Supporting Formal and Informal Learning through Domain Ontologies

Paola Monachesi, Thomas Markus, Eline Westerhout, Petya Osenova, and Kiril Simov

***Abstract*—One of the objectives of the Language Technology for LifeLong Learning project is to develop services that facilitate learners and tutors in accessing formal and informal knowledge sources. We show that ontologies enriched with social tags play a crucial role in achieving this goal. Ontologies are central to the knowledge discovery process and they facilitate reuse of course material through semantic search of annotated learning objects, supporting thus formal and informal learning.**

***Index Terms*—Elearning, knowledge discovery, ontologies, social media, semantic annotation.**

I. Introduction

There is an increasing demand on individuals and organizations for knowledge and skills while the availability of information and content is growing exponentially. It is crucial to identify new ways to acquire, contribute and exploit knowledge and thereby facilitate learning.

The objective of the Language Technologies for LifeLong Learning (LTfLL) [1] project is to create next-generation *support services* to enhance competence building and knowledge creation in educational and organizational settings.

The theoretical foundations of the project are based on the idea that knowledge is a socially mediated product. Learners, therefore, build knowledge collaboratively and then internalize it in a personal knowledge building process [1], [2]. As a result, learners become skilled members of a Community of Practice, mastering the learning domain speech genre.

More specifically, one of the goals of the project is to support formal and informal learning. In formal learning, learners follow a pre-defined curriculum. For each topic in the curriculum, there is a set of learning materials, provided by a tutor. In contrast to this, the curriculum in the informal learning process is not obligatory, and the role of the tutor is either absent or not obvious. The learners thus exploit different types of content from different sources to achieve their learning goals. During this process, learners often need guidance and they might benefit from using a more structured way to access content.

Domain ontologies allow for a highly structured way of interlinking different knowledge sources, providing thus the necessary means to support formal and informal learning. They can guide and support the learner in the learning process since they provide a formalization of the knowledge of a domain approved by an expert. Ontologies, however, might be incomplete or might not be compatible with the representation of the domain knowledge of a learner. The vocabulary of the learner (especially beginners) might be different from that of domain experts and could be more sensitive to evolving terminology or less specialized terms.

We have thus developed an ontology enrichment methodology that complements a domain ontology on the basis of relevant tags that emerge from social media. This approach integrates the vocabulary from the communities the learner is part of with the expert view on the domain. New, socially relevant, concepts are thus included into an existing domain ontology.

In this paper, we show that ontologies provide added value for eLearning. They play a crucial role in the knowledge discovery process by helping the learner in identifying relations among concepts and in finding their way among available resources. In addition, ontologies can be used for semantic annotation of the learning material highlighting in this way the most relevant domain concepts in a text. Semantically annotated documents can be more easily retrieved facilitating reuse and supporting thus tutors in the development of course material.

The paper is organized as follows. In section II, we discuss the lexicalized domain ontology that we employ in the project. Section III presents the ontology enrichment pipeline that is used to automatically enrich the domain ontology with socially relevant concepts. Section IV focuses on the role that the enhanced ontology plays in supporting knowledge discovery and a preliminary evaluation by learners in the context of a learning task. Section V covers semantic annotation of learning material, which is performed using the ontology and sections VI and VII illustrate that semantically annotated documents are a crucial element in supporting tutors in positioning the learner and in developing courses within the curriculum. Finally, section VIII contains some concluding remarks.

Manuscript received February 20, 2012; revised March 05, 2012. This work was supported by European project " Language Technologies for Lifelong Learning", 212578.

P. Monachesi is with the Utrecht University, the Netherlands and University of Malta, Malta (e-mail: p.monachesi@uu.nl).

T. Markus is with the UiL-OTS, Utrecht University, the Netherlands (e-mail: thomas.markus@phil.uu.nl).

E. Westerhout is with the with the UiL-OTS, Utrecht University, the Netherlands (e-mail: e.n.westerhout@uu.nl).

P. Osenova is with Sofia University and IICT-BAS, Bulgaria (e-mail: petya@bultreebank.org)

K. Simov is with IICT-BAS, Bulgaria (e-mail: kivs@bultreebank.org).

[1] http://www.ltfll-project.org/

II. Lexicalized Domain Ontologies

A domain ontology represents a formalization of the most relevant concepts of a domain. Therefore, it can be of high value in the learning process since it can guide the learner in acquiring new knowledge. The ontology provides an overview of the concepts the learner should master to be an expert in a given area and the way the various concepts are related to each other. In addition, through semantic annotation (on the basis of a domain ontology) of learning material, it is possible to highlight the most relevant concepts in a text and facilitate thus learning.

In the LTfLL project, we have chosen the domain of *computing* and we have employed an ontology manually developed in a previous project (i.e. Language technology for eLearning – LT4eL),[2] as starting point. It was developed on the basis of the most relevant terms extracted from a corpus of learning objects. Based on these terms, relevant concepts have been created which constitute the backbone of the domain ontology.

The domain ontology has been mapped to an upper ontology (DOLCE – [3]) with the positive consequence that relations already encoded in the upper ontology are inherited in the domain ontology. The mapping to the upper ontology involved OntoWordNet [4] that is a version of WordNet restructured in accordance to DOLCE.

In the literature, various approaches have been proposed to carry out the mapping task between concepts and terms [5]. Most of them consider the multilingual lexicons as a starting point and then try to establish connections to concepts. We have assumed an alternative approach to link the ontology to the lexicons that is very close to the LingInfo mode [6]. The lexicalized ontology contains thus information about the relation between a given concept and its various lexicalizations (i.e. terms or lexicalized phrases) in a given language.

The ontology contains 1002 domain concepts, 169 concepts from OntoWordNet and 105 concepts from DOLCE Ultralite.

III. Enhancing Ontologies with Social Tagging

Ontologies can play an important role within eLearning applications. They can facilitate (multilingual) retrieval and reuse of content for course creation as well as mediate access to various sources of knowledge.

Ontologies, however, might be too static since they only model the knowledge of the domain at a fixed point in time. Ontologies might therefore become outdated or might not correspond to the current representation of the domain knowledge available to the learner.

In the LTfLL project, we envisage a solution to these shortcomings by merging the dynamic knowledge provided by tagging, which is available through social media applications (i.e. Delicious) with the static formal knowledge provided by domain ontologies. Thus, we can include not only the expert view of a given domain, which might be shared by advanced learners, but also the view of beginners who are probably using a less specialized terminology. In addition, we are able to enrich ontologies automatically, which is an important condition for eLearning applications to be scalable and maintainable.

A. Ontology Enrichment Pipeline

In the LTfLL project, we have developed an ontology enrichment pipeline that can automatically enrich a domain ontology using data extracted by a crawler from social media applications (step 1), similarity measures (step 2) , DBpedia knowledge base [7] (step 3) and several heuristics (step 4), as illustrated in Fig.1.

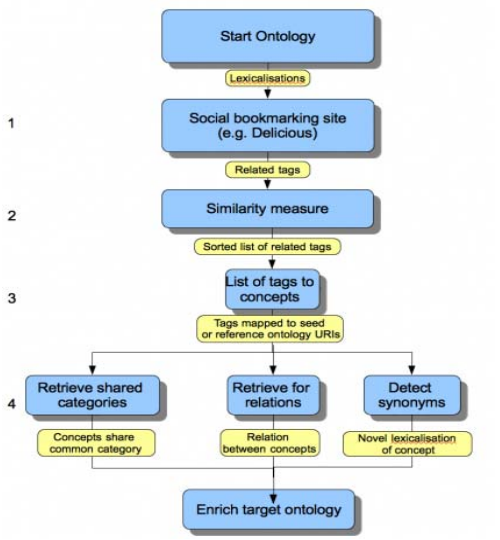

Fig. 1. Ontology enrichment pipeline.

Similarity measures have been used as first step in the ontology enrichment process. Tagging systems provide us with a domain vocabulary that represents the common knowledge of a specific community. The ontology provides the seed lexicalisations that are used by the similarity measure to select possible, socially relevant, lexicalizations of (new) concepts. We have used the resource coocurrence measure with asymmetric normalization in our system for efficiency reasons and wide use in the literature [8].

The next step consists in identifying the appropriate relationships that exist between the terms generated by the similarity measure and the concepts in the domain ontology. To this end, several heuristics are employed, which rely on the use of a large background knowledge base such as DBpedia. DBpedia is a community effort to extract structured information from Wikipedia and to make this information accessible on the Web.

For example, we employ DBpedia to assess whether a related tag can be considered a new concept or a lexicalization of an existing one. The rdf:type assertion between a DBpedia resource and a resource from some other ontology can be used to infer that the DBpedia concept is actually a sub-concept of the object of that statement and should be added as such to the seed ontology. DBpedia also contains a category structure and a list of all the DBpedia concepts and other categories present in such a category hierarchy. We can automatically calculate the closest shared categories for two concepts and return them to estimate its taxonomic position.

[2] http://www.lt4el.eu

A disambiguation component has been integrated in the ontology enrichment process in order to improve its precision. The context of the term is considered in order to associate the appropriate concept with the terms. It is thus possible to enrich the ontology with concepts that have ambiguous lexicalisations such as 'python' and 'java' amongst others.

These term-concept pairs are stored using the MOAT ontology [9]. It becomes possible to differentiate between global meanings as the list of all meanings that could be related to a tag and personal meanings related to a specific user or Community of Practice (CoP).

We have carried out an evaluation of the enrichment pipeline by comparing an ontology produced through a manual enrichment process carried out by an expert and the automatic enrichment process based on social tagging. It shows that the overlap between the two ontologies is minimal. The latter includes the vocabulary of the community of users, while the former includes very specialized tags provided by an expert. It is exactly this complementarity that we want to achieve by embedding concepts extracted from tags into an existing ontology and that we want to exploit in eLearning applications. We refer to [10] for further details.

IV. Knowledge Discovery

The ontology enriched with social tagging presented in the previous section constitutes the basis for a knowledge discovery service that has been developed within the LTfLL project. The domain ontology, which can be browsed by the learner, helps him to identify the main concepts in a domain, the relations among them and relevant resources (texts, videos, slides) that are associated with these concepts. It currently supports informal learning because it provides a more structured way to access content emerging from social media applications such as SlideShare, YouTube, Delicious and Bibsonomy. However, the service can also be employed to guide the learner through a repository of formal learning objects that are relevant for a given curriculum.

Fig. 2 shows the interface of the knowledge discovery service. The ontology is displayed as a graph and depicts the concepts and their relations to each other, allowing thus for an understanding of the meaning of the concepts.

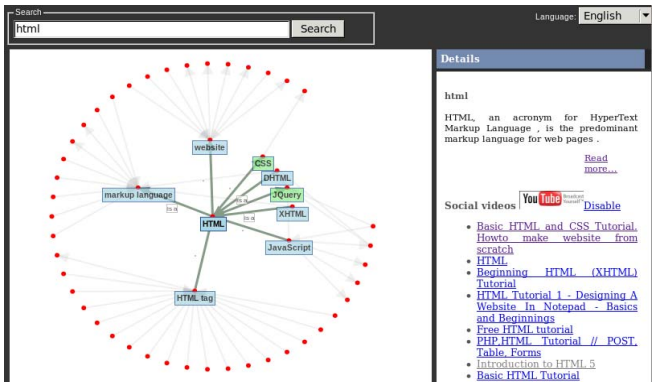

Fig. 2. The knowledge discovery component.

We also provide the definition of a concept which we extract from Wikipedia. We distinguish between concepts coming from an expert-validated domain ontology (blue concepts) and concepts that have been automatically added (green concepts). This transparency in the origin of concepts supports the learner in balancing completeness with correctness and relevancy. Searching is performed either by clicking a concept or making a query: a list of resources best matching those concepts are retrieved.

A. Evaluation

An evaluation of the ontology driven knowledge discovery component has been carried out with 12 beginners (Arts students) and 7 advanced learners (Computer Science students) in the context of a learning task (i.e. creating a website to present results from a research project). The knowledge discovery service was used by the learners to discover which techniques and tools might be useful for creating their website.

The knowledge discovery component was considered very useful for getting acquainted with a domain, since it helps the learner to easily identify how concepts are related to each other and what they mean. The advanced learners appreciated this component more than the novice learners.

The domain ontology is central in the knowledge discovery component while the definitions and learning materials offer the learner support by providing additional information on the selected concept. The ontology fragment assisted the learners in completing the learning task. Both the novice and the advanced learners agreed that concept browsing is useful in carrying out their task, the advanced learners being even significantly more positive than the novice learners. We refer to [11] for further details.

V. Semantic Annotation

In a formal learning context, learners should be able to identify and master the relevant domain concepts that are present in the learning materials, which are part of the curriculum. In the LTfLL project, we have developed a semantic annotation service that supports the learner in this task. The understanding of the conceptual relations attested in a learning object is facilitated when they are annotated semantically. In addition their retrieval is improved through semantic search.

The enriched ontology, presented in section III, constitutes the basis for the automatic annotation of the learning objects. The domain concepts in the ontology are linked to the relevant lexicalizations in the text, making thus the implicit knowledge in the text, explicit. Formal learning resources are annotated with concepts from the domain ontology at either document, paragraph or sentence level.

The semantic annotation is implemented as a language pipe performing tokenization, POS tagging, lemmatization, concept annotation and co-referential annotation. The concept annotation is performed by means of regular expressions that assign the appropriate concept label to a sequence in the text. The grammars are created semi-automatically on the basis of the terms in the lexicalized ontology. Disambiguation rules have been implemented in order to solve the problem of ambiguity of the lexical items by relying on the context of their usage in the texts.

A problem with this annotation methodology is the sparseness of the annotation (about 2 domain concepts per

sentence). This is not sufficient for extracting a conceptual network from a text, and for evaluating the concept distribution. We have thus decided to enhance the annotation pipeline with co-reference relations. The typical cases of co-reference which we encountered in the learning objects are: (1) co-reference between a term and a pronoun in which case the concept assigned to the term is shared with the pronoun; and (2) two terms are connected in which case the more general term receives as an annotation the concept assigned to the more specific term. The co-reference annotation module makes the concept annotation more detailed and allows for an extended coverage of the text (15% improvement).

VI. Positioning the Learner within Curriculum

The learner's positioning with respect to a curriculum is of a great importance in a lifelong learning context. In an informal learning context, the learner needs to achieve a certain level of competency and it is necessary to determine his prior knowledge. Similar needs arise in a formal context in the case of student mobility due to exchange programs. One way to assess the position of the learner within the curriculum is by means of grading of pre-course questionnaires by a tutor. This grading is tedious and time-consuming work.

The semantic annotation service presented in the previous section supports the tutor in the positioning task. It is possible to assess the level of knowledge of the learner by comparing the basic concepts in the curriculum and those present in the answers to the questionnaire.

More specifically, we assume that a curriculum consists of a set of topics related to the content of a course. Each topic is then associated with various learning materials – lectures, tests, descriptions of expected answers, etc. The learner needs to acquire at least two kinds of knowledge in his learning path: the domain knowledge and the skills necessary to apply the domain knowledge in a community of practice.

Our services focus on assessing the level of competence with respect to the domain knowledge, to this end, the curriculum needs to be annotated semantically. This is achieved by annotating all the learning objects related to the curriculum with concepts from the domain ontology, described in section III. In addition, the tutor creates a questionnaire that reflects the domain knowledge of the curriculum. Each question is also annotated with appropriate concepts and this is the case also for the answers provided by the learner. The annotation of the curriculum provides the level of competence to be covered in the course, while the annotation of the questionnaire's answers provides evidence for the competence of the learner.

A comparison is made to check whether the answers to the questions meet the requirements imposed by the curriculum. In this way, it is possible to assess automatically the level of competence of the learner. The tutor makes the final decision about the positioning of the learner on the basis of the provided information.

An added value of this approach is that some conceptual or terminological gaps and inconsistencies might be discovered in the curriculum itself.

A. Evaluation

In order to evaluate the performance of the positioning services, 10 questions in the *computing* domain were given to Bulgarian students. About 10 answers per topic were considered and they were given to two tutors to grade them. First, we compared the concepts, present in the answers, to those required by the curriculum description. Then, we compared the automatic grading suggested by the service to the grading provided by the tutors.

The preliminary evaluation showed that pure automatic comparison might underestimate learner's knowledge; pure tutor grading ignores some aspects of learner's knowledge while putting more weight to others (i.e. verbose answers are positively evaluated even when they are not closely related to the curriculum). Our conclusion was that it could be useful to provide the tutor with the intersection list of concepts from the curriculum and the learners' answers. In this way, the tutor can analyze the concepts from the curriculum, which were mentioned in the answers as well as the list of the ones that were not mentioned. We refer to [12] for further details.

VII. Support in Course Creation

In the LTfLL project, services have been developed to help tutors creating courses in the *computing* domain. Often, tutors have to prepare their courses from scratch, without any assisting tools. Developing courses in this way takes a lot of time and it lacks continuity among the various learning tasks.

The services we have envisaged provide support to structure the content on a specific topic, to add related learning materials to the structure and to create a glossary.

The knowledge discovery process, presented in section IV, in combination with the semantic annotation of formal learning objects, discussed in section V, constitute the basis for these services. The domain ontology, which is a crucial component of the knowledge discovery process, represents the expert knowledge. Semantic search in the repository of annotated learning material is triggered by the ontology. Concepts from the ontology lead the user to those documents that are appropriate for his query either through semantic search or through ontology browsing. In this way, the reuse of learning objects in course creation is facilitated.

A. Evaluation

The services have been evaluated by two groups of users: a group of five teachers and a group of three managers.

We evaluated the following functionalities: relevance of the learning material with respect to a domain topic; relevance of the retrieved material; suitability of the ontology for structuring a lecture in the chosen domain; visualization of the interface; combination between ontology browsing and semantic search.

We assumed that the combination of semantic search and ontology browsing would support the tutors in the process of finding the relevant materials and structuring the course in a better way than without our services.

We expected differences between the two groups in their preferences. For some tutors, semantic search would be more

intuitive, while others might prefer ontology browsing. On the other hand, the managers would look at the facilities from a different perspective, such as how this supporting system would communicate with the adopted Learning Management System platform.

Two formats were adopted to test our hypothesis: the think-aloud strategy for the tutors, and the interview for the managers. The results confirmed our hypotheses. Visualization was the most commented functionality, because the teachers had different ideas on how it should be modified. The main worry of the managers was the requirement of a training course to be able to use the services.

VIII. Conclusions

One of the objectives of the LTfLL project is to develop services that facilitate learners and tutors in accessing formal and informal knowledge sources in the context of a learning task. We have shown that ontologies mediate between these two complementary learning paradigms.

Ontologies guide the learner through the resources in a knowledge discovery process. They also support annotation of learning material making thus possible to identify in the text which are the relevant domain concepts. In addition, they allow for semantic search and ontology browsing facilitating reuse of learning material.

More generally, ontologies support learners in determining their own learning agenda, they make possible direct *access* to knowledge in whatever sequence makes sense to the situation at hand, they allow for a *personalization* of the learning experience since a learner can search for learning material customized for his needs. Finally, there is no need for content to be centralized in libraries or repositories since they allow for retrieval of static and dynamic content.

References

[1] G. Stahl, *Group Cognition: Computer Support for Collaborative Knowledge Building*, Cambridge, MA: The MIT Press, 2006.

[2] L. Vygotsky, *Mind and society: The development of higher mental processes*, Cambridge, MA : Harvard University Press, 1978.

[3] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, A. Oltramari and L. Schneider, "The WonderWeb Library of Foundational Ontologies," *WonderWeb Deliverable D17,* August 2002.

[4] A. Gangemi, R. Navigli, and P. Velardi, "The Onto WordNet Project: extension and axiomatization of conceptual relations in WordNet," in *Proceedings of ODBASE03 Conference*, R. Meersman *et al*., Ed. Springer , 2003.

[5] P. Vossen, Ed., "EuroWordNet," General Document. Version 3, Final, July 19, 1999.

[6] P. Buitelaar, *et al*., "LingInfo: Design and Applications of a Model for the Integration of Linguistic Information in Ontologies," In: *Proc. of OntoLex06*, a Workshop at LREC, Genoa, Italy, May 2006.

[7] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, R., Z. Ives, *Dbpedia: A nucleus for a web of open data. The Semantic Web*, 2008, pp. 722-735.

[8] C. Cattuto, D. Benz, A. Hotho, and G. Stumme, "Semantic grounding of tag relatedness in social bookmarking systems," in *Proceedings ISWC 2008*, LNCS, Karlsruhe, Germany, 2008.

[9] A. Passant and P. Laublet, "Meaning of A Tag: A collaborative approach to bridge the gap between tagging and Linked Data," in *Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW 2008*), Beijing, China, 2008.

[10] P. Monachesi and T. Markus, "Using social media for ontology enrichment," in *The Semantic Web: Research and Applications*, Springer , 2010.

[11] E. Westerhout, P. Monachesi, T. Markus, and V. Posea, "Enhancing the Learning Process: Qualitative Validation of an Informal Learning Support System Consisting of a Knowledge Discovery and a Social Learning Component," in *ECTEL 2010*, Springer, Berlin Heidelberg, 2010.

[12] P. Osenova and K. Simov, "Semantic Annotation for Semi-Automatic Positioning of the Learner," in *Proceedings of the LREC Workshop on Supporting eLearning with Language Resources and Semantic Data*, Malta, 2010.