

Language Resources for Semantic Document Annotation and Crosslingual Retrieval

Petya Osenova¹, Kiril Simov¹, Eelco Mossel²

¹Bulgarian Academy of Sciences, Acad. G. Bonchev St. 25A, 1113 Sofia, Bulgaria

²University of Hamburg, Department of Informatics, Vogt-Kölln-Str. 30, 22527 Hamburg, Germany

E-mail: petya@bultreebank.org, kivs@bultreebank.org, mossel@informatik.uni-hamburg.de

Abstract

This paper describes the interaction among language resources for an adequate concept annotation of domain texts in several languages. The architecture includes domain ontology, domain texts, language specific lexicons, regular grammars and disambiguation rules. This is considered the preparatory phase for the integration of a semantic search facility in Learning Management Systems. The implementation and performance of this search are discussed in the context of related work as well as other types of searches. Also the results from some preliminary steps towards evaluation of the concept-based and text-based search are presented.

1. Introduction

Given the huge amount of static and dynamic contents created for eLearning tasks, the major challenge for their wide use is to improve their accessibility within Learning Management Systems (LMS). The LT4eL project¹ tackles this problem by integrating semantic knowledge to enhance the management, distribution and retrieval of the learning material (Monachesi & Lemnitzer & Simov, 2006).

The semantic annotation has already become a key ingredient of Semantic Web. There is already a vast quantity of literature and initiatives, which approach this topic from various perspectives. For example, there was SAAW 2006 - the First Semantic Authoring and Annotation Workshop devoted on tools, standards and practice of semantic annotation. In this paper, we present a model of the relation of a domain ontology to text. In order to facilitate this relation we need to construct corresponding language resources including lexicons and grammars. Here we discuss the nature of these resources. Also we describe how they can be used for mono- and multilingual search.

The paper is structured in the following way: first we present our model on the relation between a domain ontology and domain text; then we present the ontology search based on this relation; the next section reports on a comparison between text-based search and ontology search; the last section concludes the paper.

2. Domain Ontology and Semantic Annotation

The ontology-based querying for content retrieval has

¹ <http://www.lt4el.eu/> – the LT4eL (Language Technology for eLearning) project is supported by the European Community under the Information Society and Media Directorate, Learning and Cultural Heritage Unit.

been actively explored in recent years. Here we will mention the OntoQuery Project² among others. The differences with our project are as follows: OntoQuery contributed to the issues of the onto-based search in general. We focused on this kind of search for learning purposes. OntoQuery was designed mainly for Danish-speaking users, while our search aims at multi- and crosslingual retrieval. At the moment, our process involves previous semi-automatic processing (i.e. requiring some user intervention).

The domain of the learning corpus in the LT4eL Project is “Computer Science for Non-Computer Scientists”. It covers topics like operating systems; programs; document preparation – creation, formatting, saving, printing; Web, Internet, computer networks; HTML, websites, HTML documents; email, etc. The main application of the ontology is: the indexing of these domain documents with concept information and interconnecting the same information across different languages.

The initial stages of the ontology creation were supported by a core set of manually annotated keywords in the eight languages of the project (Bulgarian, Czech, Dutch, English, German, Polish, Portuguese, Romanian). In the next development some middle-placed concepts were added and some classes were expanded (e.g. various types of text editors). Let us explain these steps in more detail: the process of choosing keywords in a text is more or less subjective. We tried to handle this problem by exploring texts in various languages. Another problem is the usage of too specific or too broad terms. Thus, the middle ones were often left out (e.g. *human activity, resource, symbol, architecture, organization*). But they seem to be very important for tracing the connection with the top part of the ontology. The annotated keywords from the other languages were translated into English. Then by search on the Web we collected definitions for the keywords. The set of definitions for a keyword either highlight the various aspects of the meaning of the keyword or the relations between its meaning and other concepts. In the

² <http://www.ontoquery.dk/index.php>

ontology we keep the most representative definition and we keep the rest of the definitions in an additional corpus in order to consult them during the formalization of the concepts in the ontology. We have preferred definitions, which reflect basically the *is-a* relation. The other relations were encoded in the ontology (*part-of, used-for* etc.). The definitions were also translated into the other languages. This is important for the user who browses the ontology. After the determination of the keyword meanings we created concepts corresponding to them. These concepts became the backbone of the domain ontology.

The next step of the ontology development was to map the domain concepts to an upper ontology (in our case we used DOLCE – (Masolo, C. et al., 2002a), (Masolo, C. et al., 2002b)) in order to inherit some knowledge already encoded in the upper ontology (relations, for instance) and to ensure the correct concept classification with respect to concept metaproperties defined in the ontology creation methodology – OntoClean (Guarino, N. & Welty, C., 2002). The mapping was facilitated by OntoWordNet (Gangemi, A. Navigli, R. & Velardi, P. 2003). The relations, inherited from the upper part, are very abstract. However, they were specified further with respect to the domain needs, or were used for consistency checks. Additionally, the ontology was extended with concepts from other sources like terminological lexicons and Wikipedia. At the moment, the domain ontology contains over 1000 domain concepts, about 50 concepts from DOLCE and about 200 intermediate concepts from OntoWordNet.

In order to use the ontology for semantic search over documents, we had to establish a connection between the ontology and the texts of the documents. We established this connection by three types of language resources and tools. The first type is the language specific lexicons aligned to the ontology. Each lexicon contains lexical items grouped by their meaning which is represented in the ontology. There exist various attempts to approach this mapping task. Most of them start from lexicons already existing for several languages, and then try to establish a connection among the concepts defined in these lexicons. Such initiatives were WordNet (Fellbaum, 1998), EuroWordNet (Vossen, 1999), SIMPLE (Lenci et al., 2000). In spite of the fact that we employ the experience from these projects (mapping to WordNet and Pustejovsky’s ideas in SIMPLE), we also suggest an alternative in connecting the ontology to the lexicons. Our model is very close to the LingInfo model (Buitelaar et al. 2006) with respect to the mapping of the lexical items to concepts, but also with respect to the other language processing tools we connect to the ontology – the concept annotation grammars and concept disambiguation tools. Thus, the other two language resources are (1) partial grammars which facilitate the mapping from the lexical items to their realization in the texts; and (2) disambiguation rules which solve the problem of ambiguity of the lexical items on the basis of the context of their usage in the texts. The partial grammars consist of

regular expressions that assign the appropriate concept label to a string sequence. The grammars were created semi-automatically on the base of the expressions in the lexicons. They were applied to the lemmas. Here is an example of such a rule for Bulgarian:

Regular Expression:

```
<"кодиране">,<" на">,<"знак">
```

Return Markup:

```
<Concept>\w<conl>
<c>lt4el:CharacterEncoding</c>
</conl>
</Concept>
```

The regular expression presents the sequence of the Bulgarian string “coding”, “of”, “sign”. The return markup expression assigns to it the domain concept *CharacterEncoding*. The disambiguation of the ambiguous cases was performed also semi-automatically. The implementation of the annotation grammars and disambiguation rules were implemented within the CLaRK System (Simov et al., 2001). A constraint was prepared to stop on the ambiguous cases only. Then a human expert differentiated among various possibilities. Here are some examples: The English term “help” is ambiguous between the concepts *HelpButton* or *HelpCommand*. The Bulgarian term “вмъкване” (inserting) might correspond to three concepts: *Import*, *InsertKey* and *InsertMenuItem*. Within the basic Bulgarian lexicon the ambiguous cases are about 5 % from all the mappings (62 rules with ambiguity out of 1092 altogether). In the next additions there were no such cases, because the concepts became more and more specific, and hence – less inclined to ambiguity.

The model is graphically depicted in the next picture:

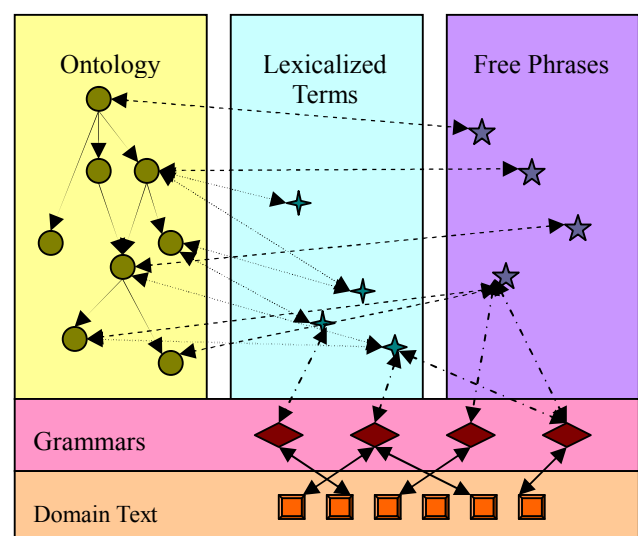


Figure 1: Model of ontology-to-text interrelations.

These mappings ensure the path from the ontology concepts to their string counterparts in the text. Let us

explain the presented architecture more explicitly. The concepts from the ontology were presented in the natural language either as lexicalized terms, or as freely combined phrases. Needless to say, one concept might correspond to only a lexicalized term, only a free phrase or both. All the language expressions were encoded in regular grammar rules, which mapped the string to the appropriate concept(s). The grammars were applied to the domain texts. In cases of ambiguity a human expert took a decision. The ontologically annotated texts became the base for the semantic search facility. Thus the relations, outlined in Figure 1, are a basis for (1) monolingual search in which the ontology inference is used for query expansion; and for (2) multi-language search in which, in addition to query expansion, the ontology is used as a mediator between the various languages.

3. Ontology-Based Semantic Search

Based on three resources as described above, i.e. the term-concept lexicons, the ontology and the concept annotation of the documents, we developed an ontology-based search engine having the following main characteristics, in accordance with the goals of the project LT4eL:

- 1) Domain concept matching: the accessibility to documents in a Learning Management System is improved by retrieving learning materials that have domain concepts in common with the search query. First, the lexicons are used to find domain concepts in the ontology; subsequently, the ontology is used for query expansion, which further improves recall.
- 2) Multilinguality: the functionality is multilingual - one implementation is used for all of the eight languages of the LT4eL project.
- 3) Crosslinguality: the search engine enables users to find documents in several languages at the same time, while using search terms or an ontology representation in one language according to the user's choice.

A presupposition for using the search is the availability of annotated learning material and a lexicon in at least one (or for crosslingual search at least two) of the languages the user knows. In the LMS, the user's choice of languages is a part of her profile. Furthermore, a requirement is that the topics of the documents are (at least partly) covered by the ontology.

The basic idea of the ontology-based semantic search is that concepts from the ontology lead the user to those documents that are appropriate for her query. The search will be most precise when the user directly selects concepts from the ontology. Some other approaches like Finkelstein et al. (2001), Lim et al. (2005) and Song et al. (2005) use a lexical ontology in an automatic way for query expansion, while the retrieval takes place on the basis of the expanded textual query.

Our approach is different in two respects. First, we allow two modes of ontology usage: an automatic mode, where users type search words, and a non-automatic ontology

navigation mode, where documents are retrieved after manually selecting concepts. Second, the ontology, together with the concept annotation of the documents, is used as an intermediate level between query and documents; no step back to a textual representation is involved. This allows for retrieval in multiple languages with one ontology, as well as for crosslingual retrieval.

Related approaches that also make use of concept annotation and retrieval by concepts are described by Kiryakov et al. (2003) and Vallet et al. (2005).

Search Procedure

The search procedure takes the following parameters:

- Language(s) of search query (determines which lexicons to use for concept lookup)
- Retrieval languages (for which the user wants to see available documents)
- Search terms (entered by user), or concepts (selected by user)
- Method for combining the concepts (AND/OR)
- Threshold for ontology-based query expansion

Learning objects are retrieved by means of the following steps:

- The search words are looked up in the lexicons of the chosen languages. Search words are normalized orthographically before lookup. In case the OR option is selected, combinations for multi-word terms are created using several concatenators (e.g. if the words "computer" and "screen" are entered, the combinations "computerscreen", "computer screen" and "computer-screen" are created and looked up, in addition to the individual words "computer" and "screen"). With AND-search, this step is skipped, since it would unintentionally restrict the search: it would require the result document to contain not only the concepts computer and screen, but also "computer screen".
- If lexical entries are found in the lexicon, their denoted concepts are taken as search concepts. These concepts are also used as starting points for ontology navigation. Alternatively, concepts directly selected from the ontology are the basis for the search.
- Documents in the desired languages are retrieved, based on the set of found or selected concepts, while taking into account the AND/OR parameter. If the number of retrieved documents is less than the threshold, the set of search concepts is expanded with their respective subconcepts and the search is repeated. Thus, documents are added that treat a topic at a more detailed level than was specified by the user. We did not experiment with dynamic ways of query expansion like Bonino et al. (2004) propose, where the number of levels and the ontology navigation direction (more general/more specific) is made dependent on the number of results. In our envisaged context, Learning Management Systems, the number of available documents can be very small and the collection biased, so for some queries, a short result list can be a "correct" result. In that case, the learner

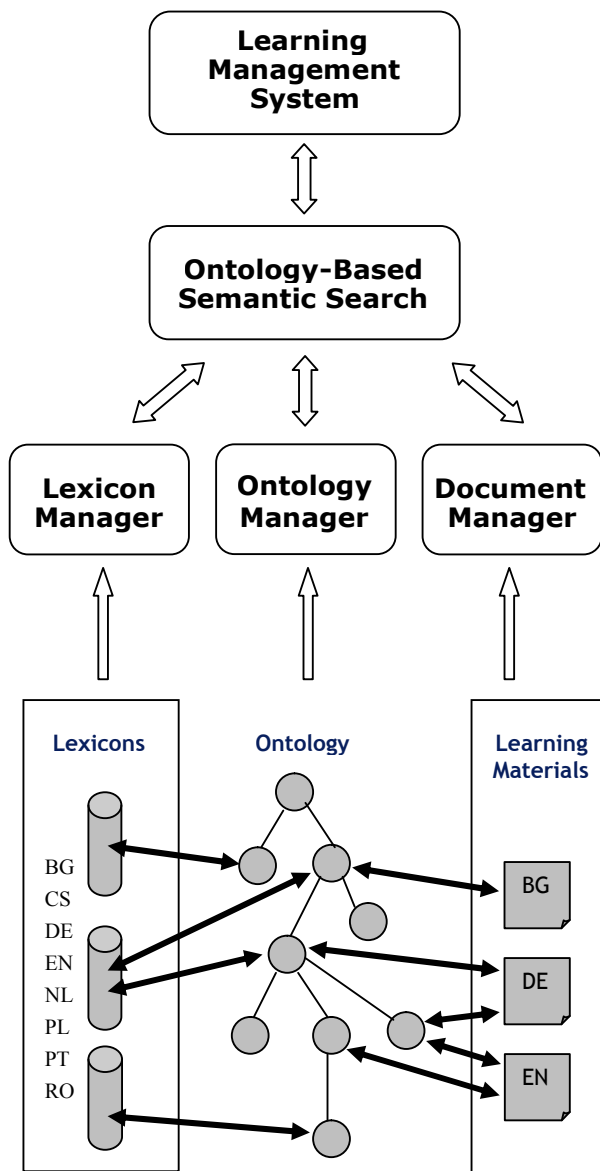


Figure 2: The architecture in which the search functionality is integrated. The lower part of the diagram shows the relationships between terms, ontology and documents.

with foreign language knowledge can look at relevant material in other languages.

- For each retrieved document, the following information is provided, so that the LMS can present it to the user in an appropriate way:
 - Matching concepts: all concepts that were the basis for search *and* match the document. This is a subset of all the concepts that relate to the document, and can include *main search concepts* (concepts that are found on the basis of the entered terms, and concepts directly selected from the ontology), but also super and subconcepts of those in case concept query expansion was invoked.
 - Relevance score, based on the occurrences of matching concepts.
 - Snippet: a part of the text selected around the

annotations of matching concepts.

The LMS can then display this information, together with some meta information such as the title of the document, its language and assigned keywords.

The next three subsections give more information on the use of the ontology in search, the relevance scores and the snippets, respectively. Figure 2 gives an overview of the architecture. The search functionality is being integrated into the Learning Management System ILIAS, in which the collected learning materials are stored, while Figure 3 shows an example of the user interface to the search functionality.

Ontology Interaction

Although it is our assumption that the search is most focused towards the desired topic if concepts are directly selected from the ontology, the reasons to allow a free-text query to initiate semantic search are two-fold. First, we assume that the users, who are probably familiar with Google, want their results fast, with not too many intermediate steps. The simplest case is to type search words and click on search. This procedure is also used for full-text search in our system, and users might avoid semantic search if they think it is much more complicated than full-text search. Second, we use the entered search words to find a good starting point in the ontology, so that the user does not have to click his way through the ontology starting at the root.

In the list of retrieved documents, the concepts that match the query as well as the document are presented. By clicking on a matching concept, the user can switch to the screen for manual ontology navigation starting from that concept, and then select related concepts as the input for a more precise search. It is also possible to start immediately from the ontology navigation view; in this case, no domain concept can be offered to the user as a starting point, so navigation will start from the root of the ontology: the most general available concept.

Relevance score

For each document, a relevance score is calculated, by which the retrieved documents are sorted. It is a value between 0 and 1 that can be presented to the user as a percentage, to indicate the estimated relevance. The value is an aggregation of two scores, reflecting the following aspects:

- The number of different *main search concepts* that match the document (excluding concepts that were automatically added by query expansion). This reflects how well the document matches the query;
- The *occurrence frequency* of the matched concepts: if they occur more often, they play a more important role in the document. The frequency is normalized for document length, to compensate for the fact that a short document cannot mention the concept as often as a long document but can still be very relevant. For this score, also the matched inferred (super/sub) concepts are taken into account, but with a lower weight than the main search concepts. Thus, the

second part of the score reflects the relevance of the concepts to the document.

We opted for a relevance score per document that reflects the correspondence between the query and the document independently of the other retrieved documents or the total available set of documents. It might look logical to set the document with the highest annotation frequency (relative to the document length) at 100% and the rest proportional to it, but we saw that in this way, the scores of the other retrieved documents were often too low and also too much influenced by changes in the repository. Instead, we base the score on an experimental *expected concept-token-ratio* per document. We use 0.005 (one matching concept per 200 tokens) at the moment. Obviously, this can result in a score above 1 for certain documents with very high annotation frequencies. To correct this, we do not cut them to 1, but rather use the following formula, that maps values between lower boundary B and infinity to values between B and 1:

$$\text{corrected} = B + (1-B) * (S-B) / (S-B + 1-B)$$

where B is supposed to be a value between 0 and 1, and S is the score to be corrected. We are currently using B = 0.7, which gives the following corrected scores (examples):

ORIGINAL	CORRECTED
0.8	0.775
1.0	0.850
1.5	0.918
3.0	0.965

Of course, the expected concept-token ratio is chosen such that the values are not above 1 in most cases. The corrected score is a factor in the final relevance score, which is a weighted average where the other factor is the normalized (between 0 and 1) number of main search concepts.

In the version of the search system that is currently being developed in LT4eL, users are able to choose which of the types of search they want to use simultaneously. The semantic search results are joined with the results of full-text search and keyword search³, which also have a relevance score.

Similar to what Vallet et al. (2005) report about their combination of semantic and full-text search, we found that the final score can be undesirably low when only one of the search methods returns a good relevance score for a relevant document, especially in our case where three methods can be combined. In our opinion, a low score does not necessarily mean that the document is not relevant, but rather that there is no evidence for relevance, while a high score is based on positive evidence from the text or meta data. Therefore, we use a weighted average that favors the best score:

$$0.6 * \text{highest} + 0.3 * \text{middle} + 0.1 * \text{lowest}$$

³ In LT4eL, by *keyword search* we mean matching with words that are assigned as keywords to the documents.

Snippets

Comparable to the presentation of results in Google, we select a snippet of the text, which serves as a preview of the document. It is a small fragment of the document (or two discontinuous fragments, connected by three dots), selected around occurrences of the matching concepts. The annotation itself is removed, but the words that were annotated by one of the search concepts are marked with tags so they can be highlighted when displaying. If there are multiple matching concepts, the ones that occur more frequently are preferred. Furthermore, occurrences of different concepts close to each other in the text are preferred. The idea behind this is that terms (or the concepts underlying them) describing a topic are likely to co-occur in one sentence or passage. Both Park et al. (2002) and Google (<http://www.google.com/technology/whyuse.html>) use this notion for ranking (a document is more relevant if it contains such a passage) while Hearst (1995) and Drori (1998) apply it for passage retrieval. In our approach, it allows for selection of a snippet that is representative as a preview of the document to the user.



Figure 3: User interface making use of semantic search. Starting from the top, it shows the field for the search terms, OR/AND option, language of search terms, retrieval languages, search methods, and found documents. For the found documents, the interface shows: keywords, language, relevance score, matching concepts, snippet.

4. Evaluation

Within the project two types of evaluation were conducted: user-oriented and search-oriented. The user-oriented one referred to tests with students and tutors, based on various scenarios. In these tests all the languages were involved. The participants had to use different types of searches within LMS in order to perform their tasks. The general

conclusion was that, when handled properly, the semantic search was appreciated for its preciseness and fast results. On the other hand, the text search was the most familiar approach. For that reason, we performed an initial evaluation, based on text-search-like and concept-search like query.

In other words, we tried to evaluate the search function, comparing simple text search and semantic search. The basic task for this evaluation was as follows: two terms (which were also lexical entries in the lexicon of the language under investigation) have been chosen as parts of a query (the equivalents of the terms “program” and “slide” were taken in each language⁴). The combination of these two terms was used under the presupposition that they had unambiguous meanings in the domain texts. We encoded the question as two queries: a simple full text search (including just the lemmatized forms of the terms and the list of terms is extended with related terms which we suppose are known by the learner), and as a semantic search using the ontology to expand the query with all the subconcepts and just one superconcept. For example, the text query for English includes strings like: *program; software; editor; slide*. These basic forms are matched to the lemma annotation within the annotated documents. As it can be seen, this set does not include a sophisticated query expansion on this level. We kept the simple level, because we wanted to be closer to the search possibilities within the LMS⁵. However, in a more general plan, we envisage to make an experiment with advanced text queries. The semantic search query included more information, since it took all subconcepts of *Program* (among which some more specific - *NotePad, CorelDraw*, and some more general - *TextEditor, SortProgram*). Altogether the expanded queries consist of 96 concept names. These concept names are matched to the conceptual annotation in documents. Both types of queries were run over the document sets. This resulted in two sets of paragraphs, one for the text search and one for the semantic search. The conceptual annotation has been used to identify the paragraphs in these documents. The conjunction of the two result sets has been investigated by a researcher and each paragraph was rated as either relevant or irrelevant to the search. The retrieval results of both methods (text search and semantic search) have been weighted against the set of relevant paragraphs with the well-known measures of recall and precision. Both values have been combined in an F-measure. Similarly to the approach by which CLEF (<http://nlp.uned.es/clef-qa>) initiatives handle recall, we assumed that the sum of all found results equals the recall measure (i.e. that all the relevant paragraphs are among the retrieved ones). Of course, it is not quite true, but it serves well for our goal. The experiment was run for six languages: Bulgarian, Dutch, English, German, Polish and Portuguese. The

⁴ The interpretation of the query is “Which program do you use for the preparation of slides?”

⁵ We thank Pavel Smrř who pointed to us that there is a more advanced text search which could be used as a baseline. We are currently working on a new evaluation.

F-measures for both text search and semantic search are presented in Table 1. The gain is due to improvements in both recall and precision. It is significant for all languages. The gain is the lowest for Portuguese, because there were only a small number of returned documents. Also, there is visible variation between the languages.

Language	Text Search	Semantic search
Bulgarian	56,25	91,30
Dutch	47,50	94,12
English	27,96	79,42
German	36,00	59,26
Polish	12,50	50,00
Portuguese	28,67	33,33

Table 1: F-measures for full text search and semantic search in six languages.

Another factor that played a role in the results was the context: the narrower the context (e.g. sentences), the better the results, and vice versa. As has been said before, the conceptual search produces results only in those cases where the search words are in the lexicon and thus matched to concepts in the ontology. This has been the case in the evaluation example. In the case where the search word did not match a lexical item, the text search as well as the keyword-based search was used as a fallback strategy.

After repairing and extending the ontology, it is a subject to further user-centered evaluations to estimate how well the semantic search performs in the context of the Learning Management Systems real tasks and alternative search methods (text, keyword, definition). This part of the evaluation is still on-going.

5. Conclusion

In this paper, we described a complex architecture for relating the domain ontology to a multilingual collection of texts. We also presented the employed resources and their usage for the ontology-based search. The very first steps towards the evaluation of the search functionality were outlined, which indicates that the ontology-based search significantly improves the retrieval for several languages.

At the moment we believe that the semantic search outperforms the simple text search when querying with more general terms (in general when inference over ontology is used for query expansion and/or answer evaluation). In the case of queries based on the specific terms (which do not allow usage of inference) practically there is no difference between the two types of search. We envisage to test the semantic search against more advanced types of text search.

6. Acknowledgements

This work has been supported by LT4eL (Language Technology for eLearning) – European project of the European Commission under the Information Society and

Media Directorate, Learning and Cultural Heritage Unit (FP6-027391).

We would like to thank also the three anonymous reviewers for their valuable remarks and suggestions.

7. References

- Bonino, D., Corno, F., Farinetti, L., & Bosca, A. (2004). Ontology Driven Semantic Search. *WSEAS International Conference on Automation & Information (ICAI 2004)*, Venice, Italy.
- Buitelaar, P., Declerck, T., Frank, A., Racioppa, S., Kiesel, M., Sintek, M., Engel, R., Romanelli, M., Sonntag, D., Loos, B., Micelli, V., Porzel, R., Cimiano, P. (2006). 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*.
- Drori, O. (1998). The User Interface in Text Retrieval Systems. In: *SIGCHI bulletin*, ACM, New York, July 1998, 30 (3), pp. 26--29.
- Fellbaum, C. (1998). (ed). WORDNET: an electronic lexical database. MIT Press.
- Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., Ruppin, E. (2001). Placing Search in Context: The Concept Revisited. In: *Proceedings of the 10th international conference on World Wide Web (WWW10, May 2001)*, Hong Kong. pp. 406--414.
- Gangemi, A., Navigli, R. & Velardi, P. (2003). The OntoWordNet Project: extension and axiomatization of conceptual relations in WordNet. Meersman R, et al. (eds.), *Proceedings of ODBASE03 Conference*, Springer, 2003.
- Guarino, N. & Welty, C. (2002). Evaluating Ontological Decisions with OntoClean. *Communications of the ACM*, 45(2): 61-65.
- Hearst, M. (1995). TileBars: Visualization of Term Distribution Information in Full Text Information Access. In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI '95)*, pp. 56--66.
- Kiryakov, A., Popov, B., Ognyanoff, D., Manov, D., Kirilov, A., & Goranov, M. (2003). Semantic Annotation, Indexing, and Retrieval. In: *Proceedings of the Second International Semantic Web Conference (ISWC 2003)*, Sanibel Island, FL, USA.
- Lenci, A., Busa, F., Ruimy, N., Gola, E., Monachini, M., Calzolari, N., Zampolli, A., Guimier, E., Recourcé, G., Humphreys, L., von Rekovsky, U., Ogonowski, A., McCauley, C., Peters, W., Peters, I., Gaizauskas, R. & Villegas, M. (2000). SIMPLE Work Package 2 - Linguistic Specifications, Deliverable D2.1. *ILC-CNR*, Pisa, Italy.
- Lim, S., Park, S. & Lee, S. (2005). Document Retrieval Using Semantic Relation in Domain Ontology. In: P.S. Szczepaniak et al. (Eds.): *AWIC 2005, LNAI 3528*, Springer Verlag, Berlin Heidelberg, pp. 266--271.
- Masolo, C., Borgo, S. & Gangemi, A. & Guarino, N. & Oltramari, A. & Schneider, L. (2002a). The WonderWeb Library of Foundational Ontologies. WonderWeb Deliverable D17, August 2002. <http://www.loa-cnr.it/Publications.html>.
- Masolo, C., Borgo, S. & Gangemi, A. & Guarino, N. & Oltramari, A. (2002b). Ontology Library (final). WonderWeb Deliverable D18, December 2003. <http://www.loa-cnr.it/Publications.html>.
- Monachesi, P., Lemnitzer, L. & Simov, K. (2006). Language Technology for eLearning. In *Proceedings of EC-TEL 2006, Innovative Approaches for Learning and Knowledge Sharing, LNCS 0302-9743*, pp. 667-672.
- Park, E., Moon, S., Ra, D., & Jang, M. (2002). Web Document Retrieval Using Sentence-query Similarity. In: *Proceedings of the 11th Text Retrieval Conference (TREC-11)*.
- Simov, K., Peev, Z., Kouylekov, M., Simov, A., Dimitrov, M., Kiryakov, A. (2001). CLaRK - an XML-based System for Corpora Development. In: *Proc. of the Corpus Linguistics 2001 Conference*. Lancaster, UK.
- Song, M., Song, I., Hu, X., & Allen, R. (2005) Semantic Query Expansion Combining Association Rules with Ontologies and Information Retrieval Techniques. In: A. Min Tjoa and J. Trujillo (Eds.): *Proceedings of the 7th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2005)*, LNCS 3589. Springer Verlag, Berlin Heidelberg, pp. 326--335.
- Vallet, D., Fernández, M., & Castells, P. (2005). An Ontology-Based Information Retrieval Model. In: Gómez-Pérez, A., Euzenat, J. (eds.): *The Semantic Web: Research and Applications: 2nd European Semantic Web Conference (ESWC 2005). Lecture Notes in Computer Science, Vol. 3532*. Springer Verlag, Berlin Heidelberg, pp. 455--470.
- Vossen, P. (1999). (ed). EuroWordNet General Document. Version 3, Final, July 19, 1999, <http://www.hum.uva.nl/~ewn>.