



Project no. 027391

Project acronym: LT4eL

Project title: Language Technology for eLearning

Instrument Specific Targeted Research Project

Thematic Priority Information Society Technology

### **D6.2b Exploitation plan and report**

Due date of deliverable: 30-11-2007

Actual submission date: 21-12-2007

Start date of project: 1-12-2005

Duration: 30 Months

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

Revision [1]

<b>Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)</b>		
<b>Dissemination Level</b>		
<b>PU</b>	Public	x
<b>PP</b>	Restricted to other programme participants (including the Commission Services)	
<b>RE</b>	Restricted to a group specified by the consortium (including the Commission Services)	
<b>CO</b>	Confidential, only for members of the consortium (including the Commission Services)	

# D6.2 - Exploitation plan and report M24

## Outline

### Contents

- 1 Executive Summary
- 2 Introduction
  - 2.1 Exploitable results
  - 2.2 Potential users
  - 2.3 Main exploitation activities
  - 2.4 IPR issues
  - 2.5 Division of tasks
- 3 Exploitable results
  - 3.1 Know-how
  - 3.2 Corpus of learning objects
    - 3.2.1 Description
    - 3.2.2 Use and exploitation
  - 3.3 Tools for metadata generation: keyword extraction, glossary candidate detector
    - 3.3.1 Description
    - 3.3.2 Use and exploitation
  - 3.4 Domain specific ontology and language specific vocabularies
    - 3.4.1 Description
    - 3.4.2 Use and exploitation
  - 3.5 ILIAS integrated functionalities
    - 3.5.1 Description
    - 3.5.2 Exploitation and Use
  - 3.6 Validation methodology
    - 3.6.1 Description and exploitation
- 4 Exploitation plan
  - 4.1 eLearning, LT, NLP and SW community
  - 4.2 The LT4eL project and the Technology Enhanced Learning (TEL) community
  - 4.3 Commercial organizations
  - 4.4 Higher education institutions
  - 4.5 The user panel
  - 4.6 Contacts with other (EU) projects which are potential users of our results
  - 4.7 New projects
  - 4.8 PhD projects and Master theses
- 5 Overview table

## 1 Executive Summary

According to the annex, the aim of work package 6 is to present and disseminate results of the project, to foster the transfer of knowledge and a broad take-up of programme results. The following groups will be the special targets:

- end users at universities: learners and teachers
- scientific community

- national educational authorities

The exploitation plan has to be delivered in M12 (D6.2), M24 (D6.2b) and M30 (D6.2c). This report is the M24 exploitation plan (D6.2b). According to the revised list of deliverables, the M24 exploitation plan and report has to contain:

- an exploitation plan and updates;
- a usability study of the project developments in TEL field;
- information on the user panel;
- future planning activities.

The goal of the report is to provide a plan and a report of the way the resources developed are and can be exploited.

The purpose of deliverable D6.2b is to present the exploitation plan of the LT4eL project at the end of the second year of the project (M24). More specifically, D6.2b describes the consortium's strategy for exploiting project outputs during and after the second year of the project. The purpose of this document is to describe how the results of the projects will be used or exploited by the LT4eL partners and by other potential users during the remaining months of the project as well as after the project. The document is organized as follows:

- The first section identifies the exploitable results, the potential users, and the main exploitation activities.
- The second section contains a description **per exploitable result** on how it can be exploited or used in further research.
- The third section contains a more detailed description of the exploitation channels we use for the exploitation of the results described in the second part.

## 2 Introduction

This introductory section identifies and briefly describes the exploitable results, the potential users, the main exploitation activities and IPR protection measures.

### 2.1 Exploitable results

- Know-how in Language Technology (LT), Natural Language Processing (NLP), Semantic Web and eLearning technologies;
- Corpus of learning material of at least 1000 pages for eight of the languages represented in the consortium, except for Maltese (WP1);
- Tools (methodology) for metadata generation (key word extractor, glossary candidate detector) (WP2);
- Developed domain specific ontology with at least 1000 concepts, language specific domain vocabularies, ontology-based search engine and semantic search engine (WP3);
- ILIAS integrated functionalities (keyword extractor, glossary candidate detector and multilingual retrieval via ontologies) (WP4);
- Validation methodology (WP5).

### 2.2 Potential users

- LT4eL partners who will use the collected knowledge and resources for their research activities and the developed functionality integrated in the ILIAS system for their didactic activities.
- Scientific community (LT, NLP, Semantic Web and eLearning), including other ICT projects, who will use the collected resources, the developed concepts and methodology for their research activities.

- LMS developers who will use the developed concepts and methodology to integrate new functionalities within their systems.
- LMS end-users who will use the new functionalities integrated in the ILIAS or other systems.
- Local authorities (e.g. ministries of education), and higher education institutions (e.g. universities) who will become aware of the potential of eLearning.

## 2.3 Main exploitation activities

The project results will be marketed in different communities and in different ways.

Communities:

- eLearning community
- LT community
- NLP community
- SW community
- Commercial institutions
- Higher education institutions
- EC-TEL community
- Contacts with other projects
- PhD and MA projects

A more detailed description of these communities and the ways in which we use them, can be found in section 4 of this deliverable.

Dissemination of the results has been and will be ensured in different ways:

- through the multilingual Web portal ([www.lt4el.eu](http://www.lt4el.eu));
- by participation at scientific events, notably eLearning, LT, NLP and Semantic Web conferences and workshops;
- by means of reports;
- by means of mailing lists;
- via the user panel.

In order to promote the use of the LT4eL functionalities, and in general of open source Learning Management Systems (LMS), we will establish contacts with local educational authorities (ministries of educations, universities and schools) in particular in the new member states, for example via the organization of special meetings or awareness events on a national level.

The technical results of the project will be made available in well-documented releases. The new functionalities as well as the enhanced LMS will be freely available under the GNU General Public Licence (GPL) and distributed also through the ILIAS channels, as well as through the “Language GRID” initiative in which the University of Tübingen participates.

For a more detailed description of our planned dissemination activities see deliverable D6.1b.

## 2.4 IPR issues

All the IPRs (Intellectual Property Rights) concerning publications and other know-how resulting from the project will be handled in agreement with all partners. A budget is foreseen in case it is necessary to cover IPR for the learning objects collected. The IPR status of these documents has been investigated in detail. We have assessed that material used for validation is in general IPR free, but this is not the case for the whole corpus. However, we are coming across more sources of archives for sharing content thus we might consider add new material to our existing corpus given that the annotation

process is rather straightforward at this point. It is the task of the WP-managers that all issues related to the intellectual property will be respected. Basically our project will deal only with freely available resource and tools. However, sources and authors of all resources, tools and platforms will be mentioned in the web inventories and in the reports. The project deliverables will be publicly available. Reports of the consortium management boards will be available only for the project members.

## **2.5 Division of tasks**

Exploitation activities are part of WP6 with Utrecht University as WP leader. All partners however are committed to the widespread dissemination and active exploitation of the results of the project.

## **3 Exploitable results**

This section contains a list of all exploitable results delivered by the LT4eL project. For each of the results, a description of the result and a plan for its use or exploitation are included.

### **3.1 Know-how**

Through the LT4eL project, each partner acquires knowledge and skills in state-of-the-art technologies in LT, NLP, Semantic Web and eLearning. It is only natural for all partners to exploit this know-how for their future research and for the acquisition and execution of future projects.

### **3.2 Corpus of learning objects**

#### **3.2.1 Description**

This corpus consists of at least 200000 words (1000 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. The complete annotated corpus has been delivered at month 12 (December 2006). During the second year of the project, an extra annotation layer was added to the objects with ontological annotation of concepts. The semantically annotated objects are also stored in the corpus.

#### **3.2.2 Use and exploitation**

The relevance of this corpus lies essentially in its use for the development of the new functionalities within the project. However, a subcorpus can be created with IPR free documents in order to make it publicly available in the ILIAS library, but also on the project portal. As potential users we identify ILIAS end-users (teachers and students) interested in the content of the documents, but also the whole scientific community, notably other eLearning projects like iCamp (<http://www.icamp-project.org/>), as well as NLP and SW projects. The latter can use the multilingual corpus - which, if necessary, could be easily annotated with other levels of annotation - as a data base for further experimental research, for example, in language technology and natural language processing. Knowledge about the availability of this corpus will be disseminated via the usual channels (project webportal, user panel, presentations at scientific events and mailing lists such as the corpora list).

### **3.3 Tools for metadata generation: keyword extraction, glossary**

## **candidate detector**

### **3.3.1 Description**

The aim of the project is to improve the retrieval and accessibility of content through the identification of the learning material by means of descriptive metadata. To this end, we have employed available language technology tools and resources to develop two functionalities that facilitate the semi-automatic generation of metadata. The two functionalities are:

- A keyword detector and extractor, i.e. 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.
- A glossary candidate detector, i.e. 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. Again the input of this tool are annotated texts (and background resources). The output of this tool is a set of candidates for definitions. Again the user decides on the inclusion of these candidates into the metadata. The tool has been tested on eight languages.

We assume that the addition of these functionalities will improve the retrieval of the learning objects in terms of their content. First validation results point at that direction. The tools have been fully documented. The documentation is essential not only for a proper integration within the ILIAS system, but also for the integration within other LMSs that will be interested in adopting them. The second version of the tools is delivered at month 24 (December 2007). The final version will be delivered at month 30 (May 2008). Documentation can be found on the lasi portal ([http://consilr.info.uaic.ro/uploads\\_lt4el/](http://consilr.info.uaic.ro/uploads_lt4el/)).

### **3.3.2 Use and exploitation**

The developed functionalities have been integrated into the Learning Management System ILIAS. To favor further exploitation, they are offered independently as web services so that they can be used on the internet by other applications as well. Full integration into other LMSs is thus also possible. To foster implementation into other systems, a document describing the integration process into ILIAS (a step-by-step integration manual) has been provided targeted at developers of other LMSs, and a workshop on the integration process will be organized in 2008. The consortium further plans to participate to eLearning or other LMS-related exhibitions to demonstrate the developed tools and methodology as well as to create direct contact with other LMS developers. Beyond direct exploitation in the eLearning industry, we foresee for these results an impact on the scientific community. These functionalities use state-of-the-art LT and NLP technologies and therefore can provide useful feedbacks to these communities and be the starting point for utilization of these technologies into other domains. To further promote the use of these tools in eLearning and other fields, three workshops have been organized in this reporting period, cf. deliverable 6.1b).

## **3.4 Domain specific ontology and language specific vocabularies**

### **3.4.1 Description**

The domain specific ontology (i.e. computing) is developed in a language independent way and comprises round 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 all the languages of the consortium. The domain specific ontology was delivered at month 12 (December 2006), and the language specific vocabularies at month 18 (May 2007). In month 24 we deliver a new version of the ontology which contains additional 200 concepts and related lexica.

The ontology is being used to structure, query and navigate through the learning objects that are part of a LMS. 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 this part of the project 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.

### **3.4.2 Use and exploitation**

From the perspective of our project, the relevance of these semantic web technologies lies essentially in their use within an LMS to improve the organization and the retrieval of the learning objects across languages. Potential users in this respect will be LMS (notably ILIAS) end-users, but mostly LMS developers, open source as well as commercial.

Documentation of the ontology development and a report for its integration within an LMS has been provided to facilitate its utilization in any eLearning platform. However, again, we expect also a scientific impact of these results. The use of ontologies for multilingual retrieval is in its infancy. We expect to contribute to the development of these techniques (to this aim we are in contact with the APOSDLE (<http://www.aposdle.tugraz.at/>) project to use common resources) and to provide important feedbacks to the semantic web community on the potential of their technology in this field. Further contacts have been established with OU-UK and the KMI groups as well as the Laboratory for Applied Ontology in Rome. To further promote the use of semantic knowledge in eLearning we organized a workshop on this topic at the EUROLAN 2007 which has been held in Romania in the summer 2007, and we have been and will be present at various Semantic web and eLearning events. In addition this issue has been the topic of the other two workshops we have organized this year. The developed ontologies, and in particular the methodology to link these to language specific vocabularies, will be further disseminated via the usual channels to be used by the scientific and semantic web community for further research. We have already written several papers on this topic which have been presented at conferences (cf. Deliverable 6.1b). We believe that the use of ontologies in eLearning is one of the most innovative aspect of our project and very much appreciated as appears from the first validation results as well as the peer review comments that we receive for submitted papers on the subject. It was particularly this aspect that was found innovative and a motivation for the best scientific paper award we have received at EC-TEL 2007.

## **3.5 ILIAS integrated functionalities**

### **3.5.1 Description**

The tools described above (keyword extractor, glossary candidate detector and ontologies for multilingual retrieval) have been implemented within the ILIAS system. 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.

### **3.5.2 Exploitation and Use**

The main users of these integrated functionalities are ILIAS end-users (teachers and students). Use cases are described in the WP2 and WP3 deliverables they are related to the tasks performed within ILIAS that use the tools provided by the LT4eL project. The scenarios produced within WP5 have also been developed having these users in mind, cf. deliverable 5.1b for further details.

To promote the use of the newly developed functionalities, the ILIAS English user manual will be extended with detailed descriptions of these functionalities. The user documentation will be distributed online via the ILIAS website as well as via the project website. To distribute ILIAS with the newly integrated functionalities, the project is also maintained at two open source communities, namely SourceForge.net and Ohloh.net. Clearly the functionalities could be integrated in other systems and applications since they have been implemented as web services.

## **3.6 Validation methodology**

### **3.6.1 Description and exploitation**

A suitable validation methodology has been developed and applied to the validation of the new functionalities as well as to their integrated set into ILIAS. 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. We expect the methodology developed within the LT4eL project to be at least a first step towards this missing standard. To promote dissemination, and thus exploitation, papers on the subject will be written and presented at conference and the current network of eLearning project which we have built will be employed to exchange results and best practice.

## **4 Exploitation plan**

As can be seen from the previous section, the LT4eL project has delivered and will deliver different types of results, ranging from tools to scenario descriptions. For each of the individual results, some ways in which they can be exploited in future are already shown in the previous section. The first section mentioned the exploitation channels that are used for the exploitation of the project results (e.g. research institutions and commercial institutions). In this section we provide a more detailed description of the channels via which the project results are exploited and the communities we address.

For each of the channels and communities, a description of (i) the channel and (ii) the way

in which it is or will be used for exploitation are provided in this section. For the communities, we provide a description of the way in which we foresee the exploitation of our results within this community and the activities we have undertaken already.

#### **4.1 eLearning, LT, NLP and SW community**

The functionalities developed within the project have been integrated in the ILIAS Learning Management System (LMS). We will however market our approach with other open source or commercial systems as well, such as Moodle, OLAT, ML3, ATutor, Sakai, Claroline, Blackboard, WebCT. To this aim a research about content formats and metadata support used in other LMSs has been already carried out and published in the project internal webpage. To create interest of potential users and prospects, as well as direct contact with other LMS developers, the consortium participated to various eLearning events or other LMS-related exhibitions and demonstrated the developed tools and methodology.

To further favor exploitation, the developed functionalities will be also offered independently as web services, so that they can be used on the internet by other applications as well. Besides, the project decided to make the source code available under an open source licence and to host it on the SourceForge.net portal (<https://sourceforge.net/projects/lt4el/>) and the Ohloh.net portal (<http://www.ohloh.net/projects/9785?p=LT4eL>). This makes the results immediately available to the general public and gives everyone the opportunity to join and collaborate with the project.

OU, ZHW and the ILIAS team act as link with the eLearning community. For example, the ILIAS team is part of a German network of eLearning developers called Campussource (<http://www.campussource.de/org/>), and the eLearning Centre of the ZHW has direct contact with Moodle, OLAT and ML3. Both channels will be used as starting point to establish collaborations.

To exploit our results further into this community, we have planned several activities:

- As soon as we have our first official release (M24) packaged and available, ETHZ would like to feature it in the LMS communities (forums) and send e-mails to leading LMS developers (such as Moodle, Olat, WebCT).
- The new functionalities will be explained and advertised in the next ILIAS newsletter.
- At the Langtech 2008 conference (<http://www.lang-tech.org/>), dissemination material on the project will be distributed. Besides, a demo will be given of ILIAS with the new functionalities developed within LT4eL.
- We are going to cooperate more closely with the SHARE project ([share.uni-koeln.de](http://share.uni-koeln.de)). Within the SHARE project, they seek to promote several freely available tools that teachers can use for producing, sharing and re-using learning content and metadata issues. On the website of this project, a description will be added of our project and the tools we developed.
- We will present a paper on the IMCL conference. The IMCL conference is part of an international initiative to promote technology-enhanced learning and online engineering world-wide.

#### **4.2 The LT4eL project and the Technology Enhanced Learning (TEL) community**

Our project results have been presented at several conferences, including eLearning conferences. The work done within our project is appreciated and considered to be innovative by the eLearning community. The fact that our paper on the EC-TEL 2007 conference won the best paper award, is already a strong indication that the work done within the LT4eL project is considered to be valuable by the TEL community. EC-TEL stands for 'European Conference on Technology Enhanced Learning'. The subtitle of the 2007 conference was 'Creating new learning experiences on a global scale'. The conference

provides "a unique forum for all research related to technology-enhanced learning, including its interactions with knowledge management, business processes and work environments. This is a competitive and broad forum for technology enhanced learning research in Europe and world-wide through specialized workshops and the main conference. EC-TEL 2007 provides unique networking possibilities for participating researchers throughout the week and includes project meetings and discussions for ongoing and new research activities supported by the European Commission" ([1] (<http://www.ectel07.org/>)).

### **4.3 Commercial organizations**

- UU has further established contacts with Giunti Labs who is monitoring the project in order to assess potentiality for commercialization of the functionalities developed.

### **4.4 Higher education institutions**

The project is now ripe to organize meetings with higher education institutions and university to show the potential of eLearning. Attempts have been made in this direction at Utrecht University but it is rather impossible to have an impact on the commercial policy with respect to eLearning tools that Universities have adopted against Open Source products. Nevertheless, some more attempts in this direction will be made in the last phase of the project.

### **4.5 The user panel**

The user panel consists of representatives of the different fields addressed within the LT4eL project. We aimed at an equally representation of eLearning, Semantic Web and Language Technology experts. Besides, the panel contains both representatives from academic and commercial institutions. The list with members of the panel can be found in the dissemination report.

The panel has been informed on the activities and results of the project on a regular basis the last year. Besides, some of them were in June present at the workshop in Prague organized by the project. They provided us with feedback and ideas on the basis of their expertise in their research areas. The results of the project can be exploited in different ways by the users of the panel. Whereas the Language Technology experts are more interested in the technologies used within the keyword extractor, the commercial representatives are interested in the use of the tools within their own products.

### **4.6 Contacts with other (EU) projects which are potential users of our results**

The It4eL project is committed to liaising with other EU projects, particularly those engaged in the same subject area, as well as with other projects and partners which are relevant for the activities carried out in the project.

We have had collaborations and established new collaborations with the following projects:

- I\*teach (<http://i-teach.fmi.uni-sofia.bg/>). We have started adding documents produced by the I\*teach project to our corpus. These documents are particularly suitable for our purposes because they constitute a mini-corpus in a domain which is very close to ours (use of computers in education) of material translated in almost all languages of our consortium (mainly: Bulgarian, Dutch, English, German, Polish, Romanian). In the second year, we added more documents produced by the I\*teach project to our corpus.
- ELENA (<http://www.elena-project.org/en/>). We have exchanged research results on metadata extraction with the ELENA project (finished in May 2005).

- CALIMERA (<http://www.calimera.org>). We have used the documents collected within this European project to assemble our corpus of learning objects.
- PROLEARN (<http://www.prolearn-project.org/>). All LT4eL electronic meetings are organized using their developed application Flashmeeting (<http://flashmeeting.open.ac.uk/>).
- iCamp (<http://www.htk.tlu.ee/icamp/>). The iCamp project is planning to use our corpus of LOs for their research.
- APOSDLE (<http://www.aposdle.tugraz.at/>). We collaborate with the APOSDLE project with a constant exchange of information and results on the common issue of the use of ontologies in eLearning.
- BAZAAR (<http://www.bazaar.org/>). We plan to use the Bazaar community to disseminate our project results. Bazaar supports a community for teachers and trainers by providing information and services for teachers and trainers in the development, creation, exchange and use of e-learning materials.
- SHARE (<http://www.share.uni-koeln.de/>). We plan to get in touch with this project through Matthias Kunkel since they might be interested in our tools for metadata generation.

More in particular, contacts with the following projects/individuals have been established or reinforced:

- A visit has been organized by Paola Monachesi to AUP in Paris and Claudia Roda (member of the user group) in combination with a visit to Luxembourg. Claudia Roda is a partner in the FP6 AtGentive eLearning project (<http://www.atgentive.com/>) in order to discuss the role that language technology can play in determining the attention of learners. The main aim was to assess possibility to exploit LT4eL results in this context. In addition, feedback to the LT4eL project has been given especially with respect to validation methodology.
- Paola Monachesi has established contacts with the Open University in The Netherlands (Peter van Rosmalen, Jan van Bruggen), who is coordinator of the TENCompetence (<http://www.tencompetence.org/>) IP project aiming to create an infrastructure to support individuals, groups and organisations in lifelong competence development. Also in this case, the aim was to investigate the role of language technology in relation to Life Long learning services, this has lead to a new collaboration and a new STREP project which has been accepted within FP7, called Language Technology for LifeLong Learning (LTfLL) in which results of the LT4eL projects will be employed. Three LT4eL partners will participate in the new project, that is Utrecht University, University of Tuebingen and Bulgarian Academy of Science. In addition, exchange of methodology with respect to validation activities in eLearning context has been performed.
- Exchange of results and informal collaboration has been established by Paola Monachesi with the Open University UK (KMI group) in particular, Enrico Motta, Mathieu d' Aquin (member of the user panel) and Marta Sabou. The aim was to investigate the use of Semantic Web techniques and tools, especially Watson which is an ontology search engine, to improve the developement of domain ontologies. Mathieu d'Aquin was invited to present results in this respect at the workshop we have organized in Prague while discussions have been carried out with Enrico Motta during the ASOS workshop, that we have organized at the EUROLAN summer school in order to exchange project results.
- A visit at CNR Rome - Laboratory of Applied Ontology (Aldo Gangemi, Alfio Gliozzo - user panel member) has been organized by Paola Monachesi in combination with the Malta meeting, in order to discuss issues related to ontology development, especially with respect to DOLCE, the use of ontologies in eLearning and possibilities for project

results exchange. We have also assessed that there are common interests wrt. the exploitation of semantic knowledge in eLearning, especially wrt. activities aiming at integration of the Mediterranean Countries through eLearning.

- The dissemination meeting organized during the Malta meeting was also particularly relevant to assess the expertise of University of Malta in the field of eLearning and mobile learning. Contacts have been established with an SME in Malta (Across Limits - Angele Giuliano) in the context of re-use of project results wrt. activities aiming at integration of the Mediterranean Countries through eLearning. AcrossLimits is partner in another FP6 eLearning project, that is Arise (<http://www.arise-project.org/>)
- Results have been exchanged and contacts have been reinforced with the FP6 APOSDLE eLearning project (<http://www.aposdle.org/>) in the area of keyword extraction. Viktoria Pammer and Stefanie Lindstaed have given a talk in this respect at the workshop we have organized in Prague.
- We have reinforced our collaboration with Paul Buitelaar (member of the user panel) who was invited to the Prague workshop and results have been exchanged between the LT4eL project and the SmartWeb project (<http://www.smartweb-projekt.de/>) in the area of ontology development.
- Iryna Gurevich (member of the user panel) was invited at our RANLP workshop and she has provided useful feedback to the LT4eL project especially in the area of semantic knowledge for eLearning and ontology evaluation. We plan to exchange results in the context of her semantic information retrieval project (<http://www.ukp.tu-darmstadt.de/projects/sir>).
- In the context of exploitation and commercialization of the LT4eL results, further contacts have been established by Paola Monachesi with Giunti Labs and Fabrizio Giorgini during the EC-TEL conference in Crete. They will monitor the development of the project. This is also the case for Tjerk van Dijk (ProVer) who has been developing a multilingual semantic based search engine for the province Friesland and could benefit from our results.
- The ILIAS team is part of a German network of eLearning developers called Campussource (<http://www.campussource.de/org/>) and can use this channel for dissemination of results.
- Partner ZHW has established contacts with different e-learning centers (e.g. the eLearning Centre of the ZHW <http://elearning.zhwin.ch>) and with the Competence Service Production Center (CSPC) in St Gallen, Chur and Rapperswil (cf. <http://www.e-learning.zfh.ch> or <http://elearningfho.fh-htwchur.ch/>).
- Partner UTU participated in a meeting of a newly launched Special Interest Group "NLP and eLearning" under the auspices of the German "Gesellschaft für Linguistische Datenverarbeitung", 12 participants. There, they presented Lt4EL and its current results.
- Contacts have been established by Paola Monachesi with the Bazaar project in order to disseminate the results of the LT4eL project to that community of teachers and trainers.
- Contacts have been established through Matthias Kunkel with the Share project, that might be interested in the tools developed in the LT4eL project to develop metadata semi-automatically.

## 4.7 New projects

We believe that a concrete way to exploit the results of the LT4eL is that we can participate to two new related projects.

- We plan to carry out activities in collaboration with the CLARIN ([www.clarin.eu](http://www.clarin.eu)) project aiming at building an archive of language resources for the Humanities. Most partners of the LT4eL project are part of this project and eLearning could be considered an interesting application. Utrecht University of the coordinator of CLARIN facilitating thus the connection.
- The results of the LT4eL project will be employed in the new EU strep project 'Language Technology for LifeLong learning' which will be carried out in collaboration with OU-NL. The project was very positively evaluated (15/15) and selected on a pool of 196 projects (only 4 strep projects had been accepted). Three of the LT4eL partners will 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 (a remark made by the reviewers which can thus be explored in a subsequent phase).
- We are also investigating the impact that Language Technology can have for attention in eLearning as well as for facilitating the integration among Mediterranean countries on the basis of eLearning.

#### **4.8 PhD projects and Master theses**

The project has been the starting point for three PhD theses. They focus on different aspects of the project.

- Rosa Del Gaudio, University of Lisbon, Portugal: As a follow up of the project Rosa Del Gaudio asked a grants to do a PhD. The grant was asked to "Fundação para a Ciência e a Tecnologia" (FCT), the Portuguese research council. The main focus of the research project is the automatic extraction of definition for Portuguese language. This work includes the construction of a system for the automatic extraction of definition as well as its integration in some applications where it can play a key role, in particular in the context of a question answering system, and for semi-automatic extraction of ontologies. The starting point will be the glossary candidate detector developed during the project, that will be improved trying to exploit deep linguistics analysis and machine learning algorithms. Furthermore, the corpus collected for the project will be used, at least in a first phase, to validate new versions of the system.
- Ionut Pistol, UAIC, Romania: Ionut Cristian Pistol is working on a Ph.D. at the "Alexandru Ioan Cuza" University in Iasi, under the supervision of prof. Dan Cristea. The Ph.D. thesis will be completed at the end of 2008 and will be named "The Automated Processing of Natural Language Discourse" and will involve the development of ALPE, a LP meta-system improving on several aspects existing systems, such as GATE and UIMA. Since our project requires the employment of several linguistic processing chains as a functionality of the ILIAS management system, it was decided in a workshop held in June 2007 that ALPE will be used as a management system for the processing chains which will be deployed. These processing chains will take any kind of input documents and bring them automatically to the format required for the KWE and GCD tools used in the ILIAS system. The work on integrating part of ALPE in ILIAS will be done by Ionut, as a testing and demonstration stage for several of ALPE's functionalities. The final integration will benefit both LT4eL and Ionut Pistol's Ph.D. work.
- Eline Westerhout, Utrecht University, The Netherlands: The precision obtained with the

Glossary Candidate Detector (GCD) developed within the LT4eL project is not very high. Therefore, the Dutch team started experimenting with Machine Learning techniques to try to improve the results. In her PhD project, Eline will go on with this work and investigate in which way the best results can be obtained. The final GCD will be evaluated in different contexts, in the first place the context of semi-automatic glossary creation, this is the way in which the tool is used within the LT4eL project. Furthermore, she will investigate to which extent the extracted definitions can be used for the semi-automatic creation of ontologies by extracting relations from them. In her PhD project Eline will use many of the results of the project: the corpus created in WP1 for the validation of the GCD, the work on the GCD performed within WP2, the ontology created within WP3, and the results of the validation (WP5) to get an idea of what kind of results users appreciate.

Two other persons are writing their Master thesis related to the project:

- Claudia Borg, UOM, Malta: Within the LT4eL project, we look at definition extraction using a rule-based approach. This thesis explores the possibility of using Genetic Algorithms and Genetic Programming to rank definitions and learn new grammatical patterns which could be then included within the rule-based approach. These learning algorithms will be fed the manually written patterns within the project as the initial seed to their population, so as to develop new improved rules with the aim of classifying definitions correctly. The aim of this work is to compare this technique to other techniques used in definition extraction. This work will use the manually annotated definitions from within the English corpus in the LT4eL project as part of the training definitions for the learning algorithms.
- Jantine Trapman, UU, The Netherlands: In the developmental phase, the domain of the LT4eL ontology and the lexicons is restricted to Computing. This thesis describes how other domains can be included in a semi-automatic way using semantic web techniques. A new domain ontology is created using algorithms described in (Basili et al, 2007). It yields concepts from WordNet via which we can map the ontology easily to LT4eL. At the same time, also using the WordNet identifiers, the required terms of the domain can be extracted from the Dutch lexical database Cornetto Database in order to add them to the Dutch lexicon in LT4eL. The additional concepts and the lexical information originating from Cornetto can be used to improve the search and for word sense disambiguation.

## 5 Overview table

<b>Exploitable Material</b>	<b>Sector of Application</b>	<b>Timetable for use</b>	<b>Patent or other IPR protection</b>	<b>Owners and other partner involved</b>
Corpus of LOs	Research & eLearning industry	2007	IPR free	all partners
key word extraction & glossary candidate detector	Research & eLearning industry	first version: 2006; second version: 2007 & final version: 2008	IPR/patent free	all partners
ontology & language specific vocabularies	Research & eLearning industry	2007	IPR/patent free	all partners

ILIAS integrated functionalities	eLearning	first version: 2007; final version: 2008	IPR/patent free	all partners
Validation methodology	Research + eLearning industry	2008	IPR/patent free	all partners