THE SOCIAL ONTOLOGY BUILDING AND EVOLUTION (SOBE) PLATFORM

Daniela Angelucci[1], Alessia Barbagallo[2], Tania Di Mascio[1] and Francesco Taglino[1]
[1] *Isitituto di Analisi dei Sistemi ed Informatica "A. Ruberti" (IASI-CNR), Viale Manzoni 30, 00185 Rome, Italy*
[2] *TXT e-solution, Via Frigia 27, I-20126 Milan, Italy*

Keywords: Ontology building, Knowledge extraction, Social participation, Service oriented architecture.

Abstract: In this paper, the Social Ontology Building and Evolution (SOBE) method and the corresponding software platform for cooperative ontology building in the context of a cluster of enterprises is presented. The SOBE is characterized by three main aspects: (i) automatic knowledge extraction from unstructured documents; (ii) social validation, involving a community of domain experts; (iii) a step-wise approach which goes through five main steps which produce incremental results towards the construction of a domain ontology.

1 INTRODUCTION

Knowledge is one of an organization's most valuable assets. For this reason, it is crucial to establish methodologies and technologies to manage the creation, capture, and sharing of knowledge and information at and among all levels of the organization. In particular, developing enterprise ontologies to support structuring and sharing of enterprise information is an emerging practice in enterprise knowledge management.

In this paper, we present the SOBE (Social Ontology Building and Evolution) platform for supporting ontology building in the context of an enterprises cluster. SOBE is strongly rooted on three main aspects: (i) automatic knowledge extraction from unstructured enterprise documents which are pregnant with enterprise knowledge; (ii) social participation of a community of experts which are enabled and requested to validate, discuss about (to find a consensus), and enrich the results of the automatic extraction; (iii) a step-wise approach that goes through intermediate results: lexicon, glossary and taxonomy.

Developing ontologies in general is not a project undertaken by a single person, but rather it is a large project with numerous participants; this approach is known as social participation. A collaborative approach for solving terminological problems is presented in (Campbell et al., 1998). Some years later, a collaborative approach for ontology building was introduced by (Holsapple and Joshi, 2002). ased on the Delphi method (Lindstone and Turoff, 1975), this approach proposes collaborative development of ontologies in Agentcities (Ceccaroni and Ribiere, 2002) was carried out through both face-to-face meetings and remote communication among several partners of the EU Agentcities RTD project. Another contribution on consensus building techniques applied to ontology engineering is presented by (Karapiperis and Apostolou, 2006). The proposed methodology includes the definition of ontology design criteria, the development of an initial ontology, the iterative process of ontology evaluation and evolution, and the ontology application. Moreover, (Tempich et al., 2007) proposes the use of argumentation theory, and (Walton, 2009) proposes the support of ontology engineering. With respect to (Holsapple and Joshi, 2002), (Karapiperis and Apostolou, 2006), SOBE deals with the complete ontology evolution process, the detailed collaborative process (i.e., debating, and voting) and the needed steps and relative milestones. Furthermore, in SOBE, we introduce automatic tools trying to integrate at best human and software-based activities. With respect to (Campbell et al., 2010) and (Holsapple and Joshi, 2002), only dealing with the terminological level, we focus on a substantial extension and enrichment of an existing ontology rather than on its simple revision. Finally, in (Ceccaroni and Ribiere, 2002), the type of

Angelucci D., Barbagallo A., Di Mascio T. and Taglino F..
THE SOCIAL ONTOLOGY BUILDING AND EVOLUTION (SOBE) PLATFORM.
DOI: 10.5220/0003101204160419
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2010), pages 416-419
ISBN: 978-989-8425-29-4

communication is inadequate with respect to the SOBE community dimension.

The objective of this paper is to present the whole work, by also adding with respect to (Barbagallo et al., 2010) an overview of the platform architecture and a case study in the context of the COIN European project. The rest of the paper is organized as follows. In Section 2, we describe the social ontology evolution process on which SOBE is based. In Section 3, we present the platform architecture and the modules that perform the various steps of the process. Section 4 is dedicated to the presentation of a case of study in which we carried out an experimentation of the SOBE platform in the ICT domain. Finally, in Section 5, we present conclusions and future works.

2 THE SOBE PROCESS

The *social ontology building and evolution* (SOBE) process (see Figure 1) exploits the UPON methodology (De Nicola et al., 2009), and the ontology learning methodology defined in (Velardi et al., 2007), enriching them with social participation aspects. In particular, UPON is characterized by an incremental nature, reflected by the outcomes produced in the different phases of the process: first the relevant terms in the domain are identified and gathered in a *lexicon*; then the latter is progressively enriched with definitions, yielding a *glossary*; adding to it the specialisation relationships allows a *taxonomy* to be produced, until further enrichments and a final formalization produces the sought *domain ontology*. The SOBE process exploits this step-wise approach and enriches it through an automatic support for knowledge extraction from existing digital resources, and social participation aspects for consensus reaching among the community of experts that participates to the ontology building. The automatic knowledge extraction support aims at reducing the workload of the people involved in the ontology building, and at reusing the amount of knowledge contained in any type of existing documental resources (e.g., technical papers and reports, standards specifications, etc.), and structured resources (e.g., dictionaries, thesauri, ontologies). In accordance with the UPON methodology, SOBE firstly addresses the terminological aspects, i.e. the lexicon. The start up consists in processing a corpus of documents, related to the addressed domain, for automatically extracting terms that are considered relevant in that domain. This extraction phase is based on natural language processing techniques, statistical analysis, and contrastive analysis against a pre-defined corpora of documents related to different domains. The extracted terms are referred in Figure 1 as *E-Lexicon*. In the case of enrichment of an existing ontology, the *E-Lexicon* is filtered out of the terms that are already in the lexicon of the current ontology (*O-Lexicon*). Then, the *E-Lexicon* is validated by the community of experts to reach an agreement on the new terms to be included in the ontology (*N-Lexicon*). After the identification of the *N-Lexicon*, terms have to be enriched with natural language definitions in order to build the desired glossary. As in the philosophy of SOBE, definitions are firstly extracted from existing dictionaries or ontologies (e.g., Google Define, WordNet), yielding the *E-Glossary* which has to be humanly validated. Glossary validation is performed by voting extracted definitions. Any potential conflict due to lack of agreement or terms with no definitions are managed by opening discussion forums about glossary entries. The result of the glossary validation step is gathered in the *N-Glossary*. The following step is the categorization of the *N-Glossary* entries by associating to each of them a *kind* (i.e., Object, Process, Actor) in accordance with the OPAL framework (D'Antonio et al., 2007). Terms with definitions and associated kinds represent the new concepts to be inserted in the ontology. Starting from the newly acquired concepts definitions, natural language processing techniques allow an automatic proposal of their hypernyms, producing a set of micro-taxonomies: *Eμ-Taxonomies*. An *Eμ-Taxonomy* is a specialization hierarchy between concepts. In the case of ontology enrichment the *Eμ-Taxonomies* are merged with the taxonomy of the existing ontology (i.e., *O-Taxonomy*) producing the *N-Taxonomy*. In the last step, the taxonomy is enriched with other relationships (e.g., part of, attributes) producing the final *N-Ontology*. The human validation phase involves three types of actors, who play specific roles in validating the results of the automatic extraction tools: the Ontology Master (*OM*), Participants (*P*s) and Moderators (*Ms*). The *OM* is responsible for the whole ontology enrichment process and is in charge of managing and supervising all its different phases. People involved as *P*s play an active role in the validation of the extraction tools results. They are in charge of validating, adding, modifying terms and definitions, and proposing new ones. This represents the SOBE social participation aspect which is realized through three different mechanisms: voting, discussing and proposing. Voting enables the

classical validation represented by accepting or discarding terms and definitions proposed by the automatic knowledge extraction services. Discussing enables a more active participation of *P*s with the aim of finding a shared agreement on content both in the lexicon and glossary building steps. The system provides several mechanisms for managing users' validations and processing their results. As an example, after the glossary validation, two scenarios are possible: (1) all the terms have at least one definition and all of them have been accepted with a reasonable consensus (more than 60% of *P*s): in this case the *OM* is suggested to select accepted definitions; (2) there are some terms with no definitions or with definitions with low acceptance rate: in this case the *OM* is invited to open a discussion forum for each of these terms. For each forum a *M* is designated by the *OM* from a list of *Ps*. S/he is responsible for managing comments and conversations and for proposing a final definition for each term at the end of the discussion phase. The system supports *Ms* assignment displaying each *P*'s reliability, obtained by calculating the precision of his/her own validation with respect to the accepted list of terms.

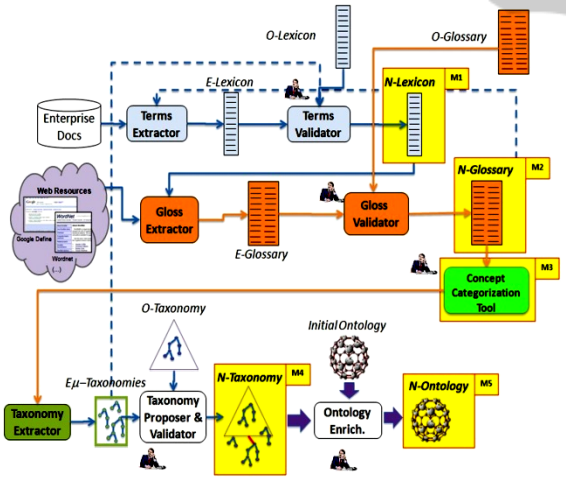

Figure 1: Overview of the SOBE process.

3 THE SOBE ARCHITECTURE

The SOBE platform is designed as a web application with a classical three layer architecture where: (i) the *Data Layer* contains the different knowledge bases, that are in accordance with the intermediate results of the SOBE process; (ii) the *Logic Layer* contains the functional modules; (iii) the *Presentation Layer* manages the graphical user interface and incoming requests via web services from external client applications. In particular, the Logic Layer presents two main components: the *Extraction Subsystem* which provides the automatic support in terms of *Lexicon*, *Glossary* and the *μ-Taxonomies* extraction; the *Validation Subsystem* which supports the social participation of the SOBE users for the *Lexicon* and *Glossary* validation. Furthermore, the Extraction Subsystem is organized in accordance with an open service oriented paradigm: it is service oriented, because the core of the knowledge extraction tasks is provided by external web services (for instance, in the current implementation of the SOBE platform, we integrated the TermExtractor and GlossExtractor (Velardi et al., 2008) web services); it is open in the sense that it intends to be flexible and ready to allow the linkage of additional extraction services in order to: i) integrate results from different knowledge extraction services to enrich the automatic support of the SOBE platform; ii) to exploit different extraction criteria, capabilities and performances of different services for using the most suitable ones depending on the dimension, addressed domain and further features of the analyzed corpus of documents.

4 CASE STUDY: ONTOLOGY BUILDING IN ICT DOMAIN

In this section we introduce a preliminary case study of SOBE in building an ontology about the competencies and skills of a cluster of enterprises in the ICT domain. The involved enterprises refer to the IVSZ (Hungarian Association of Information Technology Companies, http://english.ivsz.hu) cluster. The IVSZ cluster gathers almost 300 enterprises, however, in order to test the SOBE platform we involved 8 enterprises, which represent a small, but real, subset of the cluster itself. The involved team was composed by 4 people: 1 *OM* and 3 *P*s. The overall ontology building process took place approximately in 10 days. After the terms extraction step, an *E-Lexicon* of 102 terms resulted. The *OM* opened a 3-days social validation phase and after global terms validation, the *N-Lexicon* was composed by 21 terms of which: 10 had an acceptance rate of 100%, 8 had an acceptance rate of 66% and 3 had an acceptance rate of 33%. In glossary extraction phase, the average number of automatically extracted definitions per term was 4, while the maximum was 8. The deadline for social validation over the extracted definitions was of 4 days. After the first stage of glossary validation, for 14 terms a definition was identified due to the high

acceptance rate of the *P*s. For the remaining 7 terms a forum was opened. This was due to the fact that in 6 cases no definition had reached a sufficient acceptance rate while for 1 term no definition was found through glossary extraction and no one was proposed during the validation phase. After having entitled the *Ms*, the *OM* set a deadline of 3 days for allowing discussions in forums. After the categorization of the couples term-definition, we obtained 3 concepts identified as *Actor*, 9 as *Process*, and 9 as *Object*. Referring to the taxonomy phase, in 18 cases the system was able to automatically extract at least one hypernym. Since the number of automatically extracted hypernyms accepted by the *OM* was 15, in 6 cases hypernyms had to be manually defined. At the end of the use case we interviewed the system users to receive some feedback. They found the system very easy to use thanks to its user friendly interface and its use articulated in incremental steps. The social aspects of the platform granted a distribution of responsibilities among all the system users and a consequent drop of the stress level in performing the tasks. This led to a more participative and propositive attitude of the users in building the ontology.

5 CONCLUSIONS

In this paper, the Social Ontology Building and Evolution process and platform have been presented.

The SOBE platform has been experimented in the construction of an ontology on ICT competencies for a cluster of enterprises. The results have been reported step by step. Our aim for future works, is to evaluate the ontology generated through the SOBE process by using it for creating semantic profiles for the cluster enterprises, i.e. tagging each enterprise document(s) with concepts from the ontology, and asking people belonging to the reference domain to judge the ontology accuracy in representing the real enterprise competencies. Moreover, we intend to reinforce the μ-Tax extraction module for the identification of micro taxonomies and suggestions on how to attach such micro taxonomies to the specialization hierarchy of the existing ontology in the case of ontology evolution. Furthermore, we intend to address the automatic identification of synonyms among extracted terms, with the aim of having, for each concept in the ontology, one preferred term and a list of terminological expressions referring to the same concept.

ACKNOWLEDGEMENTS

This work has been partly funded by the European Commission through ICT Project COIN: Collaboration and Interoperability for networked enterprises (No. ICT-2008-216256). The authors wish to acknowledge the Commission for its support.

REFERENCES

Barbagallo A., De Nicola A., Missikoff M., eGovernment Ontologies: Social Participation in Building and Evolution. In the proceedings of the 43rd *Hawaii International Conference on System Sciences (HICSS2010)*. Koloa, Kauai, Hawaii. January 05-08 2010. ISBN: 978-0-7695-3869-3

Campbell K., Oliver D., and Shortliffe D., The Unified Medical Language System: toward a collaborative approach for solving terminologic problems. www. ncbi.nlm.nih.gov/pubmed/9452982. 1998

Ceccaroni L. and Ribiere M.; Experiences in Modeling Agentcities Utility-Ontologies with a Collaborative Approach. In *Proceedings of the workshop AAMAS 2002 – W10: Ontologies in Agent Systems (OAS2002)*, Bologna, Italy, 2002.

D'Antonio F., Missikoff M., Taglino F., Formalizing the OPAL eBusiness ontology design patterns with OWL, In *Proc. of I-ESA 2007 conference*.

De Nicola, A., Navigli, R., Missikoff, M., A software engineering approach to ontology building, Information Systems, 34 (2), 258-275, 2009.

Holsapple, C. W., Joshi, K. D. A collaborative approach to ontology design. In Communications of the ACM, 45 (2), 42-47, 2002.

Karapiperis S. and Apostolou D. Consensus building in collaborative ontology engineering processes*, Journal of Universal Knowledge Management 1 (3)*, 2006.

Lindstone, H. and Turoff M., The Delphi Method: Technology and Applications. *Addison-Wesley, Reading*, MA, 1975.

Noy F. N., Chugh A., Liu W., Musen M. A., A Framework for Ontology Evolution in Collaborative Environments. LNCS 4273, pp. 544-558, 2006.

Velardi P., Cucchiarelli A., Petit M.; A taxonomy learning method and its application to characterize a scientific web community *IEEE Transactions on Knowledge and Data Engineering*, 2007

Velardi P., Navigli R., D'Amadio P.; Mining the Web to Create Specialized Glossaries. *IEEE Intelligent Systems*, 2008

Walton D. Argumentation Theory: A Very Short Introduction. In "Argumentation in Artificial Intelligence", *SPRINGER*, 2009.