This can be seen in figure 3.

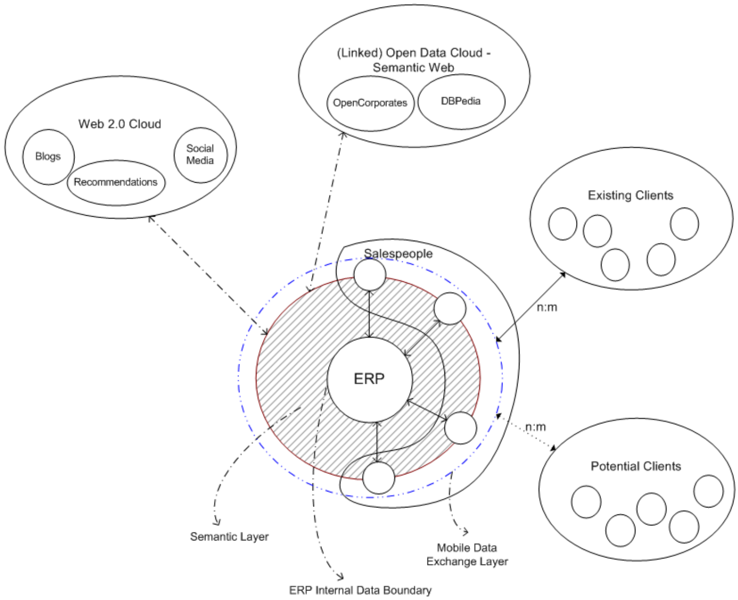

Fig. 3. Abstraction of the information layers

Information is shared within the dashed layer throughout the extended system. Practically, the data that exists within the information layer is in RDF form, stored in the dedicated quad store. The mobile data is exchanged at the boundary of the semantic layer. Finally, external information that stems from web sources such as social media and (linked) open data is exchanged through the boundaries of the semantic layer as well. Within these, data is stored in aggregation, under common semantic structure.

Web Interface

A web interface has been created for demonstration. This is a web environment where employees with managerial status are notified when price alerts are triggered. The implemented version includes a browsing environment of employees, clients and products/services, with respect to various sets of restrictions. Furthermore, the user can read descriptions and metadata of incoming alerts, get automatically-generated handling strategies, assign employees to clients and products, see communication profiles of employees and clients and read several data-centric analyses of cross-patterns between the variables.

a) Alert Notifications

Upon request of this webpage, the server is queried in order to return a list of the unhandled alerts, along with a subset of their non-functional properties. It is assumed that the alerts are created from external sources, mainly social media. However, the creation of such alerts lies beyond the scope of this paper.

By selecting an incoming alert notification, the web environment shows the user the alert's profile. As each price alert is created in reference to a particular product with a particular business function[1], the alert profile page will go on to provide an initial analysis of the alert based on these constants. Based on this design, an alert profile page offers department, employee and client rankings with respect to the alert.

b) Employee, Client and Product/Service profiles

Employee and Client views provide insights to the associated agent, including their static as well as dynamic information profiles. Employee and Client profiles are similar for the most part, with the differences being in the types of associations that are produced. Their conjunction is made up of the following information:

- Static information that defines the employee/client as an ERP entity, including contact details drawn from the Contacts Level. In the case of employees, their department associations are further provided. In the case of clients, a link to data drawn from OpenCorporates is also present.
- Top products/services associated with the particular agent, along with processed information. This includes transaction statistics, rankings by business function, prominent communication strategies and communication type statistics, paired employee-client rankings and so on.
- Rankings of communication types associated with the agent, including communication statistics by type or supertype, (e.g. email vs. text messaging in general), communication patterns that arise, preferred methods of communication by product or by client and so on.
- Rankings of business functions independently of the product/service they refer to, including total number of transactions and total income. This highlights the top-selling business functions for each employee/client/good.

c) Threat handling

Incoming threat alerts are ranked and classified based on their description. The system offers its users the possibility to assess the threat automatically, by suggesting the best means to handle it, based on predefined rule sets. Given that incoming threats are always in reference to a particular product/service, it is easy to identify the threatened clients and outline the context. This way we can evaluate historical data and suggest communication strategies based on the identified context. The overall alert resolution process can be seen in figure 4.

Evaluation of incoming alerts is followed by resolution. The alert metadata helps classify and categorize it, and assign it to appropriate handlers. Technically this can

[1] *Business functions* have to do with the different ways a product/service can be commercialized, e.g. Sell, Repair, Maintain etc. They are formally represented with GoodRelations.

be achieved as follows: the web environment is user-dependent (controlled with dynamically triggered ontology rules based on the user and their profile/history) and identification of the user restricts the query results accordingly.

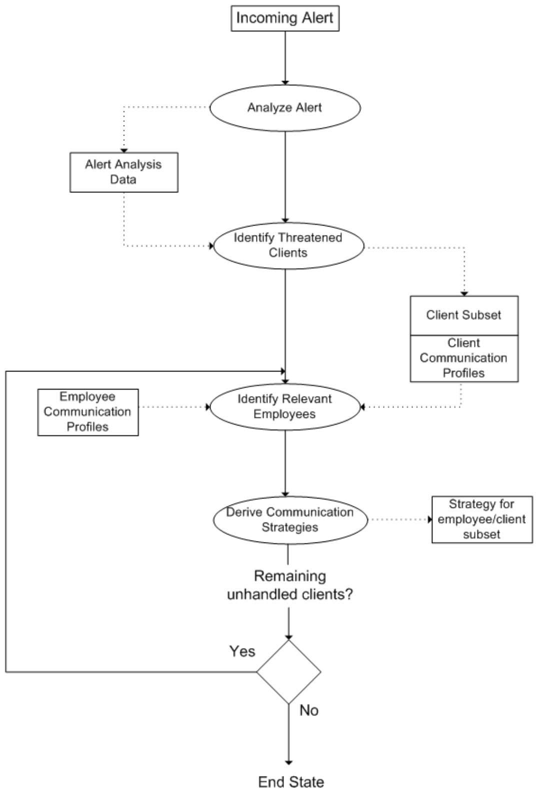

Fig. 4. Alert handling process

5 Evaluation

The main purpose of this work is to highlight the richness of information that business communications have to offer, and this is shown through a corporate use-case scenario. Furthermore, it is demonstrated how various components can be deployed both server-side (ERP extensions) and client-side (Android apps) in order to convert and publish data as RDF, with the use of ontologies. In this section we will focus on qualitative evaluation of the conceptual approach, as well as discuss implementation implications.

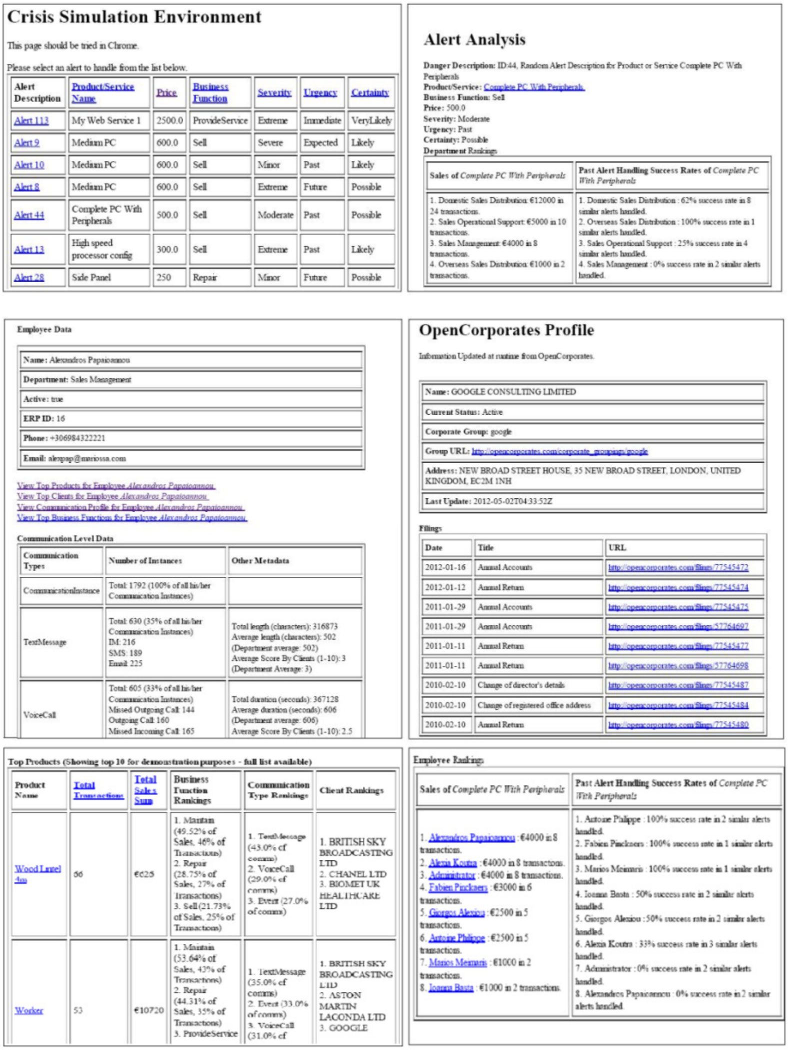

Fig. 5. Screenshots from demonstration interface

5.1 Communications Ontology

Our ontology captures the knowledge that is related to communications. These stem from mobile phones, landlines, the world wide web, and so on, but they all belong to the same communications taxonomy. As smartphones and modern computing solutions are able to capture metadata of specific communication instances, we show how this can be performed on the fly and in an ontology-based manner.

Heterogeneous sources are aggregated under a common denomination, creating value as super-layers of information are brought forth as a result of this mash-up. For instance, combining the knowledge of how a particular employee usually handles (communication-wise) clients of a particular nature, along with the knowledge of how successful in terms of sales this employee has been, we outline market conditions and client behaviors on the fly. This is particularly useful in environments where employees need to handle threats with minimal response times, as in the scenario. Combining this information helps companies in more than one ways. To name a few:

- Improved HR management: the communication profiles of employees create an in-house network of nodes and edges that show how employees communicate with each other. Sinks and sources can be identified, case-specific or overall.
- Educated threat-handling: when clients are threatened, we can identify the best strategy to approach the case at hand, depending on their communication profiles.
- Unknown market analysis: when entering new markets, communication patterns of goods in known markets are consulted and approach strategies are adapted accordingly.

5.2 Semantic Business Information Management

The added functionality from the superimposing of diverse data lies within the derivation of new information. The new information, because of the use of knowledge-based technologies, creates dynamic and on-demand rule-driven derivations that help classify, categorize and annotate entities on custom classifications. Knowledge about communications becomes a manageable and tractable asset and ontology revisions can be performed without any downtime.

Sharing and exchanging chunks of data within the company or with the outside world is subject to rules that trigger dynamically. For instance, depending on the status and history of an employee, she can have access to past historical data concerning the threat at hand. The temporality of the access is dependent on the context under which she consumes the requested information. SPIN [15] is an example of how ontology rules can be managed and used programmatically.

With the use of semantic web standards, the system is suited for integration with external semantic web resources. In the business environment, web services are emerging as ways of incorporating knowledge exchange in the markets, making the sharing of information, both inbound and outbound, a task that can be handled by automated software agents.

5.3 Implementation

Even though the implemented system serves to demonstrate some of the functionalities of the conceptual approach, several of the limitations that were imposed during the development reflect issues that should arise in commercial design and development. We will focus on the extraction, conversion and storing of internal data.

Quad Store

It was necessary to deploy a dedicated quad store that would act as the centralized store for all platforms. Being able to deploy such a store is a strong assumption and not a viable commercial solution. The need for two separate stores is a weakness, but necessary for platform-independent interoperation. Quad stores also offer ready-to-deploy data endpoint functionality (e.g. SPARQL) that are customizable (security-wise) and web-accessible. However, ERP systems that are designed to function on top of RDF databases will exhibit the benefits of this approach, and have built-in RDF creation without having to depend on conversion extensions.

Mobile Data

With this approach, mobile data extraction and conversion is inexpensive and decentralized, as each corporate phone is responsible for converting and uploading its own data. Because android phones cannot yet act as quad stores with SPARQL endpoints, RDF are stored in files in the phone and deleted after uploading. However, a more proper mechanism for storing should be deployed, without having to rely on the phone's file system.

ERP Data

The component for converting ERP data to RDF is deployed server-side, running in parallel to the ERP services and acting as a bridge between the ERP internal store and the central quad store. The conversion process is not costly and is triggered by SQL transactions with the use of event handlers at run-time.

6 Conclusions

This work provides a conceptual framework and a demonstrative implementation[2] for taking advantage of the business communication layer of information, which has grown to be a diverse environment of heterogeneous sources and causality relationships. Communications cover a large range of information flows from and to companies, both internal and external.

The proposed extensions form a unification layer around the company's data core that can exchange data on demand and under rule-based control. The implications as far as business management improvements are concerned, have to do with the following points.

1. Minimization of response time
2. Enrichment of internal context to that of an *open business environment*
3. Shared meaning and data, creation of *super-layers* of information that create value bigger than the summation of their constituents

---

[2] The code was part of the author's MBA thesis and can be found in `https://github.com/mmeimaris/android-rdfizer`.
It is our intention to keep it updated.

4. Exploitation of mobile information, as mobile devices provide an abundance of usage data and metadata, thus defining contexts (and sub-contexts) on their own.

References

1. Herschel, R., Yermish, I.: Knowledge Management in Business Intelligence. In: King, W.R. (ed.) Knowledge Management and Organizational Learning, Annals of Information Systems, vol. 4, pp. 131–143. Springer US (2009)
2. Choi, H.R., Kim, H.S., Park, B.J., Park, N.-K., Lee, S.W.: An ERP approach for container terminal operating systems. Maritime Policy & Management 30(3), 197–210 (2003)
3. Smaizys, A., Vasilecas, O.: Business Rules Based Agile ERP Systems Development. Informatica 2009 20(3), 439–460 (2009)
4. Tjoa, A.M., Shayeganfar, F., Anjomshoaa, A., Karim, S.: Exploitation of Semantic Web Technology in ERP Systems. In: Tjoa, A.M., Xu, L., Chaudhry, S. (eds.) Research and Practical Issues of Enterprise Information Systems. IFIP, vol. 205, pp. 417–427. Springer, Boston (2010)
5. Graf, H., Tellian, N.: Smartphones as enabler of Supply Chain Event Management. In: Grzybowska, K. (ed.) Management of Global and Regional Supply Chain – Research and Concepts, pp. 133–143. Publishing House of Poznan University of Technology, Upper Austria University of Applied Sciences, Poznan (2011)
6. Davies, A., Ghahramani, Z.: Language-independent Bayesian sentiment mining of Twitter. In: Workshop on Social Network Mining and Analysis (2011)
7. Frischmuth, P., Klímek, J., Auer, S., Tramp, S., Unbehauen, J., Holzweißig, K., Marquardt, C.M.: Linked Data in Enterprise Information Integration (2012)
8. Gruber, T.R.: Toward Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal of Human Computer Studies 43.5, 907–928 (1995)
9. Hepp, M.: An Ontology for Describing Products and Services Offers on the Web. In: Gangemi, A., Euzenat, J. (eds.) EKAW 2008. LNCS (LNAI), vol. 5268, pp. 329–346. Springer, Heidelberg (2008)
10. Brickley, D., Miller, L.: FOAF vocabulary specification 0.98. In: Namespace Document 9 (2010), `http://xmlns.com/foaf/spec/`
11. Carroll, J.J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A., Wilkinson, K.: Jena: implementing the semantic web recommendations. In: Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers & Posters, pp. 74–83 (2004)
12. Pinckaers, F., Gardiner, G.: A Modern Approach to Integrated Business Management (2009)
13. Herzog, T.: A Comparison of Open Source ERP Systems (2006)
14. Mironov, V., Seethappan, N., Blonde, W.: Benchmarking triple stores with biological data. In: Semantic Web Applications and Tools for Life Sciences (SWAT4LS), Berlin, Germany (2010)
15. Fürber, C., Hepp, M.: Using SPARQL and SPIN for Data Quality Management on the Semantic Web. In: Abramowicz, W., Tolksdorf, R. (eds.) BIS 2010. LNBIP, vol. 47, pp. 35–46. Springer, Heidelberg (2010)