



Project no. 027391

Project acronym: LT4eL
Project title: Language Technology for eLearning

Instrument Specific Targeted Research Project

Thematic Priority Information Society Technology

D1.2 Wrap-up report on difficulties in the corpus and tools collection

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: Al. I. Cuza University of Iasi (UAIC)

Revision [1]

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

D1.2 Wrap-up report on difficulties and IPR issues in the corpus and tools collection

1. Maltese Language Resources

Introduction

The feasibility of addressing Maltese has been a topic of discussion during the negotiations of the LT4eL project. Since the situation for this language was unclear at that stage, it was decided that Malta should carry out work wrt. English and if resources would be available, it would also carry out the work wrt. Maltese, the assumption being that developing the LT4eL functionalities for English would be less costly than for the other languages due to the availability of resources and attested research. In practice, the work carried out to develop the functionalities for English has been equivalent to that of the other languages.

In this document we nevertheless provide a brief overview of available language resources for Maltese. The document begins with a sketch of the language itself, followed by a description of a project (MLRS) whose main aim is the creation of infrastructure for language resources. We conclude with an assessment of possible strategies for Maltese within LT4eL.

Some Background on the Maltese Language

Maltese is the national language of Malta and, together with English, one of the two official languages of the Republic of Malta. . Altogether there are about 1 million native speakers, 400,000 of whom live on the island. Its use beyond the shores of the Maltese islands is limited to emigrant communities in Canada and Australia, but within the geographical confines of Malta, the language is used for the widest possible range of types of interaction and communication, including education, journalism, broadcasting, administration, business and literary discourse. We should note that its use in education tends to be bypassed for scientific areas of study. Hence the difficulties we have experienced obtaining LOs in Maltese for LT4eL.

One of the most challenging aspects of the language is its so-called 'mixed' nature, with a substrate of Arabic, a considerable superstrate of Romance origin (especially Sicilian) and, to a much more limited extent, English. The Semitic (Western/Maghrebi Arabic) element is evident enough to justify considering the language a peripheral dialect of Arabic. Its script, codified as recently as the 1920s, utilizes a modified Latin alphabet. This is just one of the peculiarities of Maltese as compared to other dialectal varieties of Arabic, more important ones being its status as a 'high' variety and its use in literary, formal and official discourse, its lack of reference to any Qur'anic Arabic ideal, as well as its handling of extensive borrowings from non-Semitic sources.

The morphology is still based on a root-and-pattern system typical of Semitic languages. For example, from the trilateral root consonants ħ-d-m one can obtain forms like: ħadem work (verb); ħaddiem worker; ħidma work (noun); nħadem be worked (verb passive); ħaddem caused to work.

Most of these forms are based on productive templates of which Maltese has a subset of those in Classical Arabic. One other typical feature shared with Semitic languages is "broken plural" formation as opposed to so-called sound plural. A few examples are qamar (moon); qmura (moons) tifel/tifla; (boy/girl); tfal (children), which involves a change in consonant/vowel pattern. The "sound plural" formation on the other hand involves affixation of suffixes such as -i, very common with words of Romance origin, -iet or -a as in: karozza (car) karozzi (cars) ikla (meal) ikliet (meals) haddiem worker haddiema worker.

Maltese has taken on a very large number of Romance lexical items and incorporated them within the Semitic pattern. For example, pizza, a word of Romance origin, has the broken plural form pizez (compare Italian pizza/pizze), and ċippa, a very recent borrowing from English (computer chip) has a broken plural form ċipep. In certain cases, one gets free variation between the broken plural e.g. tappet (carpet) tappiet/twapet (carpets).

Verbs are also often borrowed and fully integrated into the Semitic verbal system and can take all of the inflective forms for person, number, gender, tense etc. that any other Maltese verbs of Semitic origin can take. For example: spjega (explain) from (It. spiegare) yields jispjega (he explains), nispijegaw (we explain), spjegat (she explained, spjegajt (I explained).

The vigour and productivity of these processes is attested to by the fact that one keeps coming across new loan verbs all the time (increasingly more from English), both in spoken and in written Maltese, without the language having any difficulty in integrating them seamlessly into its morphological setup.

Within the verbal system complex inflectional forms can also be built through multiple affixations. For example, the word bghatthielux (I didn't send her to him), contains the suffixes -t (for 3rd person singular masculine subject (perfective)), hie for 3rd person singular feminine direct object, -lu for 3rd person singular masculine indirect object, and -x for verb negation. This ready potential for inflectional complexity is another Semitic feature of Maltese which applies across the board, whatever the origin of the verb. It also raises interesting questions concerning the nature of lexical entries, the relationship between lexical entries and surface strings, and the kind of morphological processing that is necessary to connect the two together.

Many of the linguistic issues that could help to resolve these questions are themselves unresolved for lack of suitably organized language resources (like the lexicon itself!). For this reason, we see the design/implementation of the lexicon, the development of language resources, and the evolution of linguistic theory for Maltese as three goals which must be pursued in parallel.

It is these considerations that motivated MLRS (Maltese Language Resource Server), a project funded by the national research and technical development initiative (RTDI) that started in 2005 and is being implemented by the University of Malta in collaboration with an SME that is responsible for the implementation of national IT policy on behalf of the Government.

MLRS

The aim of the project is the creation of electronic resources for Maltese that are crucial for the language to flourish within the information society. This main aim is being achieved through subsidiary objectives that are of direct relevance not only to LT4el but to future projects involving computing with the Maltese language. These are:

- ◆ the development of a Maltese National Corpus of language data.
- ◆ the realisation of a computational lexicon
- ◆ implementation of a portal offering related services

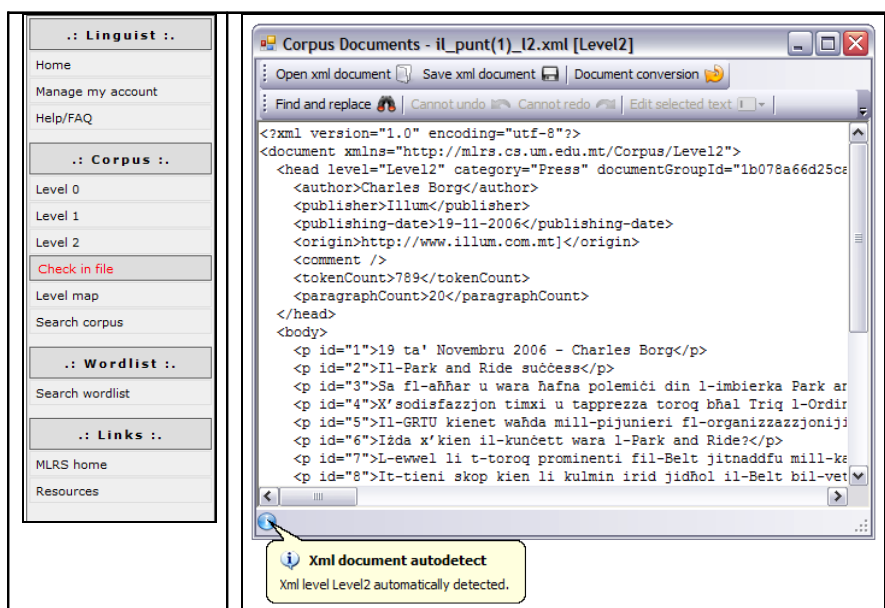
More details are available from the project website at <http://mlrs.cs.um.edu.mt>

Corpus

At the outset of MLRS, which began in 2005, the corpus consisted of about 1M words of news text, collected on an opportunistic basis primarily from Maltese language newspapers. The texts represented a broad spectrum of the types of articles found in these media. Submissions were managed by hand using an ftp server and kept in a directory. Download was also managed by ftp. After two years of development, the situation has changed somewhat.

Corpus Server

The most important change to date is the availability of a corpus server which handles the submission and maintenance of content, including meta-data and allows different levels of access by different categories of user. Most of the functionality is implemented using web services, so no special procedures are necessary to use it apart from user registration. The main functions are shown on the left-hand panel of the figure below. The figure also illustrates level 2 annotations at the text structure level which is computed automatically from level 1.



Corpus website

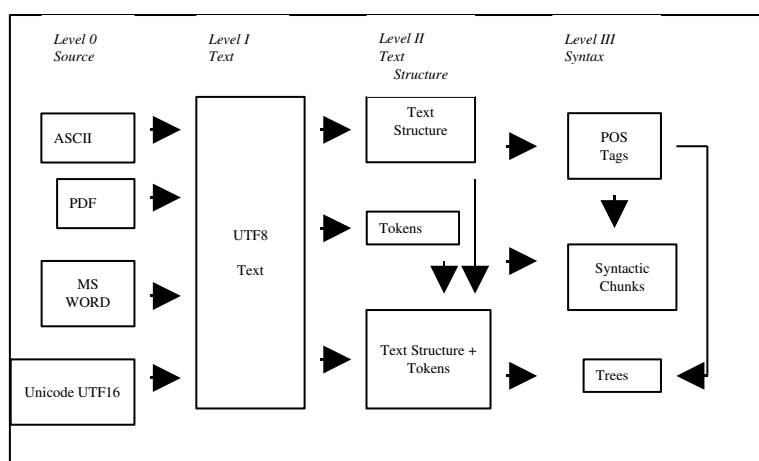
Corpus Contents

Contents are now more systematically classified, being divided into the following main categories, some of which are subdivided. The numbers in parentheses refer to the number of files currently available. However, these are constantly increasing.

- **Academic (12):** academic and scholarly papers.
- **Blog (35):** weblogs
- **Chat (0):** multi-party text based interaction over the internet.
- **Education (1)**
- **Email (0)**
- **Europe (6).**
- **Government (29).** Government documents, legal notices and publications. Will probably subdivide into different government offices. EU documents will also find their place here.
- **Law (1871).** Legalistic documents including laws and judgements.
- **Leaflet (23).** Leaflets, circulars.
- **Literature (57):** Literary works of all kinds. Subdivided into poetry, prose, and drama.
- **Parliamentary Debates (16).**
- **Press (198).** Newspaper text from daily and weekly newspapers. The directory is currently subdivided by newspaper (Nazzjion; Il-Mument; Kull-Hadd; Lehen)
- **Religion (1).** Religious and biblical texts
- **Speeches (16):** speech data (currently empty)
- **Spoken (0)**
- **Wordlist (0)**

Corpus Levels of Representation

Documents are stored at different levels of representation as summarized in the following figure, with semi-automatic migration of text between the levels.



Level 0. source level - resources received in the ftp directly from the originator as being at Level 0, sometimes referred to as the source level.

Level I. The text level is the first level at which we are in a position to define a standard. Concretely, level 1 is encoded in UTF8 and includes a lightweight TEI-style

header identifying and certain key features of the source text including title, author, date.

Level II. Text structure level including (i) tokenization with type information (i.e. with distinction between words, punctuation, number) and (ii) sentence and paragraph boundaries.

Level III. Linguistic annotation is introduced in order identify chunk structures (cf. Abney 1991) or trees.

Finally, higher levels of annotation, involving semantic information, such as named entities coreferencing etc, are foreseen at higher levels.

Corpus: Progress and Prospects

The availability of raw materials suitable for the corpus has improved a great deal. To qualify, such materials have to be in sufficient quantity; free from copyright; easily accessible and convertible to machine readable form and in good quality Maltese. As mentioned in a previous document, these comprise

- A. Legal materials of various kinds e.g.
 - Laws of Malta
 - Legal Judgments (i.e. individual cases)
 - Acquis Communautaire
- B. News-like articles
 - European parliament (europarl.europa.eu)
- C. Religious texts
 - the bible
 - other theological materials
- D. Literary and Linguistic works

There is nothing to stop any of these materials from being included. However, there are still some issues with respect to the human resources required to ensure that texts entered are correct. Representation at the higher levels is also lacking at the time of writing.

Computational Lexicon

The starting point for lexicon construction is a wordlist, and the system therefore provides a web services for extraction of words from documents, and maintenance of wordlists. The functionality of the service is implemented as follows:

1. User submits text, files or page URLs.
2. These resources are scanned and the words extracted from them and displayed.
3. User edits the resulting lists of extracted words manually
4. User submits final version for incorporation into the wordlist database.

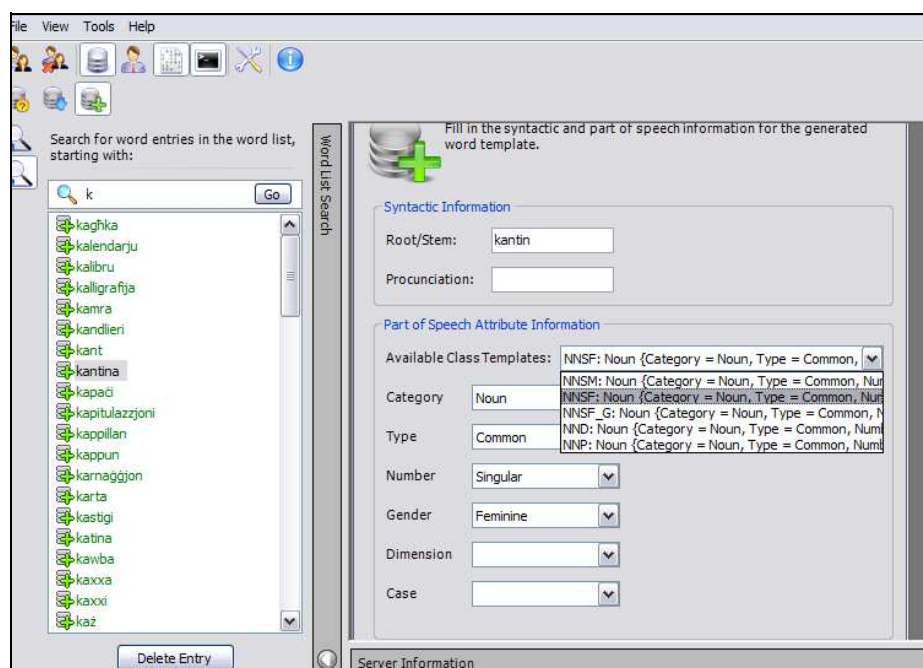
The wordlist is then used as a source for the lexicon itself, which is implemented as a database of entries which provide information about words. In order not to prejudge the kind of information that might be represented, we assume that,

conceptually at least, it takes the form of attribute value pairs. This works for morpho-syntactic information, and subsequently, we hope to extend the approach to deal with the semantic and phonological properties of words.

The current implementation is structured around a general purpose database table containing information about the word itself, its stem or root, and associated morpho-syntax. The latter property was implemented as a long integer bit field, since this is extremely efficient for the purpose of when queries are made. The system currently includes a web-based Lexicon Editor (Led) whose main function is to enable a linguist to create and maintain lexical entries.

At start-up, the lexicon editor parses a configuration file containing all the class, rule and macro definitions. Then, it modifies its word-entry interface to reflect the classes found in the ODL file, together with its attributes and values each one can take.

The figure below shows a screen-dump of Led in operation. The word to be entered (kantin (cellar)) appears in the first text box. The user then chooses which part or parts of speech apply to that word. The figure shows the process just after having decided that it is a noun.



Lexicon editor in operation

Lexicon Progress and Prospects

The current wordlist comprises about 100,000 words. Unfortunately the list contains errors and, so, cannot be directly incorporated into the lexicon as empty entries. The current challenges are therefore:

- ♦ Orthographic correction of original texts or existing wordlists.
- ♦ Creation of new wordlists, possibly based on corrected corpus materials.
- ♦ Population of the lexicon. This involves the act of selecting candidates from amongst correct wordlist items for inclusion in the lexicon.

- ◆ Addition and revision lexical information of existing lexical entries.

Unfortunately, all of these tasks require significant level of manual intervention, some of which involves specialist linguistic knowledge. On the other hand, the entries that are present (currently about 1000 words) are correct, and all the mechanisms for increasing that number lexicon are in place. For these reason we anticipate some delay until sufficient human resources can be marshaled to complete the entries.

Other Tools

Minimum requirements for participation in projects like LT4eL at a level which is comparable to that of the other languages is a tokeniser and part-of-speech tagger. For Maltese a tokeniser is available but a tagger is not, although work to develop one is ongoing within MLRS.

It is important to underline that the technical infrastructure for building a tagger is there. The bottleneck is a sufficient quantity of tagged texts for training purposes. These have to be prepared manually by specialists with a certain level of linguistic expertise.

At the time of writing, a tagger is available but its accuracy is in the region of 60%. We are confident that this level of performance will be improved over the next year subject to the availability of resources to allow corrective feedback to be produced.

Conclusion: What is Feasible in Maltese within LT4eL

The overall picture painted in this report is that, as regards the medium-term availability of language resources for Maltese, a lot of the groundwork has actually been carried out. What is lacking is the actual deployment of human resources to the problem of correcting data which is imperfect, hand-tagging of texts, selection and creation of new lexical entries.

We had considered various solutions to the problems for LT4eL created by the lack of LOs and tools described above. We have to bear in mind that any proposed solution has fulfill three criteria: (a) be within the scope the project (b) yield results that are compatible project results for other languages and (c) fit within the financial and temporal constraints imposed by available budget and the work program.

We still believe that, subject to human resources being available, a feasible and achievable goal for Maltese lies in CLIR. Specifically, this would imply that Maltese should figure in the relevant part of D3.2 which concerns development of the lexicon so that every concept in the ontology has a Maltese translation.

This, in principle, will clearly enable queries expressed in Maltese to be semantically profiled according to the concepts they mention, and hence to be related to semantically indexed documents. However, for such a task to be incorporated into the project systematically, an appropriate scenario within WP5 has to be planned and executed using available resources. Given that the Maltese partner is still heavily

engaged in activities for English, it is not yet clear whether such resources would be available in time.

In short, the full implications carrying out CLIR for Maltese need to be quantified in terms of resources and squared against current commitments. We propose to carry out this exercise during the coming weeks.

2. Description of problems in corpus collection due to IPR issues, solutions adopted and problems unsolved

IPR issues with respect to our corpus of learning objects are difficult to solve. As a first step, the partner ETHZ has investigated our specific situation and created a Wiki page that lists questions and answers on this issue. Copyright protection and copyright infringement are very complex issues in a case like ours, with many national states involved, and with copyrighted documents displayed on a web server. This is due to the fact that internet is non-territorial by nature, so the principle of territoriality is difficult to apply when there is uncertainty about the countries which in fact are involved by diverse users. ETHZ has contacted the EU funded IPR helpdesk at EC IPR Helpdesk¹ for the clarification of these issues. The feedback of the IPR helpdesk made us considering the option of consulting a lawyer who is specialized in intellectual property issues in order to obtain a detailed assessment of the copyright legislations which will apply in our case. Independently of such consultation, we are also trying to create a legal situation which can avoid copyright infringements. For this purpose, we have tried to obtain the prior written consent of each copyright holder for the planned use. Even though each partner had specific problems mainly due to the national legislation, we provided for each LO the IPR status, following the next classification, with small variations:

- Free to use; must cite **author**, copyright obtained from author
- Free to use; **permission** obtained
- **GPL** License
- **Government** source
- **Copyrighted**
- **unknown**, because it was impossible to contact the author (but the respective LO is valuable)

Generally, the LT4eL partners have tried to obtain from authors their agreement to use the documents for the purposes of the project. Most of the documents that specified copyrights have been removed from the collection. Notwithstanding this, it is still possible that some partners have IPR issues not completely solved.

The IPR status of the LOs for each language is illustrated in the following table.

¹ <http://www.ipr-helpdesk.org/index.html>

Language	No. of LOs	IPR type					
		author	permission	gov	GPL	copyrighted	unknown
Bulgarian	54	-	40	-	-	-	14
Czech	60	24	23	2	10	-	1
Dutch	77	25	3	-	-	-	49
English	104	16	22	1	8	24	33
German	47	7	8	-	-	9	23
Maltese **	9	7	-	-	-	-	2
Polish	87	-	12	19	16	21	19
Portuguese	62	25	15	-	1	10	11
Romanian	55	51	-	2	-	-	2

Note **: For Maltese see section 1 in this report.

3. IPR issues for LOs used in the scenarios

The IPR issues for the LO used in the experiments are summarized in the tables below.

Language: Bulgarian

Experiments	used LO	IPR
WP2 Scenario		
WP2 Tutor 'New Content'	An Idea on JavaScript, CGI and Java	free for research
WP2 2 Student 'Paper Synopsis'	An Idea on JavaScript, CGI and Java	free for research
WP2 Student 'Lecture slides & quiz'	NOT DONE	-
WP3 Single Language Scenarios		
WP3 SL Tutor 'Course Pre-requisites'	1. An Idea on JavaScript, CGI and Java 2. Creating Web Pages 3. Creating Web Site 4. Design Web Pages 5. HTML document 6. Inserting Objects in WebPage 7. Tendencies in HTML 8. WWW (World Wide Web)	1. free for research 2. free for research 3. free for research 4. free for research 5. free for research 6. free for research 7. free for research 8. free for research
WP3 Student ' Multiple-choice Quiz'	1. An Idea on JavaScript, CGI and Java 2. Calimera_part3_text2 3. Calimera_part3_text7 4. Colors and Images 5. Creating Web pages	1. free for research 2. Gov 3. Gov 4. free for research

	6. Creating Web site 7. Forms 8. HTML document 9. Inserting Objects In Web Page 10. Introduction into XML technology 11. Introduction to Power Point 12. Language Characteristics 13. Setting up and working with email 14. Text formatting 15. The model Client-Server 16. WWW(World Wide Web)	5. free for research 6. free for research 7. free for research 8. free for research 9. free for research 10. free for research 11. free for research 12. free for research 13. free for research 14. free for research 15. free for research 16. free for research
WP3 Multi-Lingual Scenarios		
WP3 ML Student Hanspeter Scenario	1. Connecting the computer to Internet 2. Document Formatting I 3. HTML document 4. Searching in World Wide Web 5. Setting up and working with email	1. free for research 2. free for research 3. free for research 4. free for research 5. free for research
WP3 ML Student Rachel Scenario	1. Adding a Touch of Style in HTML, (English LOs) 2. Tutorial XML (English LOs) 3. SQL XML XQuery and Native XML (English LOs) 4. A Technical Introduction to XML (English LOs) 5. Extensible Markup Language (XML) 1.0 (English LOs) 6. Getting Started with HTML (English LOs) 7. More Advanced Features HTML (English LOs)	1. copyright 2. free 3. unknown 4. copyright 5. Can use according to restrictions www.w3.org/Consortium/Legal/2002/copyright-documents-20021231 6. copyright 7. copyright
WP3 ML Tutor Pre-requisites	1. Introduction to the Internet 2. Technical and technological aspects of Internet	1. free for research 2. free for research

Language: Czech

Experiments	used LO	IPR
WP2 Scenario		
WP2 Tutor 'New Content'	cze Using Power Point	author
WP2 2 Student 'Paper Synopsis'	cze Using Power Point	author
WP2 Student 'Lecture slides & quiz'	NOT DONE	-
WP3 Single Language Scenarios		

WP3 SL Tutor 'Course Pre-requisites'	cze HTML_web publishing web pages	Permission
	cze Using Javascript	Permission
WP3 Student ' Multiple-choice Quiz'	cze Bash scripting	author
	cze Using Power Point	author
WP3 Multi-Lingual Scenarios		
WP3 ML Student Hanspeter Scenario		English LOs, most of them free for research
WP3 ML Student Rachel Scenario	Introduction to word	English LOs, most of them free for research
	More on word processing	
	Open office writer guide	
	Open office getting started	
	The Office drawing tools	
WP3 ML Tutor Pre-requisites	Creating a World Wide Web site	English LOs, most of them free for research

Language: Dutch

Experiments	used LO	IPR
WP2 Scenario		
WP2 Tutor 'New Content'	dutXML in 10 points	W3C document, using document is permitted, provided that a link is included to the original W3C document and a copyright notice is added
WP2 2 Student 'Paper Synopsis'	dutIntroductie HTML	unknown
WP2 Student 'Lecture slides & quiz'	NOT DONE	-
WP3 Single Language Scenarios		
WP3 SL Tutor 'Course Pre-requisites'	dutFrontpage	unknown
WP3 Student ' Multiple-choice Quiz'	dutInternet introduction	unknown
WP3 Multi-Lingual Scenarios		
WP3 ML Student	dutIntro Internet	Author

Hanspeter Scenario		
WP3 ML Student Rachel Scenario	dutIntro Internet	Author
WP3 ML Tutor Pre-requisites	dutInternet and e-mail	Author

Language: English

Experiments	used LO	IPR
WP2 Scenario		
WP2 Tutor 'New Content'	Guidelines for Writing a Scientific Paper	unknown
WP2 2 Student 'Paper Synopsis'	Introduction to NLP	Free for research
WP2 Student 'Lecture slides & quiz'	NOT DONE	-
WP3 Single Language Scenarios		
WP3 SL Tutor 'Course Pre-requisites'	1. Nothing specific, depends on the search results – here the results mainly were: 2. The Unix and Internet Fundamentals How To 3. Introduction to the Internet 4. Introduction to computer hardware and software	1. GNU 2. Free for research 3. Free for research 4. Free for research
WP3 Student ' Multiple-choice Quiz'	Various documents returned	
WP3 Multi-Lingual Scenarios		
WP3 ML Student Hanspeter Scenario	NOT DONE	-
WP3 ML Student Rachel Scenario	NOT DONE	-
WP3 ML Tutor Pre-requisites	NOT DONE	-

Language: German

Experiments	used LO	IPR
WP2 Scenario		
WP2 Tutor 'New Content'	Excel Schulungsunterlagen	Free for research
WP2 2 Student 'Paper Synopsis'	NOT DONE	-

WP2 Student 'Lecture slides & quiz'	Folien Officepakete	Free for research
<i>WP3 Single Language Scenarios</i>		
WP3 SL Tutor 'Course Pre-requisites'	no LOs mentioned in scenario	English/German LOs, most of them free for research
WP3 Student ' Multiple-choice Quiz'	no LOs mentioned in scenario	English/German LOs, most of them free for research
<i>WP3 Multi-Lingual Scenarios</i>		
WP3 ML Student Hanspeter Scenario	user is asked to select 2 Los	English/German LOs, most of them free for research
WP3 ML Student Rachel Scenario	user is asked to select 2 LOs	English/German LOs, most of them free for research
WP3 ML Tutor Pre-requisites	no LOs mentioned in scenario	English/German LOs, most of them free for research

Language: Polish

Experiments	used LO	IPR
<i>WP2 Scenario</i>		
WP2 Tutor 'New Content'	<u>Konstrukcja publikacji naukowej</u>	Unknown
WP2 2 Student 'Paper Synopsis'	<u>Co to jest e-learning</u>	copyrighted
WP2 Student 'Lecture slides & quiz'	NOT DONE	-
<i>WP3 Single Language Scenarios</i>		
WP3 SL Tutor 'Course Pre-requisites'	Samouczek Emacsa	GNU GPL
	Teksty wielojezyczne w edytorze GNU Emacs	GNU GPL
	Łagodne wprowadzenie do TeXa	GNU GPL
	Podstawy obsługi edytora tekstu Microsoft Word	No information, no contact to an author
	<u>Internetowy podrecznik edytora tekstu</u>	No information

WP3 Student ' Multiple-choice Quiz'	Duch03 - Reprezentacja informacji w komputerze	Author's agreement to use it
	Wszystko, co musisz wiedziec o wyszukiwarkach	No changes allowed; free distribution and copying allowed
	Podstawy obsługi edytora tekstu Microsoft Word	No information, no contact to an author
	HTML dla początkujących	No information, no contact to an author
WP3 Multi-Lingual Scenarios		
WP3 ML Student Hanspeter Scenario	-	-
WP3 ML Student Rachel Scenario	<u>Samouczek Emacsa</u>	GNU GPL
	<u>Teksty wielojezyczne w edytorze GNU Emacs</u>	GNU GPL
	Łagodne wprowadzenie do TeXa	GNU GPL
	<u>Internetowy podrecznik edytora tekstu</u>	No information
	<u>Podstawy obsługi edytora tekstu Microsoft Word</u>	No information, no contact to an author
WP3 ML Tutor Pre-requisites	<u>Samouczek Emacsa</u>	GNU GPL
	<u>Konstrukcja publikacji naukowej</u>	No information, no contact to an author
	<u>Podstawy obsługi edytora tekstu Microsoft Word</u>	No information, no contact to an author
	<u>Teksty wielojezyczne w edytorze GNU Emacs</u>	GNU GPL

Language: Portuguese

Experiments	used LO	IPR
WP2 Scenario		
WP2 Tutor 'New Content'	porIntroducao a internet	Free to use for the goals of the project; must cite author, copyright obtained from author for non commercial purposes
WP2 2 Student 'Paper Synopsis'	porUm modelo baseado em XML para suporte da dinimica processual de negocio	Free to use for the goals of the project
WP2 Student 'Lecture slides & quiz'	NOT DONE	-
WP3 Scenarios		

All scenarios use the same group of documents

Apostilha XML	Free to use for the goals of the project; must cite author, copyright obtained from author for non commercial purposes
Introducao ao XML	Free to use for the goals of the project; must cite author, copyright obtained from author for non commercial purposes
porArmazenamento de XML em bases de dados	Free to use for the goals of the projec
porXSL	Free to use for the goals of the project; must cite author, copyright obtained from author for non commercial purposes
porComandos dos XML	Free to use for the goals of the project; must cite author, copyright obtained from author for non commercial purposes
porXSchema	Free to use for the goals of the project; must cite author, copyright obtained from author for non commercial purposes
porXML e DTD: Uma introducao	Free to use for the goals of the project; must cite author, copyright obtained from author for non commercial purposes
porUma Perspectiva historica das linguagens de marcacao	Free to use for the goals of the project
porLinguagens de marcao e XML	Free to use for the goals of the project

Language: Romanian

Experiments	used LO	IPR
WP2 Scenario		
WP2 Tutor 'New Content'	Microsoft word notiuni introductive	Free to use; must cite author, copyright obtained from author
WP2 2 Student 'Paper Synopsis'	Navigare pe Internet	Free to use; must cite author, copyright obtained from author
WP2 Student 'Lecture slides & quiz'	NOT DONE	-
WP3 Single Language Scenarios		
WP3 SL Tutor 'Course Pre-requisites'	Utilizarea tehnologiilor Web-ului semantic in E-Learning	Free to use; must cite author, copyright obtained from author
WP3 Student ' Multiple-choice Quiz'	Introducere in e-learning Medii virtuale de invatare Web semantic_ XML_ XHTML_ RSS_ Weblog	All Los are free to use; must cite author, copyright obtained from author
WP3 Multi-Lingual Scenarios		
WP3 ML Student Hanspeter Scenario	HTML Proiectarea siturilor web Web semantic-xml-xhtml-rss-weblog Posta electronica Comunitati online Html manual online	Free to use; must cite author, copyright obtained from author
WP3 ML Student Rachel Scenario	HTML Proiectarea siturilor web Web semantic-xml-xhtml-rss-weblog Posta electronica Comunitati online Html manual online	Free to use; must cite author, copyright obtained from author
WP3 ML Tutor Pre-requisites	Proiectarea siturilor web – partea 4	Free to use; must cite author, copyright obtained from author

Generally, the Learning Objects used in the experiments are free to use. The authors have been contacted and the copyright was obtained for non commercial purposes.

All Romanian, Bulgarian, Czech, German and Portuguese documents used in testing are free to use if author is mentioned. For the multilingual scenarios, English documents are retrieved also, with various IPRs.

For Dutch, most of the LOs are free, with the exception of WP2 2 Student 'Paper Synopsis' scenario and WP3 SL Tutor 'Course Pre-requisites' Scenario, for which the author was contacted, but no information was received yet, and for WP3 Student 'Multiple-choice Quiz' scenario, where the document containing most of the answers is no longer available via internet and contacting the author resulted in no response.

For English, all the LOs used are free to use for non-commercial purposes, except for the LO in WP2 Tutor 'New Content' scenario, where no information was received from the author.

Similarly, for Polish, most of the LOs are GNU GPL, with the exception of the LO used for WP2 Tutor 'New Content', where no information has been found.