

Semantic Audit Application for Analyzing Business Processes

Ildikó Szabó^(✉) and Katalin Ternai

Department of Information Systems, Corvinus University of Budapest,
Fővám tér 13-15, Budapest 1093, Hungary
iszabo@informatika.uni-corvinus.hu,
katalin.ternai@uni-corvinus.hu

Abstract. Standard regulations are used to assess the compliance of business operations by auditors. This procedure is too time-consuming and Computer Assisted Audit Tools lack the feature of processing documents semantically in an automatic manner. This paper presents a semantic application which is capable of extracting business process models in the shape of process ontologies from business regulations based on reference process ontologies transformed from process models derived from standard regulations. The application uses ontology matching to discover deviations of a given business operation and creates a transparent report for auditors. This semantic tool has been tested on one of the Internationalization processes in the respect of Erasmus mobility.

Keywords: Audit · Business process management · Text mining · Process ontology · Ontology matching

1 Introduction

Auditing information systems and business processes provide a valuable feedback about business management. Constraints derived from regulations, guidelines, standards provide compliance requirements. In the light of them, auditors have to investigate information reflecting business operations. This information can be extracted from transaction data stored in operational databases, data warehouses or resides in internal regulations, handbooks, event logs, waiting for their interpretation.

Computer Assisted Audit Tools and Techniques (CAATs) support auditors by reports resulted from data analysis. Their main functions are to investigate internal logic of an application directly with testing transactional data produced by an IT application after feeding it with real or dummy data or by executing parallel simulation. Moreover there are indirect methods like Generalized Audit Systems and embedded audit modules for scrutinizing the compliance of an application. The latter focuses on monitoring transactions within the application. GAS are used to extract and analyze data. Two GAS systems – ACL and IDEA – were widely used among participants of the study conducted by Braun and Davis [1].

The lack of CAATs that they do not investigate the compliance of business processes directly and use only fact data provided by an information system. Evaluation based on processing business regulations is missing from them.

But regulations hide processes per se which can be extracted from them with creating process models or executing text mining. Process models extracted from business and standard regulations provide a basis to discover deviations between actual business operation and requirements articulated in regulations. Discovering deviations requires structural investigation of these processes within a context-dependent environment. Ontology-based approach using text mining is one way to fulfil this requirement. Because the aim of creating ontologies is to specify a conceptualization of a given domain and text mining can help to build ontologies in a semi-automatic manner, hence ontologies can reflect contexts and their concepts unambiguously. Process ontologies preserve business process elements from the models in a unified way. Ontology matching can enhance structural and semantic investigation of process ontologies.

These techniques – creating process ontologies with text mining, running ontology matching and interpreting its result – can underlie a tool which is capable of processing business regulations, standard regulations and revealing discrepancies and similarities between them, focusing on the meaning of these documents. This feature is missing from the computer assisted audit tools, hence this application can be regarded as a solution to fill this gap.

This paper presents an application that melts these techniques down and provides a transparent report from discovered knowledge for auditors. Several countries have adopted quality audits in their higher education system. The process of quality audit in the education system focuses primarily on procedural issues rather than on the results or the efficiency of a quality system implementation. Internationalization is part of the quality culture of a higher education institution, as well. The institutions are getting increasingly motivated to participate in an internationalization audit, that's why this domain was selected as a use case of this application.

The aim of this paper is to show a semantic application using text mining which is to assist institutions with valuable feedback about their international activities and to improve the chance of their international accreditation. Section 2 presents the above-mentioned semantic techniques (process ontology, text mining on this field, ontology matching). The process along this application works is showed in Sect. 3. The results of this implemented application are explained by the one of the Internationalization processes in the respect of Erasmus mobility in Sect. 4.

2 Semantic Theoretical Background

All organizations represent their processes based on general and specific characteristics which are conditioned to locations, ranges or antecedents. The boundary conditions and restrictions turn up environmental makings maybe regulatory elements or best practices from the levels of business process maturity. Business process management (BPM) provides support for managing the processes of organizations and facilitating their adaptation to dynamic changing environment. BPM encompasses methods, techniques, and tools to design, enact, control, and analyze operational business processes involving humans, organizations, applications, documents, and other sources of information [2]. Modern BPM suites are evolving to automate the modeling, monitoring and

redesign of complex processes, although there are still many open issues to be addressed. Conceptual model captures the semantics of a process through the use of a formal notation, but the descriptions resulting from conceptual model are intended to be used by humans. The semantics contained in these models are in a large extent implicit and cannot be processed. With the web-based semantic schema such as Web Ontology Language (OWL), the creation and the use of the conceptual models can be improved, furthermore the implicit semantics being contained in the models can be partly articulated and used for processing [3] Ontologically represented process models allow querying on a relatively high level of abstraction.

The usage of Semantic Web technologies like reasoners, ontologies, and mediators promises business process management a completely new level of possibilities. This approach is known as semantic business process management (SBPM). [4] When a new regulation is established the business process has to comply with this regulation. In SBPM the business processes as well as the new regulation are defined in a way that a machine is able to understand, therefore no manual work is needed to verify that the business processes comply with the new regulation.

2.1 Process Ontologies

Process ontologies are created in order to describe the structure of a process, whereas organization related ontologies provide a description of artefacts or actors that are utilized or involved in the process. The domain ontologies provide additional information specific to an organization from a given domain.

Process ontologies have no precise definition in the academic literature. Some refer to process ontology as a conceptual description framework of processes [5]. In this interpretation process ontologies are abstract and general. Contrary, task ontologies determine a smaller subset of the process space, the sequence of activities in a given process. In our approach the concept of process ontologies is used, where ontology holds the structural information of processes with multi-dimensional meta-information partly to ground the channeling of knowledge embedded in domain ontologies. We present an approach for representing business processes semantically, by translating them into process ontology that captures the implicit and explicit semantics of the process model. We have also implemented a translation tool to convert business process model to its OWL representation, serving as a basis for further analysis.

We elaborated a method for extracting a business process in the shape of a process ontology with using semantic text mining from documents. Two process ontologies are served as a basis for detecting deviations in business processes. Ontology learning and matching techniques were integrated into our application.

2.2 Ontology Learning

The objective of ontology learning is to generate ontologies with using natural language processing and machine learning techniques. Text mining techniques like similarity measures, pattern recognitions etc. are used to extract terms and their relationships to build ontology [6].

Methontology is one of the most known methodologies for ontology construction, supplying a set of reference tasks necessary to build an ontology. It is a general, domain independent methodology, which defines the main activities of the ontology construction process and specifies the steps for performing them. [13] Some approaches construct domain ontologies that reflect the domain covered by the input texts, and not top level, highly abstract ontologies or lexicalized ontologies (WordNet). A couple of realizations for example: Text2Onto – Text2Onto [14] OntoLearn – TermExtractor, WCL System [10], and SPRAT – SPRAT [11].

The relevant techniques for automating, the conceptualization are:

1. Build glossary of terms (term extraction),
2. Build concept taxonomies,
3. Identify ad-hoc relations.

Other techniques cover also other tasks for example: Describe rules [12].

Term extraction is usually supported by linguistic and statistical of technique used jointly. For building concept taxonomies structural and contextual of techniques are often used. For identify ad-hoc relations, pattern based techniques are often used. Relevant tools combine several techniques resulting a hybrid method.

2.3 Ontology Matching

Alasoud et al. [7] define ontology matching problem identifying semantic correspondences between the components of the entities of ontologies.

Element-level matching techniques focus on matching entities and its instances without any information about their relationships with other entities and instances. Structure-level matching techniques address to scrutinize not only matching entities but their relations with other entities and instances as well [8].

To build our semantic audit application requires an ontology matching tool that fulfill the following criteria:

- It must be customized to adapt the changes of new or improved process models into the audit report.
- It must be integrated with other components to ensure that ontology building and matching procedures can work together.
- Technical report provided by this tool must be structured texts in order to process them automatically.
- It must handle different languages of process ontologies (RDF, XML, OWL).

Ontology Alignment Evolution Initiative contest inspires researchers to develop new ontology matching tools. LOGMap, Yam++ and Protégé 4 OWL Diff were investigated based on these criteria in [15]. Though Protégé 4 OWL Diff is capable of just running structure-level investigation but it provides libraries to handle different ontology languages, provide open source codes and well-structured technical reports. This tool was used to develop this semantic audit application.

3 Semantic Audit Application

Auditors have to collect evidence that operations of companies comply with requirements articulated in guidelines, standards, policies etc. These evidences are settled in document or data provided by information systems. As we have seen in Sect. 1, CAATTs tools are usually not capable of processing documents semantically. The main functional requirements of this semantic audit application are the following ones. It must be capable of:

- Processing organizational documents in an automated manner
- Focusing on semantic contents of these documents
- Interpreting these contents in the respect of the requirements extracted from reference documents
- Comparing semantic contents of organizational and reference documents
- Presenting the result of this comparison into an interpretable and transparent report.

The process of this semantic audit application is presented in Fig. 1.

The **first phase** is to create an Adonis process model from the standard regulation (1) and transform it into the Reference Process Ontology (RPO) (3) with using XSLT transformation (2) [16].

ADONIS is a graph-structured Business Process Management language. Its main feature is the method independence. Our approach is principally transferable to other semi-formal modeling languages. The semantic annotation for specifying the semantics of the tasks and decisions in the process flow explicitly is important in our method.

For conceptualization, several parameters have to be set or defined when modeling a business process. Vertically, we can specify operational areas only, or process areas, process models, sub processes, activities, or even deeper; the algorithms. Horizontally, extra information can be modeled within the business process: organizational information can be specified in an organogram; the roles can be referred in the RACI (Responsible, Accountable, Consulted, Informed) matrix of the process model, the input and the output documents in the document model and the applied IT system elements can be added to the IT system model as well.

In the second step (2), the mapping of the conceptual process models to process ontology concepts will be shown. The transformation procedure follows a meta-modeling approach. The links between model elements and ontology concepts have been established. The process ontology describes both semantics of the modeling language constructs as well as semantics of model instances.

In order to map the conceptual models to ontology concepts, the process models are exported in the structure of ADONIS XML format. The converter maps the Adonis Business Process Modeling elements to the appropriate Ontology elements in meta-level. The model transformation aims at preserving the semantics of the business model. The general rule followed by us is to express each ADONIS model element as a class in the ontology and its corresponding attributes as attributes of the class. This conversion is performed with an XSLT script and results the Reference Process Ontology (RPO).

```

...
<Declaration>
  <Class IRI="#Actor"/>
</Declaration>
<Declaration>
  <Class IRI="#Process_step"/>
</Declaration>
<Declaration>
  <Class>
    <xsl:attribute name="IRI">#<xsl:value-of select="$baseClassName"
/></xsl:attribute>
  </Class>
</Declaration>
<Declaration>
  <ObjectProperty IRI="#belongs_to"/>
</Declaration>
<Declaration>
  <ObjectProperty IRI="#followed_by"/>
</Declaration>
...

```

To represent the business model in the ontology, the representation of ADONIS model language constructs and the representation of ADONIS model elements have to differentiate. ADONIS model language constructs are created as classes and properties and the ADONIS model elements can be represented through the instantiation of these classes and properties in the ontology. The linkage of the ontology and the ADONIS model element instances are accomplished by the usage of properties.

The **second phase** is to build the Organizational process ontology (OPO) with the help of the RPO from organizational documents within the process ontology building

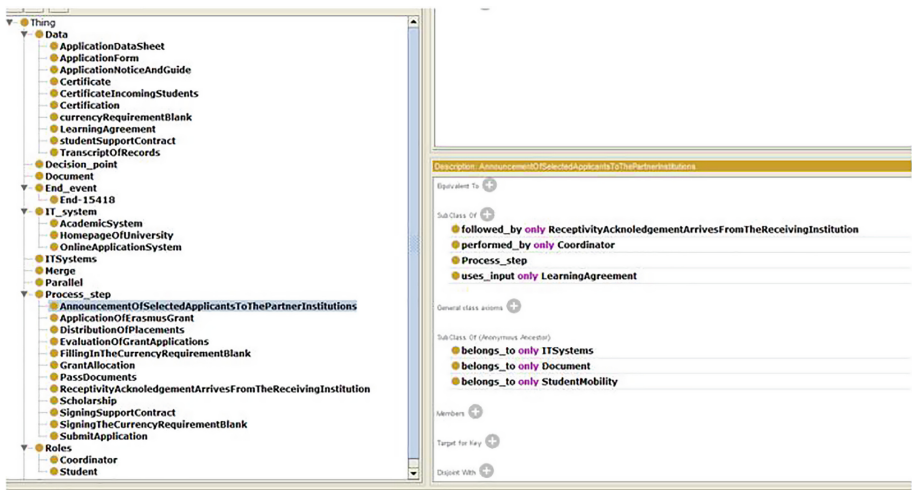


Fig. 1. Reference process ontology

component (4). The first step of its algorithm is to identify process element of the RPO (like *Student*, *Coordinator* as a specific *Role*) - excluding process steps – or discover new ones within a given document and add them to the OPO as subclasses of the appropriate super classes like *Role*, *Document* etc. New process elements are discovered with the help of semantic text mining. The algorithm focuses on finding patterns shaped into open queries. Relations are regarded as ordered pairs. The algorithm assumes that certain expressions can represent a given relation within the document e.g. *produces_output* (*Process_step*, *Document*) relation suggests that something must be happen with a document e.g. it is submitted or signed. That’s why the algorithm wants to find *x submit y* pattern within the document, where *y* is a document. It seeks “submit” term and collects few words after that. It adds this expression as subclass of the *Document* class to the OPO.

The second step in the building of the OPO is to identify the process steps of RPO within the document and connect them to the nearest process elements already existed in the OPO. The algorithm seeks every terms of a given process step within each sentences of the document and counts the hits. If the number of the hits is greater than a given threshold, the identified process step will be added to the OPO as subclass of the *Process step* class and it will be connected to other process elements identified nearby (namely in a given radius of words) within the text.

Having created the OPO (5), its full or tailored version as a result of a DL Query will be compared to the appropriate version of the RPO within the ontology matching component. Its technical report is processed by a report generator to create a transparent report for auditors which contains information about the number of tasks, filtered roles, missing, unnecessary or common organizational process elements. Hence auditors can discover areas requiring deeper investigations in the next phase when leaders are interrogated by them (Fig. 2).

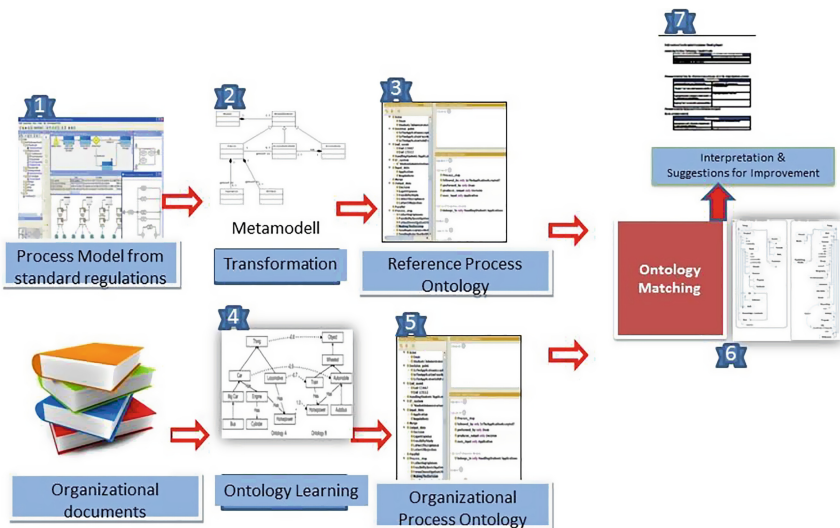


Fig. 2. Processes of the semantic audit application

This application is a Java application that uses the libraries of OWL API, DLQueryExample and the SVN Repository of Protégé 4 OWL Diff. The following section presents how this system works on the field of Internationalization of higher education institutions.

4 Its Implementation on the Erasmus Mobility Field

An example run of this application will be showed in this section. It uses Student Application process as standard process from the Erasmus Mobility Handbook. Erasmus mobility calls represent the organizational documents. The following questions were investigated during an audit procedure conducted in the life of some Hungarian higher education institutions.

How effective are the current mechanisms? What kind of communication channels exist between the various levels managing internationalization activities? How effective are they? What are the missing functions of the internationalization units? Which units are less efficient? Why? [9].

In the modeling phase reference process models have been formalized from Erasmus+ Programme Guide¹ and process models were implemented by using BOC ADONIS modeling platform². The reference business process model detailed with the above mentioned parameters can be seen in Fig. 3.

The Reference Process Ontology transformed from this model and an Erasmus call of a Hungarian higher education institution were used to present the applicability of this solution in the respect of the next audit questions. These questions aim at investigating the effectiveness of the current mechanisms.

Audit Question 1: Are the same role responsible for performing this process?

The answer requires filtering the ontology by Roles. The report created from the technical report provided by the Protégé 4 OWL Diff ontology matching component presents that the University was mentioned as role in the organizational document instead of the Coordinator. In the Erasmus mobility call, it was mentioned that “Qualified applicants will be invited to take the entrance examination organized by the university”. It reveals such problem that who is the responsible person for organizing this entrance examination (Fig. 4).

But we can state that Student was mentioned on both sides, hence we can try to investigate the next audit question.

Audit Question 2: In what measure are the tasks performed by the same role overlapped?

¹ http://ec.europa.eu/programmes/erasmus-plus/sites/erasmusplus/files/files/resources/erasmus-plus-programme-guide_en.pdf.

² <http://www.boc-group.com/products/adonis/>.

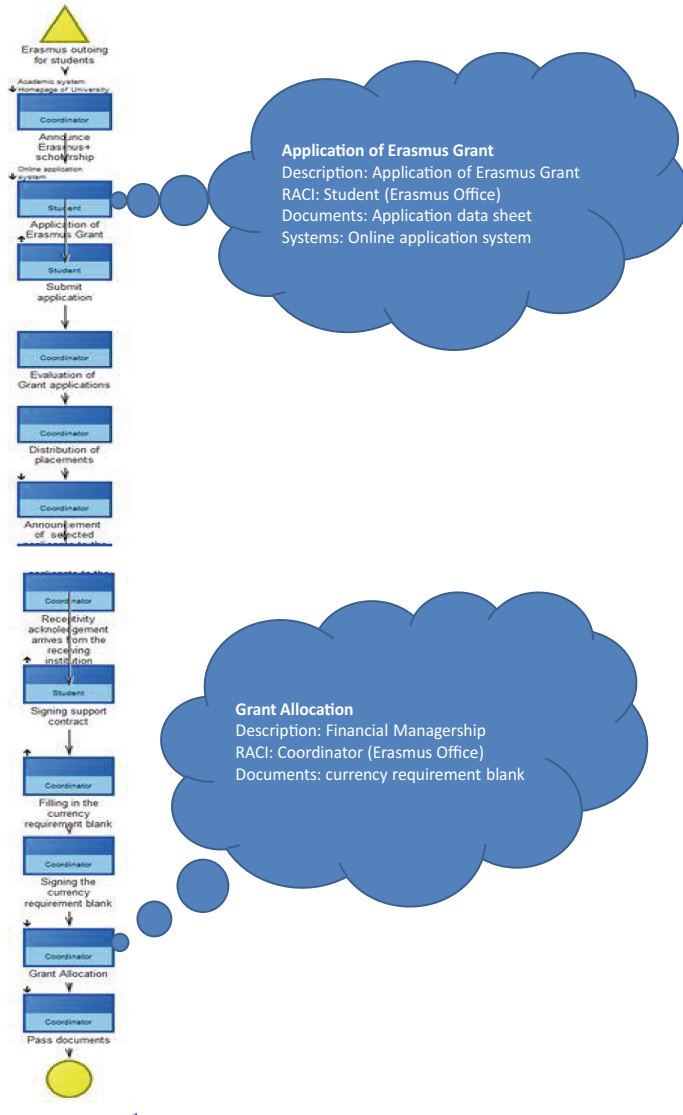


Fig. 3. The reference process model

To answer it, the ontologies were filtered by the ‘performed_by only (Student)’ DL Query. The result is presented on Fig. 5.

This report shows that to sign of support contract is not a task of a student or it is not mentioned in the organizational document. We found that the latter event happened.

Draft version of the Semantic Compliance Checking Report

created by Corvinno Technology Transfer Center

Actual business process	Reference business process
Actual business process	Reference business process
Number of elements: 2	Number of elements: 2
Role:	Role:
Report generated at 27-4-2016	

Process elements from the reference document are not in the actual business process

Process elements	
Coordinator	

Process elements that are not in the reference document

Process elements	
University	

Same process elements

Process elements	
Student	

Fig. 4. Report of the role investigation

Please type a class expression in Manchester Synt

`performed_by only (Student)`

QUERY: `performed_by only (Student)`

SuperClasses.....

`performed_by only (Student)`

QUERY: `performed_by only (Student)`

SuperClasses.....

Thing

Thing

EquivalentClasses.....

[NONE]

SubClasses.....

SubmitApplication
ApplicationOfErasmus

Instances.....

[NONE]

Process elements from the reference document are not in the actual business process

Process elements	
SigningSupportContract	

Process elements that are not in the reference document

Same process elements

Process elements	
ApplicationOfErasmusGrant	SubmitApplication

Fig. 5. Report about not mentioned tasks

These reports revealed that a role and a task were missing in this Erasmus mobility call, so this Student Application process do not comply with the requirement of the Erasmus mobility handbook. The process is not effective, because students do not know about their responsibility for signing the contract, so they will be informed in a latter phase of this process which makes this process more slow. The auditor has to investigate that the source of this problem is a document that does not reflect well the process or the process itself.

5 Conclusion and Future Work

Nowadays Campus Mundi projects are to improve higher education processes in Hungary. The audit guideline elaborated for investigating compliance checking of Internationalization activities wants to detect “how the current mechanisms are effective”. Our semantic audit application can help to compare institutional processes with standard processes articulated in the Erasmus Mobility Handbook. The Student Application process was used to test this application. Erasmus mobility calls represent the organizational documents. This test was executed on ten different sources.³ The first chart presents that the algorithm identified at least one roles within each organizational documents. These roles were mostly interpretable (like Student, Coordinator, and University etc.) except only in the 6th case. We can state that the “by the” semantic rule being responsible for identifying rules can be applied, because its false discovery rate is low.

The second chart shows that there is a notable differentiation between tasks performed by students within the institutions and according the handbook. It implies that our algorithm identifies these tasks not well, so we have to improve it. Or different institutions obligate students to perform such kind of tasks that are not mentioned in the handbook, maybe these tasks belong to another role. This is the problematic of segregation of duties that must be investigated by the auditors (Fig. 6).

This application can be used to process business regulations semantically instead of manually, that spares time for auditors. It provides a report that shows deviations within managing business, if they exist. Auditors can use this knowledge to seek information with more focus or ask managers relevant questions during the next phase of the audit procedure.

We can use more metrics to test the precision of the algorithm of this application. The organizational process ontology (OPO) stores several text parts used to identify the above-mentioned process elements. These text parts can be used to calculate hit rates like false or true positive/negative rates. Based on these information we can improve this algorithm.

The complexity of process models and the granularity of organizational documents influence the scalability of this system. The ten above-mentioned materials were processed within similar time period, but they were not too large documents. Testing the scalability of this system is a future work.

³ <http://erasmus.yasar.edu.tr/student-mobility/outgoing-students/>,
<http://www.ciim.ac.cy/outgoing-students>,
<https://www.international.tum.de/en/going-abroad/students/erasmus/...etc>.



Fig. 6. Results of this application

Acknowledgements. The authors wish to express their gratitude to Dr. András Gabor, associate professor of the Corvinus University of Budapest, for the great topic and the powerful help provided during the development process.

“This work was conducted using the **Protégé** resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.”

References

1. Braun, R.L., Davis, H.E.: Computer-assisted audit tools and techniques: analysis and perspectives. *Manag. Auditing J.* **18**(9), 725–731 (2003)
2. van der Aalst, W.M.P., ter Hofstede, A.H.M., Kiepuszewski, B., Barros, A.P.: Workflow Patterns. *Distrib. Parallel Databases* **14**(1), 5–51 (2003)
3. Hepp, M., Cardoso, J., Lytras, M.D.: *The Semantic Web: Real-World Applications from Industry*. Springer, New York (2007). ISBN: 0387485309
4. Koschmider, A., Oberweis, A.: Ontology based business process description. In: *Proceedings of the CAiSE*, pp. 321–333 (2005)
5. Hepp, M., Roman, D.: An ontology framework for semantic business process management. In: *Proceedings of Wirtschaftsinformatik, Karlsruhe, 28 February–2 March 2007* (2007)
6. Maedche, A., et al.: Ontology learning part one — on discovering taxonomic relations from the web. In: Zhong, N., et al. (eds.) *Web Intelligence*, pp. 301–319. Springer, Heidelberg (2003)
7. Alasoud, A., Haarslev, V., Shiri, N.: An effective ontology matching technique. In: An, A., Matwin, S., Raś, Zbigniew, W., Ślęzak, D. (eds.) *ISMIS 2008. LNCS (LNAI)*, vol. 4994, pp. 585–590. Springer, Heidelberg (2008). doi:[10.1007/978-3-540-68123-6_63](https://doi.org/10.1007/978-3-540-68123-6_63)
8. Otero-Cerdeira, L., et al.: Ontology matching: a literature review. *Expert Syst. Appl.* **42**(2), 949–971 (2015)
9. Temesi, J.: Nemzetköziesítés a magyar felsőoktatási intézményekben, *Audit materials* (2013)
10. Velardi, P., Cucchiarelli, A., Pétit, M.: A taxonomy learning method and its application to characterize a scientific web community. *IEEE Trans. Knowl. and Data Eng.* **19**(2), 80–191 (2007)
11. Maynard, D., Funk, A., Peters, W.: SPRAT: a tool for automatic semantic pattern-based ontology population. In: *Proceedings of the International Conference for Digital Libraries and the Semantic Web* (2009)

12. Buitelaar, P., Cimiano, P., Magnini, B.: Ontology learning from text: an overview. In: *Ontology Learning from Text: Methods, Applications and Evaluation*, pp. 3–12. IOS Press (2005)
13. Corcho, O., Fernández-López, M., Gómez-Pérez, A., López-Cima, A.: Building legal ontologies with METHONTOLOGY and WebODE. In: Benjamins, V., Richard, Casanovas, P., Breuker, J., Gangemi, A. (eds.) *LNCS (LNAI)*, vol. 3369, pp. 142–157. Springer, Heidelberg (2005). doi:[10.1007/978-3-540-32253-5_9](https://doi.org/10.1007/978-3-540-32253-5_9)
14. Cimiano, P., Völker, J.: Text2Onto—a framework for ontology learning and data-driven change discovery. In: Montoyo, A., Muñoz, R., Métais, E. (eds.) *NLDB 2005. LNCS*, vol. 3513, pp. 227–238. Springer, Heidelberg (2005). doi:[10.1007/11428817_21](https://doi.org/10.1007/11428817_21)
15. Szabó, I.: Future development: towards semantic compliance checking. In: Gábor, A., Kő, A. (eds.) *Corporate Knowledge Discovery and Organizational Learning. KMOL*, vol. 2, pp. 155–173. Springer, Heidelberg (2016). doi:[10.1007/978-3-319-28917-5_7](https://doi.org/10.1007/978-3-319-28917-5_7)
16. Ternai, K.: Semi-automatic methodology for compliance checking on business processes. In: Kő, A., Francesconi, E. (eds.) *EGOVIS 2015. LNCS*, vol. 9265, pp. 243–256. Springer, Heidelberg (2015). doi:[10.1007/978-3-319-22389-6_18](https://doi.org/10.1007/978-3-319-22389-6_18)