



Project no. 027391

Project acronym: LT4eL
Project title: Language Technology for eLearning

Instrument Specific Targeted Research Project

Thematic Priority Information Society Technology

Final Activity Report

Period covered: from 1-12-2005 to 31-05-2008

Date of preparation: 05-08-2008

Start date of project: 1-12-2005

Duration: 30 Months

Project coordinator name: Paola Monachesi

Project coordinator organisation name: Utrecht University (UU)

Organisation name of lead contractor for this deliverable: Utrecht University (UU)

Revision [1]

Final activity report

Contents

- 1 Executive Summary
- 2 General project information
 - 2.1 Introduction
 - 2.2 LT4eL objectives
 - 2.3 LT4eL methodology
 - 2.4 Contractors involved
 - 2.5 Co-ordinator contact details
- 3 Project execution: work performed to reach the objectives
 - 3.1 Integration of Language Technology resources and tools in eLearning
 - 3.2 Integration of semantic knowledge in eLearning
 - 3.3 Supporting multilinguality
 - 3.4 Development of enhanced LMS open source ILIAS Learning Management System
 - 3.5 Validating Language Technology enhanced LMS
 - 3.6 Knowledge transfer of the consortium
 - 3.7 Awareness raising
- 4 Project results
 - 4.1 LT4eL final results
 - 4.2 Impact of the project on its industry or research sector
 - 4.3 Lessons Learned
 - 4.4 Publishable results
 - 4.4.1 Corpus of Learning Objects
 - 4.4.2 Keyword Extractor
 - 4.4.3 Glossary Candidate Detector
 - 4.4.4 Ontology & language specific vocabularies
 - 4.4.5 ILIAS integrated functionalities
 - 4.4.6 Validation methodology

1 Executive Summary

This document is the publishable final report of LT4eL (Language Technology for eLearning, contract no. 027391), a Specific Targeted Research Project supported by the European Community under the Information Society and Media Directorate, Learning and Cultural Heritage Unit. The project started on December 1st, 2005 and lasted 30 months. The consortium is composed of 12 partners. The document introduces the objectives, the project partners, the degree of achievement, the main project results and their impact at the scientific and industrial level as well as the lesson learned. Further information can be found on the multilingual project Web site at <http://www.lt4el.eu>



2 General project information

2.1 Introduction

In the Language Technology for eLearning project (LT4eL), we have addressed one of the major problems users of ever-expanding Learning Management Systems (LMS) will be confronted with: how to retrieve learning content from an LMS. We have tackled this problem from two different but related angles: from the content end and from the retrieval end.

On the content side, the fast growing content cannot be easily identified in the absence of systematic metadata annotation. Providing metadata is a tedious activity and the solution we offer is to provide Language Technology based functionalities, such as a keyword extractor. It allows for semi-automatic metadata annotation on the basis of a linguistic analysis of the learning material. In addition, we provide a glossary candidate detector which allows for the extraction of definitions from the learning objects and the creation of glossaries related to them. The aim was to provide these functionalities for all the nine languages represented in our project, that is Bulgarian, Czech, Dutch, English, German, Maltese, Polish, Portuguese and Romanian.

On the retrieval side, the standard retrieval systems, based on keyword matching, or full-text search, only consider the query terms. They do not really take into account the systematic relationships between the concepts denoted by the queries and other concepts that might be relevant for the user. In the LT4eL project, we use an ontology as an instrument to express and exploit such relationships, which should result in better search results and more sophisticated ways to navigate through the learning objects. Furthermore, by linking the ontology to language specific lexica, cross-lingual retrieval is made possible. In the context of the internationalization of the learning process, cross-lingual retrieval of learning objects should be a key feature of any LMS.

The functionalities developed within the LT4eL project could be integrated in any open source LMS, however, for validation purposes the ILIAS Learning Management System (www.ilias.de) has been adopted. A validation methodology has been developed.

The main contribution of the project consists thus in the introduction of new functionalities based on Language Technology which enhance the adaptability and the personalization of the learning process through the software which mediates it.

2.2 LT4eL objectives

In order to achieve the goals of the project, the following scientific and technological objectives which had been stated at the beginning of the project had to be met:

1. Integration of Language Technology resources and tools in eLearning - the project employs Language Technology resources and tools, such as corpora, lexica, taggers, lemmatizers and chunk parsers which have been produced in the context of other projects. They are used for the development of new functionalities that allow for the semi-automatic generation of metadata for the description of learning objects in an LMS and for the creation of glossaries related to the learning objects. This will result in two modules: a) a keyword extractor; b) a glossary candidate detector. These functionalities are implemented as web services and they can be integrated into any Learning Management Systems. Within the project, the functionalities will be integrated into ILIAS.

2. Integration of semantic knowledge in eLearning - the project integrates the use of ontologies, a key element in the Semantic Web vision, to structure and retrieve the learning material within the LMS. The learning objects are annotated on the basis of the ontology which is used to browse the learning material and also for the cross-lingual retrieval of the learning objects.

3. Supporting multilinguality - the project supports the multilingual character of the enlarged Europe. Special attention is given to the commercially less attractive languages which are represented in the consortium. The new functionalities are developed taking into account nine of the languages represented in the consortium. Furthermore, cross-lingual retrieval of learning objects is implemented (by linking the ontology to vocabularies of the various languages), the focus is on the languages of the New Member States as well as on the language families represented in the consortium (Romance, Germanic, Slavic).

4. Development of enhanced LMS open source ILIAS Learning Management System - the new functionalities are integrated in the existing open source ILIAS Learning Management System, and validated in a realistic learning environment. However, we assume a modular approach so that the functionalities developed within the project will be compatible with different open source platforms. We focus on learning material in the area of computing and eLearning. We collect a corpus of at least 1000 pages (i.e. 200 000 running words) for nine of the languages represented in the consortium.

5. Validating Language Technology enhanced LMS - the project examines the extent to which the ILIAS eLearning management system is improved by the addition of the new functionalities. The investigation looks at how the new functionalities will impact upon usability, pedagogy and virtual learner communities. To do this, the project develops a validation methodology suitable for investigating the impact of the new functionalities on the learning experience in a multilingual context.

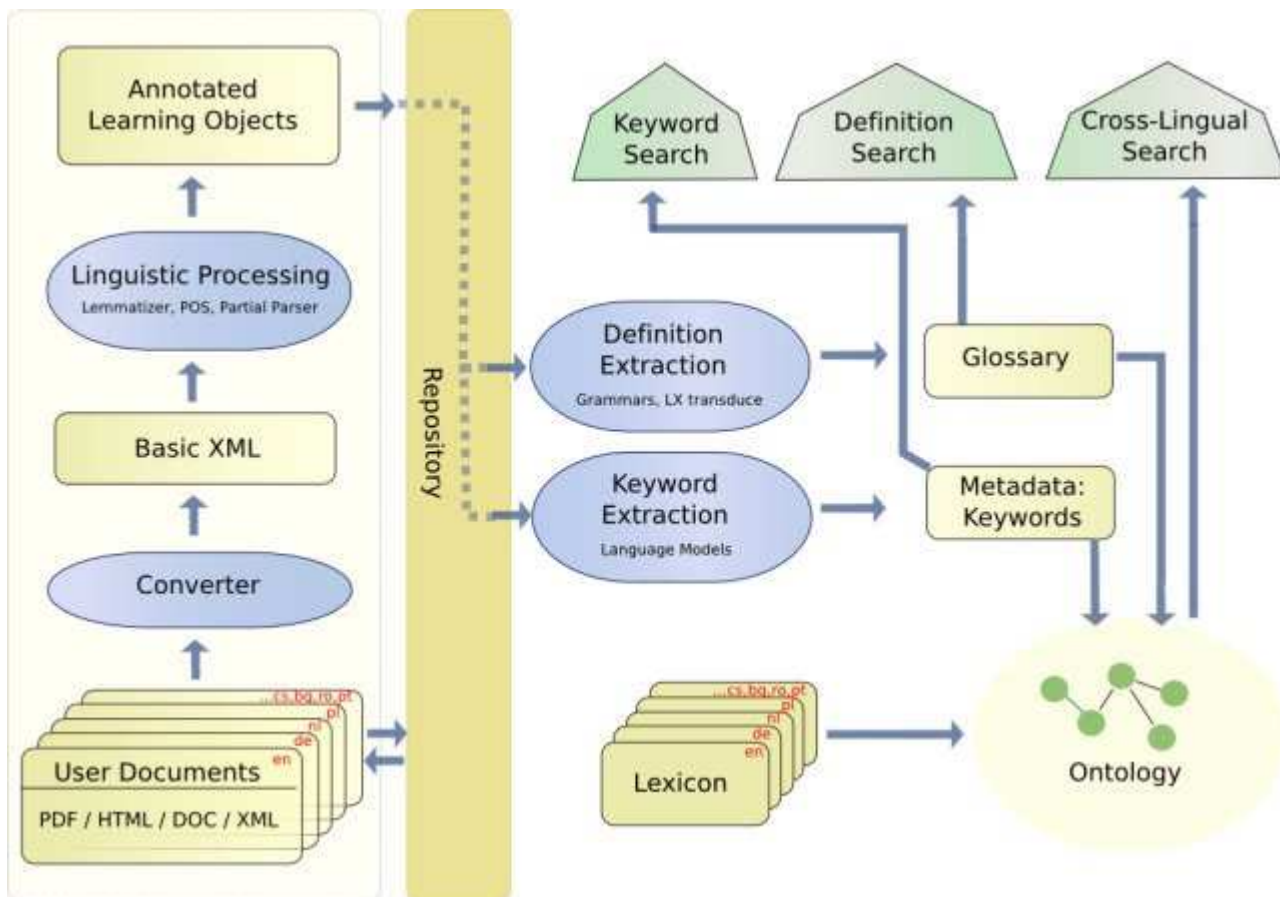
6. Knowledge transfer of the consortium - the project should encourage the flow of knowledge among the members of the consortium. Partners of the NMS actively collaborate with the leaders of the core Workpackages to develop the relevant functionalities ensuring thus transfer of expertise within the NMS. Furthermore, we stimulate, through the participation at relevant workshops and conferences the information flow outside the parties of the consortium strengthening thus the integration of the IST Research in Europe.

7. Awareness raising - the project draws attention, within the eLearning community to the significant potential of Language Technology and emerging technologies offered by the Semantic Web initiative. The approach we develop in the project is marketed within this community and eventually extended to other open source LMS. Furthermore, we address ministries of education in order to stimulate support for the development of learning material in the national languages.

8. Exploitation plan - information about activities and results of the project are disseminated through the multilingual Web portal (www.lt4el.eu). The technical results of the project are available through the webportal, via the sourceforge webpage as well as in documented releases and through participation and organization of relevant events.

2.3 LT4eL methodology

The picture below shows the architecture of the project and the methodology adopted to reach the objectives described in the previous section.



In the left hand side of the picture, we see that the learning objects which are uploaded into the Learning Management System, can be in any format (i.e. pdf, html, doc or xml) and are converted into a basic XML format. They can thus undergo the necessary linguistic annotation by means of a part-of-speech tagger, a lemmatizer and a morpho-syntactic analyser. A linguistically annotated document in an XML format which is derived from the XCESAna standard for linguistically annotated corpora is produced and uploaded into the repository of the LMS.

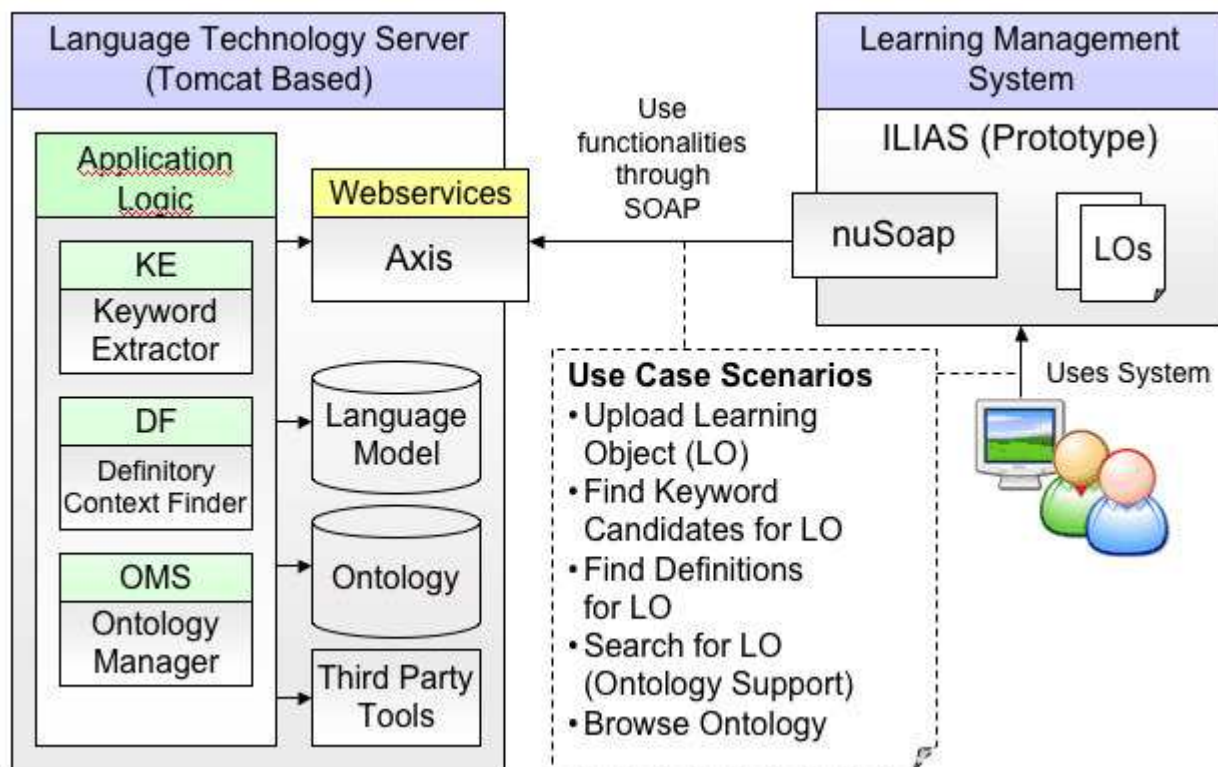
In the LT4eL project, a keyword extractor has been developed that accepts linguistically annotated learning objects as input and outputs a list of suggested keywords to be included in the Learning Object Metadata (IEEE LOM). We use the same algorithm for all the eight languages under consideration while we encode the language specific differences in the language model. In particular, the linguistic information, which is extracted from the corpus of learning objects, is added to the language model for the specific language which consists of three parts: lexical units, word form types and documents. The list of candidate keywords is ranked by their saliency and to determine it an approach based on frequency has been adopted. In our project, we have implemented and evaluated the appropriateness of four statistical measures, that is TFIDF, Residual IDF and Adjusted RIDF and Averaged Reduced Frequency. Learning objects are thus annotated with appropriate keywords and it is possible to retrieve them by means of keyword search.

Within the LT4eL project, we have also developed a glossary candidate detector that extracts definitions from the learning material in order to create glossaries semi-automatically which are related to a given learning object. In our approach, we have combined Natural Language Processing and Machine Learning techniques to extract definitions from our learning objects. In the first step, local grammars, based on manually selected definitions, have been created for all the languages and have been employed for the detection of the definitions within the learning objects. The grammars are used in combination with the lxtansduce tool which allows for the matching of grammar rules with subtrees within XML-annotated documents. In addition, for some languages, we have experimented with machine learning techniques to filter incorrect results in order to improve the precision of the system. The definitions extracted can be selected to create glossaries and can also be searched independently.

One of the aims of the LT4eL project is to improve the retrieval and the usability of (multilingual) learning material within a Learning Management System. In order to achieve this objective, an ontology-based search functionality has been developed which allows for cross-lingual retrieval of the learning objects. It is

based on a (language independent) ontology of about 1300 concepts that includes an upper ontology (i.e. DOLCE), a domain ontology in the area of computing based on the keywords extracted from the learning objects as well as appropriate definitions. In addition, we have produced a lexicon for each of the languages addressed which comprises words or phrases that are mapped to concepts attested in the ontology as well as a collection of learning objects annotated on the basis of the concepts attested in the domain ontology.

The functionalities described above have been integrated in a Learning Management System, that is ILIAS in the case of the LT4eL project. The picture below shows the major components of the integration setup. The language technology server on the left of the picture provides the functionalities developed within the project, that is the keyword extractor, the definitory context finder and the ontology management system. The tools themselves are developed using the Java programming language and are hosted on a Java web server. The functionalities can be accessed by the learning management system through the web service interface. The web service interface separates the technologies used on both sides.



2.4 Contractors involved

The consortium of the LT4eL project has included 12 partners mentioned below. Additional details about the institutions involved, the role they have played in the project and the people who have taken part in it can be found on the multilingual project website, that is <http://www.lt4el.eu>

- Utrecht University (UU), The Netherlands
- University of Hamburg (UHH) Germany
- University "Al.I.Cuza" of Iasi (UAIC) Romania
- University of Lisbon (FFCUL) Portugal
- Charles University Prague (CUP) Czech Republic
- Institute for Parallel Processing, Bulgarian Academy of Sciences (IPP-BAS) Bulgaria
- University of Tübingen (UTU) Germany
- Institute of Computer Science, Polish Academy of Sciences (ICS-PAS) Poland
- Zürich University of Applied Sciences Winterthur (ZHW) Switzerland
- University of Malta (UOM) Malta
- Open University (OU) United Kingdom
- International Relations and Security Network, Swiss Federal Institute of Technology (ETH) Switzerland

2.5 Co-ordinator contact details

Dr. Paola Monachesi

Utrecht University, Uil-OTS, Trans 10, 3512 JK Utrecht The Netherlands

Phone: +31-(0)30-2536065 Fax:+31-(0)30-2536000

Paola.Monachesi@let.uu.nl

<http://www.let.uu.nl/~Paola.Monachesi/personal>

3 Project execution: work performed to reach the objectives

In the rest of this section, we give a brief overview of the work which has been carried out in the project to reach the various objectives which we stated at the beginning of the project.

3.1 Integration of Language Technology resources and tools in eLearning

In order to reach this objective a corpus of learning objects of 1000 pages has been created for eight of the languages of the consortium in two areas: Computing and eLearning. The choice of the domains is motivated on the need for having a domain in which we have sufficient expertise within the consortium (Computer Science) and where we can easily find students and tutor for validation.

The documents selected for the corpus which were in different formats (i.e. pdf, doc, html) have been normalized and converted into XML (via an html step in order to retain layout information). The multilingual corpus has been annotated with Part of Speech tags automatically by means of available taggers. In addition, relevant keywords (1000) have been annotated manually for each language as well as definitions (300). They constitute our gold standard for the quantitative evaluations of the functionalities developed within the project (i.e. keyword extractor and glossary candidate detector). A keyword extractor has been developed which relies on different statistical measures that is TFIDF, Residual IDF, Adjusted RIDF and Averaged Reduced Frequency.

The manually annotated definitions have served as basis for the development of the language specific grammars which are being employed for the detection of the definitions within the learning objects. The grammars are employed in combination with the lxtransduce tool which allows for the matching of grammar rules with subtrees within XML-annotated documents. The tool has been developed by the University of Edinburgh and adapted for our purposes (<http://www.cogsci.ed.ac.uk/~richard/lxml2/lxtransduce-manual.html>).

Even though we employ standard NLP and information extraction techniques in the development of our tools, we couldn't simply apply standard evaluation methods because they would not be appropriate for our application. Therefore, we have developed an appropriate quantitative evaluation methodology that takes into account the eLearning application we are dealing with. The results of our evaluation are available in a dedicated report as part of our deliverables and show that our tools perform at the state of the art level.

We have taken pains at achieving a proper embedding of the tools in the learning process and to reflect the real needs of the users in this process. Specific scenarios have been developed in this respect to validate the tools within real learning tasks.

On the basis of the iterative quantitative verification of the Keyword Extractor and the Glossary Candidate Detector and validation of these tools in scenarios, we have improved the tools in response to evaluation and validation results obtained.

For the Keyword Extractor, we have improved the run-time performance of the tool, provided modifications to the linguistic models for each language, improved the processing and presentation of keyphrases and added more weight to keyphrases to give them more prominent places in the list of keyword candidates, non-distributional information has been employed for the weighting and ranking of keywords. Furthermore, we have implemented a voting mechanism which combines the various measures of keywordiness into one measure to improve results further.

As for the Glossary Candidate Detector, we have: divided the definitory contexts into types, established a baseline for monitoring the development of the tool and we have focussed on Machine Learning to improve precision of the developed tool. Experiments have been carried out for 4 languages, that is Dutch, Polish, Portuguese and in limited measure for English. We have experienced with the following settings:

- machine learning applied to only the most important and frequent types of definitions (Dutch, Portuguese) vs. machine learning applied to all types of definitions (Polish)
- machine learning applied in combination with local grammars, post-processing the results (Dutch, Portuguese) vs. machine learning applied as the sole method (Polish)

Several machine learning methods were used, among them, and with the most promising results, Naive Bayesian classifier and Balanced Random Forests. In general, the results have been very positive with a significant increase in precision but maintaining a good recall. Machine Learning filtering has been made available for use within ILIAS for Polish.

To summarize: in order to achieve this objective, which has resulted in the development of two modules: a) a keyword extractor; b) a glossary candidate detector, the following tasks have been carried out:

- created a corpus of 200.000 words (1000 pages) of learning objects for eight languages
- carried out a quantitative evaluation and profiling of our corpora
- normalized and converted the corpus in XML;
- annotated the multilingual corpus with PoS, lemmata and inflection features;
- developed a keyword extractor based on different statistical measures;
- carried out a three-step quantitative evaluation of the keyword extractor (for eight languages);
- integration of the keyword extractor in an LMS and embedding into learning processes which reflect users needs;
- carried out a two step scenario based evaluation of the keyword extractor;
- the evaluation and validation has provided feedback for an improvement of the tool, more specifically:
 - improvements in the run-time performance of the tool;
 - improvements in the linguistic models for each language which are used to filter the words which are inappropriate as keywords;
 - improvements in the processing and presentation of keyphrases;
 - use of non-distributional information into the weighting and ranking of keywords;
 - integration of different measures into a combined weight to improve performance;
 - implementation of a voting mechanism which combines the various measure of keywordiness into one measure;
 - added more weight to keyphrases to give them more prominent places in the list of keyword candidates;
- developed a glossary candidate detector based on:
 - already existing tools for xml transduction (Itransduce)
 - new language specific grammars for definition extraction, developed for eight languages, with Itransduce
- carried out a two-step quantitative evaluation of the glossary candidate detector (for eight languages);
- integration of the glossary candidate detector in an LMS and embedding into learning processes which reflect users needs;
- carried out a two step scenario based evaluation of the glossary candidate detector;
- the evaluation and validation has provided feedback for an improvement of the tool, more specifically:
 - classification of different types of definitory contexts;
 - identification of a baseline for monitoring the development of the tool;
 - use of machine learning techniques in a post-processing step to increase the precision of the tool.

The tools developed and their documentation can be downloaded from the multilingual web portal (<http://www.lt4el.eu>) while a description of the methodology adopted in their development can be found in the reports produced for the various deliverables which can be also found on the multilingual web portal.

3.2 Integration of semantic knowledge in eLearning

In order to achieve this objective, a methodology for ontology creation has been developed and adapted to our purposes. In particular, an ontology in the domain of Computing has been created. We have ensured

that the ontology covers the keywords manually annotated in the selected learning objects as well as the relevant general terms of the domain. The ontology has been linked to an upper ontology in order to ensure consistency with respect to the general ontology development methodology, as upper ontology DOLCE (<http://www.loa-cnr.it/DOLCE.html>) has been selected. In a subsequent phase, the upper ontology has been simplified by compacting the unary branches in the hierarchy

Furthermore, definitions have been collected for the concepts present in the ontology and a formalization of the definitions in OWL-DL has been carried out. In additions, concepts have been mapped to synsets in OntoWordNet, which is a version of WordNet 1.6 mapped to the DOLCE ontology. The mapping was performed via the two main relations equality and hypernymy. The resulting ontology contains 1002 domain concepts, 169 concepts from OntoWordNet and 105 concepts from DOLCE Ultralite. It also contains 107 object properties.

The ontology has been linked to domain specific lexica which have been developed for nine of the languages of the consortium. These lexica have been mapped to the concepts present in the ontology. Grammars have been developed for the semi-automatic concept annotation of the learning objects in the eight languages. Our multilingual corpus has been thus annotated. Word sense disambiguation has been performed manually. A formal comparison of ontology-based semantic search vs. textual search has been carried out, showing the advantages of semantic search. Furthermore, the ontology has been extended and revised on the basis of the feedback coming from the concept annotation and from the validation process resulting also in a revision of the lexica and the annotation grammars.

The ontology has been integrated into ILIAS. ILIAS users can thus carry out semantic search and ontology-browsing search both within one language and across languages. The integrated ontology and the cross-lingual search have been validated in two steps using scenarios.

This setup has been revised in several aspects to encode user feedback after validation. In particular, we have added the possibility to have and/or search, snippets of the documents in which the relevant concepts appear are extracted and presented to the users, a relevance score to the selected documents has been added, the possibility to combine various searches has been implemented while the speed of ranking has been improved.

To summarize, in order to achieve this objective, the following tasks have been carried out:

- developed an ontology in the area of computing of 1002 domain concepts + 169 concepts from OntoWordNet and 105 concepts from DOLCE Ultralite
- linked the domain ontology to OntoWordNet
- linked the domain ontology to the DOLCE upper ontology
- developed grammars for semi-automatic concept annotation of multilingual learning objects
- developed domain specific lexica in nine languages which have been linked to the ontology
- annotated language specific corpora on the basis of the ontology
- developed an ontology-based search engine and integration within ILIAS;
- formal evaluation of ontology-based semantic search vs. textual search;
- two steps qualitative evaluation of the semantic search on the basis of scenarios which has lead to:
 - a revision of the ontology, the lexica, the annotation grammars and the annotation of the learning objects;
 - revision of the ontology-based search engine which has been integrated within ILIAS. In particular:
 - possibility for AND/OR search
 - availability of snippet extraction
 - availability of relevance score
 - added possibility to combine various searches
 - improved speed of ranking

The tools developed and their documentation can be downloaded from the multilingual web portal (<http://www.lt4el.eu>) while a description of the methodology adopted in their development can be found in the reports produced for the various deliverables which can be also found on the multilingual web portal.

3.3 Supporting multilinguality

Multilinguality has played an important role in the development of the Language Technology based functionalities. Therefore, in the creation of the keyword extractor and the glossary candidate detector and obviously in the cross-lingual search, we have considered the multilingual dimension. The aim of the project was to develop the relevant tools for all the nine languages represented in the consortium Bulgarian, Czech, Dutch, English, German, Maltese, Polish, Portuguese and Romanian. This goal has been achieved except for Maltese. It was not possible to collect a corpus in this language for the lack of relevant documents in our domain. Similarly, the Language technology based functionalities could not be developed for lack of relevant resources which were not yet available for this language during the realization of the project. However, it was possible to include Maltese in our cross-lingual search since we were able to create a lexicon for this language.

More specifically, for the creation of the keyword extractor, corpora have been collected in eight languages and language models have been assembled for the various languages. Similarly, for the development of a glossary candidate detector, grammars have been developed for eight languages for the extraction of definitions from the learning objects. As for cross-language retrieval, the ontology was created considering terms emerging from all the languages represented in the consortium and lexica have been created for all languages.

For the validation of our functionalities, we have designed and refined (in two phases) multilingual scenarios addressing both tutors and students. In addition, the tools developed have been validated in a multilingual setting since trials have been carried out by all partners. Even though we were not able to show outright that the ability to retrieve in more than one language supports learning of students studying in a language other than their native language. We were, however, able to show that explicit inclusion of multi-lingual support for search developed in the LT4eL project increases the use of documents in a language other than the language of the course. Students will use the facility if it is offered. Since we established that locating sources is conducive to learning, it is reasonable to suppose this would also apply in a multi-lingual context. Qualitative results clearly feel that multi-lingual search is useful for students in their studies.

To summarize, in order to achieve this objective, the following tasks have been carried out:

- developed the relevant functionalities taking eight languages into account;
- developed language specific grammars for the detection of the definitions in the learning objects;
- developed the ontology starting from the extracted keywords in all languages;
- developed domain specific lexica in eight languages to be linked to the ontology;
- developed a cross-lingual search engine and integration within ILIAS;
- designed, instantiated and refined scenarios in two stages involving teachers and learners in search focused tasks in multi-lingual settings in order to assess impact of functionalities to fulfil users needs;
- carried out validation trials with learners and tutors in eight languages.

3.4 Development of enhanced LMS open source ILIAS Learning Management System

In order to reach this objective, the functionalities developed in the project, that is the keyword extractor, the glossary candidate detector, the semantic search and the ontology browsing both allowing for cross-lingual search have been integrated into ILIAS. In particular, it is now possible to generate keywords automatically and fill a metadata form related to a learning object, generate a glossary automatically on the basis of extracted definitions and link it to the related learning object. Various types of search have been integrated, that is definition search, full text search, keyword search, semantic search and search based on ontology browsing. The personal desktop has been revised to allow users to organize selected learning objects in sections and subsections. In addition, a linguistic processing chain has been implemented in order to allow for the linguistic annotation of new learning objects. Three versions of integrated ILIAS have been developed, the first one contained the integrated functionalities, the second one has included several revisions due to feedback from quantitative and qualitative evaluation and the third one is the final version which includes the revisions after the last validation phase. The system has extensively been debugged by all partners.

To summarize, in order to achieve this objective, the following tasks have been carried out:

- developed functionalities have been integrated into ILIAS, more specifically:
 - possibility to generate keywords automatically and fill a metadata form related to a learning

- object and revisions;
- possibility to generate a glossary automatically on the basis of extracted definitions and link it to the related learning object and revisions;
- implementation and revisions of definition search;
- implementation of full text search;
- implementation and revisions of keyword search;
- implementation and revisions of semantic search;
- implementation and revisions of search based on ontology browsing
- revisions of the personal desktop to organize selected learning objects in sections and subsections;
- linguistic processing chain has been implemented for Czech, Dutch, English, Polish, Portuguese and Romanian
- debugging of the system;
- development of three versions of ILIAS with integrated functionalities:
 - ILIAS with integrated functionalities;
 - revised version on the basis of feedback from quantitative and qualitative evaluation (first phase);
 - final version with modification due to quantitative and qualitative evaluation (second phase);
- installed the final version of ILIAS with integrated functionalities on a server of Utrecht University to ensure access after the end of the project.

The enhanced ILIAS and the relevant documentation with respect to the integration process as well as the ILIAS user manual can be downloaded from the multilingual web portal (<http://www.lt4eL.eu>) while a description of the methodology adopted in the integration process can be found in the reports produced for the various deliverables which can be also found on the multilingual web portal. The source code is also available under an open source licence and hosted it on the SourceForge.net portal for open source projects.

3.5 Validating Language Technology enhanced LMS

In order to achieve this objective, a validation methodology has been developed as well as scenarios that could take into account real learning situations. The scenarios have been localized on the basis of the partners' available content and have been used to run trials at all the various partners' locations. The feedback of the scenarios has been used to carry out modifications of enhanced ILIAS as well as to carry out improvements on the quality of the developed functionalities. The methodology and the scenarios have undergone revisions. In the last phase of the project, the hypotheses to be tested have been more clearly formulated. In addition, the validation has focussed on two scenarios (one for students and one for tutors) that have been developed on the basis of the formulated hypotheses. The scenarios have been localized by the various partners trying to neutralize the differences which could arise from the participation of different institutions and cohort used for testing. Suitable content has been identified and new content has been selected and adapted for the validation, including thus also revisions of the ontology and the lexica. The statistical significance of the cohort adopted has been investigated and concluded that the available number was in almost all cases statistically relevant. Results emerging from the trials have been analyzed and matched against the hypotheses formulated. The results indicate that LT4eL functionalities add value to aspects of learner and teacher support offered by the ILIAS learning environment, and over and above what is available from mainstream tools supporting internet search and information sources. In particular, with respect to learners, our experiments indicate that the added functionalities developed in the LT4eL project increase the retrievability of learning objects that are relevant to a specific task (that of answering a quiz), and that they increase the effectiveness of learners in locating such learning objects. Those who used LT4eL in correctly answering a question were also faster in giving the correct answer without relying on the additional functionalities. In addition, the results do show that tutors find that LT4eL tools make the retrieval of second language documents easier. Our results show that tutors using the keyword extractor produce more consistent keywords compared to ILIAS, and take less time to do so (although they do not assign fewer keywords). The experimental methodology we were asked to adopt carried high risk and involved a great number of variables for which we could not control. In spite of this several quantitative hypotheses were verified. Qualitative feedback from all users is uniformly positive.

To summarize, in order to achieve this objective, the following tasks have been carried out:

- developed a validation methodology to test the functionalities in ILIAS;
- identified quantitative and qualitative methods to test functionalities;
- designed, instantiated and refined seven scenarios involving teachers and learners in search focused tasks, in both mono- and multi-lingual settings (first cycle);
- carried out trials for 3 students scenarios with at least six students for each partners (i.e. 144 student feedback forms) (first cycle);
- carried out trials for 3 tutor scenarios with at least 3 tutors for each partners (i.e. 72 tutor feedback forms) (first cycle);
- developed feedback questionnaires related to the scenarios;
- localized scenarios for teachers and learners in eight languages on the basis of available content;
- analyzed results which have fed into the revised implementation of ILIAS platform, revised functionalities and revised scenarios;
- designed, instantiated and refined two scenarios involving teachers and learners in search focused tasks, in both mono- and multi-lingual settings;
- carried out trials for 1 student scenario with at least 24 students for each partner (i.e. 184 student feedback forms)(second cycle);
- carried out trials for 1 tutor scenario with at least 12 tutors for each partner (i.e. 84 tutor feedback forms) (second cycle).

A description of the validation methodology developed and an analysis of the validation results can be found in the reports produced for the various deliverables which can be downloaded from the multilingual web portal (<http://www.lt4el.eu>).

3.6 Knowledge transfer of the consortium

The LT4eL project didn't only have scientific objectives but also had societal objectives in fact it was one of the aims of the project to encourage the flow of knowledge among the members of the consortium especially between MS and NMS. In the project, we strongly encouraged a collaboration in the development of the functionalities and dedicated special budget for working visits. In particular, partners of the NMS have actively collaborated with the leaders of the core WPs (in case the leader was not from a NMS) to develop the relevant functionalities ensuring thus transfer of expertise from MS to to the NMS and vice versa. An additional side effect is that expertise wrt. relevant tasks has been available at at least two sites within the consortium preventing problems which might have been caused by the absence of one of the partners, a strategy which turned out to be especially valuable. Furthermore, we have stimulated, through the participation at relevant workshops and conferences the information flow outside the parties of the consortium strengthening thus the integration of the IST Research in Europe. We have aimed at organizing meetings at the NMS sites and usually organized public events to promote and disseminate the knowledge arising from the project. Young researchers working in the project have been particularly stimulated to participate in exchanges among partners and to attend conference and workshops. This has resulted in new Ph.D. and Master projects which have been financed by the relevant institutions.

To summarize, in order to reach the objective we have ensured that:

- partners of the NMS have actively collaborating with the leaders of the core WPs to develop the relevant functionalities ensuring thus transfer of expertise between MS and the NMS;
- stimulated, through the participation at relevant workshops and conferences, the information flow outside the parties of the consortium strengthening thus the integration of the IST Research in Europe;
- young researchers working in the project are particularly stimulated to participate in exchanges among partners and to attend conferences and workshops.

3.7 Awareness raising

The project has drawn attention, within the eLearning community to the significant potential of Language Technology and emerging technologies offered by the Semantic Web initiative. The approach developed in the project has been marketed within this community and eventually extended to other open source LMS. We have also shown to the Language Technology community that eLearning constitutes an interesting application as proven by the massive participation of the project at the Language Technology and Evaluation conference in 2008. Furthermore, we have addressed ministries of education in order to

stimulate support for the development of learning material in the national languages. We would like to organize an awareness event in the Netherlands to address ministries of education and higher education institutions.

4 Project results

4.1 LT4eL final results

We provide below an overview of the final results achieved by the LT4eL project. They are in accordance with what stated in the description of work, the only exception being that it was not possible to develop the Language Technology based functionalities for Maltese due to the limited resources for this language. However, it was possible to include Maltese among the languages that allow cross-language retrieval of learning objects within ILIAS. The resources and tools developed in the project and their documentation are freely available from the multilingual web page: <http://www.lt4el.eu> (under tools) as well as the deliverables describing their development (under results). In addition, they can be downloaded via the SourceForge.net portal for open source projects at <https://sourceforge.net/projects/lt4el/>. The new functionalities as well as the enhanced LMS are freely available under the GNU General Public Licence (GPL). Demo videos of the functionalities developed and their integration within the LMS can be viewed on the project website.

Overview of the final results:

- corpus of learning material of at least 1000 pages for eight languages: Bulgarian, Czech, Dutch, English, German, Polish, Portuguese and Romanian;
- development of a linguistic processing chain for linguistic annotation of learning objects for Czech, Dutch, English, Polish, Portuguese and Romanian.
- a key word extractor for semi-automatic selection of keywords within Learning Object Metadata. The tool is available to extract keywords in eight languages (i.e. Bulgarian, Czech, Dutch, English, German, Polish, Portuguese, Romanian) and can be used stand-alone or within the ILIAS Learning Management System.
- a glossary candidate detector for the extraction of definitions and the creation of glossaries to be attached to learning objects. The tool is available to extract definitions in eight languages (i.e. Bulgarian, Czech, Dutch, English, German, Polish, Portuguese, Romanian) and can be used stand-alone or within the ILIAS Learning Management System.
- language specific grammars for the extraction of definitions in eight languages (i.e. Bulgarian, Czech, Dutch, English, German, Polish, Portuguese, Romanian);
- an ontology with 1002 concepts in the domain of computing, 169 concepts from OntoWordNet and 105 concepts from DOLCE Ultralite. It also contains more than 100 object properties.
- language specific vocabularies which are linked to the ontology in nine languages (i.e. Bulgarian, Czech, Dutch, English, German, Maltese, Polish, Portuguese, Romanian);
- development of a search functionality which has been implemented as web service and which includes:
 - full text search;
 - keyword search;
 - definition search;
 - cross-lingual search engine based on the ontology and the lexica;
- integration of functionalities within a LMS (ILIAS is chosen in the context of this project)
- a methodology for a quantitative evaluation of developed functionalities in an eLearning application;
- methodology for a qualitative validation of the enhanced LMS in a multilingual setting.

4.2 Impact of the project on its industry or research sector

Significant research has been carried out in the area of Language Technology and the Semantic Web: the aim of the LT4eL project was to enhance eLearning with these technologies in order to develop innovative applications for education and training. Several initiatives have been launched in the area of Language Technology and within the Semantic Web vision both at the national and international level aiming at the development of resources and tools in the areas of Parsing, Tagging, Corpus Linguistics and ontology development. However, their integration in enhancing eLearning platforms has not been fully exploited yet. We believe that language resources and tools can be employed to facilitate tasks which are typically

performed in an LMS such as facilitating the search for learning material in a multilingual environment, semi-automatic metadata development and generating glossary items or definitions of key terms. Our project has been very innovative in this respect because even though research in the Language Technology area is developing fast, there are virtually no projects that try to integrate the results obtained within eLearning.

Our project has been very active in disseminating its results and has had an impact on the scientific community through its participation at scientific events and publication of scientific papers. In particular, we have participated at 45 scientific events, that is 19 conferences, 19 workshops, 2 summers schools, 1 summit, 1 symposium and 3 other type of events with either an oral presentation (52) or a poster (9) and we have organized 5 workshops and one seminar. In addition, we have written 38 papers.

The work done within our project is appreciated and considered to be innovative by the Technology Enhanced Learning (TEL) community. The fact that our paper has won the best scientific paper award at the EC-TEL 2007 conference, is already a strong indication that the work done within the LT4eL project is considered to be valuable by the TEL community. It was motivated on the basis of the innovative character of the functionalities developed in our project and the basis we are laying for further development within the eLearning community. We have taken pains to favour exploitation of our results in the TEL community. In particular, the results of the LT4eL project will be employed in a new EU strep project 'Language Technology for LifeLong learning' which will be carried out in collaboration with the Open University in The Netherlands, that is coordinator of the TENCompetence (<http://www.tencompetence.org/>) IP project. The new STREP project has started in March 2008. Three of the LT4eL partners participate in the new project (i.e. Utrecht University, University of Tuebingen and IPP-Bulgarian Academy of Sciences) which will combine the results and expertise gained within the LT4eL project, the TenCompetence project and the iCamp project. In particular, the new project will offer the possibility to explore the role of ontologies and networks in eLearning further, as well as the possibility to create ontologies semi-automatically on the basis of available techniques. In addition, it will be possible to explore the use of more lightweight knowledge representations such as MindMaps.

The tools and methodology developed within the LT4eL project can have an impact and can be exploited also in the context of other languages and cultures. We have investigated the impact that Language Technology can have for facilitating the integration among Mediterranean countries on the basis of eLearning. To this end contacts have been established in several Mediterranean countries which might lead to the creation of new collaborations and projects. In particular, initial contacts have been established with the responsible for the national eLearning center in Egypt (<http://www.nelc.edu.eg/>) to investigate the organization of eLearning standards for Egyptian universities. These are needed given the exponential growth of students and the increasing availability of good digital infrastructure. Similar contacts have been established with members of the Jordan Education Initiative (<http://www.jei.org.jo>) while the possibility to use our eLearning infrastructure in the Palestinian Admn. Area which would allow for easier access to education is being investigated.

In addition several contacts have been established or reinforced during the project to disseminate our results among other TEL (European) projects.

On the basis of the experience we have achieved so far, we do believe that eLearning can benefit significantly from Language Technology and we have taken pains at disseminating our results among both the Language Technology and the eLearning community to make them aware of the common benefits that could be gained from a closer collaboration. Our massive presence at the Language Technology and Evaluation (LREC 2008) conference shows that there is an interest in that community for eLearning as possible application and we have gained an excellent opportunity to market our results among the members of this well attended conference.

Within the Semantic Web area, the situation is slightly different with respect to the results obtained in the last year, specifically with respect to ontology development. Several projects are arising which try to exploit ontologies for eLearning purposes. We have established collaboration in this respect with European Projects within the Technology Enhanced eLearning area such as APOSDLE (<http://www.aposdle.tugraz.at/>), as well as with researchers who are carrying out projects very similar to ours. However, we believe that the potential of ontologies in the area of multilingual retrieval is not sufficiently exploited and there are still few enterprises in this direction. Our project remains groundbreaking

in this respect. In fact, it is the multilingual character of the project which is the most relevant for the state of the art also with respect to the results being achieved in the Language technology area. In addition, the resources developed within the project in the Semantic Web area such as the ontology and the lexica, are being distributed in this community through engines such as Watson and Ontoselect.

eLearning applications are very much an emerging field, and there are no standard, general methodologies that can be used to validate effectiveness of the learning process in our specific context. The validation methodology that we developed within the project which merges both quantitative and qualitative approaches allowed us to contribute significantly to the state of the art research in this area.

Within the LT4eL project, language technology based functionalities have been integrated into the ILIAS LMS and validated. A crucial issue is whether these functionalities will be offered as part of ILIAS in the immediate future. In order to accomplish this, we need to ensure a certain maturity of the functionalities that are included within the core system. Within the LT4eL project, it has been possible to create a prototype but integrating the tools within a stable ILIAS release would need improvements in various aspects. We can conclude that the LT4eL project proved that the developed functionalities can be useful for eLearning. To make them mature, will take another one to two years development and additional funding which are not available at the moment.

Another issue would be whether the functionalities we have developed could be commercialized. Our initial contacts we have had through the presentation of our results at conferences show that indeed the functionalities we have developed could be easily integrated within commercial products and would be very much welcomed. This is especially the case because we are concerned with languages such as those of NMS which have been less studied from a theoretical point of view but are of crucial importance for commercial applications. However, this will involve resources on the side of commercial institutions which are usually not available. More generally, it seems difficult to commercialize the results of this kind of European Projects. This is a difficulty that cannot be overcome even if an interested SME is part of the project. Usually the lack of funding is a big obstacle for the further development of the results obtained into a commercial enterprise. The SME cannot commercialize the product while it is being developed and after the project has finished it often doesn't have the funding to develop the prototype into a product.

We believe that a solution to this problem might be the creation of venture capital companies which could constitute a bridge between industry and research institutions. They would employ the results obtained in such projects in a way that they could be used by the commercial enterprise. To this end, we are exploring possibilities to create a spin-off of the University of Utrecht which could have this task. The spin-off should be in a position to translate the research results and the prototypes into a format that could be commercialized with little effort.

4.3 Lessons Learned

We believe that the LT4eL project has been quite successful since it has managed to keep in line with the objectives set up at the beginning of the project and to produce the expected results which have been disseminated extensively at various scientific events. In addition, the project has been a very positive experience for all its partners which has widened horizons and expertise. We have learned quite a lot from the challenge to work with so many languages, and it has been very interesting to be confronted with different techniques and methods. The exchange of researchers and the constant virtual meetings have contributed enormously to a close contact among teams and to a wide dissemination of knowledge between MS and NMS.

As expected, the development of the project helped to gather experience and knowledge that could not be available beforehand. When looking back at the whole project, we find certain aspects that could be addressed differently. Below a short overview of the lessons learned.

Corpora collection It would have been desirable to have learning objects of better quality. Our learning material was mainly composed of manuals and papers and in certain cases not really matching the standard definition of learning objects. This had some impact when having to run the scenarios in trials with learners and tutors and also on the quality of the keywords extracted. However, it is not trivial to find appropriate learning material which could satisfy our various requirements concerning the domain of expertise, the free property rights and multilinguality. Even though more and more open content is getting available, it is often

only in English and difficult to adapt to the specific needs of the project.

Evaluation of the functionalities This aspect has required more effort than initially expected. Even though keyword extraction and definition extraction are tasks widely carried out in the Natural Language Processing domain, their evaluation is still an issue. We have noticed that standard evaluation techniques were not appropriate given our eLearning application. It has been necessary to spend some resources to produce an appropriate evaluation methodology that could assess the performance of our functionalities in an eLearning context. On the positive side, we believe our work has contributed to the research in the area of evaluation given also the acceptance rate of our papers at the Language Resources and Evaluation conference (LREC2008).

Ontology development The experience related to the ontology building reveals that there is a lack of available resources. Many of the ontologies available on the web are in fact toy examples. There are few ontologies of reasonable size which are freely available. Similarly with respect to the tools related to their visualization which are often unsupported, without documentation, or with not all the functionalities available. We conclude that a better methodology for ontology creation is necessary. The methodology has to be supported by the corresponding tools. Clear documentation is also necessary.

Semantic annotation of texts The methodology developed for the semantic annotation of texts was rather successful and the ontology, lexicons and grammars produced are of good quality. They are relevant resources that could be very useful in applications related to semantic annotation. We plan to extend them to cover additional semantic features in the LTfLL project.

Conceptual representation We have realized that there is a gap between the human understanding of the conceptual information and what is represented in the ontology. The limitation of the current ontologies and the technologies related to them are still not well-understood by the average user. More work is necessary in order to allow a better exploitation of the conceptual information attested in the ontology by users. We will work in this direction in the LTfLL project.

Multilinguality We believe that the most important achievements of our project wrt. multilinguality are the use of the ontology as interlingua mediated by the lexicons and the methodology developed for a validation of the tools in a multilingual environment. On the other hand, the choice of the domain of 'computing' even though a familiar one for us to work with, didn't allow us to exploit the potential of the ontology in a multilingual setting fully. Future projects exploring the eLearning/multilinguality frontier should be wary of concentrating on learning domains where English tends to be used as a lingua franca or where there is a preponderance of English terminology.

Validation of the functionalities It would have been useful to validate the enhanced ILIAS in a real setting during a course with students and tutors being able to familiarize with the functionalities over time instead of in a very short session. A longer period of familiarization and use of the tools would perhaps demonstrate learning improvement more significantly. Users would be better able to select specific LT4eL functionality appropriate to search requirement. The subject domain may have been too far removed from the experience and interests of some students, but this was inevitable given the need to demonstrate learning. It would have been useful to carry out frequent small scale user testing throughout the project. This could provide useful feedback to the project and may also create a group of 'experienced' testers for the final experiments to compare against non-experienced. It might have been useful to have started the whole validation process half year earlier or to have another half year to dedicate to it. On the positive side, the design of pre-test, quiz and post-test worked well in terms of attempting to assess learning in a single test session. The Quiz method is a good way to encourage researching content and lends itself to various search methods. However, it was not realistic to compare searches against Google and other internet resources given the difference in content.

Integration of the functionalities in ILIAS The decision to use web services as a link between the language technology tools and the LMS has been proved to be a good one. The communication between the Java server and the PHP based LMS was stable and the development costs stayed at the estimated level. On the presentation side, the integration of the keyword extractor and the glossary candidate detector into the user interface of ILIAS was straightforward. This is due to the fact that they both deliver lists of items that are not very complex (keywords and term/definition pairs). Integrating the new search capabilities and the ontology into ILIAS was more complex. The first attempt to merge the ontology with the search results (including

found concepts) resulted in a user interface that was not expected by the users using a search functionality. Users are used to simple tools like Google. Summing up, the integration of the web services and the integration of the keyword extractor and the glossary candidate detector into a LMS user interface are both possible in affordable time. The integration of the search extensions and the ontology need more time and depend on the user interface guidelines of the target LMS. Open issues are the performance and the scalability of the language technology tools. If the tools should be usable in mid-size to large LMS environments, these aspects needs to be improved greatly. In addition, more attention should be given to the design of the interfaces to the functionalities.

4.4 Publishable results

For each exploitable result, this section indicates:

- Result description (product(s) envisaged, functional description, main advantages, innovations): a short description is provided in the table. Below the table, a more detailed description is given for each of the exploitable results.
- Possible market applications (sectors, type of use ..) or how they might be used in further research (including expected timings)
- Stage of development (laboratory prototype, demonstrator, industrial product...)
- Collaboration sought or offered (manufacturing agreement, financial support or investment, information exchange, training, consultancy, other)
- Collaborator details (type of partner sought and task to be performed)
- Intellectual property rights granted or published
- Contact details

More information on the tools can be also found on the LT4eL web page:
<http://www.lt4el.eu/index.php?content=tools>

Result description (see section 3.1 - 3.6 for description of results)	Possible applications	Stage of development	Collaboration sought or offered	Intellectual property rights granted or published	Contact details
Corpus of Learning Objects	Education, learning systems, distance teaching, eLearning	Prototype/demonstrator available for testing	Information exchange/Training, Available for consultancy	part of the corpus is IPR free	Dan Cristea, dcristea@info.uaic.ro, Paola Monachesi, Utrecht University, paola.monachesi@let.uu.nl
Keyword extractor	Research, Education, learning systems, distance teaching, eLearning	Prototype/demonstrator available for testing	Further research or development support, Laboratory prototype, Information exchange/Training, Available for consultancy	IPR / patent free	Lothar Lemnitzer, University of Tuebingen, lothar@sfs.uni-tuebingen.de, Paola Monachesi, Utrecht University, paola.monachesi@let.uu.nl
Glossary candidate extractor	Research, Education, learning systems, distance teaching, eLearning	Prototype/demonstrator available for testing	Further research or development support, Laboratory prototype, Information exchange/Training, Available for consultancy	IPR / patent free	Lothar Lemnitzer, University of Tuebingen, lothar@sfs.uni-tuebingen.de, Paola Monachesi, Utrecht University, paola.monachesi@let.uu.nl
Ontology & language specific vocabularies	Research & eLearning industry	Prototype/demonstrator available for testing	Further research or development support, Laboratory prototype, Information exchange/Training, Available for consultancy	IPR/patent free	Kiril Simov, Bulgarian Academy of Sciences, kivs@bultreebank.org, Paola Monachesi, Utrecht University, paola.monachesi@let.uu.nl

ILIAS integrated functionalities	Learning systems, distance teaching, eLearning	Prototype/demonstrator available for testing	Laboratory prototype, Information exchange/Training, Available for consultancy	open source software under the GNU General Public License	Alex Killing, Eidgenössische Technische Hochschule Zürich, alex.killing@gmx.de, Paola Monachesi, Utrecht University, paola.monachesi@let.uu.nl
Validation methodology	Learning systems, distance teaching, eLearning	Validation methodology used within the LT4eL project for the validation of the ILIAS LMS with integrated Language Technology functionalities	Information exchange/Training	IPR/patent free	Anne de Roeck, Open University, Milton Keynes, A.Deroeck@open.ac.uk, Paola Monachesi, Utrecht University, paola.monachesi@let.uu.nl

4.4.1 Corpus of Learning Objects

The LT4eL corpus consists of more than 200,000 words (1,000 pages) of learning objects for each language represented in the consortium. The broad domain of this collection is computing. More specific subdomains include teaching academic skills, creating webpages, basic computer skills. All material has been linguistically annotated at least up to the level of part-of-speech tagging and morphological analysis. Besides, it has been marked up with (i) key words, (ii) definitory contexts and (iii) ontological annotation of concepts.

More information on the LT4eL project can be found at: <http://www.lt4el.eu>.

4.4.2 Keyword Extractor

To improve the retrieval and accessibility of content through the identification of the learning material by means of descriptive metadata, available language technology tools and resources have been employed within the LT4eL project to develop a keyword extractor that facilitates the semi-automatic generation of metadata. The keyword extractor is a tool that supports authors and content managers in selecting, in the chosen learning objects, the keywords that best represent the topic(s) of these learning objects. The tool analyses a set of annotated documents and returns the best keyword candidates for each learning object. The user of this functionality decides on the inclusion of these candidates into the metadata. The tool has been tested on eight languages (i.e. Bulgarian, Czech, Dutch, English, German, Polish, Portuguese, Romanian) and it can be used stand-alone or within the ILIAS Learning Management System (<http://sourceforge.net/projects/lt4el/>).

The tool has been fully documented. The documentation is essential for a proper integration within each LMS that will be interested in adopting it.

More information on the LT4eL project can be found at: <http://www.lt4el.eu/index.php>. For information on the keyword extractor, check the tools section on the project web page: <http://www.lt4el.eu/index.php?content=tools>.

4.4.3 Glossary Candidate Detector

Dictionaries constitute an important support to the learning process. One of the assumptions of the LT4eL project is that they can be especially useful if they are constructed on the basis of the definitions extracted from a given learning object. To this end, we have employed available language technology tools and resources to develop a glossary candidate detector.

The glossary candidate detector is a tool that supports authors and content managers in creating glossaries by identifying definitory contexts in a text that contains the term to be defined and its definition. The input of this tool are annotated texts (and background resources). The output of this tool is a set of candidates for definitions that are extracted by means of pattern based grammars. The tool has been developed for eight languages (i.e. Bulgarian, Czech, Dutch, English, German, Polish, Portuguese, Romanian) and it can be used stand-alone or within the ILIAS Learning Management System.

The tool has been fully documented. The documentation is essential not for a proper integration within each LMS that will be interested in adopting it.

More information on the LT4eL project can be found at: <http://www.lt4el.eu/index.php>. For information on the glossary candidate extractor, check the tools section on the project web page: <http://www.lt4el.eu/index.php?content=tools>.

4.4.4 Ontology & language specific vocabularies

The domain specific ontology on computing is developed in a language independent way and comprises more than 1000 concepts. An English vocabulary is mapped to the concepts and the relations within the ontology. In addition, language specific domain vocabularies have been developed and linked to the ontology for eight languages (Bulgarian, Czech, Dutch, German, Maltese, Polish, Portuguese and Romanian). The ontology contains 1002 domain concepts, 169 concepts from OntoWordNet and 105 concepts from DOLCE Ultralite. It also contains more than 100 object properties.

The ontology has been used to structure, query and navigate through the learning objects that are part of a Learning Management System. The ontology can play two major roles:

- Classification of learning objects. Each learning object is connected to a set of concepts in the ontology. This classification allows ontological search, i.e. search based on concepts and their interrelations within the ontology.
- Multilingual search for learning objects. In this case the ontology plays the role of Interlingua between the different languages. Thus the user might specify the query in one language and get learning objects in other language(s).

The innovative aspects of the use of the ontology consist firstly in the application of semantic web technologies (ontologies) to facilitate learning processes and, secondly, in their use (linked to language specific vocabularies) to address problems of multilingual nature, in particular multilingual search.

The ontology in combination with the lexica and the learning objects have been used for ontology browsing and semantic search. The software developed to this end has been fully documented. The documentation is essential for a proper integration within each LMS that will be interested in adopting it.

More information on the LT4eL project can be found at: <http://www.lt4el.eu/index.php>. For information on the ontology and the language specific vocabularies, check the tools section on the project web page: <http://www.lt4el.eu/index.php?content=tools>.

4.4.5 ILIAS integrated functionalities

Within the LT4eL project, three tools have been integrated into the LMS ILIAS (<http://www.ilias.de>):

- keyword extractor;
- glossary candidate detector;
- ontologies for multilingual retrieval.

ILIAS is a web-based learning management system and allows users to create, edit and publish learning and teaching material in an integrated system with a normal web browser. Tools for cooperative working and communication are included as well. ILIAS is available as open source software under the GNU General Public License. The software development worldwide is coordinated by the team at the ETHZ.

The last release (3.5.0) of ILIAS before the project already offered content authors the possibility of annotating learning objects with metadata based on the LOM standard. Also a metadata based search is available. However, the metadata had to be provided manually by the author. This made the annotation process very time consuming and thus only a few authors provide useful metadata. Tools for semi-automatic metadata generation could help to solve this problem. Furthermore, ILIAS does not provide semantic web based functionalities, but it already offers the possibility of reusing learning objects like media objects or glossary items in the creation process of learning material. Ontology based retrieval of learning objects will considerably improve the task of reusing learning objects since ontologies will allow for intelligent searching and navigation in huge amounts of data. Metadata annotation and ontology driven search and navigation allow for individual content assemblance for learners. Learners will be able to build individual learning paths by entering key terms of concepts they need to learn.

ILIAS with integrated LT4eL functionalities and documentation, can be downloaded from SourceForge:
<http://sourceforge.net/projects/lt4el/>.

More information on the LT4eL project can be found at: <http://www.lt4el.eu/index.php>. For additional information on ILIAS with integrated LT4eL functionalities, check the tools section on the project web page: <http://www.lt4el.eu/index.php?content=tools>.

4.4.6 Validation methodology

As part of the LT4eL project, a suitable validation methodology has been developed and applied to the validation of the functionalities developed within the project as well as to their integrated set into ILIAS. Within the LT4eL project, the following three tools have been integrated into the LMS ILIAS (<http://www.ilias.de>):

- keyword extractor;
- glossary candidate detector;
- ontologies for multilingual retrieval.

eLearning applications are very much an emerging field, and there are no standard, general methodologies that can be used to validate effectiveness of the learning process in our specific context. The methodology developed within the LT4eL project can be at least a first step towards this missing standard.

More information on the LT4eL project can be found at: <http://www.lt4el.eu/index.php>. For additional information on the validation methodology, check the tools section on the project web page: <http://www.lt4el.eu/index.php?content=tools>.