Isabel de Fátima Silva Azevedo

Semantic and Pragmatic Characterization of Learning Objects

a thesis submitted for the Degree of Doctor of Philosophy
supervised by

Prof. Doutor Eurico Manuel Carrapatoso
Prof. Doutor Carlos Miranda Vaz de Carvalho

U. PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

University of Porto
Faculty of Engineering
Doctoral Program in Informatics Engineering

Porto, Portugal
October 2011

To David
To my family

ACKNOWLEDGEMENTS

This thesis is submitted to the Faculty of Engineering at the University of Porto, Portugal, in fulfilment of the requirements for the Degree of Doctor of Philosophy. The work reported was carried out at the Graphics, Interaction and Learning Technologies (GILT) research group. Many people and organisations contributed to the successful completion of this work.

I am deeply indebted to my supervisors, Professor Eurico Carrapatoso and Professor Carlos Vaz de Carvalho, for the invaluable guidance in the shaping of this work. The discussions we have had and the feedback I have received during this time have been doubtless significant for this work, as well as all the comments suggested during the review of all versions of this document.

The Monitoring Group designated to follow my work provided valuable comments and suggestions which allowed me to improve my research. Thus, I thank all the individuals in that group, including my supervisors, the members of the Scientific Committee of the Doctoral Program in Informatics Engineering, and Professor Jaime Villate and Professor Martin Llamas Nistal.

I would like to thank GILT research group for providing me with all the conditions to conduct the research described in this document.

I need also to recognise the valuable support given by the Portuguese Foundation for Science and Technology (FCT), through the PhD grant SFRH/BD/43437/2008, and the Ministry of Science, Technology and Higher Education, through the PhD grant SFRH/PROTEC/49362/2009. This latter grant allowed me to be entirely dedicated to my research for two semesters.

I would also like to thank Instituto Superior de Engenharia do Porto and Instituto Politécnico do Porto, in the person of their Presidents (Dr. João Rocha, Dr. Rosário Gamboa and Dr. Victor Santos) for the partial financial support of the fees at the University of Porto and for supporting my application for FCT (SFRH/BD/43437/2008) and PROTEC (SFRH/PROTEC/49362/2009) grants. Also, I would like to express my gratitude to Dr. Ana Almeida, director of Informatics Engineering Department, for always trying to facilitate my schedule during the service distribution each semester at ISEP, and for her suggestions regarding my grants' applications and later doubts concerning their internal rules.

Another financial assistance from the FCT is hereby acknowledged, as the research described in this document in an early phase conducted to a project submitted in 2006, later approved, that sponsored this research beyond November 2007: the FCT project PTDC/EIA/65387/2006 'CASPOE – CAracterização Semântica e Pragmática de Objectos Educativos' ('Semantic and Pragmatic Characterization of Learning Objects', in English). Under this project two fellowships supported the work of Adela Ortiz and Rui Seiça, who I would like to acknowledge, as they worked in the development of many parts of this PhD work.

I am also extremely grateful to Christopher for providing me valuable feedback in all my English doubts and for all his pieces of advice.

I also desire to thank all anonymous referees for their critical comments and useful hints on papers submitted under the work carried out and described in this document. The assistance of many authors who provided me draft versions of their papers or clarified some points of their research was also valuable to my research.

My thanks also go to Paulo Taveira, who developed the component related to the use of IMS LD templates, and Luciano Santos, who implemented the component that uses DBPedia, as part of their final courseworks.

I am indebted to many colleagues for substituting me in some classes or other teaching activities while I was presenting my PhD work in conferences, such as Anibal Oliveira, António Rocha, António Silva Pereira, Nuno Escudeiro, Nuno Silva, Rosa Reis, and Rui Coentro.

For all the support and encouragement they gave me, I thank mainly my family, but also several people, including Alexandre Gouveia, Ana Barata, Constantino Martins, David Gouveia, Dulce Mota, Fabricio Santos, Filipa Pimenta, Luís Paupério, Nuno Morgado, Paula Escudeiro, Paulo Matos, and Ricardo Almeida.

ABSTRACT

Currently there are a great number of available repositories that store resources used for learning and teaching. However, the simple characterisation of the resources as typically done in libraries is not adequate. The wide availability of learning resources has made it relevant to deal with the topic of how to contextualise those in order to make it easier for teachers and learners to find what they need for their purposes.

In developed countries, many educational materials used in tertiary education are in digital format, with a tendency to increase and reach other levels of education over time. This format enables the automation of several processing steps for the characterisation of learning resources and other forms of regarding their usage.

Using a common repository platform, the main objective of this PhD work was to state how advanced functionalities that attend to the semantics and the pragmatics of the resources could be designed.

This thesis presents contributions to the area of Learning Objects Repository, which can improve how learning materials are represented and used in repositories, and thus:

- It provides a model to semantically and pragmatically characterise learning resources;
- It offers methodological guidelines to obtain semi-automatically some key aspects of learning materials that are included in the IEEE Learning Objects Metadata standard;
- It delineates how knowledge can be considered in a repository, also proposing methodological steps for its acquisition, with emphasis on ontology reuse;
- As learning objects are strongly related to learning and teaching practices, and pedagogical information cannot be disregarded, it also proposes how a description linked to the practical utilisation of learning objects can be provided through the use of a learning technology specification;
- It provides guidelines to capture users' opinions about resources through tagging in a pragmatic approach, which can be used for various reasons: to provide feedback to providers related to the resources they submitted to the repository, or even to highlight resources that might require an in-depth revision.

To sustain the proposed approach, the related functionalities were implemented in a working prototype in compliance with the relevant specifications.

RESUMO

Atualmente há um grande número de repositórios disponíveis e utilizados para fins de ensino e aprendizagem. No entanto, a simples caracterização dos recursos, habitual em bibliotecas, por exemplo, não é suficiente. A ampla disponibilidade de recursos de aprendizagem tornou importante o tópico de como contextualizá-los, de forma a que seja possível a professores e alunos encontrar o que precisam mais facilmente.

Nos países desenvolvidos, muitos materiais pedagógicos utilizados no ensino superior estão em formato digital, com tendência para aumentar e se atingir outros níveis de ensino ao longo do tempo. Esta situação permite a automação de várias etapas de processamento quando os recursos estão a ser caracterizados e também outras formas de considerar a sua utilização. Recorrendo a uma plataforma de repositório comum, o objetivo principal deste trabalho de doutoramento foi mostrar como funcionalidades avançadas que consideram a semântica e a pragmática dos recursos podem ser concretizadas.

Esta tese apresenta contribuições para a área de Repositórios de Objetos Educativos, que podem melhorar a representação dos materiais de aprendizagem e o seu processamento em repositórios, e, portanto:

- Propõe um modelo para caracterizar semântica e pragmaticamente recursos educativos;
- Proporciona orientações metodológicas para a obtenção semi-automática de alguns aspetos-chave de materiais de aprendizagem que estão incluídos na especificação de metadados IEEE LOM;
- Delineia como o conhecimento pode ser considerado, e também propõe passos metodológicos para a sua aquisição, com ênfase na reutilização de ontologias;
- Como os objetos educativos estão fortemente relacionados com a prática educativa, e a informação pedagógica não deve ser ignorada, propõe-se o uso de uma descrição relacionada com a utilização prática dos mesmos e expressa em concordância com uma especificação que permite a modelação de atividades educativas;
- Sugere como captar as opiniões dos utilizadores sobre os recursos através da sua marcação numa abordagem pragmática. Esta abordagem pode ser usada com finalidades distintas: fornecer feedback aos utilizadores sobre os recursos que submeteram anteriormente, ou mesmo sugerir quais os recursos que podem necessitar de uma revisão aprofundada.

Para sustentar a abordagem proposta, as funcionalidades relacionadas foram implementadas num protótipo funcional que utiliza as especificações consideradas relevantes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EXAMPLES

LIST OF ABBREVIATIONS

ABox	Assertional Box
ADL	Advanced Distributed Learning
AICC	Aviation Industry CBT Committee
AP	Average Precision
API	Application Programming Interface
ASK-LDT	ASK Learning Designer Toolkit
CanCore	Canadian Core Learning Object Metadata Application Profile
CBT	Computer-based Training
CDF	Curriculum Description Format
CELEBRATE	Context eLearning with Broadband Technologies
CEN	Comité Européen de Normalisation
CMC	Computer-Mediated Communication
CNF	Conjunctive Normal Form
CORDRA	Content Object Repository Discovery and Registration/Resolution Architecture
CoSMoS	Collaboration Script Modelling System
DAML	DARPA Agent Markup Language
DARPA	Defense Advanced Research Projects Agency
DC	Dublin Core
DCMES	Dublin Core Metadata Element Set
DCMI	Dublin Core Metadata Initiative
DL	Description Logics
DRI	Digital Repositories Interoperability
DVD	Digital Versatile Disk
EdNA	Education Network Australia
ELEONET	European Learning Objects Network

EML	Educational Modelling Language
F2F	Face-to-Face
Fedora	Flexible Extensible Digital Object Repository Architecture
FOL	First-Order Logic
GEM	Gateway to Educational Materials
GLM	Graphical Learning Modeller
GNU	GNU is Not Unix
Heal	Health Education Assets Library
ICT	Information and Communication Technologies
idf	inverse document frequency
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronic Engineers
IFID	Illocutionary Force Indicating Devices
IMS	Instructional Management Systems (when IMS started in 1997, not adopted anymore)
IMS CC	IMS Common Cartridge
IMS CP	IMS Content Packaging
IMS GLC	IMS Global Learning Consortium
IMS LD	IMS Learning Design
IMS LIP	IMS Learner Information Package
IMS QTI	IMS Question and Test Interoperability
IMS SS	IMS Simple Sequencing
IR	Information Retrieval
IRI	Internationalised Resource Identifier
ISEP	Instituto Superior de Engenharia do Porto
ISO	International Standardisation Organisation
ISSS	Information Society Standardisation System
IST	Information Society Technology
ITS	Intelligent Tutoring System
JTC	Joint Technical Committee
KB	Knowledge Base
KBS	Knowledge Base System
KPS	Knowledge Pool System
LAMS	Learning Activity Management System
LCMS	Learning Content Management System
LD	Learning Design
LLL	Life Long Learning
LMML	Learning Material Markup Language
LMR	Metadata for Learning Resources
LMS	Learning Management System

LN	Learning Networks
LO	Learning Object
LODE	Learning Object Discovery and Exchange
LOM	Learning Object Metadata
LOR	Learning Object Repository
LRE	Learning Resource Exchange
LTSC	Learning Technologies Standards Committee
MAP	Mean Average Precision
MELT	Metadata Ecology for Learning and Teaching
MP3	MPEG-1/2 Audio Layer 3
MPEG	Moving Picture Experts Group
MRR	Mean Reciprocal Rank
NDCG	Normalised Discount Cumulative Gain
NIST	National Institute of Standards and Technology
P@n	Precision at Position n
NVC	Non-Vocal Communication
OAI	Open Archives Initiative
OAI-DC	OAI Dubin Core
OAI-PMH	OAI Protocol for Metadata Harvesting
OIL	Ontology Inference Layer
OKI	Open Knowledge Initiative
OTK	On-To-Knowledge
OTKM	OTK methodology
ORSD	Ontology Requirements Specification Document
OSID	OKI Service Interface Definition
OUNL	Open Universiteit Nederland
OWA	Open World Assumption
OWL	Web Ontology Language
P@n	Precision at Position n
PA	Positive Affect
PoEML	Perspective-oriented Educational Modeling Language
QE	Query Expansion
NA	Negative Affect
RDF	Resource Description Framework
RDFS	RDF Schema
RIF	Rule Interchange Format
RLO	Reusable Learning Object
RTE	Run-time Environment
SCO	Shareable Content Object
SCORM	SCO Runtime Model

SPARQL	SPARQL Protocol And RDF Query Language
SQI	Simple Query Interface
SQL	Structured Query Language
SWD	Semantic Web Document
SWSE	Semantic Web Search Engine
SWSL	Semantic Web Services Language
SWT	Semantic Web Term
S@n	Success@n
TBox	Terminological Box
tf	term frequency
TREE	Teaching Resources for Engineering Education
TRP	Transition Relevance Place
UML	Unified Modelling Language
UNA	Unique Names Assumption
UoL	Unit of Learning
URI	Universal Resource Identifier
URL	Universal Resource Locator
W3C	World Wide Web Consortium
WRL	Web Rules Language
WSLT	Workshop on Learning Technology
XML	Extensible Markup Language
XSD	XML Schema Definition

FIRST PART: INTRODUCTION

Chapter 1

INTRODUCTION

This chapter introduces the main problems that led to this thesis and the related research questions. It summarises the objectives of the thesis and the applied research methodology.

In addition, it provides a guide to reading the thesis, presenting the different chapters and their relation to the thesis statement.

1.1 Context

The view on education has changed substantially, especially during the 20th century. A report published in the 19th century, and cited in [Huberman, 1979], shows a earlier view on education: "To give education to the poor working class ... would actually be detrimental to their morals and happiness; to learn to despise their lots in life instead of making them good servants in agriculture and other laborious jobs, that their positions in society had intended ... Education allows them to read seditious pamphlets ... and become insolent to their superiors".

Today, in the beginning of the 21st century, education is an acquired right and a duty of governments, at least in the early years of schooling and in the more developed countries. Moreover, currently people have been increasingly dealing with an excess of information, which may distract their attention from what is essential. In addition, learning is no longer seen as a necessity of only certain stages of life, usually early ones. The term 'lifelong learning' has become common.

Education is increasingly a competitive area and the vision statements of institutions necessarily address quality, also by launching more and better services to students and staff. Following this trend, many institutions have provided repositories of educational resources.

Furthermore, in a study that aimed to uncover the diverse views on what users need and what those who develop resources consider as adequate, the following obstacles to adopting digital resources were identified: how to find, manage, maintain, and reuse resources even in new contexts [Harley, 2008]. Learning objects repositories may have a role in diminishing those barriers.

In addition, Learning Objects are seen as a way to bring about increased flexibility, adaptation, updating, tracking, management and reuse of learning resources. However, to fully benefit from the use of Learning Objects it is necessary to store them in repositories that are able to meet those objectives.

Higgs et al are explicit by suggesting that a repository must have the following functionality [Higgs, Meredith et al., 2002]:

- To search and to browse;
- To request a Learning Object (LO) that has been located;
- To maintain LOs;
- To retrieve an object that has been requested;
- To control the quality (technical, educational and metadata requirements) of LOs;
- To allow the submission of LOs for storage;
- To store LOs with unique, registered identifiers that allow them to be located;
- To allow users to obtain metadata about objects in other repositories;
- To provide metadata to other repositories.

A few examples of Learning Objects Repositories developed over the last years are:

- Resource Generator (http://www.resourcegenerator.gov.au),
- Toolbox repository (http://toolboxes.flexiblelearning.net.au/repository/index.htm),
- The Le@rning Federation (http://www.thelearningfederation.edu.au/default.asp),
- iLumina (http://www.ilumina-dlib.org/),
- TeleCampus (http://telecampus.edu/),
- Multimedia Educational Resource for Learning and Online Teaching (MERLOT) ((http://www.merlot.org).

However, existing educational repositories are often under-used [Neven and Duval, 2002]. A recent study found that the reuse of learning objects varies from 11.5% to 32.2%, depending on their granularity [Ochoa, 2008]. The reuse of learning objects has been a recurrent theme in the literature, and it is obvious that many factors affect their reuse, which are deeply analysed in [McNaught, 2003].

One way to promote the learning objects reutilisation is the employment of specifications for describing learning objects, which make possible a good agreement platform, making easier their reuse.

Bennett et al. consider that “there has been considerable attention given to technical issues of repositories, with much less consideration of how to attend to the needs of those who will use them” [Bennett, Parrish et al., 2008]. In fact, repositories have been seen as providing “poor or inadequate search and discovery tools”, which constitutes a barrier to the reuse of learning objects, as if they cannot be found they cannot be reused [Philip and Cameron, 2003].

A Swedish project that ended in April 2009, called "Open Educational Resources in institutional repositories”, highlighted the need for more sophisticated use of repositories than that allowed by the institutional repositories, among other recommendations [Lindholm, Axdorph et al., 2009]. In fact, it is necessary to avoid considering learning objects stored in repositories merely as archives, not fully helping learners interested in some subjects or educators developing courses [Azevedo, Martins et al., 2006]. More efficient search and retrieval possibilities are a point to consider in order to avoid it. A semantic representation of learning contents may enable a richer recovery of learning objects.

Moreover, there has been an emphasis on the need for more efficient and systematic storage, search and retrieval of learning resources, without forgetting what characterises them: their use in learning contexts. The conceptualisation of learning objects in terms of their pedagogical value and significance, and not in terms of their neutrality, is necessary [Friesen, 2004].

Wiley considers the following point as very important: “the role of context is simply too great in learning, and the expectation that any educational resource could be reused without some contextual tweaking was either naive or stupid” [Wiley, 2006]. The author argues that there will be repositories that enclose lots of learning objects that nobody knows how to use/reuse if pedagogical and practical issues are not followed [Wiley, 2002].

Content and context are two of the most important aspects for searching learning materials, even for students [Stojanovic, Staab et al., 2001].

Thus, if semantic characterisation of learning objects can improve their contents’ description and make easier their discovery and later reuse, it is common sense that the semantic description of a learning object, even though important, is not enough to promote its reuse, but it is also necessary some information related to how to apply it in a pragmatic characterisation. A wide description of learning resources, considering the different views of the different kinds of users of repositories and a use context – an important point in pragmatics - can provide a better representation of learning objects.

1.2 Research questions

It is necessary to put semantic and pragmatic concerns in the full representation of learning resources to allow their very fine description. However, a substantial effort is necessary to consider these issues in the representation of LOs in repositories. The main research question of this PhD work is:

`How can resources be semantically and pragmatically characterised in repositories when their main purpose is to support learning and teaching?`

That research question can be subdivided into others:

- **Research question 1**: `What learning technologies specifications are appropriate for representing learning objects in repositories?` – Dealing with learning resources and repositories, it is necessary to study the specifications that can be used to characterise learning resources in repositories;

- **Research question 2**: `What is semantics and pragmatics in the context of learning contents?` – It is important to clearly state what is semantics and pragmatics, in order to have learning contents "semantically and pragmatically" characterised;
- **Research question 3**: `How semantic and pragmatic concerns can be put in use in repositories?` – Having reached definitions of semantics and pragmatics, it is necessary to analyse how to put them in use in a repository, considering specifications that can be useful for this purpose;
- **Research question 4**: `How semantic and pragmatic characteristics can benefit repository users?` – The answer to this question should be translated into added-value services for the end-users.

In other to answer these research questions, the main objective of this work was to propose and validate a model that considers semantics and pragmatics in learning objects repositories to support the broad description of learning contents and additional services, using a typical repository platform for its implementation. This last concern was related to the desire to make the envisaged solution of practical usage in other repositories as well. The work must be supported in models, specifications and standards in the field of learning technologies, and in available practices and methods for semantic and pragmatic description. That major goal can be divided in more specific objectives:

- **Objective 1. State what is semantics and pragmatics in the context of learning contents.** Many adjectives, adverbs and nouns related to semantics and pragmatics are often applied, but the concepts of semantics and pragmatics need to be deeply analysed if they are to be used in the characterisation of learning resources, assuring their proper usage;
- **Objective 2. Consider knowledge in repositories and use semantics in the representation of learning contents and in the provision of additional search and retrieval possibilities.** Semantic concerns can enrich the description of resources (what they are about), and thus their findability and retrieval, but it is important to analyse how those can be considered in a common learning objects repository using a general platform and learning technology specifications;
- **Objective 3. Present a method of knowledge acquisition to be applied in a repository.** This objective is related to the previous one, because using knowledge in a repository requires its acquisition;
- **Objective 4. Provide a way to consider the users who submit resources, the people who use resources and pragmatic concerns in the characterisation of learning resources.** This should allow capturing the view of the two main types of repository users: providers and consumers, using appropriate learning technology specifications, for a more wide and pragmatic representation of resources;
- **Objective 5. Using a typical repository platform, study how the other Objectives can be systematised and implemented.** That objective is concerned with checking if the solutions followed for each previous objective is viable in practice or not.

1.3 Research methodology

The IEEE defines a methodology as "a comprehensive, integrated series of techniques or methods creating a general systems theory of how a class of thought-intensive work ought to be performed" [IEEE, 1995].

In the research carried out, the recommendations of David Avison and colleagues [Avison, Lau et al., 1999] were considered and also the phases for software engineering research outlined by Adrion [Adrion, 1993]. Therefore, the research was conducted iteratively through the following phases:

- **Analysis of existing solutions** - The first stage of an engineering method aims to identify problems to be addressed in the research from the study of existing proposals;

- **Proposition of a better solution** - After studying possible approaches to address the limitations identified in existing solutions, it is necessary to make a proposal that overcomes these limitations;
- **Development of a new solution** - At this phase the solution proposed in the previous one is developed. From the study carried out previously, the architecture of a new system that can surmount the limitations of existing solutions is considered. It was also developed a prototype of the system in accordance with the proposed architecture;
- **Evaluation of the adopted solution** - In the last stage of the engineering method it is necessary to verify if the solution developed overcomes the problems detected in the first phase of this methodology.

The research was based on a literature survey and empirical analysis in many phases. For the initial phases the participation in several conferences whose topics were related to this research work was important and also helpful in the proposition of a model for repositories. In addition, two summer schools provided some research directions to be analysed in different fields, as well as the opportunity to discuss many research questions with prominent lecturers and other participants:

- The Sixth Summer School on Ontological Engineering and the Semantic Web (SSSW'08) (http://kmi.open.ac.uk/events/sssw08/), Universidad Politècnica de Madrid, Cercedilla, Spain, 6-12 July 2008;
- PROLEARN 2006 Summer School (http://www.prolearn-academy.org/Academy%20Events/SS06), which covered advanced topics about computer assisted learning. It was in Bled, Slovenia, from 5 to 9 June 2006.

A description of the planned PhD work was sent and accepted in both events, and for the latter a scholarship was awarded to participate in the summer school.

The development phase was carried out iteratively and incrementally in order to obtain a suitable final solution. A prototype was implemented and its components were evaluated in many experiments with the help of experts in the related areas, also assessing the perceived usefulness of the different approaches. It culminated in the TREE (Teaching Resources for Engineering Education) repository, which is introduced in Chapter 6. In addition, the followed approaches were validated regarding their functionality, but they have to be tested in a wide scope, in very well-populated repositories and also in other initiatives.

1.4 Related publications

The PhD work described in this document resulted in some papers, which are presented in this section grouped by publication type. Some of the chapters of this document are extended versions of one or more of these articles.

Journal:

- Isabel Azevedo, Rui Seiça, Adela Ortiz, Eurico Carrapatoso, and Carlos Vaz de Carvalho, "A Semantic Approach for Learning Objects Repositories with Knowledge Reuse". P. Cimiano and H.S. Pinto (Eds.): EKAW 2010, Lecture Notes in Artificial Intelligence, 6317, pp. 580–589, Springer-Verlag Berlin Heidelberg;
- Isabel Azevedo, Carlos Vaz de Carvalho, and Eurico Carrapatoso, Effective Characterization of Learning Objects, International Journal of Advanced Media and Communication. ISSN (Online): 1741-8003 - ISSN (Print): 1462-4613, 2008.

Book chapter:

- Isabel Azevedo, Eurico Carrapatoso, and Carlos Vaz de Carvalho, "Caracterização Semântica e Pragmática de Objectos Educativos em Repositórios". TICAI 2007 - TICs Aplicadas a la enseñanza de la Ingeniería, 2008.

Divulgation publication:

- Isabel Azevedo, Rui Seiça, Adela Ortiz, Eurico Carrapatoso, and Carlos Vaz de Carvalho, “Using ontologies in Learning Objects Repositories”. IEEE Learning Technology Newsletter, Volume 13, Issue 1, ISSN 1438-0625, January 2011.

Conference Proceedings:

- Luciano Santos and Isabel Azevedo, “Using DBPedia for Computation of Semantic Relatedness (Relacionamento Semântico entre Conceitos com a Dbpedia)”. 6ª Conferência Ibérica de Sistemas e Tecnologias de Informação, Chaves, Portugal, 15-18 July 2011;
- Isabel Azevedo, Rui Seiça, Adela Ortiz, Eurico Carrapatoso, and Carlos Vaz de Carvalho, "Applying and reusing knowledge in a repository". The 10th IEEE International Conference on Advanced Learning Technologies (ICALT 2010), July 5-7, 2010, Sousse, Tunisia;
- Paulo Taveira and Isabel Azevedo, “Supporting the reuse of learning objects through the use of IMS LD templates”. 9th European Conference on e-Learning, 4-5 November 2010, Porto, Portugal;
- Rui Seiça, Adela Ortiz, Isabel Azevedo, Carlos Vaz de Carvalho, and Eurico Carrapatoso, “Tree - Um Repositório de Objectos de Aprendizagem para Engenharia baseado em Software Aberto (A Repository of Learning Objects for Engineering based on open source tools)”. FINTDI (Fomento e Innovación con Nuevas Tecnologías en la Docencia de la Ingeniería - Development and Innovation with New Technologies in the Teaching of Engineering) Conference. Vigo, Spain. 14 - 15 December 2009;
- Adela Ortiz, Isabel Azevedo, Carlos Vaz de Carvalho, Rui Seiça, and Eurico Carrapatoso, “Benchmarking de Herramientas para Manejo de Mapas Conceptuales en Ingenieria (Tools Benchmarking for Concept Mapping Management in Engineering)”. FINTDI (Fomento e Innovación con Nuevas Tecnologías en la Docencia de la Ingeniería) Conference. Vigo, Spain. 14 - 15 December 2009;
- Adela Ortiz, Isabel Azevedo, Rui Seiça, Eurico Carrapatoso, and Carlos Vaz de Carvalho, “Domain Ontology Creation Based on Automatic Text Extraction for Learning Objects Characterization”. 8th European Conference on eLearning - ECEL-2009. 29-30 October, Bari, Italy;
- Isabel Azevedo, Eurico Carrapatoso, and Carlos Vaz de Carvalho, "A Framework to Scaffold the Reuse of Learning Objects". The 8th IEEE International Conference on Advanced Learning Technologies (ICALT 2008), 1- 5 July 2008, Santander, Cantabria, Spain;
- Isabel Azevedo, Eurico Carrapatoso, and Carlos Vaz de Carvalho, "The use of templates to support learning design". The 2008 Frontiers in Education conference (ASEE/IEEE), October 22-25, 2008 in Saratoga Springs, NY;
- Isabel Azevedo, Eurico Carrapatoso, and Carlos Vaz de Carvalho. "Pragmatic Characterization of Learning Objects" - Cognition and Exploratory Learning in Digital Age (CELDA 2007). 7 - 9 December 2007, Algarve, Portugal;
- Isabel Azevedo, Eurico Carrapatoso, and Carlos Vaz de Carvalho. "Effective Characterization of Learning Objects" - Intelligent, Interactive, Learning Object Repository Networks Conference (I2LOR 2007), 4-7 November 2007, Montreal, Canada (the paper was selected for journal publication and an updated version was published in the International Journal of Advanced Media and Communication);
- Isabel Azevedo, Eurico Carrapatoso, and Carlos Vaz de Carvalho, “Caracterização semântica e pragmática de objectos educativos em repositórios”. 9th International Symposium on Computers in Education, 14 - 16 November 2007, Porto, Portugal (the paper was selected for book publication as a book chapter and an updated version was published in the TICAI 2007 - TICs Aplicadas a la enseñanza de la Ingeniería, 2008);

- Isabel Azevedo, Constantino Martins, Eurico Carrapatoso, and Carlos Vaz de Carvalho, "Learning Objects Repositories: how to reinforce reuse ". Interactive Computer Aided Learning Conference (ICL 2006), 27-29 September 2006, Villach, Austria;
- Isabel Azevedo, Constantino Martins, Eurico Carrapatoso and Carlos Vaz de Carvalho, "A model for the effective reuse of learning objects in repositories". The 6th IEEE International Conference on Advanced Learning Technologies (ICALT 2006), 5 - 7 July 2006, Kerkrade, Holanda;
- Isabel Azevedo, Constantino Martins, Eurico Carrapatoso, Carlos Vaz de Carvalho. "Contextualizing Learning Objects in Repositories" - 17th EAEEIE Annual Conference on Innovation in Education for Electrical and Information Engineering (EIE). 1-3 June 2006, Craiova, Romania.

1.5 Conventions and statements

Some conventions are used throughout this document, namely:

- Examples are written in a monospace font (Courier New) `like this`;
- The symbols used in the designation of description logics are represented in the font Lucida Handwriting;
- Latin words are presented in the italic type;
- Boldface characters are used to highlight a piece of text;
- Quotation marks are used to enclose direct quotation.

All used abbreviations are listed in a section at the beginning of this document, except those frequent in the English language or in logic and related fields, such as 'i.e.' (from Latin 'id est'), 'e.g.' (from Latin 'exempli gratia'), 'etc' (from Latin 'et cetera'), or 'iff' (meaning 'if and only if').

At the beginning of each chapter, a summary and a word cloud, a graphical representation of the most used words in the chapter, are provided.

Another convention followed in this document is related to the use of the following expressions: 'learning design' and 'Learning Design'. The latter is used to designate a specific learning design that follows the IMS Learning Design specification, while the former implies a generic learning design, usually represented textually.

Other expression that could be misleading is 'learning object'. Many others terms have a similar significance and are used interchangeably. CISCO uses the expression 'Reusable Information Object' (RIO) with quite the same meaning, but the definition stands that a RIO must have assessment items [CiscoSystems, 1999]. Other common expressions are: Educational Objects, Content Objects and Training Components [CiscoSystems, 1999].

A commonly adopted definition of learning object is "any entity, digital or non-digital, that can be used, re-used, or referenced during technology supported learning" [IEEE, 2002], which could be considered excessively broad. In this work it is followed Wiley's definition that only considers digital entities [Wiley, 2002], and the terms 'learning objects' and 'learning resources' are used interchangeably.

1.6 Thesis organisation

This thesis was divided into four main parts:

- **First Part: Introduction** – It is an initial piece that facilitates the understanding of the other parts;

- **Second Part: State of Art** – This part provides background knowledge and a general literature survey of learning technologies specifications, semantics, pragmatics, knowledge representation, and information retrieval areas;
- **Third Part: Contributions** - The third part contains research extensions on precise problems that arose from the main thesis research road while dealing with the research problem and answering some of the research questions;
- **Fourth Part: References and Appendices** – It encompasses annexes and bibliographic information that are not essential to the understanding of the topics discussed in the other parts, but complements them.

These parts comprise one or more chapters. The First Part only contains a single chapter, this one, which provides detailed information about the thesis scope and objectives, as well as the dissertation structure.

The Second Part is composed of four chapters as follows:

- **Chapter 2 - Learning Technology Standards, Specifications and Models** – It mainly examines deeply the specifications that allow characterising learning resources, but also those more related to learning objects repositories;
- **Chapter 3 - Some Topics on Pragmatics** – This chapter starts by discussing the concept of Semiotics, defining Semantics and Pragmatics. The latter field is analysed in detail in accordance with Linguistics and Philosophy areas, where the subject has a tradition that has not reached other areas;
- **Chapter 4 - Knowledge Representation and Engineering: Principles, Methods and Technologies** - A semantic characterisation needs knowledge representation to state the implicated meanings, but also some steps from knowledge engineering. These subjects are covered is this chapter;
- **Chapter 5 - Information Retrieval** – Learning object metadata is important to the findability of resources, but also the Information Retrieval features used in repositories. The semantic and pragmatic descriptions of resources can lead to new IR functionalities, and thus this chapter addresses this topic.

The third part is dedicated to the contributions of this PhD work, and includes Chapters 6, 7, 8 and 9:

- **Chapter 6 - Learning Objects Repository - Model and Implementation** – This chapter presents a model for repositories that considers the semantics and pragmatics of learning objects. Additionally, it discusses some of the preliminary work in the selection of a learning object repository platform and its preparation as the basis for all the implementation;
- **Chapter 7 - Semantic Component** – This chapter presents the semantic component that was designed and implemented in the TREE repository;
- **Chapter 8 - Pragmatic Component** – It discusses how pragmatics concerns was addressed in this work and achieved in the TREE repository;
- **Chapter 9 - Conclusions** – It provides final remarks, summarises the most relevant contributions, and states future research in the field.

The fourth part is composed of three annexes as follows:

- **Annex 2 -**

- **IMS LD Projects and** Tools – It provides information about projects and tools related to one of the specifications used in this research, a complex one. This information was also decisive in its adoption. This annex complements Chapter 2;
- **Annex 2 - Emotion** – This appendix is dedicated to the topic of emotions, in a supplement to its approach in the third chapter;
- **Annex 3 - Use Cases** – It adds more details for each use case briefly described in Chapter 6.

This last part also includes the bibliographic references cited along this document.

SECOND PART:
STATE OF ART

Chapter 2

LEARNING TECHNOLOGY STANDARDS, SPECIFICATIONS AND MODELS

A research work is unlikely to succeed and produce useful results without considering relevant specifications in the pertinent areas.

Learning technology standardisation is an area with intense activity over the last years. This chapter provides a review of the current state of the art of some learning technology specifications, especially those more related to the thesis statement (characterisation of learning resources).

2.1 Introduction

Standards can be defined as "documented agreements containing technical specifications or other precise criteria to be used consistently as rules, guidelines or definitions of characteristics, to ensure that the materials, products and services are fit for their purpose" [Friesen, 2002]. Standards are approved by national standards bodies recognised by national governments or by international standards bodies officially recognised by many national governments. All other bodies may produce specifications that can later turn into standards when they achieve a mature stage.

A standard can be a:

- *De jurus* standard - It is approved by national governments or international standards bodies (they will be designated only as standard in this document);
- *De facto* standard – It is widely followed, albeit missing the regulatory and approval process.

The term 'internal standard' is sometimes used to designate a set of agreed rules followed internally in an organisation [Horton and Horton, 2003]. Other terms that are also usually used in the context of learning technology standardisation efforts are [Friesen and McGreal, 2002]:

- **Specification** – It is not as advanced as a standard. After a long time getting consensus in the e-Learning community and getting the status of a de facto standard, a broadly used specification can finally be approved as a standard;
- **Application profile** [Duval, Hodgins et al. 2002] – It captures some elements from one or more standards or specifications, eventually refining them to create a set of metadata descriptors customised to a particular application whereas preserving interoperability with the original base schemas;
- **Reference model** - A reference model can be defined as "an adapted and reduced version of a combination of standards and specifications focusing on architectural aspects of an e-learning system, definitions of parts of the system, and their interactions" [Devedzic, Jovanovic et al., 2007].

Another common expression is 'open standard'. It means that the process that led to a standard or specification was open to many parties that contributed with suggestions, influencing the final available standard or specification. The key aspect is that the final result is publicly available. So, even commercial sectors can develop tools that use them, however that does not mean that these tools will be open source.

Standards and specifications in the technological education area have started to appear mainly to promote the reuse of learning objects. Although not sufficient, the use of standards and specifications is considered essential to promote the reuse of learning and educational material. Another aim was the establishment of a common understanding, which is also desired in other areas with well-known standards like the Digital Versatile Disk (DVD), or *de facto* standards like the MPEG-1/2 Audio Layer 3 (MP3) format.

The fields of interest in the standardisation of technological education include educational metadata, learning design, runtime environments, course organisation and packaging, and student models, among others.

Standards and specifications can be grouped into two groups [Anido-Rifón, Santos-Gago et al., 2002]:

- Information model specifications covering aspects like format, syntax and semantic of data to be transferred between different systems, e.g. IEEE LOM, Dublin Core, SCORM, IMS CP and IMS LD;
- Specifications related with more technical aspects, such as architectures, software components and interfaces, like the IEEE LTSC Learning Technology Systems Architecture (LTSA).

Considering another categorisation criterion, the current specifications fall into the following groups [Muñoz-Merino, Kloos et al., 2009]:

- **Architectural specifications**. They specify architectures relating to learning components. The IEEE LTSA belongs to this group;
- **Data specifications**. They propose data models and related bindings that facilitate the sharing of information between different learning platforms;
- **Behavioural specifications**. They specify programming interfaces for educational modules, allowing communication calls between different learning systems, like the OSIDs (OKI Service Interface Definition) of OKI.

Some of the most relevant specifications are shown in Table 1. A detailed overview of these specification and others is provided by [CEN-WS LT, 2010].

Table 1. Some important specifications

Acronym	Designation	Area	Organisation	Web Address
AICC PENS	AICC Package Exchange Notification Services	Digital repositories	AICC	http://www.aicc.org/docs/tech/cmi010v1a.pdf
CEN SQI	Simple Query Interface	Digital repositories	CEN	ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/WS-LT/CWA15454-00-2005-Nov.pdf
CORDRA	Content Object Repository Discovery and Registration/ Resolution Architecture	Digital repositories	ADL, CNRI and LSAL	http://cordra.net
IMS CP	IMS Content Packaging	Educational Resource Packaging and Organisation	IMS	http://www.imsglobal.org/content/packaging/
IMS DR	IMS Digital Repositories	Interoperability	IMS	http://www.imsproject.org/digitalrepositories/
IMS LD	IMS Learning Design	Learning Design	IMS	http://www.imsglobal.org/learningdesign/
IMS LIP	Public and IMS Learner Information Package	Learner Information	IMS	http://www.imsproject.org/profiles/
IMS LODE	IMS Learning Object Discovery & Exchange	Digital repositories	IMS	http://imsglobal.org/LODE/
IMS QTI	IMS Question and Test Interoperability	Evaluation	IMS	http://www.imsproject.org/question/
IMS SS	IMS Simple Sequencing	Educational Resource Packaging and Organisation	IMS	http://www.imsproject.org/simplesequencing/
ISO MLR	Metadata for Learning Resources [ISO/IEC, 2009]	Metadata	ISO/IEC	http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?ics1=35&ics2=240&ics3=99&csnumber=50772
LOM	Learning Object Metadata	Metadata	LTSC-IEEE	http://ltsc.ieee.org/doc/wg12/LOM_1484_12_1_v1_Final_Draft.pdf

SCORM-CAM	Content Aggregation Model	Educational Resource Packaging and Organisation	ADL	http://www.adlnet.gov/Technologies/scorm/SCORMSDocuments/SCORM%202004%204th%20Ed%20V1.1/Documentation%20Suite/SCORM_2004_4ED_v1_1_Doc_Suite.zip
SCORM-RTE	SCORM Runtime Environment	Runtime	ADL	http://www.adlnet.gov/Technologies/scorm/SCORMSDocuments/SCORM%202004%204th%20Ed%20V1.1/Documentation%20Suite/SCORM_2004_4ED_v1_1_Doc_Suite.zip

The next sections (2.2 Specifications related to resources' metadata, 2.3 Content Aggregation, 2.4 Educational Modelling Languages and 2.5 Digital Repositories Specifications) provide a complete summary of learning standards and specifications related to learning objects or repositories somewhat connected to the research conducted and described in this document. SCORM is presented apart, in section 2.6, because it is a model, not a specification, and cannot be addressed in just one of the areas described in the other sections.

Subsection 2.1.1 provides an overview of the main educational standardisation bodies.

2.1.1 Educational standardisation bodies

Many organisations have been working on learning standards and specifications (see Table 2 for an overview).

Table 2. Main organisations that work on the standardisation of learning technologies

Acronym	Institution	Web Address
ADL	Advanced Distributed Learning Initiative	http://www.adlnet.gov/
AICC	Aviation Industry CBT Committee	http://aicc.org/
ARIADNE	Alliance of Remote Instructional Authoring and Distribution Networks for Europe	http://www.ariadne-eu.org/
CanCore	Canadian Core Learning Resource Metadata Initiative	http://www.cancore.ca/
CEN/ISSS	European Committee for Standardisation - Information Society Standardisation System	http://www.cen.eu/isss/Workshop/lt/
DCMI	Dublin Core Metadata Initiative	http://dublincore.org/
EdNA	Education Network Australia	http://www.edna.edu.au/
GEM	Gateway to Educational Materials	http:// www.thegateway.org/
IEEE LTSC	Institute of Electrical and Electronics Engineers Learning Technologies Standardisation Committee	http://ieeeltsc.org/
IMS	IMS Global Learning Consortium	http://www.imsproject.org
ISO IEC/JTC1 SC36	International Standards Organisation, Joint Technical Committee 1, Sub-Committee 36 - Information Technology for Learning, Education and Training	http://jtc1sc36.org/Training
JISC/CETIS	Joint Information Systems Committee - Centre for Educational Technology and Interoperability Standards	http://jisc.cetis.ac.uk/
OKI	Open Knowledge Initiative	http://www.okiproject.org/
SIFA	Schools Interoperability Framework Association	http://www.sifinfo.org/

The educational standardisation bodies responsible for the standards and specifications discussed in this chapter are:

- **IEEE Learning Technologies Standardisation Committee (LTSC)** – It is mainly devoted to the development of technical standards and recommendations. As it will be discussed in section 2.2.2, LTSC was responsible for the first standard for learning technologies, IEEE LOM;
- **IMS Global Learning Consortium** – It is a non-profit organisation that was launched in 1997 by EDUCAUSE, a consortium of North American educational institutions and their industrial partners. IMS previously stood for Instructional Management System; however the work of IMS goes beyond instructional management systems. Nowadays, the main achievements of the IMS are in the fields of metadata (IMS metadata), content packaging (IMS CP), learning design (IMS LD), test definition (IMS QTI), and student profiling (IMS LIP). All IMS specifications are compatible with IEEE LOM;
- **Advanced Distributed Learning (ADL) initiative** – It was launched by the US Department of Defence and the White House Science and Technology Bureau in 1997. Its work is coordinated with others, like the two mentioned before, for instance. It produced the Sharable Courseware Object Reference Model;
- **The Dublin Core Metadata Initiative (DCMI)** – It is an open organisation dedicated to the development of interoperable online metadata standards. DCMI's activities include work on architecture and modelling, discussions and collaborative work in DCMI Communities and DCMI Task Groups, annual conferences and workshops, standards liaison, and educational efforts to promote widespread acceptance of metadata standards and practices;
- **European Committee for Standardisation (in French: *Comité Européen de Normalisation* – CEN)** – With two exceptions, in the electrotechnology and telecommunication areas, “it is the only recognised European organisation ... for the planning, drafting and adoption of European Standards in all areas of economic activity” [CEN, 2009]. It hosts the Information Society Standardisation System (ISSS) subcommittee;
- **International Organisation for Standardisation / International Electrotechnical Commission** - International Electrotechnical Commission (IEC) and International Organisation for Standardisation (ISO) are two of three global sister organisations (IEC, ISO, ITU) that develop International Standards for the world. They can cooperate and establish Joint Technical Committees (JTC) with the necessary experts to work on standards. JTC1 is dedicated to the development of global Information and Communication Technologies (ICT) standards for business and consumer applications.

These institutions are not at the same level, as can be seen in Figure 1: IEEE LTSC is an accredited organisation, as CEN/ISSS WSLT (http://www.cenorm.be/isss/Workshop/lt/) and ISO/IEC JTC1 (http://isotc.iso.org) – this one is more global than IEEE LTSC and CEN/ISSS WSLT. While IMS GLC and ADL (as well as many others) are on the lower level, they actively contribute to accredited organisations.

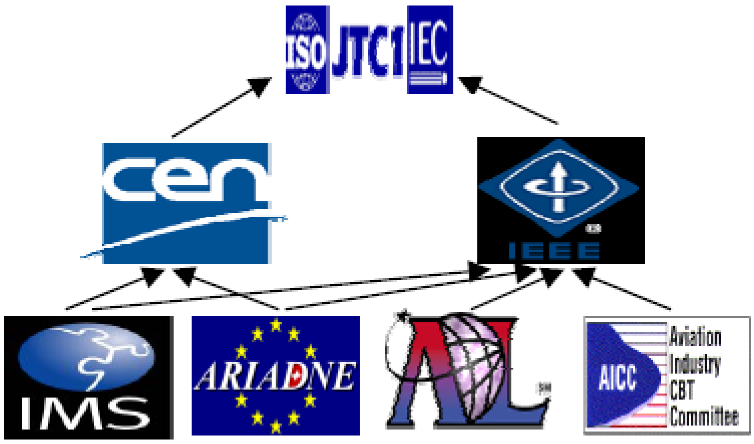


Figure 1. Learning Technology Standardisation Organisations (Source: [Duval, 2004])

2.2 Specifications related to resources' metadata

The Greek prefix *meta* means "with, across, or after", according to the Oxford Advanced Learner's Dictionary. Thus, metadata accompanies the data themselves, and may not be considered essential. It is common to talk about administrative, structural and descriptive metadata, in accordance with the metadata purpose.

This section provides an overview of two of the most used specifications about metadata for educational resources: Dublin Core (section 2.2.1) and IEEE LOM (section 2.2.2). Section 2.2.3 supplies some information about the ISO/IEC 19788 multipart standard "Metadata for Learning Resources (MLR)", which was published in 2011.

2.2.1 Dublin Core

Dublin Core (DC) has become the basis for many other initiatives [Anido, Fernandez et al., 2002]. The word 'Dublin' in the name is a reference to the city (Dublin, Ohio) where an invitational workshop took place in 1995, which aggregated people with diverse backgrounds: professionals from the field of librarianship, information technology, text encoding, and people from the museum community, among others. The word 'Core' emphasises that it encompasses just some core elements. Its simplicity is both an advantage and a disadvantage, as it is not difficult and expensive to use but really does not make it possible to precisely describe the semantics of learning objects.

Many institutions have developed DC application profiles, like the OAI-DC application profile used by all Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) data providers [Sompel, Nelson et al., 2004], and RDN-DC application profile for the Resource Discovery Network [Day and Cliff, 2009].

DC application profiles were also developed for the description of learning objects, like the Education Network Australia (EdNA) Metadata Standard, which included additional elements (EDNA.Audience, EDNA.Approver, EDNA.CategoryCode, EDNA.Entered, EDNA.Indexing, EDNA.Review, EDNA.Reviewer, EDNA.Version) and qualifiers (EDNA.Sector and EDNA.UserLevel) [Millea, 2003]; the Renardus Dublin Core Application Profile developed under a project that aimed to provide a reliable source of resources for teaching, learning, and researching purposes in higher education in Europe [Neuroth and Koch, 2001]; and the Gateway to

Educational Materials (GEM) element set developed under the GEM Project[1], which is composed by 23 Elements (15 are DC elements) [Forger, 2003]. These data are summarised in Table 3.

Table 3. Some DC application profiles

Application Profile	Number of elements	URL
Education Network Australia (EdNa)	23 (15 + 8)	http://www.edna.edu.au/edna/go/resources/metadata/edna_metadata_profile
Gateway to Educational Materials (GEM)	23 (15 + 8)	http://jesandco.net/gateway/GemTools.Asp
Renardus	12 (9 + 3)	http://renardus.sub.uni-goettingen.de/renap

Even the DCMI is working in a DC Educational Application Profile[2], initially to be finished by mid-2008[3], but still in progress.

The Dublin Core model consists of two levels, Simple Dublin Core and Qualified Dublin Core, which are discussed in the next subsections.

2.2.1.1 Simple Dublin Core

Simple Dublin Core - It is a set of 15 elements, which is known as Dublin Core Metadata Element Set (DCMES). It enables the characterisation of general resources. The fifteen elements are: 1. Title, 2. Creator, 3. Subject, 4. Description, 5. Publisher, 6. Contributor, 7. Date, 8. Type, 9. Format, 10. Identifier, 11. Source, 12. Language, 13. Relation, 14. Coverage, 15. Rights. They are not mandatory and may occur many times. The Dublin Core metadata set, which is composed by these 15 elements, has been formally endorsed by the following accredited standards (DCMI, 2004):

- ISO Standard 15836-2003 of February 2003 [ISO, 2003];
- ANSI/NISO Standard Z39.85-2007 of May 2007 [NISO, 2007];
- IETF RFC 5013 of August 2007 [IETF, 2007].

2.2.1.2 Qualified Dublin Core

In 2000 some changes to Simple Dublin Core produced what is called Qualified Dublin Core, which extends and refines the former. An additional element was considered, `audience`, which specifies to whom the work is intended to. Also, the DCMI issued a list of qualifiers. There are two classes of qualifiers: Element Refinement (see Table 4) and Encoding Scheme. The latter refers to controlled vocabularies used to restrict the possible values of some metadata elements, while the former allows the specification of additional and more specific information. For example, ISO 639-2 is one of the Encoding Schemes recommended for the Language element, which uses `por` for Portuguese.

[1] The Gateway to Educational Materials (GEM) Project was funded by the U.S. Department of Education's National Library of Education and a special project of the ERIC Clearinghouse on Information & Technology. It aimed to make finding educational resources efficient and effective.

[2] http://dublincore.org/educationwiki/DC_2dEducation_20Application_20Profile

[3] http://dublincore.org/educationwiki/DC_2dEducation_20Application_20Profile_20Task_20Group

Table 4. Qualifiers used for element refinement

DCMES Element	Qualifiers for Element Refinement
Title	Alternative
Creator	-
Subject	-
Description	Table Of Contents, Abstract
Publisher	-
Contributor	-
Date	Created, Valid, Available, Issued, Modified
Type	-
Format	Extend, Medium
Identifier	-
Source	-
Language	
Relation	Is Version Of, Has Version, Is Replaced By, Replaces, Is Required By, Requires, Is Part Of, Has Part, Is Referenced By, References, Is Format Of, Has Format
Coverage	Spatial, Temporal
Rights	-
Audience	-

More information about their use can be found at [DCMI, 1999, 2008].

2.2.2 IEEE LOM

The group Learning Technology Standards Committee (LTSC) of the Institute of Electrical and Electronics Engineers (IEEE) has worked on a metadata specification designed for education, learning and educational management, the Learning Object Metadata (LOM) standard (1484.12.1-2002), with great contribution from IMS. Since version 1.3 of IMS Learning Resource Metadata (LRM) the IEEE XML binding is used for both specifications. Actually, the term LOM can be used to refer to IEEE LOM and IMS LRM.

The IEEE approved the first version of the standard in June of 2002. LOM was renewed on 13 May 2009 by the IEEE-SA (Standards Association) Standards Board, and is in its second 5 year period as an active standard.

LOM was the first accredited standard for learning technologies. The fundamental purpose of using metadata in education contexts is not different from the objective in other areas: it enables the search and discovery of the desired learning objects without having to inspect all available ones.

The LOM standard was influenced by Dublin Core, and the Education Working Group of the Dublin Core Metadata Initiative (DCMI) is working together with the IEEE in order to harmonise their metadata developments (Dublin Core and LOM).

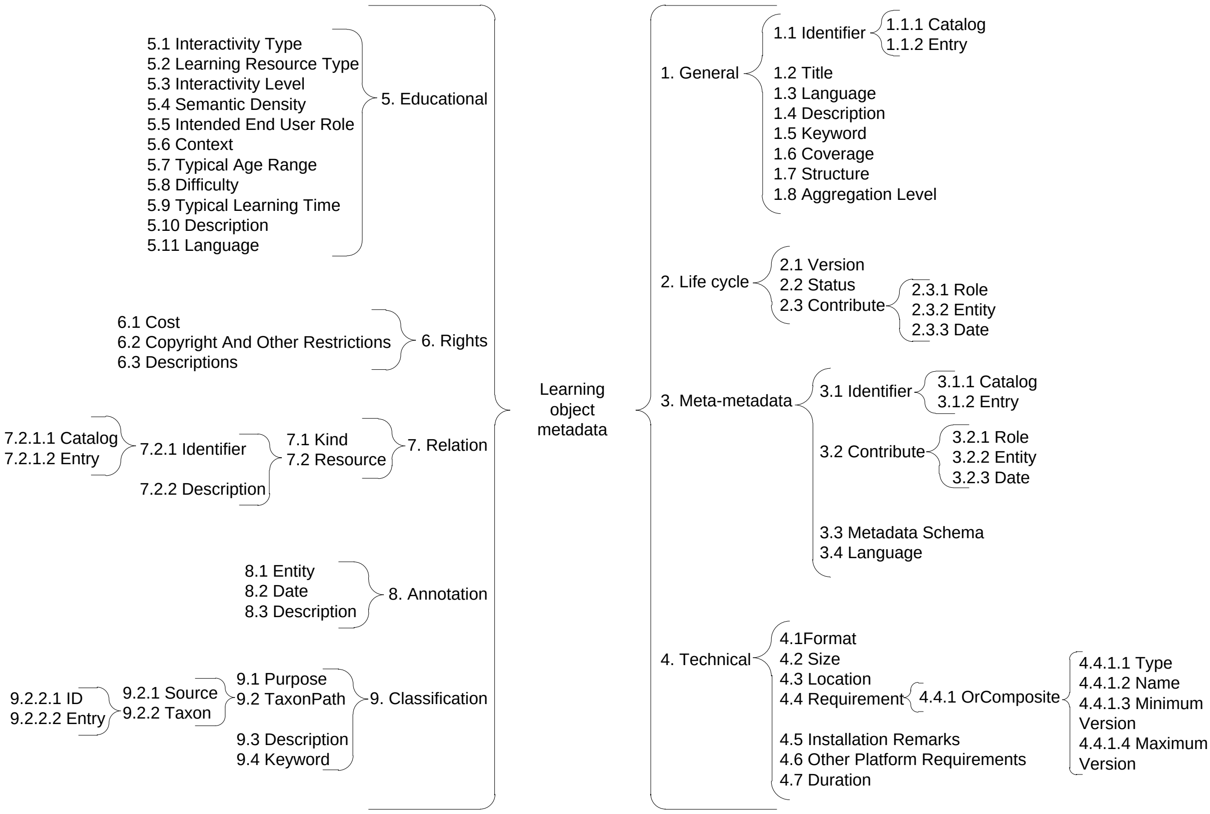


Figure 2. LOM structure

As shown in Figure 2, there are two types of elements: some elements contain others, e.g. elements 3.Meta.metadata and 7.2.1.Identifier, and other elements contain data, such as 4.4.1.1 Type.

The LOM standard contains 77 optional elements grouped into nine categories (also LOM elements), to describe a learning object:

- **1. General (10 elements)** – This category groups general information about the learning object like the author and the creation date;
- **2. LifeCycle (6)** - It groups elements that state the chain of versions and the related information;
- **3. Meta-Metadata (9)** – This category includes descriptive information about the metadata record itself, like when it was created or the provider of this kind of information (element 3.2 – Contribute);
- **4. Technical (12)** – The technical category groups the technical requirements and characteristics of the learning object;
- **5. Educational (11)** – It intends to state the educational and pedagogical characteristics of the learning object;
- **6. Rights (3)** – This class groups the intellectual property rights elements and in what conditions the learning object can be used;
- **7. Relation (6)** – This category is derived from the DC and it makes explicit the relationship between a particular learning object and another element identified by '7.2:Relation.Resource', if any exists. The meaning of these elements is given in informal

English, so it is not an unambiguous approach to state the possible relations between two learning objects[4];

- **8. Annotation (3)** – This class of elements supplies notes about the educational use of the learning object and provides information on when and by whom the notes were supplied;
- **9. Classification (8)** – It describes the learning object in relation to a specific classification system.

The lifecycle and meta-metadata categories are not related to educational purposes, but all the other ones, and not only the educational category, can have, directly or not, an instructional meaning.

LOM is a reference standard for educational systems that manage learning objects of different types. However, there have been many criticisms about some elements of the standard, but it is considered that some freedom must be given to researchers and developers in order not to restrict different approaches. For example, the granularity element of LOM can have four different values, each one encompassing the previous one, but their significance is not exactly stated.

The IEEE LOM defines some vocabularies for describing educational attributes of learning objects, specially for the characterisation of the elements of category `5: Educational`, including: `5.1 Interactivity Type`, `5.2 Learning Resource Type`, `5.3 Interactivity Level`, `5.4 Semantic Density`, `5.8 Difficulty`, and `5.10 Description`.

The IEEE LOM standard describes the `Learning Resource Type` element as "Specific kind of learning object. The most dominant kind shall be first", recommending the following vocabulary values: `exercise`, `simulation`, `questionnaire`, `diagram`, `figure`, `graph`, `index`, `slide`, `table`, `narrative text`, `exam`, `experiment`, `problem statement`, `self assessment`, and `lecture`. This suggested vocabulary is specially criticised for combining terms for physical resource type, like `slide` and `table`, and terms that suggest a pedagogical approach, like `lecture` or `self-assessment`. The vocabulary for `Learning Resource Type` has been considered inadequate by some implementers and there are a number of application profiles worldwide that have defined their own `Learning Resource Type` vocabularies [Currier and Campbell, 2005].

The IEEE LOM standard has an XML binding (IEEE 1484.12.3), approved by IEEE and ANSI, and although some work was carried to define an RDF binding (IEEE 1484.12.4) [Nilsson, Palmér et al., 2003], at this moment there is only a proposal.

In practice the description of a learning object using LOM is achieved by adding an XML metadata record for the learning object into the metadata element of a content package, so the metadata will be available always that the learning object itself is available. It is also common to have the metadata stored apart, used in indexing methods to later access the related learning object.

In 2003 an empirical analysis on the metadata instances stored in the ARIADNE Knowledge Pool System (KPS) concluded that the most filled-in optional metadata element was the granularity one (at about 92%), while others were used in about 50 % of the descriptions and the rest were rarely used [Najjar, Ternier et al., 2003].

Several application profiles of the LOM standard were developed, such as ARIADNE [Ariadne], CanCore [Cancore], Heal (Health Education Assets Library) [Heal] and UK LOM Core [UKLOM] (see Table 5). They vary not only in the number of elements considered, but also in their purpose. The UK LOM core (originally the UK Common Metadata Framework), for instance, tried to capture the common practice and was developed to be an educational reference in UK. It defines a minimum set of metadata elements and the related domain value for each of them.

On the other side, metadata in the ARIADNE Foundation is mainly used to make easy the share and reuse of learning objects [Duval, Forte et al., 2001]. The ARIADNE AP only allows one

[4] The type of relation can be described using the sub-element 'Kind', which can be filled with one of 12 predefined values based on the corresponding element of the Dublin Core Element Set: `ispartof`, `haspart`, `isversionof`, `hasversion`, `isformatof`, `hasformat`, `references`, `isreferencedby`, `isbasedon`, `isbasisfor`, `requires`, or `isrequiredby`.

instance of some data elements that in LOM may have many, such as the General/Language element.

Table 5. Some LOM application profiles

Application Profile	Number of elements	URL
ARIADNE	50	http://www.ariadne-eu.org/en/publications/metadata/index.html
CanCore	61	http://www.cancore.ca/documents.html
HEAL	52 + 15	http://www.healcentral.org
UK LOM Core	77	http://www.cetis.ac.uk/profiles/uklomcore

Besides some LOM elements, the HEAL application profile considers 15 metadata elements, which are mainly related to the health area, such as General.Extension.RadiographType, General.Extension.DiseaseProcess, General.Extension.ClinicalHistory, among others.

The CanCore application profile, for example, contains 61 elements from LOM (see Table 6 – M and O are used to denote mandatory and optional fields, respectively). This is one of the more used LOM application profile and section 2.2.2.1 describes it in more detail.

Table 6. LOM elements considered in some application profiles

IEEE LOM		ARIADNE	CanCore	HEAL	UK LOM Core
LOM number	LOM name				
1	General	M	O	M	M
1.1	Identifier		O	M	M
1.1.1	Catalog	M	O	M	M
1.1.2	Entry	M	O	M	M
1.2	Title	M	O	M	M
1.3	Language	M	O	O	M
1.4	Description	O	O	M	M
1.5	Keyword		O		O
1.6	Coverage			O	O
1.7	Structure				O
1.8	Aggregation Level	O	O		O
2	Life Cycle	M	O	M	M
2.1	Version	O	O		O
2.2	Status				O
2.3	Contribute		O	M	M
2.3.1	Role	M	O	M	M
2.3.2	Entity	M	O	O	M
2.3.3	Date	O	O	O	M
3	Meta-MetaData	M	O	M	M
3.1	Identifier		O	M	M
3.1.1	Catalog		O	M	M
3.1.2	Entry		O	M	M
3.2	Contribute		O	M	M
3.2.1	Role	M	O	M	M
3.2.2	Entity	M	O	O	M
3.2.3	Date	M	O	O	M
3.3	Metadata Scheme		O	O	M

IEEE LOM		**ARIADNE**	**CanCore**	**HEAL**	**UK LOM Core**
LOM number	**LOM name**				
3.4	Language	M	O	O	M
4	Technical	M	O	M	M
4.1	Format	M	O	M	O
4.2	Size	M	O	M[5]	O
4.3	Location	M	O	M	M
4.4	Requirement			O	O
4.4.1	Or Composite				O
4.4.1.1	Type	M		O	O
4.4.1.2	Name	M		O	O
4.4.1.3	Minimum Version	O			O
4.4.1.4	Maximum Version				O
4.5	Installation Remarks	O			O
4.6	Other Platform Requirements	O	O		O
4.7	Duration		O	O	O
5	Educational	M	O	O	M
5.1	Interactivity Type	M		O	O
5.2	Learning Resource Type	M	O	O	O
5.3	Interactivity Level	O	O	O	O
5.4	Semantic Density	O			O
5.5	Intended End User Role	M	O		O
5.6	Context	O	O	O	O
5.7	Typical Age Range		O		O
5.8	Difficulty	O			O
5.9	Typical Learning Time	M	O	O	O
5.10	Description				O
5.11	Language		O		O
6	Rights	M	O	M	M
6.1	Cost	M	O		O
6.2	Copyright And Other Restrictions	M	O	M	M
6.3	Descriptions	M	O		M
7	Relation	O	O		O
7.1	Kind	O	O	O	O
7.2	Resource		O	O	O
7.2.1	Identifier		O	M	O
7.2.1.1	Catalog		O	M	O
7.2.1.2	Entry		O	M	O
7.2.2	Description	O		O	O
8	Annotation	O	O		O
8.1	Entity	M	O		O
8.2	Date	M	O		O

[5] Except for Web pages.

IEEE LOM		ARIADNE	CanCore	HEAL	UK LOM Core
LOM number	LOM name				
8.3	Description	M	O		O
9	Classification	M	O	O	O
9.1	Purpose	M	O		O
9.2	TaxonPath		O	O	O
9.2.1	Source	M	O	O	O
9.2.2	Taxon		O	O	O
9.2.2.1	ID		O	O	O
9.2.2.2	Entry	M	O	O	O
9.3	Description				O
9.4	Keyword		O	O	O

2.2.2.1 CanCore

Canadian Core Learning Object Metadata Application Profile (CanCore) was not born from a single project but from the needs identified in a number of e-learning projects sponsored by Canadian institutions for higher education and k-12 educational domains. Its development started in 2000 and later it was supported by the eduSource project. It has been widely accepted by many institutions and projects [MacLeod, 2005].

CanCore has the same 9 categories of LOM, but only 44 'leaf' elements, all optional, even the title element, for instance. CanCore mainly considers those elements regarded as critical for supporting the discovery and reuse of learning objects. For instance, most elements of the 4th LOM category (Technical) are not adopted in CanCore (see Table 6).

However, there are CanCore recommendations available for all the IEEE LOM fields, in an attempt to significantly decrease the complexity of the LOM standard and the uncertainty in the use of some elements [Fisher, Tozer et al., 2002].

Cancore profile has been used for years by a large number of LOM implementers [Friesen, 2005].

2.2.3 ISO MLR

A promising initiative is the ISO Metadata for Learning Resources (http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?ics1=35&ics2=240&ics3=99&csnumber=50772). Since 2011-01-07, it is in the stage 60.60 (International Standard published). It is not available without charge, but from the draft versions made accessible[6], the summary provided by CEN WS_LT [CEN WS_LT, 2011b] and the discussions aroused in the Dublin Core Mailing List (https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=dc-education), it can be said that ISO MLR is based on Dublin Core, with one of its parts providing an application profile that combines DC and LOM elements.

According to CEN WS_LT, this multi-part standard is composed of the following parts [CEN WS_LT, 2011a]:

- Part 1. Framework,
- Part 2. Core Elements,

[6] Draft documents: http://isotc.iso.org/livelink/livelink/fetch/-8917700/8917725/8917726/36N1524_Text_of_ISO_IEC_CD2_19788-1%2C_ITLET_-_Metadata_for_Learning_Resources_-_Part_1._Framework.pdf?nodeid=6452873&vernum=-2 and http://isotc.iso.org/livelink/livelink/6452875/36N1525_Text_of_ISO_IEC_CD_19788-2__ITLET_-_Metadata_for_Learning_Resources_-_Part_2._Core_elements?func=doc.Fetch&nodeid=6452875

- Part 3. Core Application Profile,
- Part 4. Technical Elements,
- Part 5. Educational Elements,
- Part 6. Availability, Distribution, and Property Elements.

2.3 Content Aggregation

For the sharing of educational resources among different systems a number of specifications were proposed in the last years for their encapsulation. Some of them are:

- ADL SCORM Content Aggregation Model,
- IMS Common Cartridge,
- IMS Content Packaging,
- IMS Resource List Interoperability.

In the next subsections IMS Content Package and IMS Resource List Interoperability are explored to provide an overview of the questions addressed in this field. In addition, these are two of the earliest initiatives in the area, which also inspired some others.

2.3.1 IMS Content Package

The IMS Content Package (CP) specification allows users to describe and state the structure of learning objects in a common format [IMS, 2000a], enabling interoperability between different systems with capability to import, export, aggregate and disaggregate packages of content.

The core element of the specification is the **Package**. A Package represents anything with meaning that can be seen as a unit of usable (and reusable) content. It can be a whole course or a part of one, but with significance alone. A **Package Interchange File** is a Package incorporated into a single file for transportation (e.g., .zip, .jar, .cab file).

A content package consists of two parts (see Figure 3):

- The physical files included in the package,
- An XML file called **manifest**, which states the static sequence of the learning objects included in the package.

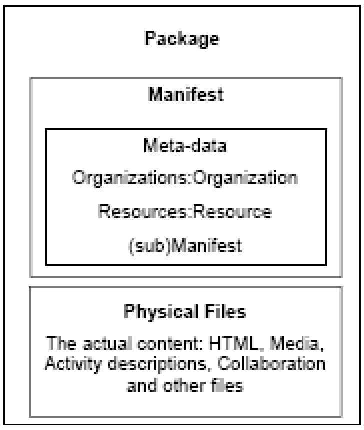


Figure 3. IMS Content Package structure (Source: [IMS, 2000a])

A manifest contains the following sub-elements:

- **metadata** (optional) - It describes the manifest using elements from the IMS Meta-Data Specification (the latest version) or from any other included via an XML namespace in a manifest's meta-data element(s);
- **organizations** (required) - An 'organization' is a hierarchy of items that forms a structured view into the resources;
- **resources** (required) - A resource may refer to a physical file within the package or to an external resource;
- **manifest** (optional) - It specifies zero or more (sub)manifests.

The IMS Content Packaging Specification 1.0 was released in June 2000. V1.1.4 is the current version and the next one, V2, is expected to be released in the near future. This version of the IMS Content Packaging specification is composed of four documents:

- IMS Content Packaging XML Binding Specification,
- IMS Content Packaging Information Model,
- IMS Content Packaging Best Practice Guide,
- IMS Content Packaging XML Schema.

IMS CP is the most used learning technology specification in the world [CEN WS-LT, 2010]. This specification is currently being considered as a new work piece for standardisation in ISO/IEC JTC1 SC36 (projects ISO/IEC FCD 12785-2 and ISO/IEC FCD 12785-3), and it may become an ISO/IEC standard.

2.3.2 IMS Resource List Interoperability

The Resource List Interoperability (RLI) specification (http://www.imsglobal.org/rli/) addresses the interoperability of resource lists between repositories and e-learning systems. It specifies how structured meta-data can be exchanged between systems that store and expose resources for the purpose of creating resource lists, such as a reading list, and systems that collect and organise those lists for educational purposes.

Although IEEE LOM per se is not sufficient to describe the reading list resources common in the library and publishing communities that the RLI specification addresses, it provides a base for the development of the Resource List Interoperability data model.

The current version (1.0) was released on 08 July 2004 and it is composed of four documents:

- IMS Resource List Interoperability Best Practice and Implementation Guide,
- IMS Resource List Interoperability Information Model,
- IMS Resource List Interoperability XML/WSDL Binding,
- IMS Resource List Interoperability Conformance Requirements.

2.4 Educational Modelling Languages

Koper and Olivier argue that some e-learning specifications like the IEEE LOM consider learning in a perspective of contents only, and that this narrow view is not adequate as the current educational practice is more complex than this [Koper and Olivier, 2004].

Many EML (Educational Modelling Languages) emerged intending to describe pedagogical models, modelling educational units, like CDF (Curriculum Description Format) [Verbert and Duval, 2004], LMML (Learning Material Markup Language) [Slavin, 1995], PALO (http://sensei.lsi.uned.es/palo/), Targeteam (TArgeted Reuse and GEneration of TEAching Materials - http://www.targeteam.org/), TML/Netquest ((http://www.ilrt.bris.ac.uk/netquest) [Brickley, 1996], the OUNL (Open University of the Netherlands) EML [Koper, 2001] and PoEML

(Perspective-oriented Educational Modelling Language) [Caeiro-Rodríguez, Llamas-Nistal et al., 2006].

A learning design is a schema of a unit of study, with instructional purpose(s), the succession(s) of activities to be followed by the learner, an explanation of the learning surroundings, the players in the learning activities, and the resources and services necessary for the learning activities.

The IMS Learning Design (IMS LD) specification [IMS, 2003c], which incorporates the OUNL EML, has emerged as the de facto standard for the representation of any learning design, independently of the pedagogical structures been applied. This specification is fully described in the next section.

2.4.1 IMS LD

IMS LD was released in February 2003 and aims to provide a notational method to specify any instructional design in a formal way, without prescribing any instructional design, separating contents from context of use: "By labelling the strategy and the components of the strategy in a common, machine-readable manner, the context of a learning opportunity can be managed separately from the content itself" [IMS, 2003b]. Its information model integrates a subset of the Educational Modelling Language (EML) of the Open University of the Netherlands, mentioned before, and it has an XML binding.

An interesting aspect of the IMS LD specification is that it is possible to use or incorporate other existing specifications and standards through the mechanisms of XML Namespaces [IMS, 2003c], namely:

- IMS Content Packaging,
- IMS Simple Sequencing [IMS, 2003d],
- IMS/LOM Meta-Data [IMS, 1999b],
- IMS Question and Test Interoperability [IMS, 2000b],
- IMS Reusable Definition of Competency or Educational Objective (RDCEO) [IMS, 2002],
- IMS Learner Information Package [IMS, 2001],
- IMS Enterprise [IMS, 1999a],
- SCORM [ADL, 2000].

For this reason, IMS LD has been considered as an integrative layer to other specifications [Paquette and Marino, 2006]. However, the best way to integrate IMS LD with some specifications has been the matter of a number of research works, namely the integration between IMS LD and SCORM [Tattersall, Burgos et al., 2006], IMS LD and IMS QTI [Miao, Vogten et al., 2007], while the mixture with some others can be considered an easy task, for example, IMS LD and IMS CP, as it will be seen later in this section. In fact, it is recommended the integration of the IMS Learning Design into an IMS Content Package to generate a 'Unit of Learning', or UoL, for short.

LD consists of three documents:

- IMS Learning Design Information Model,
- IMS Learning Design Information Binding,
- IMS Learning Design Best Practices Guide.

Additionally, there are a set of examples and learning scenarios developed in XML available at http://www.imsglobal.org/learningdesign/. None of the LD documents specify a content model, any format is allowed to model contents, even though XHTML is suggested[7].

[7] This is one of the differences between EML and LD, as the former contains a content model, so a content can be modelled 'in EML', but not 'in LD'.

IMS LD is an open learning design specification that is gathering great attention from the e-learning community. It has three levels: A (the most basic level), B and C (the most complete one). Level B adds `Properties` and `Conditions` to level A, which enable personalisation, not possible in Level A, and more sophisticated sequencing and interactions based on learner characteristics. LD Level C adds `Notification` to level B.

Each level has a separate XML schema.

The main elements of the LD specification are hierarchically organised as follows (a plus symbol means that an element may occur many times):

```
learning-design
   title
   learning-objectives
   prerequisites
   components
      roles
         learner+
         staff+
      activities
         learning-activity+
            environment-ref+
            activity-description
         support-activity+
            environment-ref+
            activity-description
         activity-structures+
            environment-ref+
      environments
         environment+
            title
            learning objects+
            services+
            environment-ref+
            metadata
method
      play+
         act+
            role-parts+
               role-ref
               activity-ref
metadata
```

2.4.1.1 Conceptual model

The conceptual model of the semantic aggregations levels in the LD specification is represented in Figure 4, in which the central concepts of LD can be observed.

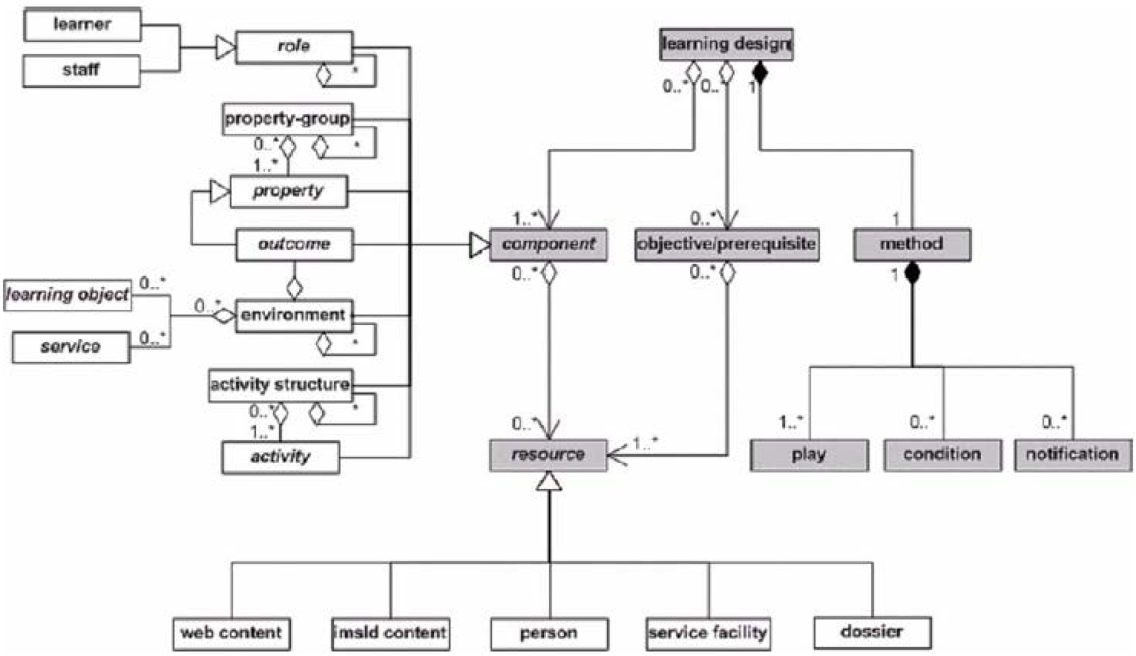


Figure 4. Semantic aggregation levels in the Learning Design Level C specification (From [IMS, 2003c])

The main elements of LD are described as follows:

- A `component` can be of one of the following types: `role`, `property group`, `property`, `activity structure`, `activity`, `environment`, or `outcome`,
- A person can have a `role` in the teaching-learning process. In a role a person works towards certain `outcomes`, performing learning and/or support `activities` within an `environment`,
- A `role` can be a `learner` or a `staff` role. A role can be specialised into sub-roles. The designer can attribute a name to the roles and sub-roles, and specify their activities. In a group-based activity, for example, learners can play different roles,
- A `resource` can be of one of the following types: `web content`, `imsld content`, `person`, `service facility`, or `dossier`. Although any of them can be referenced in a Learning Design, they are not explicitly part of the Information Model,
- `Environments` can hold two fundamental types: located `learning objects` or `generic services`. Located learning objects are usually specified by a URL with optional metadata. Generic services relate to a concrete resource available at runtime (a given URL). Examples of a Service include a discussion forum or chat rooms. In Learning Design at an abstract level it is specified how a service is activated during runtime, considering the learning design roles,
- An LD `method` contains the play and the conditions and brings together `roles`, `activities` and related `environments` making possible that learners with certain `prerequisites` attain certain `learning objectives`. A `method` may contain conditions expressed by `If-Then-Else` rules to further refine the assignment of activities and environment entities for persons and roles, which can be used for personalisation purposes,
- Related `activities` can be grouped together into `activity structures`, which can be associated with a `role` in a `role-part`. Activity structures are used to represent a range or a sequence of activities,
- Properties are used to store information about persons' roles and the UoL itself, like user profiles, user progression information or additional LO used during the teaching-learning process (e.g. reports, essays or new learning materials). A `property` can be:

  - global – It is global to the run of a unit of learning if it is accessible in any other UoL, or
  - local – Used often for temporary storage of data during the run of a specific unit of learning,

- Notifications in LD (level C) can make new activities available for a role through the use of certain outcome triggers, for example, the change of a property value. This mechanism allows modelling of adaptive tasks, often used in conjunction with conditions.

The metaphor of a theatrical play is often used to explain how the learning and education process is treated in the IMS LD specification. A spectator can see a stage, stage properties, and actors [Koper and Olivier, 2004] in a theatrical play. The actors have a script containing what they have to say or do and in what moments. Also the play may have small parts called acts. When an act ends, there are usually changes in the actors and/or the stage.

In LD the play is located in the method section and it has the same purpose as the theatrical play script. The LD play encloses the acts to be carried out in the order listed. An act defines who (which role) has to achieve which activity or set of activities. Thus, it is the method that links all the components of LD, managing the roles, activities and the environments related to the activities, and mentioning all concepts, directly or not. A role is connected to an activity through the role-parts within an act. The activity specifies what each role has to execute in what environment. Similar to an actor on the stage that can be more than one, in an act there can be more than one role active at the same instant. The activities that are at the same time performed by distinct roles are synchronised by the act, meaning that if one of the role-parts finishes an activity before the others, the next act can only turn into active when all the others of the previous act end (if it is not using features from level B or C).

Therefore, a method is formed by one or more parallel plays; a play consists of one or more sequential acts; and an act consists of one or more concurrent role-parts, and each role-part associates exactly one role with one activity or activity-structure.

LD is used to model units of learning (UoL), putting the Learning Design in a content package, such as an IMS Content Package (IMS CP). The UoL, physically, has two different parts (see the right side of Figure 5):

- The manifesto, a file called Manifest, written in XML, that describes the organisation and the learning process, as well as the connections to the learning resources and to any other XML file necessary for the execution of the global elements;
- The educative resources of any type. They also can be other files of type Web or additional writing archives in XML.

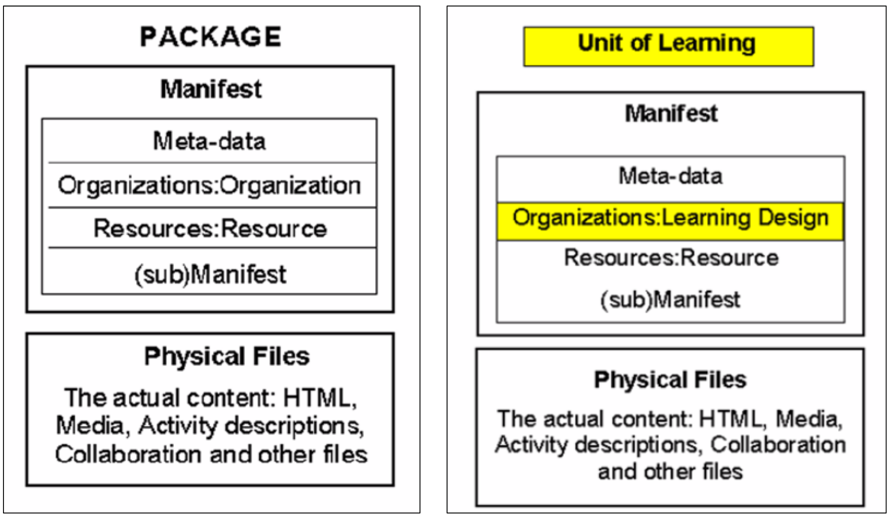


Figure 5. Relation between a CP and a UoL (Adapted from [IMS, 2003c])

The structure of a Unit of Learning is composed by including an IMS Learning Design within the part designated as 'Organizations' of IMS Content Packaging.

There are three options to make an LD available:

- distribute the LD-XML file,
- distribute the LD content package manifest with pointers to the mentioned resources,
- distribute the LD content package - a complete content packaging zip file containing also the learning objects and the LD file.

2.4.1.2 XML Schema

The `learning-design` element, which is placed in the `organizations` part of a Content Package, is complex. It contains elements that represent the conceptual model discussed in the previous section.

The LD XML schema can be represented like a tree (see Figure 6).

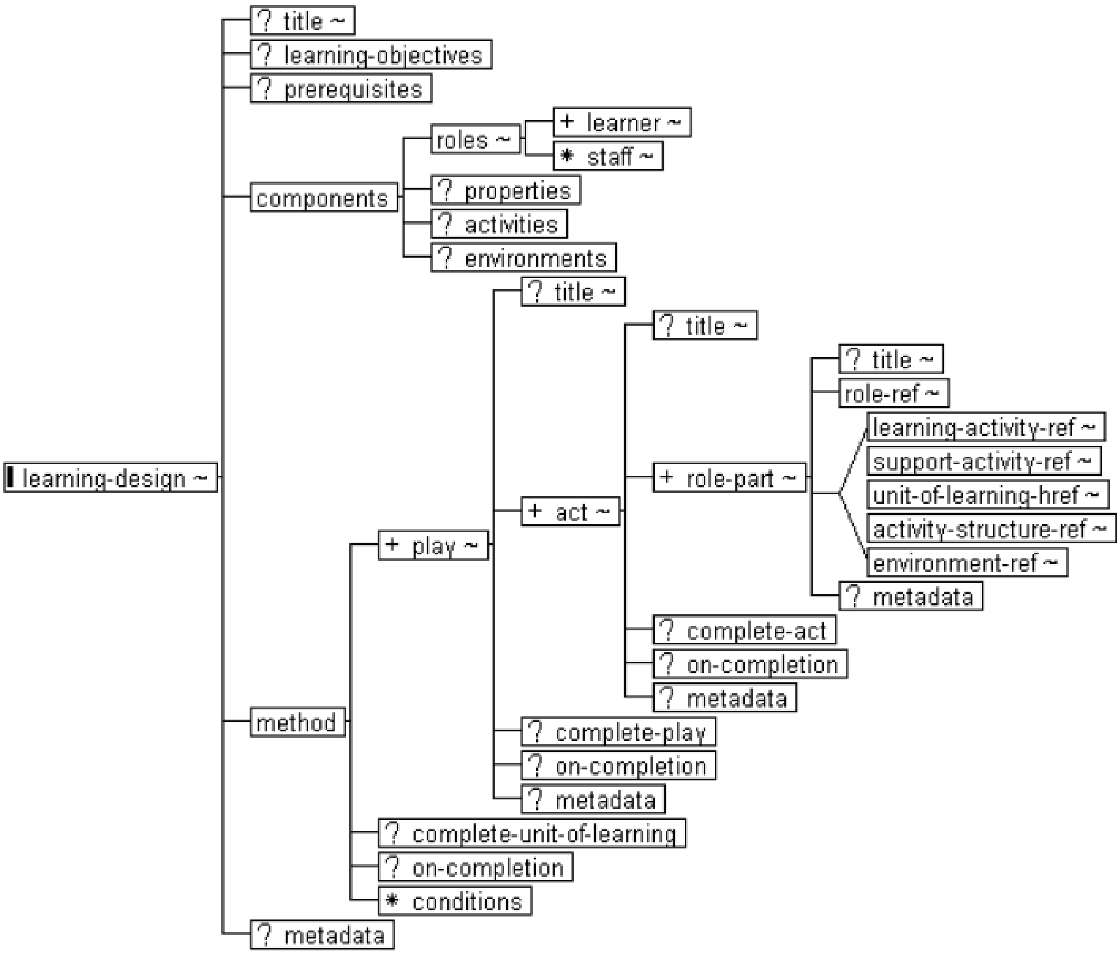


Figure 6. Representation of the LD Schema (Source: [Koper and Olivier, 2004])

2.4.1.3 IMS LD limitations

The IMS LD specification only allows the conditions of type `if-then-else` to be nested partially - an element `if-then` cannot contain an element `if-then` or `if-then-else`, although an element `else` can, as can be seen in the next figure.

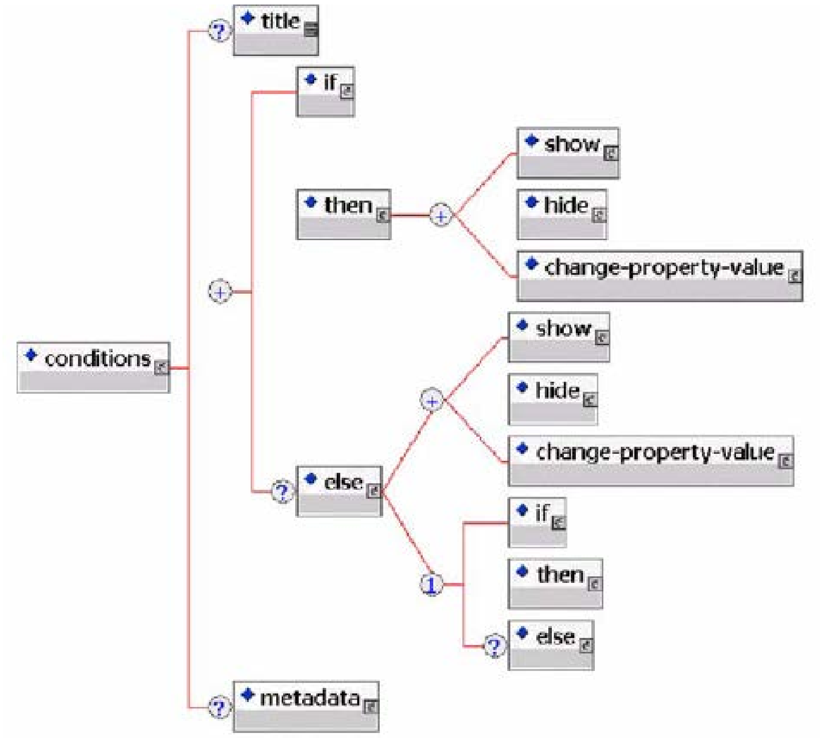


Figure 7. The model of the element conditions (Source: [IMS, 2003c])

Although IMS LD supports group and collaborative learning processes, some difficulties have been reported [Caeiro-Rodríguez, Llamas-Nistal et al., 2005; Miao, Hoeksema et al., 2005; Vignollet, Bote-Lorenzo et al., 2009].

IMS LD provides a good support for the organisation of learning activities, but the specification is not as good at dealing with fine-grained details of particular activities, since it does not allow the integration of thirty-party tools.

A wider analysis of IMS LD problems is provided in [Neumann, Klebl et al., 2010].

Because of these problems, the specification is "not considered as the definitive EML" [Caeiro-Rodríguez, Anido-Rifón et al., 2010]. In fact, a number of extensions have been proposed to deal with the specification limitations, including suggestions for new services [Leo, Perez et al., 2004; Turani and Calvo, 2006], additional assessment or adaptive learning facilities [Sitthisak and Gilbert, 2009; Monfort, Khemaja et al., 2010], and inclusion of other tools [Valentín, 2011].

2.5 Digital Repositories Specifications

E-learning systems are increasingly becoming an important part of the strategy for delivering learning contents and activities. In 2008 it was estimated the existence of more than 250 providers of commercial e-learning platforms [Al-Ajlan and Zedan, 2008] . Those numbers highlight the importance of making simple the migration of the enclosed resources from one platform to another and making services uniformly available. Thus, there has been a strong movement into the standardisation of the usage of digital repositories and related technologies.

This section provides a general description of some digital repositories specifications.

2.5.1 CORDRA

In 2003 ADL started to work in the definition of a model to search, locate and access learning objects by designers and authors of educational material and facilitate their reuse. CORDRA

(Content Object Repository Discovery and Registration/Resolution Architecture) is an attempt to define a framework for the federation of digital collections. The framework, also known as 'CORDRA specification', is "intended to be an open, standards-based model for designing and implementing information systems including registries and repositories for the purposes of discovery, sharing and reuse of information" (http://cordra.net/).

The designation CORDRA itself means three different things [Rehak, Dodds et al., 2005]:

- a model that states the way to create local federations and a global learning content infrastructure;
- a project that aims to define and document the model with sample tools and implementations;
- a working platform, i.e., a global federation of content registries.

Two organisations, Corporation for National Research Initiatives and Advanced Distributed Learning Initiative, worked together in the development of the first publicly available CORDRA implementation, ADL Registry (http://www.adlnet.gov/Technologies/adlr/default.aspx), which was launched in December 2005.

The idea behind CORDRA was to build a federated network of repositories of learning objects that can be incorporated in any entity that provides educational materials. It attempts to define a reference model based on existing standards and specifications, extending them when necessary and, in particular, determining how to combine them to reach their objectives.

As can be seen in Figure 8, the CORDRA model is based on a hierarchy of registry systems. CORDRA registries store information about educational content stored on a series of digital repositories belonging to a particular community. These registries form independent CORDRA implementations, which are registered into the federated CORDRA environment.

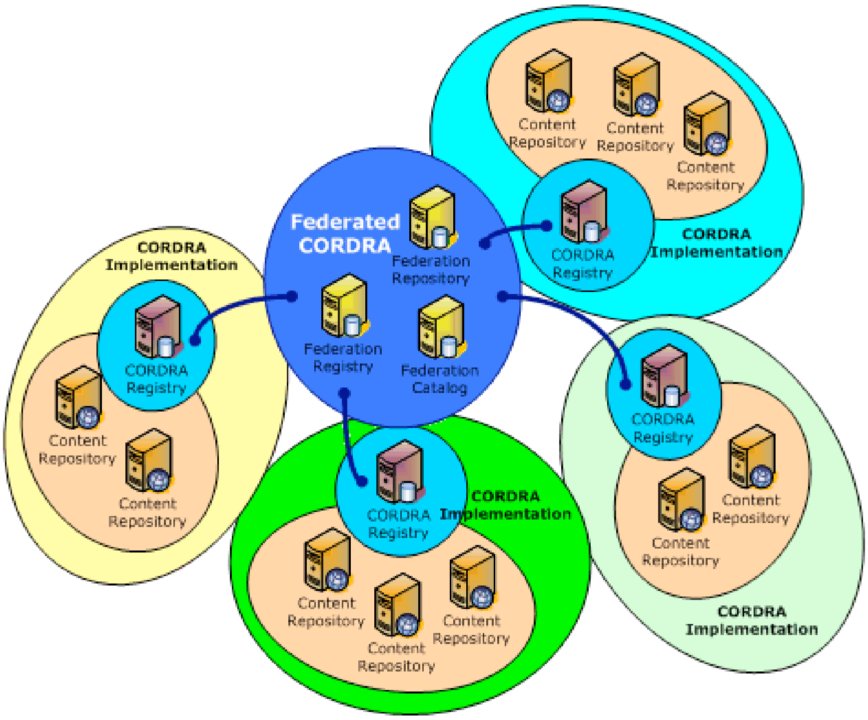


Figure 8. A model of federated CORDRA
(From http://www.cen-ltso.net/main.aspx?put=916)

The CORDRA activities have been coordinated by Advanced Distributed Learning (ADL), Corporation for National Research Initiatives (CNRI) and Learning Systems Architecture Lab (LSAL).

2.5.2 Simple Query Interface (SQI)

The Simple Query Interface (SQI) specification is supported by CEN WS-LT and presents a language impartial API for querying learning objects repositories [CEN/ISSS, 2005]. It is supported by CEN and the current version (1.0) was released on November 2005. SQI can be deployed in synchronous or asynchronous scenarios.

There are methods that can be used by a certain application to formulate initial configurations, submit queries to a learning objects repository and retrieve the results (see Figure 9).

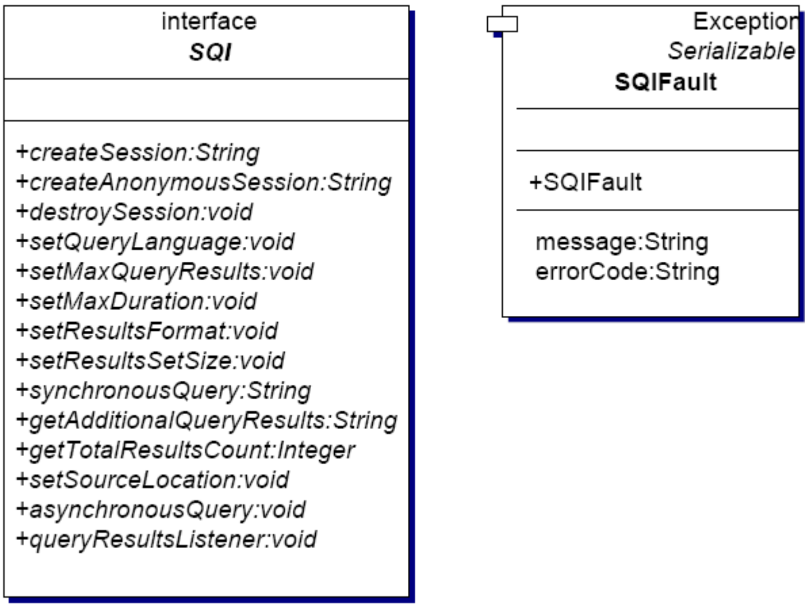


Figure 9. UML Class Diagram of the Simple Query Interface (Source: [CEN ISSS, 2005])

There are SQI methods for configuration (`SetQueryLanguage`, `SetResultsFormat`, `SetMaxQueryResults`, `SetMaxDuration`), synchronous query (`setResultsSetSize`, `synchronousQuery`, `getTotalResultsCount`) and asynchronous query (`setSourceLocation`, `asynchronousQuery`, `queryResultsListener`) purposes. All the methods are implemented at the target and called by the source with the exception of `queryResultsListener`, which is implemented at the source and called by the target.

If SQI is to be used to implement full query functionality, some conformity is required about [CEN/ISSS, 2005]:

- the attributes and vocabularies that can be employed in the query,
- the query language and its format,
- the representation of the list of the learning objects of interest, and
- the representation of discrete metadata instances on learning objects,

as these issues are out of the scope of SQI.

2.5.3 Digital Repositories Interoperability (DRI)

The IMS Digital Repositories v1.0 final specification was released on January 30, 2003. It provides some recommendations for the interoperation of the most frequent repository functions.

This specification defines an abstraction layer on a repository and thus it works with any type of implementation for the repository, which can be an SQL, an XML or a Z39.50 repository, among others.

Figure 10 shows the functional architecture, considering the four roles that can be played by users of a digital repository: Creator, Learner, Infoseeker, and Agent.

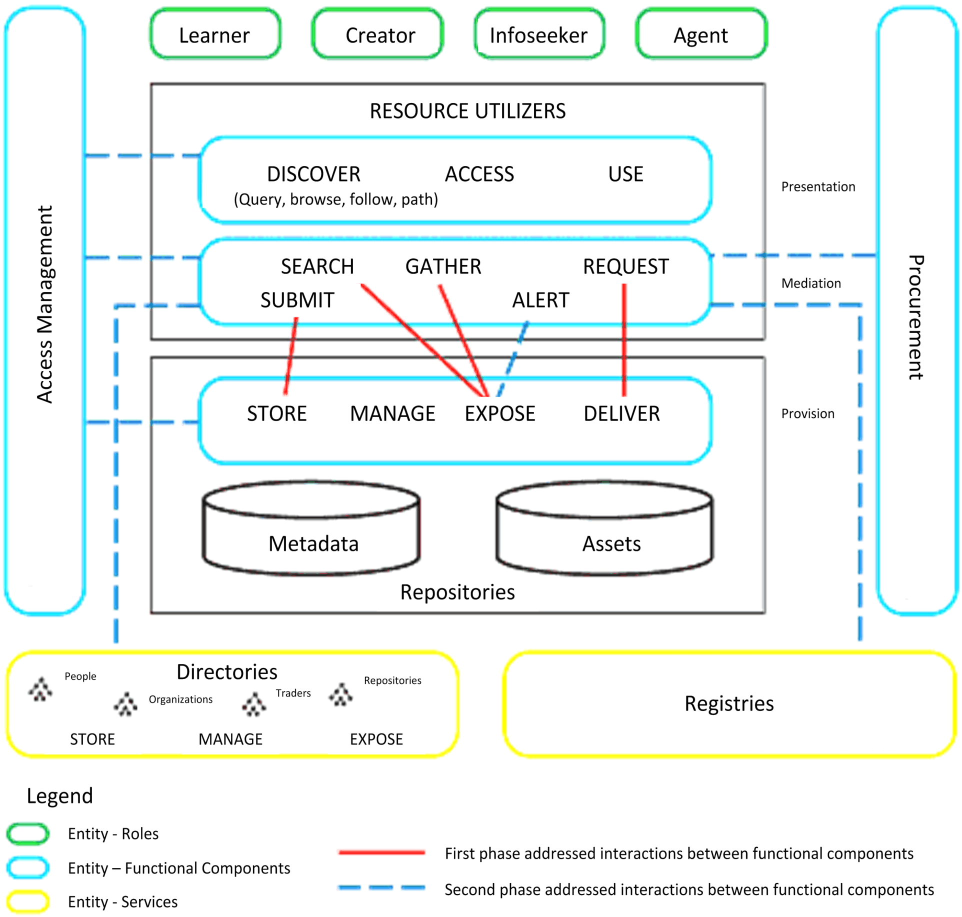


Figure 10. Functional architecture
(Source: [IMS, 2003a])

The specification defines the basic services that must be available in any system compatible with IMS-DRI. Those fundamental services are [IMS, 2003a]:

- **Search/Expose** - The Search reference model supports various configurations for performing search. In addition, it provides an optional mediation layer to permit the querying of distributed and diverse metadata sources. The two protocols supported are XQuery over SOAP, for learning object repositories, and Z39.50, for information repositories;
- **Gather/Expose** - The Gather reference model is related to the requesting and aggregation of meta-data exposed by repositories;
- **Submit/Store** - This functionality is related to the mobility of objects and how they can be transferred to a network-accessible location;
- **Request/Deliver** – Once a resource had been located using the service Search/Expose, the service Request/Deliver provides access to it;
- **Alert/Expose** – Although Alert is also a core function, it is not addressed within the current version of the DRI specification. The Alert/Expose functionality is envisioned to be available in digital repositories and intermediary aggregator services to allow the sending of messages to other systems.

2.5.4 IMS LODE

Version 1.0 of the IMS GLC Learning Object Discovery & Exchange (LODE) specification was released in March 2010. It aims to make easier the discovery and retrieval of learning resources stored in repositories.

While widely available in repositories, most of the stored learning objects are not globally discoverable, which is a point addressed by IMS LODE.

Three data models are proposed [IMS, 2010]:

- LODE Context Set for the Contextual Query Language (CQL) defines the attributes of learning objects, which can be used for searches carried out by a LODE client;
- Information for Learning Object eXchange (ILOX), which organises groups of metadata on learning objects for the purpose of exchanging data;
- Learning Object Repository Registry Data Model to be applied to discover and configure access to learning object collections.

Figure 11 highlights the connection between the IMS LODE specification and other specifications to allow a LODE Client to actually obtain the relevant learning objects.

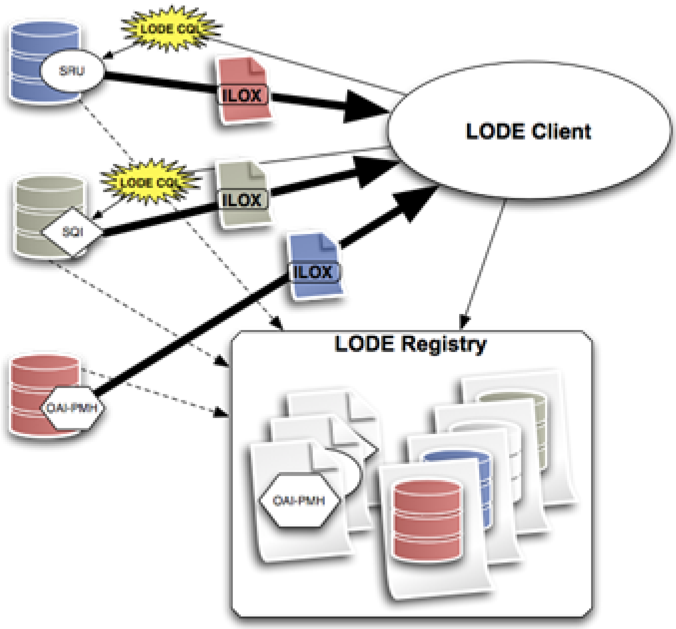


Figure 11. The idea behind IMS LODE
(From [IMS, 2010])

2.6 SCORM

The first version of SCORM was released in January 2000 [ADL, 2000]. SCORM 2004 is the latest version of the Sharable Content Object Reference Model (SCORM), consisting of a Web-based learning Content Aggregation Model (CAM), a Run-Time Environment (RTE) and a Sequencing and Navigation behaviour for learning objects. Each of these parts is described in a technical book, and there is one more, called the overview book, that provides high-level conceptual information (see Figure 12).

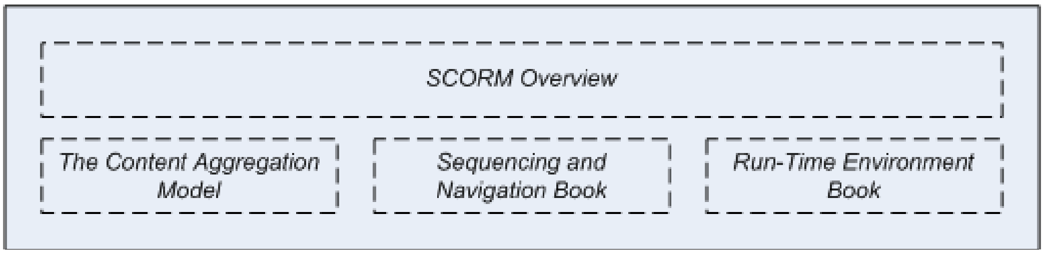


Figure 12. SCORM structure

The SCORM Model is an initiative of ADL which was established in 1997. Since then it has being much utilised. SCORM adopts the best of different e-learning standards for diverse aspects like metadata, content tracking and content sequencing and combines them with one another to build a more complete model. The organisation of the educational contents states the structure or index of the course. But different from IMS LD, SCORM does not allow multiple actors, and only a single learner can be considered.

The Shareable Content Object Reference Model defines how learning platforms must handle Web contents and make them available to users. SCORM uses specifications and guides, putting them together in a single model. It can be noted that, for the SCORM model, a Web course is a collection of content objects that are interconnected, considering multiple aspects: contents organisation, metadata, sequence of execution, among others.

In SCORM there are Shareable Content Objects (SCO) and Assets, and every SCO is composed by Assets. A Sharable Content Object is a learning object that contains one or more resources and communicates with an LMS – this is the main difference between an Asset and a SCO. It must have accessible related metadata, so it can be found in a contents repository.

A resource (asset) is any file that can be used for educational purpose, like an image, a text, a presentation, a web page – any learning object that does not communicate with an LMS.

In the metadata perspective, a package SCORM can be seen as a course or a resource (which is part of a course). The main difference is that the former (courses) contains resources and its organisation, and the latter only contains resources (SCOs and Assets), without any information about their organisation or relation.

2.6.1 IMS LD and SCORM

IMS LD and SCORM are more complex than other specifications and not easy to deeply understand and implement in tools. A comparison between these two specifications is shown on Table 7.

Table 7. SCORM and IMS LD comparison

Characteristic	ADL SCORM 2004	IMS Learning Design 1.0
Compliant e-learning platforms	Most of them, including Moodle, WebCT, LearningSpace, among others	Just a few of them: .LRN, Moodle (planned), BlackBoard[8]
Pedagogical Flexibility	Mainly devoted to self learning	Wide range of supported pedagogical models
Teacher/students interaction	Not supported	Supported
Students/contents interaction	Supported	Supported
Content/services relation	Not supported	Supported (the learning activities use learning objects and services)
Reusability possibilities	Contents	Contents, activities, learning design

Some of the differences between SCORM and IMS LD are due to the scope of the specifications. SCORM is mainly devoted to content reuse and interoperability issues, and IMS LD to learning design representation and it encompasses the learning process largely beyond what is possible with the use of SCORM. The main drawback of the LD specification is the lack of widely support from e-learning platforms.

2.7 Summary

At a very early stage of this research, several specifications were analysed in order to decide which one(s) could be valuable to fulfil its objectives. A deep work in any area must ensure its compliance with the standards and specifications suitable for its purposes. That conformity makes possible to reach a good agreement platform, for example to enable the interoperability of Learning Objects Repositories [Ternier, Massart et al., 2008], or to facilitate its broad use.

After an introductory section, which presented some educational standardisation bodies, section 2.2 discussed specifications related to resources metadata. Some metadata specifications and associated application profiles were analysed. It is interesting to note that the application profiles based on Dublin Core are more likely to use additional elements, and those based on the LOM standard usually do not consider all its elements or extra ones. ISO/IEC MLR was published in January 2011, and at the time of writing no AP was defined.

Section 2.3 discussed some specification for encapsulating learning resources. Section 2.4 addressed educational modelling languages, which are related to the definition of activities that use learning resources.

Section 2.5 briefly examined some initiatives related to digital repositories interoperability and services, which are not applied for the characterisation of learning objects. The specifications described in that section are devoted to the definition of access services for learning objects repositories.

Finally, section 2.6 discussed the Sharable Content Object Reference Model, which is a collection of standards and specifications for e-learning widely used.

From all the specifications studied, mainly two were found suitable for characterisation of learning objects: IEEE LOM and IMS LD, due to their characteristics. IEEE LOM is largely accepted for the description of learning resources and IMS LD is the only specification that allows the modelling of learning activities, considering many actors.

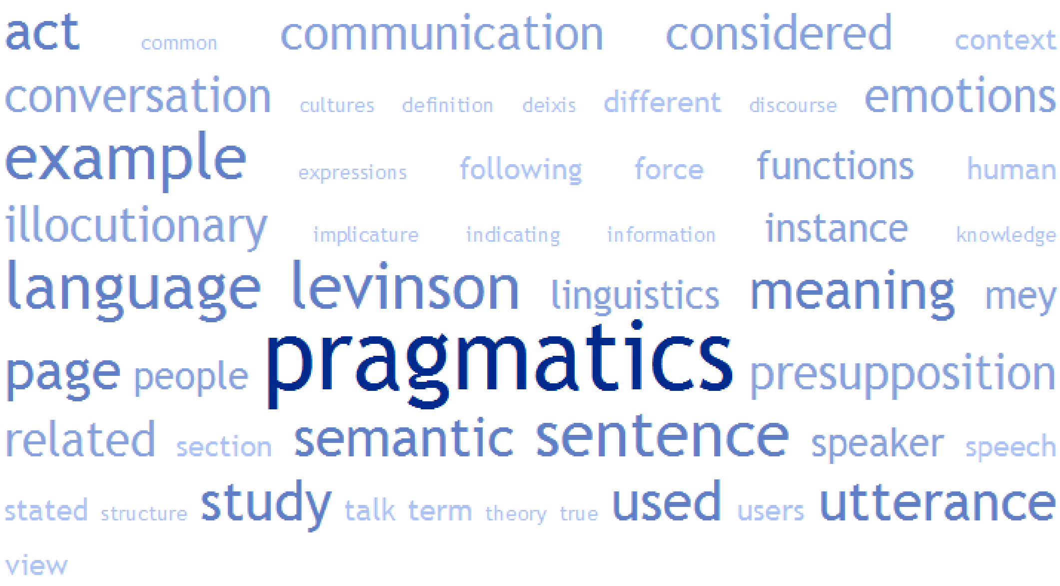
[8] It is under development the integration of OUNL's e-Learning player, Edubox, with the Blackboard e-learning platforms (http://investor.blackboard.com/phoenix.zhtml?c=177018&p=irol-newsArticle_Print&ID=519762&highlight=).

Chapter 3

SOME TOPICS ON PRAGMATICS

This chapter discusses the concept of Pragmatics and its delimitation, based mainly on its roots in the Philosophy of Language and Linguistics, as opposed to the concept of Semantics.

The notions of deixis, implicature, presupposition, speech acts, and aspects of discourse structure are introduced. Some related subjects, such as metapragmatics, cross-cultural pragmatics and cyberpragmatics, are also presented. The main findings are summarised at the end of this chapter.



3.1 Introduction

Pragmatics has the common meaning of ‘people with practical concerns’ or ‘pragmatic people’, as shown in the following extract from a debate in the European Parliament: “Actually, it divided us into ideologists and pragmatics...”[9].

Pragmatic is an adjective with the subsequent signification: “solving problems in a practical and sensible way rather than by having fixed ideas or theories”, and *realistic* is a synonym for that word, according to the Oxford Advanced Learner’s Dictionary. Some examples of that use of the word are as follows:

- “Some analysts see an ideological sibling rivalry at work, with the idealistic Fidel Castro trying to reign in recent policy shifts by his more-pragmatic brother”[10],
- “He's taking a very pragmatic approach and understands that it's a year away...”[11].

The formal definition of Pragmatics is stated in the next section. In the past few years there has been a growing interest in pragmatics. In the area of language studies there are a number of international conferences and seven well and widely known international journals (*Journal of Pragmatics*, *Pragmatics*, *Pragmatics and Cognition*, *Journal of Historical Pragmatics*, *Intercultural Pragmatics*, *International Review of Pragmatics* and *Pragmatics and Society*[12]) devoted to this theme. Furthermore, the notions of pragmatics have reached out into cognitive science, psychology and sociology[13].

It is worth to note that the concepts related to pragmatics are hard to be objectively studied, a point that was highlighted by Yale: “The advantage of studying language via pragmatics is that one can talk about people’s intended meanings, their assumptions, their purposes or goals, and the kind of actions (for example, requests) that they are performing when they speak. The big disadvantage is that all these very human concepts are extremely difficult to analyse in a consistent and objective way” [Yule, 1996] (page 4).

This chapter provides an overview on pragmatics, its definition (section 3.2) and delimitation (section 3.3). It also discusses some related topics, such as metapragmatics (section 3.4), pragmatics across cultures (section 3.5) and the pragmatics of digital discourses (section 3.6), finalising with some conclusions and closing remarks (section 3.7).

Annex 2 complements this chapter by providing some definitions related to the topic of emotions.

3.2 Some definitions

Pragmatics is a term that was popularised by Charles Morris. He defined it while outlining a science of signs: **semiotics**[14] [Morris, 1938]. The word semiotic comes from the Greek *sêma* (sign). The Oxford Advanced Learner’s Dictionary defines Semiotics as “the study of signs and symbols and of their meaning and use”.

Morris identified three divisions of semiotics:

- **Syntactics** (or syntax) – The study of “the formal relation of signs to one another”,

[9] http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+CRE+20080618+ITEMS+DOC+XML+V0//EN&language=EN

[10] A piece of news entitled “Fidel Castro back in spotlight, calls congressional special session” from CNN and available at http://edition.cnn.com/2010/WORLD/americas/08/06/cuba.castro.assembly/index.html?iref=allsearch#fbid=La8Lm2C0Ohz&wom=false-.

[11] A piece of news from CNN entitled “Petraeus wants to set record straight on July 2011” from CNN and available at http://politicalticker.blogs.cnn.com/2010/08/12/petraeus-wants-to-set-record-straight-on-july-2011/?iref=allsearch.

[12] Provided by order of their launch date.

[13] Wharton (2009) suggests also a connection to non-human animal communication.

[14] Morris adopted the term semiotic, not semiotics.

- **Semantics** - The study of "the relations of signs to the objects to which the signs are applicable",
- **Pragmatics** - The study of "the relation of signs to interpreters".

Thus, semiotics is a science of signs and their syntax, semantics and pragmatics [Culler, 1981] [Peirce, 1955]. Mey suggested the use of the terms 'messages' and 'language users' in Morris' definition of pragmatics, to emphasise the focus of pragmatics on the use of the language by humans [Mey, 1993]. In his opinion "pragmatics comprises everything that characterises people as users of language" (page 6), even if this can be considered a very broad usage of the term [Levinson, 1983].

Also, handling users in their social context is essential for a real pragmatic consideration [Mey, 1993] (page 6). Linguistically, semantics is related to what is said, while pragmatics is concerned with what is implicated.

The semiotic triangle [Ogden and Richards, 1923], represented in Figure 13, emphasises three parts: the **symbol** that we use to refer to an object itself, which is the **referent**, and also the understanding meaning (**thought** or reference) of the referent or its symbol.

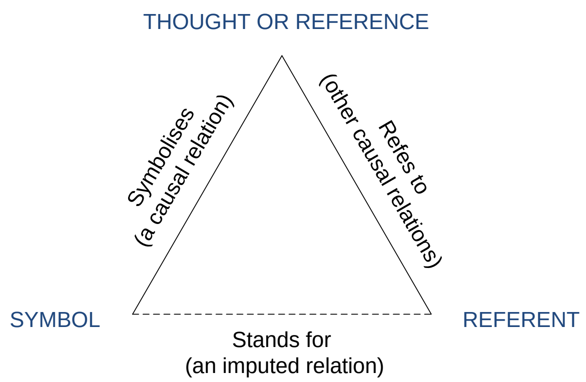


Figure 13. The Ogden and Richards semiotic triangle

Human languages and formal languages, such as logics, are also considered by semiotics. A language is a system of previously accepted spoken or written symbols used by human beings to communicate in a certain geographical area.

However, pragmatics is not just related to words, although that narrow approach is commonly followed by many linguists, but also to emotions and how they are expressed. That point is covered by Tim Wharton in his book 'Pragmatics and Non-verbal Communication': "Sentences are rarely uttered in a behavioural vacuum. We colour and flavour our speech with a variety of natural vocal, facial and bodily gestures, which indicate our internal state of conveying attitudes to the propositions we express or information about our emotions or feelings. ... Often, they show us more about a person's mental/physical state than the words they accompany; sometimes, they replace words rather than merely accompany them" [Wharton, 2009] (page 1).

In addition, according to Sabini and Silver, the social context of communication determines how an emotional experience and an emotion word are interconnected [Sabini and Silver, 2005b]. Also, they argue that the interpretation of the affective experience depends on the audience, being then subjective, and represents a communicative act. They defend that "paying attention to the pragmatic contexts of our emotion talk will provide a clearer, and in some ways, simpler view of the mental states implicated by (if not exactly referred to by) our emotion talk" [Sabini and Silver, 2005a].

Ochs and Schieffelin talk about the "pragmatic properties of affect marker" [Ochs and Schieffelin, 1989]. For instance, the use of perfect tense and quantifiers in English are considered as affect intensifiers in English. Also, they highlight that some conjunctions act as affect specifiers, like the use of 'as if' to indicate positive affect to some viable event or situation.

An important distinction to both semantics and pragmatics is the difference between **sentence** and **utterance**. According to Levinson, a sentence is "an abstract theoretical entity defined within a theory of grammar", while an utterance is "the issuance of a sentence, a sentence-analogue, or sentence-fragment, in an actual context" [Levinson, 1983] (page 18). It is common to think of an utterance as a pair of a sentence and a context (the one in which the sentence was uttered).

However, in practice it is difficult to keep the distinction between sentence and utterance. In fact, there is no consensus among linguists on the assignment of notions like presuppositions and illocutionary force, topics which are discussed later in this chapter, to sentences or utterances [Levinson, 1983].

Yule provides four different definitions of pragmatics [Yule, 1996] (page 3), each one related to an area that pragmatics is concerned with:

- "Pragmatics is the study of speaker meaning", considering the two sides of a communication process: a speaker (or writer), and what he/she said, and a listener (or reader), and what he/she understood about what was said,
- "Pragmatics is the study of contextual meaning", entailing the interpretation of what people mean in a certain context and how it has an effect on what is said,
- "Pragmatics is the study of how more gets communicated than is said", exploring how a great part of what is not said is identified as part of what is communicated,
- "Pragmatics is the study of the expression of relative distance", as how close or distant a person is, less or more is needed to be said.

Context has been a key concept in the field of pragmatics [Goodwin and Duranti, 1992] (page 1). In fact, a generic and very brief distinction between semantics and pragmatics is based on the notion of context: semantics deals with meaning independent of context, whereas contextual dependencies belong to pragmatics.

The Oxford Advanced Learner's Dictionary provides the following generic definition of context: "the situation in which something happens and that helps you to understand it". Context can be seen as frame that surrounds what is being analysed making possible its correct interpretation [Goffman, 1974].

In [Dey, 2001] a definition of context related to computer applications and their usage is supplied: "Context is any information that can be used to characterise the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves".

Another possible definition of pragmatics was provided by Stalnaker [Stalnaker, 1972], which Levinson summarised with this words: "Pragmatics is the study of deixis (at least in part), implicature, presupposition, speech acts, and aspects of discourse structure" [Stalnaker, 1972; Levinson, 1983] (page 27). It refers to some topics with significance in pragmatics, which are related to Yale's definitions presented before and are analysed from section 3.2.1 to 3.2.5.

3.2.1 Deixis

Deixis is a term from the Greek word for pointing or indicating (*deiknunai*) that means "the function or use of deictic words or expressions (= ones whose meaning depends on where, when or by whom they are used)", according to the Oxford Advanced Learner's Dictionary.

The use of demonstratives (e.g., this, that), first and second pronouns (e.g., me, mine, I, you), tense (for instance, the use of the past perfect to state that an episode took place at a time before another episode described in the past tense), specific time (e.g., now, tomorrow) and place adverbs (e.g., here, there), and many other grammatical features have a meaning that depends on the context. In sentence (a) of Example 1, 'that' is used to refer to something in the context of the communicator, while in sentence (b) the use of the deictic expression 'here' has a obvious meaning for the people present, but not for those not tied to the specific context:

(a) `What is` **`that`**`?`

(b) `Isabel is` **`here`**`.`

Example 1. Use of deictic expressions

Philosophers usually use the expression 'indexical expressions', or just 'indexicals', to talk about the topic of deixis. Deictic expressions can be used to denote people, e.g., 'you', 'us' (**person deixis**); time, e. g., 'now', 'later' (**temporal deixis**); and location, e.g., 'here' (**spatial deixis**).

Levinson highlights that most deictic phenomena could be considered semantic as semantics is taken to consider all conventional aspects of meaning, but he concludes that "deixis belongs within the domain of pragmatics, because it directly concerns the relationship between the structure of languages and the contexts in which they are used" [Levinson, 1983] (page 55).

3.2.2 Implicature

Implicatures are fundamental examples of more being communicated than is said. Consider the following example and the use of the adjective blue:

(a) `The t-shirt is blue.`

(b) `The t-shirt is blue and brown.`

Example 2. Implicature

It can be suggested that blue is ambiguous, as in Example 2(a) it apparently means 'only or wholly blue', but that same meaning is not allowed in (b), where blue can only mean 'partially blue'. The semantic core of many natural language expressions is also frequently context-specific.

Consider the following example of conversational implicature from [Yule, 1996] (page 40):

Charlene: `I hope you brought the bread and the cheese.`

Dexter: `Ah, I brought the bread.`

Example 3. More being said by conversational implicature

Dexter must intend Charlene to infer that the cheese was not brought. Using `b` for bread, `c` for cheese and `+>` for an implicature, we can express Example 3 as can be seen in Example 4.

Charlene: `b & c?`

Dexter: `b                    (+> NOT c)`

Example 4. Previous example revised with the additional conveyed meaning

Grice's theory of implicature is essentially a theory about how people use language. Grice suggests a set of assumptions guiding a conversation. The assumption of cooperation in a conversation can be stated as a principle of conversation ('cooperative principle'): "Make your contribution such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged". That principle is elaborated in four maxims known as "maxims of conversation" [Grice, 1975]:

- **Maxim of Quality**:

a) Do not say what you believe to be false,

b) Do not say that for which you lack adequate evidence.

- **Maxim of Quantity**:

c) Make your contribution to the conversation as informative as necessary,

d) Do not make your contribution to the conversation more informative than is required.

- **Maxim of Relevance**:

e) Make your contributions relevant.

- **Maxim of Manner**:

f) Avoid obscurity,

g) Avoid ambiguity,

h) Be brief,

i) Be orderly.

These maxims can be violated. For instance, if a speaker tells a lie, the maxim of quality is broken.

Conversational implicature is a special kind of pragmatic inference. As noted by Levinson, these inferences cannot be seen as semantic because "they are based squarely on certain contextual assumptions concerning the co-operativeness of participants in a conversation, rather than being built into the linguistic structure of the sentences that give rise to them" [Levinson, 1983] (page 167).

3.2.3 Presupposition[15]

The ordinary usage of the word presupposition and its usage within linguistics are different. The common language notion of presupposition is related to any kind of background assumption against which something makes sense or is considered rational. Some examples of that usage of the term are as follows:

a. `Teachers often presuppose a great level of knowledge by their students,`

b. `Effects presuppose causes,`

c. `I gave him a book, presupposing he could read.`

Example 5. Ordinary usage of the term presupposition or derivatives

[15] There is a great amount of literature available about the topic of presupposition, and many works follow different approaches that might contradict each other. In this section the vision expressed in [Levinson, 1983] about the subject is followed.

Presupposition is often regarded as a relationship between two propositions. Consider the sentence (a) from Example 6, which contains the proposition `p`, and the sentence (b), which contains the proposition `q`.

a. `Paul’s son is intelligent.        (p)`

b. `Paul has a son.                   (q)`

c. `p>>q`

Example 6. The notion of presupposition in pragmatics

If the symbol `>>` is used to mean ‘presupposes’, the relationship between *p* and *q* can be expressed as in (c) from Example 6. However, it is interesting to note that if the sentence (a) of that example is negated (see Example 7(a)), the same proposition *q* remains to be presupposed by `NOT p`:

a. `Paul’s son is not intelligent.   (NOT p)`

b. `Paul has a son.                   (q)`

c. `NOT p>>q`

Example 7. The negation in a presupposition

Thus, the presupposition of a statement remains true even when that statement is negated. That property is known as **constancy under negation** and is due to Freges’s and Strawson’s works[16] [Strawson, 1952] [Frege, 1952], embodied in **Strawson’s definition of presupposition** that can be stated as follows:

```
A statement A presupposes another statement B iff:
If A is true, then B is true.
If A is false, then B is true.
```

Considered the following definition of **semantic entailment** as provided by Levinson: “A semantically entails B (written A ||- B) iff every situation that makes A true, makes B true (or: in all worlds in which A is true, B is true)” [Levinson, 1983] (page 174), the view of **semantic presupposition** can be stated as follows:

```
A sentence A semantically presupposes a sentence B iff:
A ||- B
~A ||- B
```

For many reasons, such as the ones provided by Boer and Lycan in the book ‘The myth of Semantic Presupposition’ [Böer and Lycan, 1976], but also others [Stalnaker, 1974] [Kempson, 1975], semantic theories of presupposition have been to a great extent dropped, and the notions of pragmatic presupposition have been advanced.

However, if semantic presupposition is necessary to assert the truth or falsity of statements, the failure of a pragmatic presupposition only highlights the unsuitability of a given utterance in a certain context [Stalnaker, 1974] [Karttunen, 1974]. Levinson provides the following definition of

[16] Frege, a philosopher, is considered the architect of modern logic.

pragmatic presupposition: "An utterance A pragmatically presupposes a proposition B iff A is appropriate only if B is mutually known by participants"[17] [Levinson, 1983] (page 205).

An entailment logically follows from what is asserted in an utterance. Sentences have entailments (not the communicators), while presuppositions are at the level of the communicators, the sentences itself does not have presuppositions. In the following example:

`Paul's son has two cats.`

Example 8. Analysing the difference between presuppositions and entailments

The presuppositions that there is a person whose name is Paul and that he has a son are expected. But we can say that the sentence in Example 8 has the entailments that Paul's son has something, has two animals, has two cats, has one cat, and other similar ones.

However, if presuppositions are preserved under negation, entailments are not. Consider Example 9.

a. `David is a human being.` (p)

b. `David is a mammal.` (q)

c. `p ||- q`

Example 9. Entailment

It is interesting to note that if the sentence (a) of this example is negated (`David is not a human being`), the proposition q does not remain to be entailed. Thus, it is not truth that `NOT p ||- q`.

Levinson considers that a "theory that predicts presuppositions from the semantic specification of linguistic expressions" is needed and he concludes that "presupposition remains, ninety years after Frege's remarks on the subject, still only partially understood, and an important ground for the study of how semantics and pragmatics interact" [Levinson, 1983] (page 225).

3.2.3.1 The projection problem

The idea that the meaning of a whole sentence is a combination of the meaning of its parts is common. This belief was held in [Frege, 1952], for instance. Langendoen and Savin suggested that this principle was also valid for presuppositions and that the set of presuppositions of a complex whole is the simple sum of the presuppositions of the parts [Langendoen and Savin, 1971]. Thus, they defended that being S_0 a complex sentence containing the sentences S_1, S_2, ... S_N as constituents, the presuppositions of S_0 are the presuppositions of S_1 plus the presuppositions of S_2 ... plus the presuppositions of S_N.

What is known as the **projection problem** derives from the fact that the meaning of some presuppositions does not remain valid in some complex sentences, contradicting the idea defended by Langendoen and Savin. In fact, presuppositions survive in situations where entailments cannot, and presuppositions disappear in contexts where one might expect them to remain, and where entailments would, which is explained in detail in [Levinson, 1983] (pages 191-198).

[17] The notion of mutual knowledge is of great importance in pragmatics. According to Levinson (1983), the sender's communicative intention becomes mutual knowledge to sender (S) and receiver (H) if "S knows that H knows that S knows that H knows (and so ad infinitum) that S has this particular intention" (page 16).

3.2.4 Speech acts

One of the main phenomena that "any general pragmatic theory must account for" is speech acts [Levinson, 1983] (page 226).

The action performed by producing an utterance consists of three related actions [Austin, 1970]:

- **Locutionary act** – It is the fundamental act of utterance or making a linguistic expression with significance. For instance, if someone is not a Portuguese speaker, that person might fail to produce a locutionary act in Portuguese, and saying 'Hello' might not be usually considered as locutionary act in that language;
- **Illocutionary act** – As utterances are normally made with some purpose, the function in mind when producing an utterance is related to another dimension, the illocutionary act. One might utter 'I have just made a cake', which counts as a locutionary act, for some communicative purpose, such as make an offer, a statement, or an explanation. We say that this is the **illocutionary force** of the utterance;
- **Perlocutionary act** – That last dimension is related to the intended effect when making an utterance. One might utter "I have just made a cake" believing that the hearer will recognise the intended effect (for example, apologise for the untidy kitchen or offer a piece of cake).

Of these three dimensions, the second one was the focus of Austin's interest, and the expression 'speech act' is generally applied to mean just the illocutionary force of an utterance. Hence, actions executed via utterances are called **speech acts**, one of the main phenomena that "*any general pragmatic theory must account for*" [Levinson, 1983] (page 226). They are what we "*do with our words*" [Austin, 1962] . In common English they have more specific labels, such as apology, complain, invitation or promise. For instance, imagine a boss saying to someone:

```
You don't work here anymore.
```

Example 10. A speech act

That utterance in Example 10 can be used to establish the ending of someone's employment.

Austin highlighted the difficulty of distinguishing illocutionary and perlocutionary acts. Levinson made the following remark: "A perlocutionary act is specific to the circumstances of issuance, and is therefore not conventionally achieved just by uttering that particular utterance, and includes all those effects, intended or not intended, often indeterminated, that some particular utterance in a particular situation may cause" [Levinson, 1983] (page 237).

If someone utters 'I'll visit you tomorrow', it might have different illocutionary forces, such as promise or prediction. **Illocutionary Force Indicating Devices** (IFID) and **felicity conditions** are topics related to how a speaker assumes that the illocutionary force will be recognised by the hearer.

3.2.4.1 Illocutionary Force Indicating Devices

Some expressions have the purpose of indicating the illocutionary force and they are Illocutionary Force Indicating Devices (IFID). If instead of saying 'I'll visit you tomorrow', someone says 'I promise I'll visit you tomorrow', the illocutionary force of the utterance is clearly stated by the use of 'promise' as an IFID. Other verbs like *warn*, *predict*, *ask*, and *plan*, might also be used to indicate what the communicator really means.

Other devices might also be used as Illocutionary Force Indicating Devices, such as a lowered voice for a warning and the stress or intonation.

Also syntactic features can be used as IFIDs, such as punctuation, prosodic contour, word order and mood of verb. Consider Example 11, which illustrates the use of syntactic IFIDs.

(a) `Please, leave the room!`

(b) `You will leave the room.`

Example 11. Use of syntactic IFIDs
(Adapted from [Searle, Kiefer et al., 1980])

In Example 11 the use of different syntactic features implies different illocutionary acts. The first of these utterances may have the illocutionary force of an order and the second one may have the illocutionary force of a prediction.

3.2.4.2 Felicity conditions

Some sentences are not just used to illustrate states of affairs, but rather to do things actively. These sentences, and the utterances realised by them, are called **performatives**, whereas the others are called **constatives** [Austin, 1962]. Performatives cannot be true or false, unlike constatives, but they can go wrong. Austin studied the ways in which they can go wrong, or be unhappy, or infelicitous.

Austin stated some conditions which performatives must meet if they are to be 'happy'. These were called **felicity conditions**, the technical expression used to refer to certain circumstances for the performance of a speech act to be recognised as intended. They were grouped into three categories (from A to C):

- **A. (1)** There must be a conventional procedure having certain conventional effect, and **(2)** the particular persons and circumstances in a given case must be appropriate for the particular procedure invoked.
- **B.** The procedure must be executed by all participants both **(1)** correctly and **(2)** completely.
- **C.** Often, **(1)** the participants must have the necessary thoughts, feelings and intentions, as defined in the procedure, and **(2)** if consequent conduct is defined, then the pertinent participants must do so.

Based on felicity conditions, Searle proposed the existence of five basic kinds of actions that one can perform in speaking, by means of five types of utterances [Searle, 1976], as shown on Table 8 (adapted from [Levinson, 1983] (page 240) and [Yule, 1996] (page 55)).

Table 8. Types of speech acts

Type	Use	Paradigm cases	Example[18]	Function (S=speaker, X=situation)
Representative	States what the speaker believes.	asserting, stating, concluding, boasting, describing, suggesting	I am a PhD student.	S believes X
Directive	Used in order to get someone else to do something.	requesting, advising, commanding, challenging, inviting, daring, entreating	Please, close the door.	S wants X
Commissive	Used to commit the speaker to some action in the future.	promising, pledging, threatening, vowing, offering	I'll help you.	S intends X
Expressive	States what the speaker feels.	greeting, thanking, apologising, welcoming, congratulating	Happy birthday!	S feels X
Declaration	Causes a change.	declaring, baptising, resigning, firing from employment, hiring, arresting	I pronounce you husband and wife.[19]	S causes X

However, Levinson claims that that typology is not definitive or exhaustive, and other classificatory schemes were also proposed [Levinson, 1983] (pages 240-241).

3.2.5 Conversational and discourse structure

The other pragmatic concepts previously explained in section 3.2 are somewhat related to the topics of this section.

Conversation is a kind of talk in which two or more people speak together, normally outside institutional settings, such as law courts, religious services, among others [Levinson, 1983] (page 284). Many metaphors have been applied to explain the structure of a conversation. Some see it like a dance, traffic movements and control, or the workings of a market economy. Some common terms often used when talking about conversational structure are as bellow:

- **Floor** – The right to speak;
- **Turn** – To have control of the floor;
- **Turn-taking** – The attempt to have control in a conversation;
- **Local management system** – The system in which turn-taking operates, as a conversation is a social activity. It can be seen as a set of conventions for getting turns, maintaining them, or transferring them;
- **Transition Relevance Place** (TRP) – Any possible modification in who has the turn. For instance, an absence of talk in commonly associated with a TRP.

Levinson suggests two approaches to the analysis of conversation: discourse analysis and conversational analysis [Levinson, 1983] (page 286). The former might be seen as including the latter, if conversation is regarded as a special type of discourse.

For the French philosopher Michel Foucault, discourse was the practice of making sense of symbols [Foucault., 1969]. The study of the structure of discourse includes what makes a well-formed text, explicit connections between sentences, and elements of textual organisation, among others. However, the pragmatic study of discourse focusses on what was communicated, although not said or written. It is necessary to consider "some psychological concepts such as background

[18] It is considered that the context is the usual one for that kind of sentences.
[19] It is said by a priest.

knowledge, beliefs and expectations", examining "what the speaker or writer has in mind" [Yule, 1996] (page 84).

Pauses, overlaps and backchannels are important in conversational analysis. Short pauses usually correspond to hesitations. However, when a speaker turns over the floor to another one, who does not speak, we have an **attributable** (or **significant**) **silence**. When the speakers try to speak simultaneously, there is an **overlap**. It is something usually felt as embarrassing. **Backchannel signals**, or just **backchannels**, are vocal indicators that someone is listening to what is being said, such as 'mmm'. Other indicators include smiles, head nods, and some gestures, among others.

The following topics are related to background knowledge:

- **Schema** (plural: schemata) – A preceding piece of knowledge in memory that helps in the interpretation of new experiences;
- **Frame** – A piece of knowledge that is invoked to supply an inferential base for the understanding of an utterance. It corresponds to a static pattern of the schema. A frame for a school might assume components like a canteen and classrooms, for instance. That topic is also addressed in artificial intelligence and cognitive psychology;
- **Script** – A previous knowledge structure with event sequences. For instance, most people have scripts for what normally happens during a dental treatment.

Nonverbal behaviours, such as facial expressions, head movements, body motion and posture, or eye gaze, do have a role in conversational analysis and are used to repeat, contradict, substitute, complement, accent, or regulate verbal communication [Knapp and Hall, 1997].

3.3 Delimitation

We can talk about different views of linguistics: the component, the perspective and the function view [Mey, 1993] (pages 8-10), and consequently how pragmatics is considered in accordance with these viewpoints.

For the **component view**, which is mainly based on a 'modular' notion of the human mind, the grammar of a language comprises several 'components' and each of them corresponds to a different human ability. That view became popular in the middle of the twentieth century [Chomsky, 1957] [Chomsky, 1965] and remains popular among cognitive scientists.

The **perspective view** perceives the human language in its diverse angles. Thus, linguistic pragmatics "can be said to characterise a new way of looking at things linguistic [i.e., a 'perspective'], rather than marking off clear borderlines to other disciplines" [Haberland and Mey, 1977]. That way, pragmatics should not be considered another component of a theory of language, but it provides a distinct perspective [Verschueren, 1987].

However, the pragmatic perspective might be suitable as a container for the various components of linguistics and it is possible to think about a pragmatic perspective and a pragmatic component in coexistence [Ostman, 1988]. A pragmatic component can address all the pragmatic **functions** that might be assigned to a language, along with a pragmatic perspective (related to how these functions work).

The idea of 'functions of language' dates from the mid-thirties. Figure 14 shows the three functions of language (representation, expression, and appeal) advanced in [Bühler, 1934]. For him, all functions coexist in any utterance, although one could be dominant. Later, Roman Jakobson expanded this model to accommodate three more functions: code, channel and poetic quality [Jakobson, 1960]. Subsequently, Karl Popper suggested an extra argumentative function [Popper, 1949].

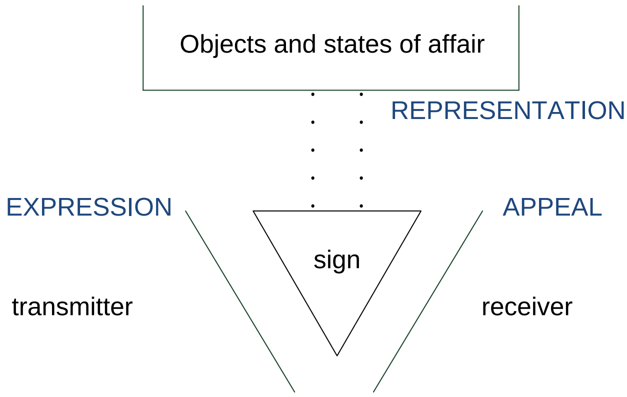


Figure 14. Buhler's organon model
(Adapted from [Hausser, 2001])

Common to these and other models is the significance of the user in the process of communication. As highlighted by Mey: "Messages are not just 'signals', relayed through impersonal channels; the human expression functions as an appeal to other users and as a means of social togetherness" [Mey, 1993] (page 10). He emphasises that this way of thinking about linguistics assuming functions of language, the **functional view**, allows the accommodation of the different views of 'componentialists', interested in technical subjects such as presuppositions and implicatures, and 'perspectivists', interested in concepts like motivations and effects.

3.3.1 The pragmatic waste-basket

Waste-basket is a term applied extensively by Mey to characterise a certain view about how the term 'pragmatics' has been used over the years [Mey, 1993], which to some extend is derived from the fact that "pragmatics is tolerant" [Caffi, 1994], by not putting apart any plausible activity that involves human language users.

The term waste-basket was formerly used in Linguistics but referring to semantics. It was applied by Yehoshua Bar-Hillel, who designated semantics as the "waste-basket of syntax" ([Bar-Hillel, 1971]. Along time linguistics has started to have more unresolved questions, which were not under syntax or semantics, and the 'pragmatic basket' has started to collect many heterogeneous problems.

A quite similar view is enunciated in the following sentence: "Semanticists can continue to use "pragmatics" as its garbage can, ... as the place where we semanticists send all the difficulties we do not want to deal with" [Manor, 2001].

Also Morgan expressed his concerns about the broad usage of the term more than 30 years ago: "The term "pragmatics" is in wide and fashionable use today in linguistics, philosophy, psychology, and adjoining fields. Its applications range from the narrow scope of the study of the meaning of deictic expressions, to the use as a catch-all category covering all aspects of communication that cannot be analysed as literal meaning, including even matters of turn-taking and social interaction" [Morgan, 1977].

3.4 Metapragmatics

The prefix *meta* comes from a Greek word that means 'change' or 'shift'. For instance, the word 'metalanguage' indicates a language used to talk about language, a step above the language itself. It studies what happens on another level (the object language). Probably, it was Jakobson who first used the word metapragmatics [Jakobson, 1960].

Thus, 'metapragmatics' is concerned with the 'object pragmatics' (the next higher level). For instance, if we discuss the various speech acts often used, we are concerned with pragmatics. But, "if we ask ourselves what principles govern the use of speech acts, and how they are related to other human communicative activities, we are touching upon metapragmatic questions" [Mey, 1993] (page 175).

We can talk about three views of metapragmatics, as pointed in [Caffi, 1994]:

- One view is devoted to the definition of pragmatics and its delimitation,
- Another view is dedicated to the study of how to act through the use of words,
- The third sense is related to reflexive, metalinguistic uses of language and common knowledge. For instance, when someone says '... as I was telling you, ...', it is being announced that there will follow a repetition or a concluding statement.

One of the reasons why metapragmatics is important is to avoid the wastebasket view (previously discussed in section 3.3.1) by clearly delimitating the field of pragmatics.

3.5 Pragmatics across different cultures and languages

Over time it has been discussed if many pragmatic phenomena are dependent or not on culture. For instance, silence as a mechanism of 'turn-taking' in conversation analysis is not regarded the same way in all cultures. The length of silence can be differently interpreted in a conversation between speakers from different cultures. For many English speakers a silence of more than one and a half seconds is regarded as a TRP, but for a speaker of an American Indian language that silence can be interpreted as a 'inter-sentence', leading to some misunderstandings in a conversation [Mey, 1993] (pages 277 and 278).

Levinson highlights that it is unknown to what extent conversational organisation is universal, as few comparative studies were carried out. Some pragmatic phenomena seem to be common in all cultures, as the existence of greeting and parting practices, but he highlights the existence of "clear pan-cultural principles governing the production of 'polite' or socially appropriate interaction" [Levinson, 1983] (pages 45-46).

Levinson distinguishes universal pragmatics from language-specific pragmatics. For instance, the pragmatics of Japanese would be significantly concerned with the grammaticalisation of the relative social ranks of participants and referents [Levinson, 1983] (page 10), enforcing the coding of social deixis.

Paul Ekman suggested the existence of many spontaneous facial expressions that reflect the internal state of people regardless of culture [Ekman, 1989, 1992; Ekman and Davidson, 1994]. He uses the term 'emblem' to refer to any gesture with a precise meaning that is invariable in all cultures.

Related to pragmatics is cross-cultural pragmatics, whose area of investigation is dedicated to the study of divergences in expectations based on cultural schemata. Schema is seen as a "cognitive construct which combines the structural and processing aspects of knowledge" [Rice, 1980].

3.6 Cyberpragmatics

Cyberpragmatics is devoted to the pragmatics of digital discourses. It was coined by Yus in 2001 [Yus, 2001]. All utterances, sent through Internet or not, inform less than the thoughts desired to be communicated with them by the speaker or sender [Yus, 2005]. The author emphasises that "context is never a taken-for-granted, static ingredient of communication, but dynamic and situation-specific information that is accessed by the human mind while interpreting verbal messages and nonverbal behaviours".

Some of the general characteristics of cyberpragmatics are [Yus, 2008]:

- On internet, "addresser users" have communicative intentions and formulate their utterances in order to make them successful;

- The "addressee users" rely on inferential strategies in order to get the most relevant interpretation using any available contextual information.

It is recognised that "addressee users" might need an additional effort when recognising the relevant interpretation in Internet-mediated communications. However Internet users also "evaluate interpretations, access contextual information, derive implications, enrich explicit content, infer emotional or attitudinal qualities attached to the messages, and identify underlying intentions" [Yus, 2008]. The author mentions the use of "textual deformation" as a way to counterbalance the lack of information arisen from nonverbal clues available in face-to-face communication, when some characters are used more than once to denote a more intense user emotional state[20], such as in *Hellooooooooo!* or *Hello!!!!!!!*.

It seems that people have found ways to deal with the limitations of computer-mediated communication, maybe dictated by the fact that humans are "fundamentally emotional and social creatures" [Immordino-Yang and Damasio, 2007]. Also, Mey highlights that "the desire to communicate, and the need to express themselves, are natural for all humans" [Mey, 1993] (page 43). For instance, people regularly use emoticons, or explicitly verbalise emotions in online communications.

There was an old assumption that "language carries information and NVC communicates emotion"[21] [Good, 1996]. However, many arguments against that position have emerged and there is an indication of the existence of a number of functions potentially associated with emotions in cognitive processes, which Good argues probably have great importance to any psychological theory of pragmatic processing and sustains that "if systems are not equipped to handle these emotional components of interaction, then they will be ... 'Prefrontal Kantians'"[22].

The importance of emotions and the body language associated to them in conversations as a complement or a modification of what is being said is recognise and led to the emergence of the use of 'smileys' and others 'emoticons' to express emotions in written messages.

Derks et al reviewed research on the subject of possible differences between CMC (Computer-Mediated Communication) and F2F (face-to-face) interactions regarding the communication of emotions. Although recognising that the characteristics of CMC might reduce spontaneity, they concluded that "emotional communication online and in F2F are surprisingly similar and that online communication even seems to reinforce rather than inhibit the expression of emotions" [Derks, Fischer et al., 2008]. Thus, often people verbalise emotions explicitly in online communications.

Justifying the need for another journal (*Pragmatics and Society*) in the field of pragmatics, in 2010 Jacob Mey, Hartmut Haberland and Kerstin Fischer said: "Pragmatics thus needs to address the affordances of interactions between humans interacting through the electronic media, such as the interactive classroom, websites, blogs, e-mail, and all sorts of internet-based information-sharing and activity-oriented fora. ...pragmatics now also needs to incorporate not only the interaction between humans, but also that between humans and computerised devices..." [Mey, Haberland et al., 2010].

3.7 Conclusions

The interest in pragmatics has grown over the years and this study area came to other fields beyond linguistics and philosophy of language. This chapter summarised many concepts related to pragmatics. Although the delimitation of this area is not consensual, from what was exposed in this chapter, we can say that 'user' and 'context' are central aspects in any pragmatic characterisation.

---

[20] In Annex 2 it is precisely characterised the notion of emotion and some others related to this subject.
[21] NVC is an acronym for 'Non-Vocal Communication'.
[22] 'Prefrontal Kantians' is part of the title of a paper ("Prefrontal Kantians. A Review of "Decartes' Error: Emotion, Reason and the Human Brain" by Antonio R. Damasio") [Sousa, 1996].

Also covered was the fact that communication does not only comprise words, but also emotions, which might contradict or reinforce the message being transmitted.

Careful consideration should be given to some issues. One of them is that pragmatic phenomena might be subject to cross-cultural variations. For instance, the Grice maxims depend on culture, with some societies Politeness is ranked above Quality, while in others the Maxim of Quality is the highest [Leech, 1983]. Another aspect is that dealing with interpretations, and how what is said is understood, pragmatics studies can be difficult to be carried out without some subjectivity.

This chapter also briefly examined the necessity of pragmatics also regard computer mediated communications.

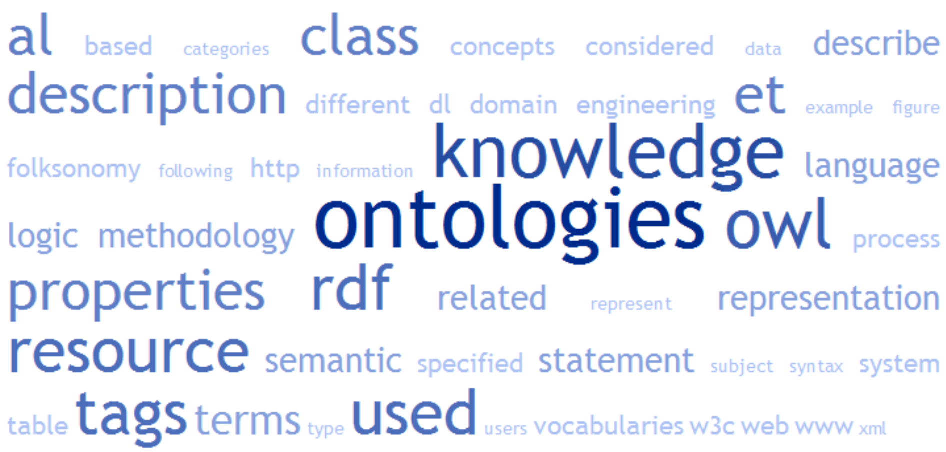
Finally, context is a key concept in pragmatics. Applied to learning objects in repositories, context can be seen as all factors that have an effect on their use or are related to their usage.

Chapter 4

KNOWLEDGE REPRESENTATION AND ENGINEERING: PRINCIPLES, METHODS AND TECHNOLOGIES

In order to use knowledge in an e-learning system, it is necessary its acquisition, the use of one or more knowledge representation formalisms to encode it using representation techniques, languages, also considering other aspects, which are part of a huge process – knowledge engineering. Knowledge representation allows for different e-learning systems to come to a common understanding of the semantics of learning objects.

This chapter is devoted to knowledge representation and engineering, and their foundations. It also provides an overview of some W3C languages.



4.1 Introduction

One possible distinction between data, information and knowledge is provided in [Morville, 2005]. Data are "a string of identified but unevaluated symbols", information is characterised as "evaluated, validated, or useful data", whereas knowledge is "information in the context of understanding". However, these concepts are related to different non-static stages in the transforming process that goes from data to knowledge [Kendal and Creen, 2006]. This last dimension of knowledge is the most related to semantics, which is connected with understanding the nature of meaning.

However, different areas have different visions of what 'meaning' is. In some fields, such as knowledge representation and artificial intelligence, "well defined syntactic structures are used to represent information or knowledge where these structures have definite semantic interpretations associated with them". Formal semantics is usually used to refer to semantics that are expressed in some well formed syntactic structure [Sheth, Ramakrishnan et al., 2005].

The signs and their relationship to what they represent have been studied extensively in semiotics (see previous chapter). In other areas, some definitions of knowledge representation have also been provided over the years. Five distinct definitions were provided in [Davis, Shrobe et al., 1993], each one is based on a slightly different perspective:

- "A knowledge representation (KR) is most fundamentally a surrogate, a substitute for the thing itself, used to enable an entity to determine consequences by thinking rather than acting, i.e., by reasoning about the world rather than taking action in it";
- "It is a set of ontological commitments, i.e., an answer to the question: In what terms should I think about the world?";
- "It is a fragmentary theory of intelligent reasoning, expressed in terms of three components: (i) the representation's fundamental conception of intelligent reasoning; (ii) the set of inferences the representation sanctions; and (iii) the set of inferences it recommends";
- "It is a medium for pragmatically efficient computation, i.e., the computational environment in which thinking is accomplished. One contribution to this pragmatic efficiency is supplied by the guidance a representation provides for organising information so as to facilitate making the recommended inferences";
- "It is a medium of human expression, i.e., a language in which we say things about the world".

According to Sowa, "knowledge representation is the application of logic and ontology to the task of constructing computable models for some domain" [Sowa, 1999](p. xii).

Methods and techniques to represent and manage knowledge in Knowledge Bases (KBs) have been used for a long time in Knowledge Base Systems (KBSs), albeit they have recognised a rise in interest with the advent of the Semantic Web in the last years.

In fact, in diverse application areas, it is necessary to capture semantic relationships in more or less structured ways. This necessity has also arisen in educational systems to formalise many aspects, such as the content, context and structure of courses [Stojanovic, Staab et al., 2001]. To semantically characterise learning objects, it is necessary to represent the knowledge of the domains of interest.

This chapter is structured as follows. Section 4.2 describes knowledge representation methods. A brief overview of some formal languages to describe knowledge is provided in section 4.3, which represent the theories and techniques underlying the Semantic Web, detailing Description Logics. The W3C languages are based on Description Logics and they are described in section 4.4. Section 4.5 discusses knowledge engineering, specially the subarea of ontological engineering and the possibilities to obtain semantic models. Finally, section 4.6 summarises this chapter with some remarks to be considered regarding formal semantic characterisation of learning objects.

4.2 Methods of knowledge representation

The representation of entities and relationships in structures that reflect knowledge of a specific domain is important in many different areas. Diverse aspects of e-learning applications or instructional materials can be described or structured through the use of dictionaries, controlled vocabulary, thesaurus, knowledge maps, and ontologies, among others. Some examples of different approaches to formalise knowledge representation in learning settings are as follows:

- Controlled vocabularies have been suggested for use with element 1.6 (keyword) of LOM (see section 2.2.2), like the GEM Controlled Vocabulary (http://www.thegateway.org/about/documentation), and ISO 3166 for country identification in element 1.7 (Coverage) of LOM. In [Currier, Campbell et al., 2005] the authors provide a detailed description of vocabularies used in education;
- A multi-lingual thesaurus is applied in the Learning Resource Exchange (LRE) platform to normalise the subject descriptions of learning materials [Jokisalo, 2010];
- An example of a common use of taxonomies is the specification of levels of learning objectives based on Bloom's taxonomy [Bloom, 1956]. The DialogPlus learning activities taxonomy [Conole and Fill, 2005] and the 8LEM taxonomy [Verpoorten, Poumay et al., 2007] are examples of taxonomies for learning activities (or events, in the 8LEM nomenclature);
- For representing the context of a course, specifying learning scenarios or facilitating the interchange of the resources, the use of pedagogical or learning scenario ontologies have been suggested [Crampes and Ranwez, 2000; Doan, Bourda et al., 2004; Rius, Sicilia et al., 2008].

In this section some common methods of knowledge representation are sketched: thesauri, taxonomy, ontology and folksonomy. Some of them are often used with controlled vocabularies, which provide the words or expressions that can be used in a given situation.

4.2.1 Thesauri

A thesaurus is a list of semantically related terms, and it is usually used for indexing. These terms are often called 'descriptors' because they are used to describe the contents of documents.

The following semantic relationships between terms can be considered:

- **Equivalence relationships** - Relationships where a term (the preferred term which is called the descriptor) has the same meaning as one or more terms (non-descriptors);
- **Hierarchical relationships** - These are based on levels of superordination and subordination between terms. The following relationship indicators can be specified: BT (Broader Term), a label for the superordinate term, and NT (Narrower Term), a label for the subordinate term;
- **Associative relationships** - These exist between terms that are closely related but not hierarchically. For instance, an author can be connected to a resource.

4.2.2 Taxonomy

A taxonomy is a hierarchy of categories used for classification purposes. It differs from a thesaurus in the way terms are related to each other. In a thesaurus there is a greater variety of relationships among the thesaurus terms.

It has its roots in the classification of species. Aristotle (384-322 BC) did an effort to classify all the kinds of animals in his History of Animals. Carolus Linnaeus[23] (1707-1778), Swedish naturalist, published 'The System of Nature', proposing the standard genus-species system of taxonomic

[23] It is the name by which he proposed the well-known classification system. His birth name is Carl Linnaeus, but he gave himself some more: Carl Linne, Carl von Linne, Carolus Linnaeus and Caroli Linnaei.

binomial nomenclature to classify animals and plants, with Kingdom, Phylum, Class, Order, Family, Genus, Species.

While in a classical taxonomy each element can only be accommodated in one branch of the hierarchical tree, in many environments and areas, such restriction is not practical, and then not followed [Varma, 2007].

4.2.3 Ontology

The word 'ontology' derives from the Greek words *ontos*, for 'being', and *logos*, for 'word'.

Ontology is a concept which has been definitely connected to Aristotle since he wrote Metaphysics in 350 BC. Nonetheless, he never used the terms ontology and metaphysics. The title 'metaphysics' was adopted later, when verified that his writings were beyond physics. In an effort to answer what is the nature of being, he surmised that all beings must have some 'thing', some attribute, which provides the property 'being' to objects.

Before the end of seventeenth century, *Metaphysica generalis* (general metaphysics), in contrast to *Metaphysica specialis*, was also called 'ontologia' (Ontology), and was related to the investigation of the most general concepts of being, existence and reality.

The use of Ontologies in Computer Science was motivated by the need in different fields (Artificial Intelligence, Software Engineering and Database) to represent knowledge [Smith and Welty, 2001].

Currently, Gruber's definition of ontology is the most used in computer science: ontology is an "explicit specification of a conceptualization" [Gruber, 1993]. It is interesting to note that an ontology always provides a taxonomy in a machine-readable form. In reality the distinction between them is not very rigid with some taxonomies being considered full ontologies [Studer, Benjamins et al., 1998]. Studer and colleagues discuss the consensus that often appears associated to ontologies, arguing that who must reach that named consensus depends on the specific context. In effect, Yahoo! Directory (http://dir.yahoo.com/), a taxonomy for searching the web, could be considered also an ontology [Lassila and McGuinness, 2001] because it provides a consensual conceptualisation of a given domain.

However, it is that consensus that distinguishes ontologies from knowledge bases according to Guarino, as the latter "may also describe facts and assertions related to a particular state of affairs or a particular epistemic state", and thus may describe facts that are not always true, which is assumed to happen in an ontology at least for a community of users [Guarino, 1997].

In an ontology there is a hierarchy of concepts with properties. The concepts are interconnected by relations. Concepts might haves instances. All these together allow knowledge formalisation in a degree that is not possible to achieve with thesaurus, for instance, where the kind of relations is rigid. Thus, using ontologies it is possible to represent the knowledge of interest. It is important to note that the concept of ontology is independent of how it is represented.

Ontologies are built using knowledge representation languages and it is not a straightforward process. Concept maps can assist in this step and be successfully used for knowledge elicitation carried by domain experts before developing ontologies [Ortiz, Azevedo et al., 2009].

Concept maps are tools for organising, representing and communicating knowledge[24]. They have been used to enable learners to construct their own knowledge models and compare them with others [Novak, 1995]. A concept map has concepts confined in circles or boxes, and relationships between concepts specified by lines connecting two concepts. In terms of nomenclature, the concepts, relations and propositions, which exist in concept maps, correspond to classes, properties, and restrictions and axioms, with more formalism and rigid hierarchical structure.

[24] Apart from this chapter was the description of Topic Maps, despite this technology was considered in an early phase of this work. It formalises the notions presented in Concept Maps, but it was not considered mature enough to be used in a robust application. Even today, there is not any standard query language.

4.2.3.1 Types of ontologies

Ontologies can be classified according to their level of generality [Guarino, 1998]:

- **Upper-level ontologies** - also known as foundational or top-level ontologies, they attempt to formalise the more general concepts in a conception of the world. Therefore, they are generally use-independent ontologies;
- **Task ontologies** – They describe the vocabulary related to a task;
- **Domain ontologies** – They define the terminology pertinent to a particular area;
- **Application ontologies** – They describe the vocabulary pertinent to a certain domain, but considered for a specific task or application.

Figure 15 depicts these kinds of ontologies. Domain and task ontologies can refine top-level concepts that appear on upper ontologies, while application ontologies can reuse concepts from both of them.

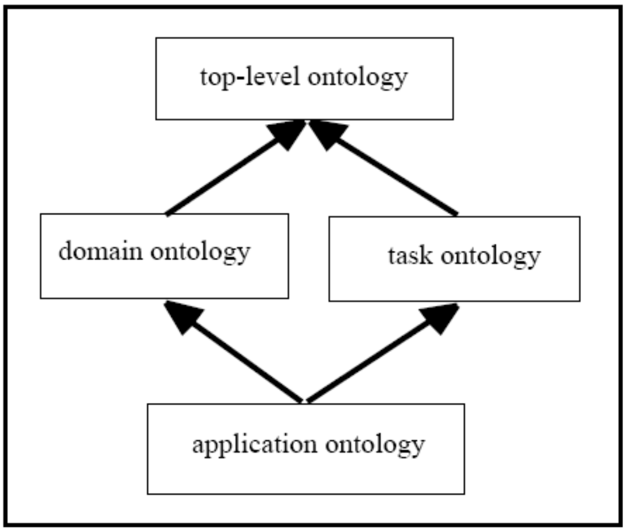


Figure 15. Kinds of ontologies according to their level of generality (From [Guarino, 1997])

Another classification was proposed in [Mizoguchi, 2003]: **heavy-weight ontology** and **light-weight ontology**. The former has a higher degree of formalisation, with a precise definition of concepts. The same idea was behind the categories proposed in [Lassila and McGuinness, 2001], with emphasis on the detail level of its internal structure, as represented in Figure 16.

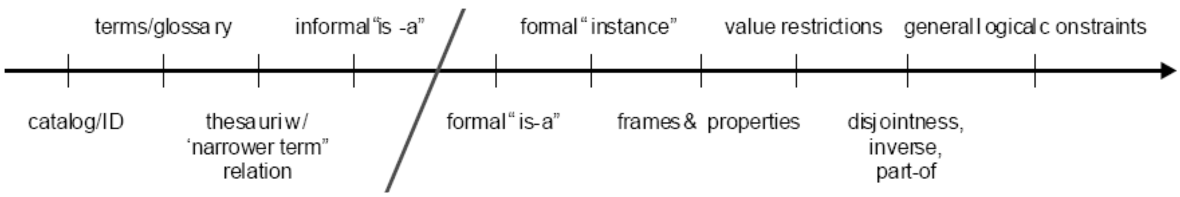


Figure 16. Lassila and McGuinness categories of ontologies (Source:[Lassila and McGuinness, 2001])

4.2.4 Folksonomy

Folksonomy is a term derived from folk and taxonomy. The name was coined by Vander Wal [Wal, 2005]. In a folksonomy plain keywords, also known as tags, are used. They are provided by users to categorise various types of resources (e.g., photos, videos, articles, scientific or educational resources).

It is a bottom-up classification system that emerges from social tagging. It is basically a quadruple F := (U; T; R; A), where U, T, R are respectively finite sets of instances of users ('taggers'), tags, and resources, and A defines a relation, the tag assignment, between these sets, that is, A is contained in U x T x R [Hotho, Jaschke et al., 2006]. It is often also needed to consider their time-stamp. As an example, we can have the tags used by `user1` for `resource2` like `{user1, {apple, macintosh, stevejobs}, resource2, November 23, 2009 10:30}`. In a few words, a folksonomy is a set of tags expressing a community vocabulary, which is dynamic.

By tagging users add their personal view about a resource, to help its later findability by themselves or others. Del.icio.us (http://del.icio.us), Flickr (www.flickr.com), 43things (www.43things.com), Furl (www.furl.net) and Technorati (www.technorati.com) are collaborative systems that use tagging to achieve enriched metadata vocabularies that can be used to perform queries, to supervise changes in areas of interest or to discover new trends.

Three different categories of folksonomies can be considered: broad, extended narrow and narrow folksonomies, and in this document their usage is consistent with the following definitions:

- **Narrow folksonomies** – These are the ones where only the authors are allowed to provide tags to resources. These tags are considered in searches carried out by the users. The folksonomy used by YouTube fits into this category;
- **Extended narrow folksonomies** – Folksonomies where tags are supplied by the authors and other users, but each tag only exists once ("set model" of folksonomies [Marlow, Naaman et al., 2006]), e. g., Flickr's folksonomy;
- **Broad folksonomies** – Tags are provided by authors and other users, allowing multiple allocation of a given tag to the same resource ("bag model" of folksonomies [Marlow, Naaman et al., 2006]). Del.ici.ous uses this kind of folksnomies.

However, many tags can be considered inappropriate for retrieving resources, because of misspellings, use of unpopular tags or personal vocabularies. But especially the latter group permits the characterisation of expressions used by only some sub-communities, which can later be used to improve the retrieval of resources [Lux, Granitzer et al., 2007]. In any way, tags add additional information to resources and can be considered valuable for retrieving resources.

Golder and Huberman's work on the collaborative tagging system del.ici.ous concerning folksonomy formation showed that each tag's percentage of use, particularly the popular tags, tends to remain the same for a certain amount of user contribution (approximately 100 tags) [Golder and Huberman, 2006]. They analysed the process of stabilising tags, when we say we have a folksonomy and the reasons for that - imitation and shared knowledge. The probability of imitating tag assignments previously made is estimated to be between 70 and 90% [Staab and Dellschaft, 2008].

Brooks and Montarez discuss three frequent aspects in tagging motivations: "annotation information for personal use, placing information into broadly defined categories, and annotating particular articles so as to describe their contents" [Brooks and Montanez, 2005].

A recent experimental research found that tagging can be a powerful mechanism for users to find related resources, and can lead to enormously useful bottom-up meanings for resources, when massively done [Xu, Dichev et al., 2009].

4.2.4.1 Tag categories

Linguistic peculiarities of tags are a research subject deeply considered at the moment, and the occurrence of regularities regarding certain forms of tags. Several studies have derived tag categories from Flickr and del.icio.us, exploring their folksonomies. Winget arranged tags from Flickr folksonomy into five categories [Winget, 2006]:

Table 9. Categories of Flickr tags

Type	Example
Date and time	'0611'
Geographical	paris
Narrative	'landscape'
Characterisations	'happy'
Individually defined tags	'mynecktie'

Golder and Huberman found seven categories of del.icio.us tags [Golder and Huberman, 2006a]:

Table 10. Categories of del.icio.us tags

Type	Description
Topics	Tags that correspond to generic descriptions, stating what (or who) the resource is about
Type	Tags related to the resource format
Adjectives	Tags which reproduce an opinion about the resource
Qualifying	Tags that refine some other category. Tags in that category can be just a number, for instance.
Ownership	Category which aggregates tags identifying who owns the resource;
Self reference	Tags that describe a relation to the tagger, usually beginning with 'my'
Task organising	Tags related to a task to be performed using the resource, like 'toread', or related to a task being carried by users like 'jobsearch'

Sen et al. collapsed these seven classes into three more general ones [Sen, Lam et al., 2006]:

Table 11. A more reduced classification of del.icio.us tags

Type	Description
Factual tags	Tags related to facts about a resource
Subjective tags	Tags which reflects the users' opinions on the tagged resource
Personal tags	Tags whose audience is the own tagger

It is interesting to note that tags like 'to read', 'todo', 'towatch', among others, describe a planned action in regard to the tagged resource, embodying a perfomative act (see 3.2.4 Speech acts). Some researchers refer to them as 'functional tags' [Kroski, 2005], 'signalling tags' [Dennis, 2006], and 'time and task related tags' [Kipp, 2007]. The tagged resources are likely to have been regarded as useful as these king of tags "express a response from the user rather than a statement of the aboutness of the document; they are intrinsically time-sensitive; they suggest an active engagement with the text, in which the user is linking the perceived subject matter with a specific task or a specific set of interests" [Kipp and Campbell, 2006].

4.3 Foundations

Logics are formal languages for representing knowledge such that conclusions can be derived, including modal logic, description logic, non-monotonic logic, intuitionistic logic, among others. The use of formal logic to represent and infer knowledge prevailed over other formalisms, specially First-Order Logic (FOL), which has proved to be the most successful forms of logic [Bagheri, Poizat et al., 2009].

Great difference exists between first and second-order logic, but "any logic of an order higher than 2 can be, without any loss of generality, considered a mere 'notational variant' of second-order logic" [Peregrin, 1996].

A logic typically has a well defined syntax, semantics and proof theory. John Sowa said about the expressivity of logic: "Everything that can be stated clearly and precisely in any natural language can be expressed in logic. There may be aspects of love, poetry, and jokes that are too elusive to state clearly. But anything that can be implemented on a digital computer in any programming language can be specified in logic" [Sowa, 2009].

To express knowledge in FOL, the simple facts of interest have to be represented as sentences in FOL. The terminology can be "extensionalised", for example, a `sister` statement can be expressed using `sibling` and `female` statement, such as

$\forall x, \forall y\ sister(x,y) \equiv sibling(x,y) \wedge female(x)$

A first-order knowledge base is a collection of sentences in FOL indicating what was known about the domain. Given a yes-no question also formulated in FOL, it is possible to know under what conditions the answer should be yes, no or unknown, and

$KB \models \alpha$ iff every interpretation fulfilling all the sentences in the KB also fulfils α.

Description Logics are subsets of first-order logic, widely used as knowledge representation formalism, and related to some of the W3C languages described in part 4.4 of this chapter. The following section briefly introduces it.

4.3.1 Description Logics

Description logics (DLs) constitute a family of languages for modelling domains of interest in terms of concepts (classes), objects (the individuals that come about in the domain), and roles (relationships) between concepts, and for reasoning about them. A domain can be specified by providing the definition of classes, and describing classes using a set of logical operators. Hence DLs are class-based and logic-based knowledge representation languages. When using DLs, one can specify the necessary and sufficient conditions that objects of a given class must observe.

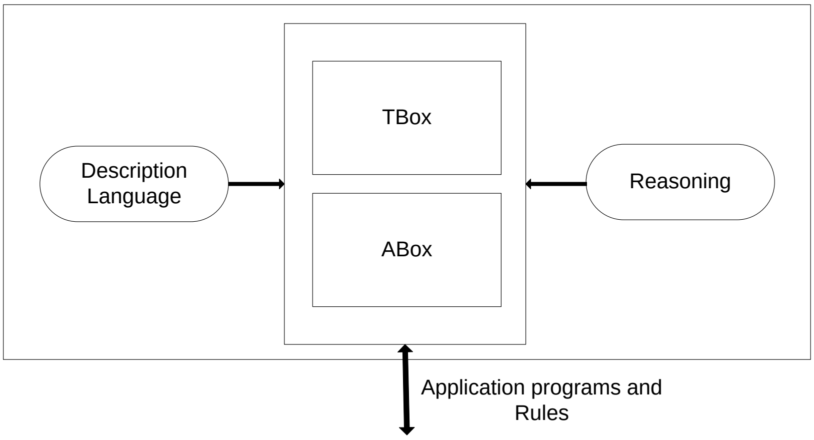


Figure 17. Architecture of a KR system based on DL (adapted from [Baader and Nutt, 2003])

Figure 17 outlines the general architecture of a KR system based on Description Logics and shows that a KB has two main elements: TBox (Terminological Box) and ABox (Assertional Box). The former specifies the vocabulary of an application domain, i.e., the concepts and roles, while the latter contains concept assertions and role assertions. But a DL system not only supplies terminologies and assertions, but also mechanisms to reason about them.

There is a naming convention for description logics and the symbols used in its designation (capital letters in a handwriting style) makes evident its expressivity (see Table 12).

Table 12. Naming convention for Description Logics

Symbol	Expressivity
$\mathcal{F}$	Functional properties
$\mathcal{E}$	Full existential qualification
$\mathcal{U}$	Concept union
$\mathcal{C}$	Complex concept negation
$\mathcal{S}$	An abbreviation for $\mathcal{ALC}$ with transitive roles
$\mathcal{H}$	Role hierarchy
$\mathcal{R}$	Limited complex role inclusion axioms; reflexivity and irreflexivity; role disjointness
$\mathcal{O}$	Nominals
$\mathcal{I}$	Inverse properties
$\mathcal{N}$	Cardinality restrictions
$\mathcal{Q}$	Qualified cardinality restrictions
$(\mathcal{D})$	Use of datatype properties, data values or data types

The designation of a Description Logics language use letters that correspond to the available constructors in the language. For instance, for the description logic $\mathcal{SHOIN(D)}$ it is immediate to know the supported constructors: $\mathcal{S}$ (Role transitivity), $\mathcal{H}$ (Role hierarchy), $\mathcal{O}$ (Nominals), $\mathcal{I}$ (Inverse properties), $\mathcal{N}$ (Cardinality restrictions), and $(\mathcal{D})$ (datatype properties, data values and types).

Some canonical DLs do not exactly fit this convention, like AL (Attributive Language). This language is the base DL.

A knowledge base represented in Descriptive Logic assumes open world semantics (OWA: Open World Assumption). This means that one cannot assume anything unless explicitly stated or derived from something else. For example, if in a knowledge base there are the statements 'Paul is the chief of Peter' and 'Paul is the chief of Carl', an inference engine cannot deduce that 'Paul has two subordinates' (unless it was explicitly stated or could be deduced from other statements), as it could have more. It may be inferred that 'Paul has at least two subordinates' because it assumes a semantic DL Unique Name (UNA: Unique Names Assumption), and therefore Peter and Carl, as they have different names, are different individuals.

As an example of the open world versus the closed world, it can be considered a database that stores student information for a certain institution I. If there is no record about student A, it is known that A is not a student at that institution. The database presumes it represents a complete knowledge model (closed world assumption). In the Semantic Web the absence of a statement about student A does not imply anything about his/her existence (open world assumption).

4.4 W3C Languages

W3C languages are Web-based languages endorsed by the W3C, also called Semantic Web languages.

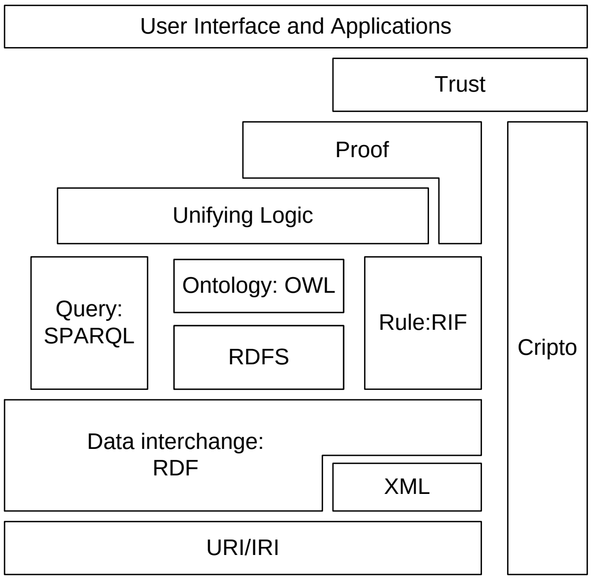


Figure 18. Semantic Web layer-cake
(Adapted from http://www.w3.org/2007/03/layerCake.png)

The main technologies of the Semantic Web fit into a set of layered specifications, often referred as 'the Semantic Web layer-cake', visible in Figure 18. Some of its components are:

- **URIs/IRIs** – Universal Resource Identifiers (URIs) and Internationalised Resource Identifiers (IRIs) are sequences of characters used to unambiguous identify Web resources, but IRIs use a wider range of characters and have some other internalisation concerns;
- **XML** – Extensible Markup Language (XML) 1.0 is a W3C recommendation since 2000. It is derived from SGML and provides a way to exchange structured data;
- **RDF, RDFS, OWL, SPARQL and RIF** – RDF, RDFS and OWL are components related to knowledge representation and are discussed later in this chapter. RIF is the W3C Rule Interchange Format for declaring executable rules. SPARQL, a standardised query language, makes possible putting together decentralised collections of RDF data;
- **Unifying Logic and Proof** - These layers refer to the use of mechanisms, such as inference engines, to make use of knowledge held by other layers, to obtain new and traceable knowledge that was not initially explicitly declared. The Semantic Web needs tools able to provide proof checking mechanisms;
- **Trust** - Using cryptography, which crosses all the other primary layers, a system must be able to rigorously check the authenticity of data.

On top of the layer-cake is the user interface and applications that use one of more of the underlying levels.

Table 13 compares some of the W3C languages (OWL Lite, OWL DL and OWL Full represent sub-languages of the OWL language) in relation to their support for some constructs. In that table an `X` in a cell means that the language in that column supports the construct in that line, `XI` means support by inheritance, `XE` means extended fully support and `X-` means support with restrictions.

Table 13. A comparison of W3C languages - supported constructs

Category	Subcategory	Construct	RDF	RDFS	OWL Lite	OWL DL	OWL Full
class	definition	class		X	XE	XE	XE
		enumerated class				X	X
		restriction			X-	X	X
		intersectionOf			X-	X	X
		unionOf				X	X
		complementOf				X	X
	axiom	subclasOf		X	XI	XI	XI
		equality			X-	X	X
		disjointWith				X	X
relation	definition	Property	X	XI	XE	XE	XE
		domain, range		X	XI	XI	XI
		subPropertyOf		X	XI	XI	XI
	axiom	(inverse) functional			X-	X-	X
		equality, inverseOf			X	X	X
		transitive, symmetric			X-	X-	X
instance	definition	type	X	XI	XI	XI	XI
	axiom	(in)equality			X-	X-	X-

RDF and RDFS are described in next section, while OWL is discussed in section 4.4.2. Sections 4.4.3 and 4.4.4 briefly describe query and rule languages, respectively.

4.4.1 RDF and RDFS

The Resource Description Framework (RDF) [Manola, Miller et al., 2004] is a data model for representing information about World Wide Web resources. It was released in 1998 as a W3C recommendation. It can be used to describe Web resources, as well as anything that can be identified on Web.

Information is stored as assertions, called RDF statements, understandable by suitable intelligent agents, search engines and browsers [Lassila and Swick, 1999] [Decker, Melnik et al., 2000].

The vocabulary of the RDF data model contains resources, properties and statements:

- Resources are objects uniquely identifiable by a Uniform Resource Identifier (URI) (Berners-Lee, Fielding & Masinter, 1998);
- Properties are attributes used to describe resources;
- Statements are structures containing a resource, a property, and a value indicating the value of the property for the specified resource. The value of a property can eventually be another resource identifier.

A particular terminology is used for talking about the various parts of statements. In RDF a `subject` (the part that identifies the thing the statement is about) is connected to an `object` (the part that identifies the value of that property) by some `predicate` (the part that identifies the property of the subject that the statement specifies). Thus, the English statement '`http://www.dei.isep.ipp.pt/~iazevedo has a date of creation whose value is 2000.01.20`', has the next RDF terms for the various parts of the statement:

- subject - `http://www.dei.isep.ipp.pt/~iazevedo;`
- predicate – `dateOfCreation`;

- object – `2000.01.20`.

These three parts constitute what is known as an RDF triple.

The RDF statements are often expressed in XML (RDF/XML syntax), which is the W3C recommendation. The subjects are referenced using the XML attribute `rdf:about` and the triples with each of these as subjects appear as subelements within these definitions.

The same statement presented before but now in RDF/XML syntax:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:dc="http://purl.org/dc/elements/1.0/">

  <rdf:Description rdf:about="http://www.dei.isep.ipp.pt/~iazevedo">
    <dc:date>2000.01.20</dc:date>
  </rdf:Description>

</rdf:RDF>
```

Example 12. A statement in RDF/XML syntax

Namespaces provide a method to qualify elements and attributes used in XML documents. This mechanism provided by XML is used to refer to where both the used properties and the RDF elements are defined (indicated by the prefix rdf and dc in Example 12). Although the particular namespace prefix used is quite irrelevant, there are some common namespace prefixes that everybody uses for clarity. Some familiar namespace prefixes are shown in Table 14. They are adopted in this document, unless otherwise stated.

Table 14. Common namespace prefixes

Prefix	URI
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
owl	http://www.w3.org/2002/07/owl#
xml	http://www.w3.org/XML/1998/namespace
xsl	http://www.w3.org/1999/XSL/Transform
xsd	http://www.w3.org/2001/XMLSchema
html	http://www.w3.org/1999/xhtml

N3 (Notation 3) is another syntax for expressing RDF statements, which is more compact than the RDF/XML notation. Example 13 shows the previous example, but now in N3 syntax. It is not evident from that, but N3 provides for a compact representation of triples that share a common subject. It uses a semicolon (;) to specify that another triple with identical subject follows and only the predicate and object need to be indicated.

```
@prefix dc: <http://purl.org/dc/elements/1.1/>.

<http://www.dei.isep.ipp.pt/~iazevedo>
dc:date "2000.01.20".
```

Example 13. A statement in N3 syntax

Also, when several triples share the same subject and predicate, N3 uses a comma (,) to separate the objects, and neither the subject nor the predicate needs to be repeated. In addition, N3 provides some abbreviations for brevity, but also to improve readability. One often employed is the word `a` to mean “`rdf:type`”.

The Terse RDF Triple Language, or Turtle, the simplest and most concise RDF serialisation, is a simplified subset of N3. It is an alternative syntax not endorsed by any official body. Currently it is published as a potential input to the W3C Process. N-TRIPLE, in turn, is a subset of the turtle format (see Example 14 for the same statement previously presented).

```
<http://www.dei.isep.ipp.pt/~iazevedo> <http://purl.org/dc/elements/1.1/date>
"2000.01.20" .
```

Example 14. A statement in N-TRIPLE syntax

In that serialisation the period at the end indicates the end of a triple and it is not necessary to fit a triple onto a single line.

RDF provides a model for capturing information but alone does not provide a way to capture the meaning of the information being modelled. It is possible to build RDF metadata descriptions by using multiple vocabularies, but not to define new vocabularies. RDF Schema (RDFS) [Brickey and Guha, 2004] has this purpose. Some of the properties often used to build vocabularies are presented in Table 15.

Table 15. Some properties employed in definition of vocabularies

Property name	Domain	Range	Comment
rdf:type	rdfs:Resource	rdfs:Class	An instance of a class.
rdfs:comment	rdfs:Resource	rdfs:Literal	A description of the subject resource.
rdfs:domain	rdf:Property	rdfs:Class	A domain of the subject property.
rdfs:label	rdfs:Resource	rdfs:Literal	A human-readable designation for the subject.
rdfs:isDefinedBy	rdfs:Resource	rdfs:Resource	The definition of the subject resource.
rdfs:range	rdf:Property	rdfs:Class	A range of the subject property.
rdfs:seeAlso	rdfs:Resource	rdfs:Resource	Recommended additional information
rdfs:subClassOf	rdfs:Class	rdfs:Class	A subclass of a class.
rdfs:subPropertyOf	rdf:Property	rdf:Property	A subproperty of a property.

With RDFS, classes and properties can be organised in generalisation/specialisation hierarchies, define domain and range possibilities for properties, state class membership, and indicate and interpret data types.

4.4.2 OWL

OWL, Web Ontology Language, is the language used to develop ontologies that describe information on the Semantic Web. OWL had some predecessors, such as OIL (Ontology Inference Layer) and DAML+OIL. OWL extends RDF and RDF Schema to provide a vocabulary of properties and classes that have associated semantics.

OWL has three sub-languages: OWL Lite, OWL DL and OWL Full. Each offers different levels of semantic expressiveness to be used according to the needs of representation and inference of ontologies to be defined. OWL Lite corresponds to the description logic $\mathcal{SHIF}$. It can be used for the purpose of defining taxonomies and expressing simple constraints. OWL DL uses 'DL' in the name to clearly state that it supports description logic capabilities. OWL DL corresponds to the description logic language $\mathcal{SHOIN}(\mathcal{D})$.

There is an equivalence between DL syntax and some OWL DL descriptions. Generally the term OWL is used to refer specifically to the sub-language OWL DL.

An OWL ontology follows the open world assumption, already described before, and the 'no unique names' assumption. It states that it cannot be assumed that resources identified by different URIs are different, unless explicitly stated otherwise.

Some of the OWL terms that can be used to describe a class are presented in Table 16. They are detailed in [Bechhofer, Harmelen et al., 2004], which also covers terms for description of annotation properties, individuals, properties, datatypes and property restrictions.

Table 16. Class description with OWL

Term	Description
owl:equivalentClass	The extensions of two specified classes are equivalent.
owl:Thing	The set of all individuals.
owl:Nothing	Its class extension is the empty set.
owl:oneOf	The members of the class is limited to the members of the specified collection of individuals.
owl:intersectionOf	The members of the class are members of all specified classes.
owl:unionOf	The members of the class are members of one or more of the specified classes.
owl:complementOf	The members of the class are not members of another specified class.
owl:disjointWith	Two specified classes does not have individuals in common.
owl:AllDisjointClasses	A class used to specify a set of classes that are pair-wise disjoint.
owl:disjointUnionOf	It specifies that the class is the union of the set of specified classes, which are pair-wise disjoint.

OWL 2 is the most recent version of OWL. OWL 2 is compatible with prior OWL 1 ontology language and provides the expressiveness of $\mathcal{SROIQ(D)}$. The new features are discussed in [Golbreich and Wallace, 2008].

The OWL profile to be used depends on how the ontology will be structured and the reasoning tasks to be carried out. There are three predefined profiles of OWL 2: OWL 2 EL, OWL 2 QL, and OWL 2 RL. $\mathcal{EL}$, a lightweight Description Logic, is the basis for the EL profile. It is possible to have reflexive objet properties with OWL 2 QL, but not with OWL 2 RL. Also the former allows some datatypes (`owl:real`, `owl:rational`, for example) that are not available in the OWL 2 QL profile. A detailed description of these profiles is provided in [Motik, Grau et al., 2009].

4.4.3 SPARQL

There have been several approaches on RDF query languages, such as RDQL[25], N3QL[26], Squish[27], and Versa[28]. However the ones with an SQL-style syntax were more popular and widely implemented. SPARQL follows this trend in style, providing a way to query RDF triples which is very familiar to those who have used SQL.

SPARQL is the W3C recommended query language for RDF since 15 January 2008. SPARQL is both a protocol and an RDF query language.

A query has a structure with two parts. In the first part an optional prefixing mechanism with the `PREFIX` keyword can be used. It associates a short label with a specific URI. Then, the label can be used in the query command, given it more readability. `BASE` is another way of abbreviating URI, defining the base URI to all following relative URIs in the query, including those specified with `PREFIX`.

In the second part a query itself is specified through a statement. SPARQL allows the following query forms:

- `SELECT` - Returns a set of pairs of a variable name with a value, a solution sequence, which usually can be seen as a table, similar to the SELECT statement used in SQL, also having an optional FROM clause;
- `CONSTRUCT` – Results in an RDF graph of triples specified by the construct clause and the data matching the conditions specified;
- `DESCRIBE` – Returns an RDF graph containing descriptions (all known properties) of matching resources;

[25] http://www.w3.org/Submission/RDQL/
[26] http://www.w3.org/DesignIssues/N3QL.html
[27] http://ilrt.org/discovery/2001/02/squish/
[28] http://copia.ogbuji.net/files/Versa.html

- `ASK` – Used when it just matters if there are or not some triples matching the expressed conditions.

There are query modifiers, which are keywords that affect the result set. Some of them are `DISTINCT`, `REDUCED`, `ORDER BY`, `OFFSET`, `LIMIT`, `FILTER`, `OPTIONAL` and `UNION`.

The current specification does not allow adding or updating RDF data.

4.4.4 SWRL

In a common syntax a rule can be expressed as follows:

```
antecedent ⇒ consequent
```

Such as in:

```
hasMother(?x1,?x2)∧ hasMother(?x2,?x3) ⇒
hasMaternalGrandmother(?x1,?x3)
```

A Semantic Web Rule Language (SWRL) proposal was submitted in 2004 by the National Research Council of Canada, Network Inference and Stanford University, in association with the Joint US/EU ad hoc Agent Markup Language Committee [Horrocks, Patel-Schneider et al., 2004]. An XML syntax was also provided by the proposers to the specification of rules with the antecedent and consequent clauses.

In 2005 there were two more proposals submitted to W3C: Web Rules Language (WRL) [Angele, Boley et al., 2005] and Semantic Web Services Language (SWSL) Rules [Battle, Bernstein et al., 2005].

None of the three mentioned languages are a standard rule language for the Semantic Web, despite the fact that from these languages, SWRL is the one with a more widely tool support.

The Rule Interchange Format (RIF) Working Group[29], constituted in 2005, has been concerned with higher level goals detailed in [Paschke, Hirtle et al., 2008], but the main objective of RIF is to "be an effective means of exchanging rules that has the potential to be widely adopted in industry and that is consistent with existing W3C technologies and specifications".

4.5 Knowledge Engineering

Knowledge engineering typically involves the following five steps [Kendal and Creen, 2006] (page 8):

- Knowledge acquisition,
- Knowledge validation,
- Knowledge representation,
- Inferencing,
- Explanation and justification.

Knowledge processes frequently entail knowledge acquisition through knowledge creation or reuse, and the methods appropriate for one, may not be convenient for the other [Davenport, Jarvenpaa et al., 1996].

As ontologies started to be widely used in Knowledge Engineering, the field of Ontological Engineering grew up. Ontological engineering is a subarea of knowledge engineering. According to Gasevic, "Ontological engineering denotes a set of design principles, development processes and activities, supporting technologies, and systematic methodologies that facilitate ontology

[29] http://www.w3.org/2005/rules/wiki/RIF_Working_Group

development and use throughout its life cycle – design, implementation, evaluation, validation, maintenance, deployment, mapping, integration, sharing, and reuse" [Gasevic, 2006](page 58).

Despite the difference in outcomes (software and ontologies), the Software Engineering and Ontology Engineering have similarities, even in the problems both fields face, such as [Bontas and Mochol., 2005]:

- Lack of agreement in the quality assurance;
- Usually the requirements are not well defined and their change may have implications in the whole process;
- The design/conceptualisation phase is more cost-intensive than the implementation phase.

4.5.1 Methodologies for ontology development

De Hoog, regarding Knowledge Based Systems (KBS) construction, differentiate between methodology and method: the former prescribes 'what', 'who' and 'when', mentioning that "methodologies refer to knowledge about methods" [de Hoog, 1998]. A methodology applies methods or techniques, as explained in Chapter One.

The first guidelines and methodologies for building ontologies were born of the experience of several authors in the development of ontologies, such as the Cyc ontology [Lenat and Guha, 1990], the Enterprise Ontology [Uschold and King, 1995], and the perceived need to follow certain steps. Also some projects advanced with strategies to be followed, as happened during the projects KACTUS [KACTUS, 1996] and TOVE[30] (TOronto Virtual Enterprise) [Grüninger and Fox, 1995].

The KACTUS approach was slightly different from the others as when an application is built, the ontology that embodies the necessary knowledge is refined, which might require searching ontologies already developed (see Figure 19). The reuse of ontologies is a step considered in the more modern methods or methodologies, even if not detailing how. In addition, using the same skeleton, a natural language-based ontology to be used for machine translation (the SENSUS ontology [Swartout, Ramesh et al., 1997]), other ontologies could be built selecting the relevant terms – the basic idea of the SENSUS-based method.

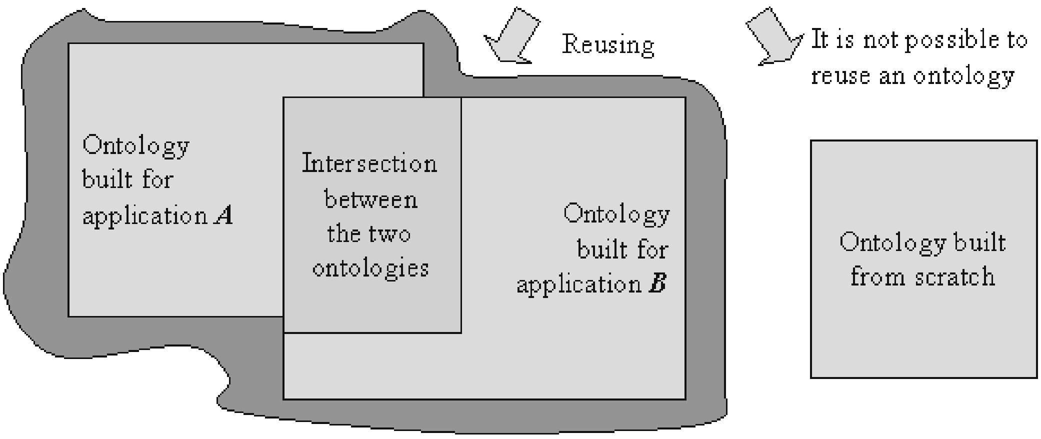


Figure 19. Overview of ontology reuse in KACTUS method (From [Gómez-Pérez, Fernández-López et al., 2004] (page 124))

Many methodologies have been suggested to be used in the development of ontologies, totally from scratch or incorporating other ontologies. Some of them adapted and extended a methodology for knowledge-based systems development (CommonKADS [Schreiber, Akkermans et al., 1999]), such as the On-To-Knowledge Methodology (OTKM) (see section 4.5.1.2).

[30] http://www.eil.utoronto.ca/tove/toveont.html

The most common tasks in all these methodologies were identified in [Siorpaes and Simperl, 2010]:

- Define domain and scope,
- Elaborate of competency questions,
- Reuse ontologies,
- Specify relevant terms,
- Put concepts into a hierarchy,
- Define properties and axioms,
- Define instances,
- Document the whole process.

The degree to which these tasks are considered and detailed in different methodologies varies. However, all of them need human input, although some of them can be partially supported by computers.

None of the methodologies for building ontologies are considered to be satisfactorily mature, but along the time they have become more complex, incorporating more details and activities in the process. In [Pinto and Martins, 2004] the authors identified three generations of methodologies for the development of ontologies, with the last generation corresponding to another level of maturity and completeness. Methontology (its current version) and On-To-Knowledge (OTK), briefly covered in the next sub-sections, are considered to belong to this last generation. They are also viewed as, respectively, the methodology which provides “the most accurate descriptions of each activity” and the methodology which “describes more activities than the other approaches” [Gómez-Pérez, Fernández-López et al., 2004] (page 153).

4.5.1.1 METHONTOLOGY

Proposed by researchers from Universidad Politécnica de Madrid, Methontology [Fernández, Gómez-Pérez et al., 1997; Fernández-López, Gómez-Pérez et al., 1999] allows the construction of ontologies, reusing or not other ontologies. Using this methodology, the **development** is made through the following steps: specification, conceptualisation, formalisation, implementation and maintenance (see Figure 20).

METHONTOLOGY also proposes **management** (schedule, control, and quality assurance), and **support** activities (knowledge acquisition, integration, evaluation, documentation, and configuration management).

The methodology specifies the steps to be executed on each activity, the techniques to be used, the results and their evaluation. It is partially supported by the ontology development tool WebODE [Arpírez, Corcho et al., 2001], but it is possible to follow this methodology using other ontology tools as well.

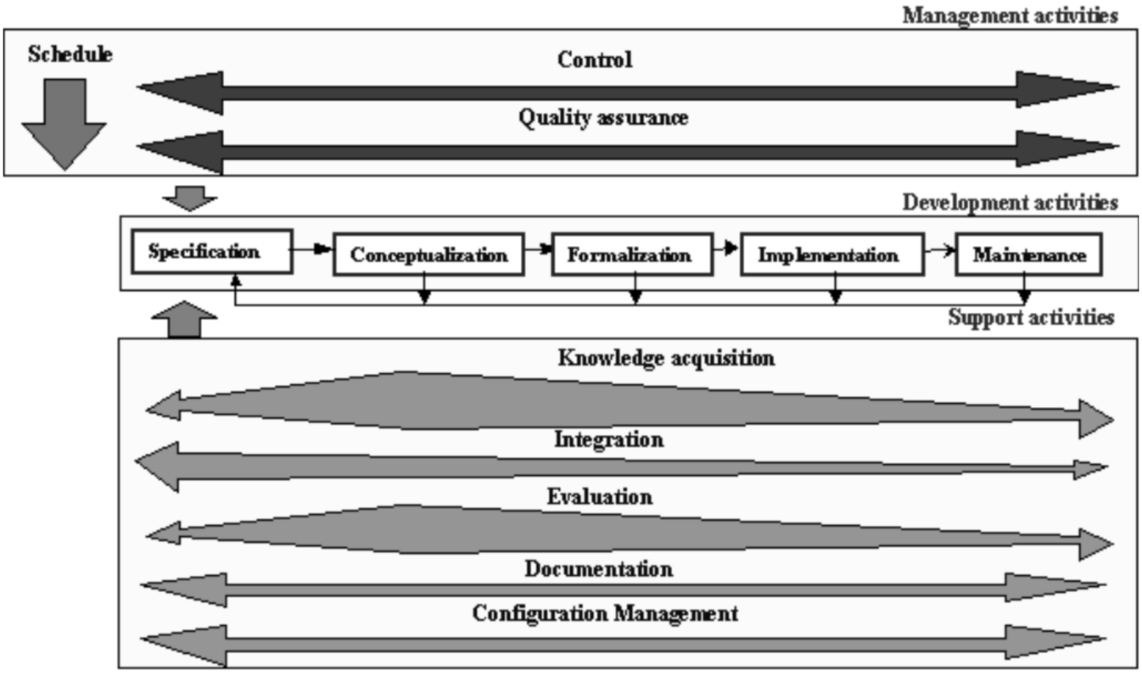


Figure 20. Overview of METHONTOLOGY (From [Corcho, Fernández et al., 2005])

4.5.1.2 On-To-Knowledge Methodology

This methodology was developed under the On-To-Knowledge project, which has some well-reputed partners, such as Institute AIFB of the University of Karlsruhe, the Vrije Universiteit of Amsterdam, and British Telecom. The On-To-Knowledge (OTK) [Staab, Schnurr et al., 2001; Sure, 2003; Sure, Staab et al., 2003] encompasses 5 processes, which are depicted in Figure 21. These high-level processes are the following ones:

- **Feasibility study** – This process precedes the development of the ontology and aims to assess its feasibility. It also identifies the most relevant focus area and target solution;
- **Kickoff** – In this phase the ontology development starts with the description of an ontology requirements specification document (ORSD). The ORSD identifies knowledge sources for the gathering of the semi-formal ontology description, done in this process. Ontology engineers and domain experts jointly decide when it is time to advance to the next phase;
- **Refinement** – The previously defined ontology description is refined. Online ontologies can be reused and also automatic document analysis can be applied in this phase. When an ontology that meets the requirements is obtained, the next process starts;
- **Evaluation** – This process comprises three different types of evaluation: (1) technology-focussed evaluation, (2) user-focussed evaluation and (3) ontology-focused evaluation. The first aspect is related to the evaluation of ontology properties and technology properties. For the user-focussed evaluation, it is necessary to verify if users are satisfied with the ontology based application. The third aspect is concerned with formal evaluation of ontologies and the OntoClean approach [Guarino and Welty, 2002] can be used for this purpose;
- **Maintenance** – This process is concerned with ontology evolution and guarantee that changes to ontologies will be successfully carried out.

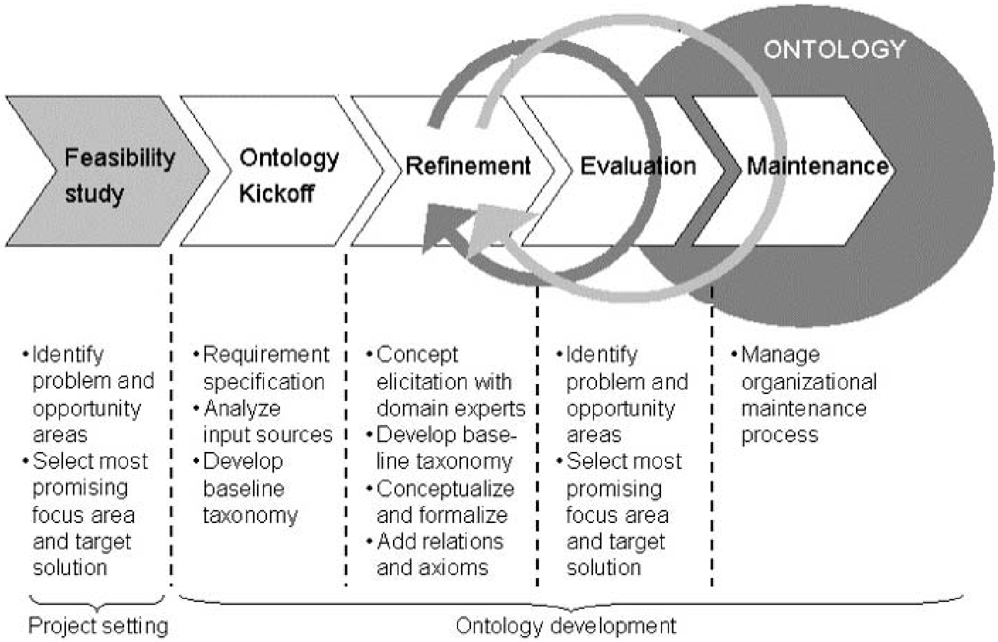


Figure 21. On-To-Knowledge overview
(From [Staab, Schnurr et al., 2001])

OntoEdit [Sure, Erdmann et al., 2002] with its plugins supports this methodology.

4.6 Conclusions

From the classification systems studied, although there are some similarities between thesaurus and ontologies, the latter can represent more relation types. Also, benefiting from the Semantic Web trend, there are a standardisation effort that supports the creation and maintenance of ontologies, with different levels of formalisation, and a great number of tools that provide standard compliance, supporting W3C languages. In addition, ontologies usually have another degree of detail in the represented knowledge.

Folksonomies do not suffer from the problem of the costs involved in the creation and maintenance of taxonomies, thesaurus and ontologies. As described, this type of classification system builds on tags provided by users, and thus the costs are mainly those necessary for the tagging system itself. The different types of tags allow capturing other aspects besides those allowed in other classification systems, such as engagement in activities and emotional responses triggered by the tagged resources.

Chapter 5

INFORMATION RETRIEVAL

How resources are characterised is intimately related to how they can be later found and retrieved. That is an important point since if the resources are poorly described in a system their findabilty will be compromised.

This chapter summarises the state of the art in Information Retrieval. It describes the classical information retrieval models, some cautions for the evaluation in the area and the developments in query operations. The last section summarises some of key points addressed in this chapter and presents some conclusions.

5.1 Introduction

Data retrieval is concerned with deciding which documents of a collection contain the keywords in the query. Information retrieval is a designation that Calvin Mooers coined in 1950 at a meeting of the Association for Computing Machinery (ACM) [Mooers, 1950]. An information retrieval system "must somehow 'interpret' the contents of the information items (documents) in a collection and rank them according to a degree of relevance to the user query" [Baeza-Yates and Ribeiro-Neto, 1999] (page 2). According to van Rijsbergen, the core objective of information retrieval (IR) is "... to retrieve all the relevant documents [and] at the same time retrieving as few of the non-relevant as possible" [van Rijsbergen, 1979].

Relevance, a central issue of IR systems, can be seen as one aspect of quality [Klein, 2002]. In fact, determining what information is relevant is a concern in many different areas, including Artificial Intelligence [Subramanian, Greiner et al., 1997], Philosophy [Schutz, 1970] and Communication [Sperber and Wilson, 1986].

In information retrieval, in particular, relevance is often applied in the sense of topical relevance – a result is more relevant if it is more related to the topics addressed in the information need. That 'topical relevance' is often criticised by not considering the user's individual characteristics (including his/her background), a specific task situation and the volatility of information needs [Park, 1994; Kagolovsky and Möhr, 2001].

Manning et al. observe that the judgement of relevance considers the information need and not the particular query used to translate it [Manning, Raghavan et al., 2008] (page 152). However, in practice, the information need is often not known and, instead, the formulated query is considered.

Many years ago the research in the area of information retrieval was often criticised mainly for two reasons:

- Lack of a solid basis;
- Lack of consistent and robust benchmarks and tests.

It is hard to argue against the first of these reasons, because of the subjectivity associated with the real task of deciding whether a document is relevant. The same problem of deficiency in theoretical basis was faced by the area of information science itself (see [Brookes, 1980a] [Brookes, 1980b] [Brookes, 1980c] [Brookes, 1981]).

However, regarding the second motive for criticism, much has been done. Some decades ago IR experimentation used relatively small collections of documents, which did not reflect many different characteristics. Also, often tests aimed at very different aspects, which did not allowed comparing different retrieval systems.

In the early '90s, the situation started to change mainly due to the work of Donna Harman from the National Institute of Standards and Technology (NIST) in Maryland. She promoted an annual conference called TREC (Text RErieval Conference), which has been dedicated to testing different IR approaches with large collections comprising millions of documents.

The field of information retrieval evolved a lot in the last years, and new interests have emerged. For instance, traditional users of information retrieval systems were not concerned about the time response but, as pointed in [Kobayashi and Takeda, 2000], Web users may have a tendency to favour performance issues, such as interactive response times and the number of valuable resources listed at first.

Currently, there are many journals and conferences on IR, such as the Journal of Information Retrieval, the ACM SIGIR International Conference on Information Retrieval, the European Conference on IR (ECIR), the String Processing and Information Retrieval Symposium (SPIRE), the Text REtrieval Conference (TREC). Hence this research area is well-established.

Related to information retrieval another research area has been rising is recent years: information interaction. Nahum Gershon is credited with coining the term 'Information Interaction' in 1995 [Morville, 2005], with the following meaning: "how human beings interact with, relate to, and process information regardless of the medium connecting the two". Later, Gary Marchionini highlighted that "the IR problem itself has fundamentally changed and a new paradigm of information interaction has emerged" [Marchionini, 2004].

Bates proposed an innovative model of searching in online and other information systems, called 'berrypicking' (see Figure 22), which she defends as being more faithful to the common information seeking behaviour. Thus, it is not expected that a single query will return the documents with interest, but a user successively modifies the queries in accordance with the results.

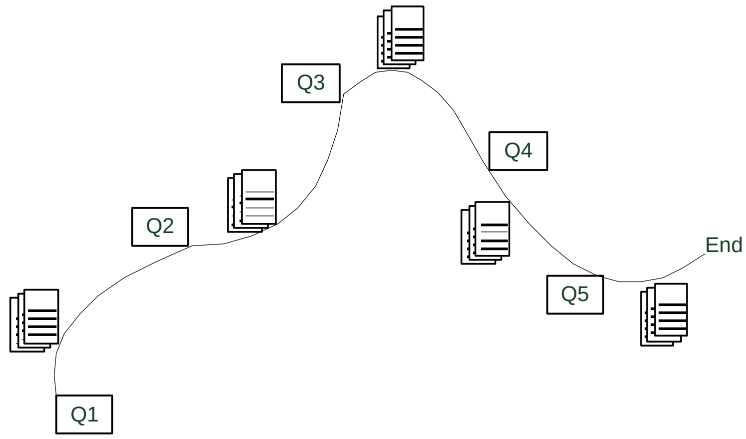


Figure 22. A Berrypicking, Evolving Search (Adapted from [Bates, 1989])

To some extent, the consideration of users'opinions is also incorporated in many IR techniques, such as query expansion methods, and there are IR models that take advantage from user relevance feedback.

In addition, an important concept is findability, a term related to IR often used in the context of the Web. Morville provides the following definitions of findability [Morville, 2005] (page 4):

- "The quality of being locatable or navigable";
- "The degree to which a particular object is easy to discover or locate";
- "The degree to which a system or environment supports navigation and retrieval".

It is possible to talk about findability at object and system levels, which are however closely related. For instance, it could be easier to find a white object in a black box than a black object in the same box. Size, colour and shape might have a play in physical environments.

In digital systems words function as labels to objects and it is common the use of keywords in queries submitted to search engines, such as Google and Yahoo!, online shop stores, like Amazon, or news sites, such as CNN, among others.

The rest of this chapter is organised as follows. Section 5.2 explains the concept of 'retrieval model', the classical models and others that have been proposed. Section 5.3 discusses how IR evaluation has been carried out, the main initiatives, as well as important collections characteristics. Query operations common in IR are explained in section 5.4. A novelty field, IR with folksonomies, is addressed in section 5.5. Finally, the last section (5.6) compiles the main findings of this chapter with potential implications in a Learning Objects Repository (LOR).

5.2 Information retrieval models

Most of the information retrieval models are based on the assumption that a resource is more relevant if there is a high degree of compliance between the terms in the query and the terms in the resource.

The various retrieval models that have been proposed vary mainly in how they determine if a resource is valuable to the user needs. An information retrieval model encompasses the following information [Baeza-Yates and Ribeiro-Neto, 1999] (page 23):

- A set of logical views for the documents in the collection;
- A set of logical views for the user information needs;
- A framework for modelling document representations, information needs representations, and how they relate to each other;
- A ranking function that provides a value for a document associated to a certain query.

Bates emphasised the formulation of a single query based on the information need in classic information retrieval models (see Figure 23) [Bates, 1989].

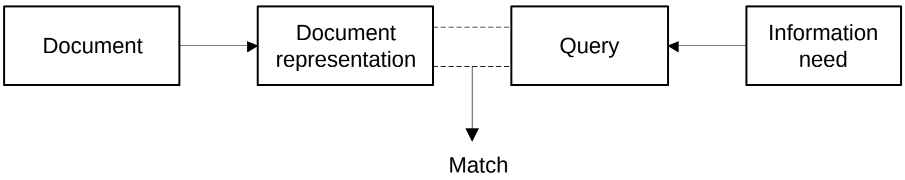


Figure 23. The Classical Information Retrieval Model (Adapted from [Bates, 1989])

The three classic retrieval models are the Boolean model, the Vector model and the Probabilistic model [Baeza-Yates and Ribeiro-Neto, 1999], which are discussed in sections 5.2.1, 5.2.2 and 5.2.3, respectively, also briefly mentioning variants of each of them.

5.2.1 Boolean model

The queries are made through Boolean expressions (using AND, OR and NOT). It is more a data retrieval model, as the documents are considered as relevant or not based on their index terms, without any kind of ranking.

A closed world assumption is held as if a certain index term is missing in a given document, it is supposed not be relevant to queries with that term. For example, if a document D_i has the following index terms $\{T_1, T_2, T_5\}$ and there is a query $Q = T_1 \text{ AND } T_3$, document D_i is not retrieved. However, many users prefer Boolean query models exactly because of this characteristic, as they have a great control on what is retrieved [Manning, Raghavan et al., 2008].

This model, without any deviation, is not adequate for situations where it is possible to have large results sets, as there is no notion of ranking: the results are simply adequate or not.

Despite the disadvantages of the Boolean retrieval model, systems implementing it were "the main or only search option provided by large commercial information providers for three decades until the early 1990s" [Manning, Raghavan et al., 2008]. However, in fact many of these implementations extended the Boolean model, considering other operators, such as a proximity operator that specifies how close two terms must be in a document to be retrieved.

Some other derived models were proposed, such as the **Mixed Minimum Maximum model** [Fox and Sharat, 1986] and the **Paice model** [Paice, 1984], both based on the fuzzy-set theory, and the **P-Norm**, whose proponents describe as a compromise between the Boolean and the Vector models through the consideration of weighted query and document terms [Salton, Fox et al., 1983]. However, none of them received a wide acceptance, especially in commercial systems.

5.2.2 Vector model

In this model both the terms of the query and the index terms associated to a document are weighted. We have two vectors, the query vector, represented as

$$\vec{q} = (w_{1,q}, w_{2,q}, \ldots, w_{t,q}),$$

being t the number of index terms in the whole system, and the document vector, represented as

$$\vec{d_j} = (w_{1,j}, w_{2,j}, \ldots, w_{t,j}).$$

The underlying assumption is that not all terms have the same importance in representing the document content (w_{ij}) or the information need expressed in a query (w_{iq}). Many approaches for the calculation of these weights have been proposed [Salton and Lesk, 1983], including the use of `tf` (term frequency) factor, the raw frequency of the a certain term inside a document, and the `idf` (inverse document frequency) factor, the inverse of the frequency of a certain term among the document in a collection `D`.

The cosine of the angle between the query vector and the document vector can be used to estimate the degree of similarity of the document d_j with regard to the query `q` (see Figure 24).

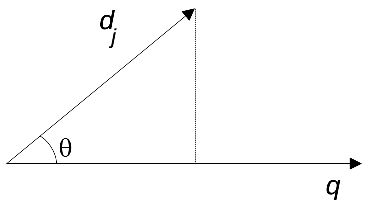


Figure 24. The similarity between a document d_j and a query q (From [Baeza-Yates and Ribeiro-Neto, 1999])

Thus, a partial matching is possible. Furthermore, these models enable the retrieval of documents considering the level of similarity and a ranking of the documents is also feasible. Other retrieval operations supported by this model are document classification and document clustering, whose ideas naturally emerge from the model, as can be seen in Figure 25 with three terms. For instance, document d_2 is closer to d_1 than d_3 is. The cosine of the angle between them (cosine similarity) can be used for that purpose.

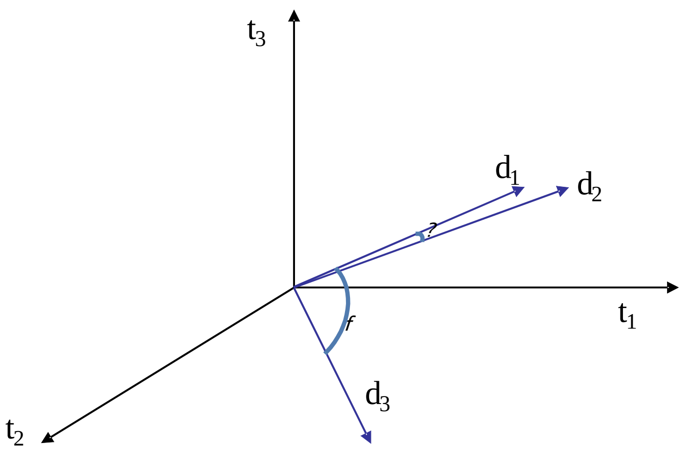


Figure 25. Documents represented in a vector model

The Vector model does not consider index terms dependencies[31], which might be seen as a disadvantage, but that consideration would affect the system performance if applied to all documents in a collection [Baeza-Yates and Ribeiro-Neto, 1999].

In the **generalised vector space model** the index term vectors are not pairwise orthogonal, although it is still assumed the linear independence between them.

Another variant of the vector model is the **Latent Semantic Index (LSI) model**, which introduces a lower dimensional space to make possible concept matching between the query terms and the documents terms. This model is described in [Deerwester, Dumais et al., 1990]. It has high computational costs [Manning, Raghavan et al., 2008].

Another alternative vector model is the **neural network model**. In this model a spreading activation method is used, simulating the way neurons in the brain activate others. In the graphical representation of a neural network there are nodes (the processing units) and edges (the synaptic connections). Each node has a certain activation level that varies over time. Moreover, each node influences the activation levels of its neighbours, resulting in a dynamic pattern of activations over all nodes. Actually not all nodes are activated at the same time.

The nodes are divided into (at least) three distinct pools [Wilkinson and Hingston, 1991]:

- **First pool** - There is one node related to each query term. The nodes in this layer send signals to the next pool (document terms nodes),
- **Second pool** - There is one node for each term in the documents. The nodes of this pool generate signals to the documents nodes (the third pool),
- **Third pool** - Each document in the system has a node associated in the network.

The connections between the nodes of the first and second pools have only one direction, but the connections between the nodes of the second and the third pools are bidirectional. After the first phase of activation, the documents nodes can send signals to the intermediary pool, which in turn can send signals to the documents nodes again. However, at each activation level the signals become weaker (the connections have weights) until the process halts.

At the end it is possible to have documents activated when they do not contain any of the query terms, a clear advantage of this model. However, Baeza-Yates and Ribeiro-Neto highlight that "there is no conclusive evidence that a neural network provides superior retrieval performance with general collections" [Baeza-Yates and Ribeiro-Neto, 1999] (page 48).

More information can be found in [Wilkinson and Hingston, 1991]. Also a detailed description of neural network models is provided in [McClelland and Rumelhart, 1986; Rumelhart and McClelland, 1986]

5.2.3 Probabilistic model

The vector model applies a similarity function between the query and each document, and not probability as it happens extensively in the Probabilistic model. The first attempts to establish this model dates from the end of the decade 1950 and beginning of the decade 1970 [Maron, Kuhns et al., 1958; Maron, Kuhns et al., 1959; Maron and Kuhns, 1960]. It was this model which established relevance ranking in IR.

If relevance is a central concept for information retrieval, the probability of relevance $P(R \mid q_i ; d_j)$ is a main aspect for this model. It is based on the idea that there is an ideal set of all relevant documents for each query. Each document is estimated as relevant or not, but this model does not clearly establish how to calculate the probabilities of relevance [Baeza-Yates and Ribeiro-Neto, 1999] (page 31), a procedure which also needs to be computational efficient.

The initial set of relevant document is rated by the user (the only one to rate in the original idea [Baeza-Yates and Ribeiro-Neto, 1999] (page 33)) or the system, and this information is used to

[31] All three classic IR models have the term independence assumption.

provide another set of documents using the similarity between documents in a repetitive process until the user is satisfied, having at least two steps.

About the vector and the probabilistic models, Manning et al. state that “the difference between “vector space” and “probabilistic” IR systems is not that great... For a probabilistic IR system, it’s just that, at the end, you score queries not by cosine similarity and tf-idf in a vector space, but by a slightly different formula motivated by probability theory” [Manning, Raghavan et al., 2008]. Nevertheless, according to the same authors, probabilistic methods always have left to be desired in performance.

In more recent years the BM25 term weighting scheme, also known as the Okapi weighting function, has being receiving great attention, derived from the good results achieved also in TREC evaluations. Further explanation about it is provided in [Jones, Karen et al., 2000]. Variations of the BM25 term weighting functions have also been proposed, such as the BM25F, which was adapted from the BM25 to score structured documents. It is described as “one of the most successful Web-search and corporate-search algorithms” in [Robertson and Zaragoza, 2009].

Two alternative probabilistic models based on Bayesian networks were proposed: the **inference network** [Turtle and Croft, 1991] and the **belief network** [Ribeiro and Muntz, 1996], which generalises the former one. In a Bayesian network there are nodes representing random variables and directional arcs between the nodes stating a causal relationship between parent and child nodes[32].

Figure 26 provides an example of a Bayesian network with 4 variables.

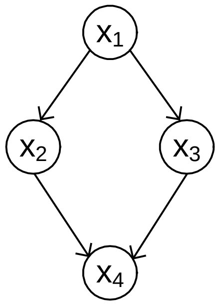


Figure 26. Bayesian network – an example

The joint probability distribution `P(x1,x2,x3,x4)` considers the dependencies stated in the Bayesian network. Thus, the local conditional probabilities can be express in the following way:

`P(x1,x2,x3,x4) = P(x1)P(x2|x1)P(x3|x1)P(x4|x2,x3)`

In the field of Artificial Intelligence the use of Bayesian networks has been a leading approach for managing probability [Campos, Fernández-Luna et al., 2004].

A main advantage of Bayesian networks in IR is that “they provide a clean formalism for combining distinct sources of evidence (past queries, past feedback cycles, and distinct query formulations) in support of the rank for a given document” [Baeza-Yates and Ribeiro-Neto, 1999] (page 48).

5.2.3.1 The Probability Ranking Principle

The order in which the results are presented to the users is the basis of the probabilistic model and is known as the Probability Ranking Principle (PRP): “If a reference retrieval system’s response to each request is a ranking of the documents in the collection in order of decreasing probability of relevance to the user who submitted the request, where the probabilities are estimated as accurately as possible on the basis of whatever data have been made available to the system for this purpose,

[32] Both neural networks and bayesian networks use directed graphs, but in the former we do not have a causal relationship between nodes.

the overall effectiveness of the system to its user will be the best that is obtainable on the basis of those data" [van Rijsbergen, 1979].

The probability of a document d_j being relevant regarding a query q depends on the document and the query, and nothing more. However, Robertson argues that the relevance or utility should be defined outside the system itself, as a criterion for the system [Robertson, 1997]. The author raises the question if the relevance or usefulness of a document affects the relevance or usefulness of another, for which he presents several arguments in favour of a positive response, including:

- If document A appears before B, B's utility is affected, for example if B provides essentially the same information as A;
- The fact that the document has already been deemed relevant by the user, may provide some indication of the possible relevance of B;
- Even if the system does not know if the document A was judged relevant, it might know that there is a correlation between the acceptance of A and B for different users;
- Two separate documents may not be relevant if analysed in isolation, but together they can be, as one may complement the other.

5.3 Evaluation in IR

It was probably the Ellen Voorhees's 2001 paper entitled "The philosophy of information retrieval evaluation", which firstly used the phrase "the Cranfield paradigm" to refer to a way to conduct IR evaluations introduced much earlier [Cleverdon, 1991]. The Cranfield collection was the first one used to provide quantitative measures of IR effectiveness.

The comparison of different approaches using test collections is a well-established practice in the area of information retrieval, with the assumption that performance observed will be maintained in operational conditions. An evaluation corpus needs documents, queries and relevance verdicts.

Besides relevance, other aspects can also be subject of evaluation in IR area in a broader perspective, including system quality and user utility. Manning et al. refer to some of the system issues that can be measured [Manning, Raghavan et al., 2008] (pages 168-169):

- How fast does a system index documents?
- How fast does a system search?
- How large is the system documents collection?
- How expressive is the system query language?

The latter is the less simple to be objectively measured. Under user utility the same authors summarise the following points (page 169):

- An indirect measure of user happiness can be the number of users that use again the same system;
- In e-commerce the portion of searchers who become buyers can be considered;
- User studies can be conducted, with participants' observation and using ethnographic interview techniques to infer user satisfaction.

Being time consuming and expensive to carry out, the latter issue is also the least straightforward.

For proper evaluation of different IR approaches, Jones and van Rijsbergen suggested the used of different styles of writing, different document types (e.g. general documents, scientific papers), different types of users with precise and non-precise requests, among others [Sparck Jones and van Rijsbergen, 1976]. They highlighted the need to consider these aspects when analysing the results.

Some other aspects to consider in an experimental IR scenario are as follows [Buckley and Voorhees, 2000]:

- **The number of descriptions of information needs** – Under TREC usually 50 requests are considered [Voorhees, 2000]. However, Sparck Jones and van Rijsbergen stated 75 as the minimum acceptable [Jones and van Rijsbergen, 1976];
- **The IR metrics** – In section 5.3.2 a number of metrics are described. The use of acceptable and appropriate evaluation measures is important. For instance, the rank at which the first relevant document is retrieved is considered a poor measure of retrieval performance [Voorhees and Harman, 1999];
- **The degree of difference** – Little differences in cross-system comparisons are negligible, particularly if there is human interaction in the process. A difference of less than 5% can be considered insignificant, and a difference between 5 and 10% is noticeable, while a difference greater than 10% is significant [Jones, 1974].

Many measures have been proposed for IR evaluation, but also many initiatives have appeared in order to provide reliable results in IR evaluation. These questions are summarised under the next sub-sections, in the traditional perspective of relevance.

5.3.1 IR evaluation initiatives

In the early '90s, the situation regarding information retrieval strategies evaluation started to change mainly due to the work of Donna Harman from the National Institute of Standards and Technology (NIST) in Maryland. Her effort was to promote an annual conference called TREC (Text REtrieval Conference), which was dedicated to tests with large document collections with millions of documents. For each TREC conference some benchmark tests are assigned, which are used by the research groups participating in the conference to compare several IR systems and methods but under the same parameters.

The organisation of the TREC conferences has been a joint responsibility of NIST (National Institute of Standards and Technology) and DARPA (Defense Advanced Research Projects Agency) as part of the TIPSTER program[33]. The first conference was held in 1992 (TREC-1) and the last was in 2011 (TREC-20). For near all conferences, the following working schema has been applied to [Voorhees, 2000; Manning, Raghavan et al., 2008]:

- There are different tracks which focus on a particular IR subproblem[34][35]. For instance, there were eight tracks in TREC-20. These tracks are not the same across all conferences. Participants choose to take part in the tracks of their interest;
- Various topics (these are questions that must be answered in TREC terminology) are created in a way that is not uniform: some are very detailed, while others are simple issues. The same collection of documents is provided to the participants who return a ranking results list with the best 1,000 documents for each topic. The same metrics are used for all systems under evaluation;
- From the participants' responses, the top 100 documents are obtained for each topic by a method described as pooling;
- Some K documents obtained in the previous phase for each of the topics are discussed by assessors (humans) who decide about the relevance of each document to the topic under concerning.

The results and an analysis of them are published for each TREC conference.

[33] It ran from 1991 to 1998 and was funded by some governmental agencies. It focused on three areas: information retrieval and extraction, and summarization. A standard architecture was proposed, which was implemented in GATE (http://gate.ac.uk/), among other systems.
[34] The consideration of different tracks started in 1994.
[35] In the beginning there was also an hoc track that was halted in 2000.

The TREC conferences have undoubtedly become a reference in the field of information retrieval and more than showing that there is nothing else to be done, they increasingly put in evidence how this area might grow.

Some other initiatives with similarities to TREC, while addressing more specific IR problems, are:

- The Cross-Language Evaluation Forum (CLEF) (http://www.clef-campaign.org/);
- The INitiative for Evaluation of XML retrieval (INEX) (http://www.inex.otago.ac.nz/).

5.3.2 Retrieval metrics

Many retrieval effectiveness metrics have been proposed over the years. Other kind of metrics were also proposed, such as satisfaction and frustration metrics [Korfhage, 1997], but they are not very well established in the area.

The classical metrics are discussed in section 5.3.2.1. They are the basis of other metrics that later appeared and that have been widely applied (see section 5.3.2.2), some of them focused on the top ranked results.

5.3.2.1 Classical metrics

Precision and **recall** are two traditional indicators of effectiveness of information retrieval systems. According to the schema suggested in [Swets, 1969] (see Table 17), we can consider the following factors:

```
a=relevant resources obtained,
b=non-relevant resources returned,
c=relevant resources that were not returned,
d=non relevant resources that were not returned.
```

The precision of a retrieval system is calculated as indicated:

```
precision = a /(a+b).
```

And the recall of a retrieval system is obtained as follows:

```
recall = a /(a+c).
```

Table 17. Partionning of records according to segments retrieved or not and relevant or not

	Relevant	Not Relevant
Retrieved	a	b
Not Retrieved	c	d

The ideal value for both recall and precision is 1, which only ideal information retrieval systems attain. The precision measures the ratio of relevant to total items retrieved by a query.

Precision and recall are two of a set of three measurements, whose formal definitions were introduced in 1955 [Kent, Berry et al., 1955]. Another one is efficiency or effectiveness, which is rarely used.

Van Rijsbergen introduced an efficiency measure, **F-measure**, which combines recall and precision in a single efficiency measure [van Rijsbergen, 1979]:

```
F = 2 * (recall * precision) / (recall + precision)
```

5.3.2.2 Other metrics

Most of the metrics in this section are applied to a subset of the documents that appear at the top of the results set, recognising the importance of relevant documents at higher ranks.

Precision at Position n (P@n) provides a value for the relevance of the top n documents in a ranked results set for a specified query. It can be determined as follows:

P@n = (number of relevant documents in top n results)/n.

If we define the binary function rel(n) as the relevance at position n, where rel(n) is 1 if the specific document at position n is relevant and zero otherwise ($1 \leq n \leq$ Number of documents retrieved), then the previous expression can be rewritten as bellow:

$$P@n = \frac{\sum_{r=1}^{n} rel(r)}{n}$$

P@n is influenced by the total number of relevant documents for a query.

The **Average Precision** (AP) considers the average of the precision at all N positions as follows:

$$\text{Average precision} = \frac{\sum_{r=1}^{N} (P@r * rel(r))}{N}$$

N can be the number of all documents retrieved or, usually, a subset of them. AP is "the most commonly quoted effectiveness measure in TREC" [Craswell, Hawking et al., 2001].

One metric very used by the TREC community is the **Mean Average Precision** (MAP), which is calculated over a number of queries submitted to a system and averaging the AP values obtained after each relevant document is retrieved. Thus, differently from the previous metrics, it is not applied to a single query.

The **Mean Reciprocal Rank** (MRR) is exactly equivalent to MAP for queries with only one correct result [Craswell and Hawking, 2004]. Both of them are standard TREC measures. It indicates at which rank the first relevant answer occurs on average. The score for an individual query q_i is the reciprocal ($1/r_i$) of the rank r_i at which the first correct result was returned or 0 if no correct answer was returned. The MRR is calculated by averaging that value across all the queries.

Success@n (s@n) is a measure that indicates the percentage of queries for which one or more good answers were found in the first n results. It can be used as a failure measure as 1 – s@n provides the proportion of queries without any good answer in the top n results [Craswell and Hawking, 2004]. It is often applied in situations where the user wants to find a resource in particular and the correct answer should be close to the top rank [Craswell, Hawking et al., 2001].

Normalised Discounted Cumulative Gain (NDCG) is the designation of a set of metrics that uses a discounting function to gradually decrease the documents score as their rank augments. The cumulated gain-based evaluation of IR techniques is based on two premises [Jarvelin and Kekalainen, 2002]:

- Highly relevant documents are more valuable than those only slightly relevant,
- If a document appears in a low rank position, it is less valuable for the user, as the probability of him/her taking notice of it gets reduced.

Akin to P@n, it is applied to a number of the top document in the results set. It is possible to have different relevance grades, and not only two (relevant or not as in binary relevance). To each grade is associated a certain gain. On the other side, a discount is attributed to every rank position. Different gain assignments and discount functions can be used.

Supposing there is a set of known relevant documents R for a given query, **R-Precision** provides the precision of the top R results. AP and R-Precision are two measures strongly correlated [Tague-Sutcliffe and Blustein, 1995; Voorhees and Harman, 1999].

5.4 Query expansion

Keyword-based search is very common in Learning Objects Repositories and many popular search engines on Internet. People are used to submitting keywords to a search engine, which in turn returns a ranked list of documents to the user. However, simple keyword-based search has a well-known associated problem, as only documents characterised with those exact keywords are retrieved.

Even if one tries to think of all words that could have been used to characterise a resource, or a document by its authors (in full-text searches), it is quite impossible to provide all the words related to a subject. Some of the common word association techniques are [Meadow, Boyce et al., 2007]:

- **Use dictionaries and thesauri** – Dictionaries can provide synonyms and thesauri can provide related words to a supplied input;
- **Apply co-occurrence statistics** – Computed the frequency of co-occurrence of each word with the others, words that co-occur with a high frequency can be considered related. This technique does not provide the nature of the relatedness and requires some computing time to maintain the frequency of co-occurrences up to date (see sections 5.5.1 and 5.5.2). In addition, it is very dependent on the corpus;
- **Use stemming and conflation methods** - 'Writ', which is the stem of the forms 'writes', 'writing' and 'written', used with a truncation symbol can admit too many related (and not that related) words, but an useful approach can be the use of a combination of roots. Conflation means the combination of different forms in a single one, and the use of write, for instance, regarding the previous example.

Many query expansion techniques have emerged in recent years. To improve precision or recall, new terms with similar meaning or somewhat related, even statistically, are appended to the original query terms specified by the users. The two major approaches to query expansion are probabilistic query expansion and expansion based on knowledge models, such as thesaurus or ontologies [Tuominen, Kauppinen et al., 2009].

Probabilistic query expansion techniques are based on the calculation of co-occurrences of terms in documents and selection of terms that are most related to query terms. They can use local or global methods [Xu and Croft, 2000]. The former might also employ global statistics.

Global techniques use statistic information extracted from the document collection, while local methods generally use information from the k documents initially retrieved.

Some interesting results have been reported on the use of ontology-based query expansion strategies [Xu, Zhu et al., 2006], in a variant on the use dictionaries and thesauri. Content relationships are explicit in an ontology and, therefore, machine-processable. GoPubMed is an example of an ontology-based literature search platform [Doms and Schroeder, 2005], which uses the Gene Ontology (GO) (http://www.geneontology.org/). However, the research works in the area mainly rely on an unique ontology specifically developed for the effect or a formally validated (by the scientific community in the area) ontology [Ali and Khan, 2008; Huang and Hsu, 2008; Lee, Tsai et al., 2008].

5.5 Information retrieval with folksonomies

The problem of the difference in the vocabulary used in searches and in indexing tasks was highlighted by Furnas et al.: "In information retrieval systems, the keywords that are assigned by indexers are often at odds with those tried by searchers. The seriousness of the problem is indicated by the need for professional intermediates between users and systems, and by disappointingly low performance (recall) rates" [Furnas, Landauer et al., 1987]. The authors found out in their experiments that the probability of two people using the same word was less than twenty percent.

In section 4.2.4 the concepts of folksonomy and tags were fully described. As tags are usually chosen by the end users, their use in IR systems can help overcoming the vocabulary problem described in the previous paragraph. In fact users can still provide different tags to the same resource, resulting in what was characterised as a "noisy tagspace" in [Begelman, Keller et al., 2006]. However, that diversity might be seen as an advantage over traditional indexing systems as possibly there are more users describing resources through tags assignments, augmenting the potential for using a specific tag. Peters states that "folksonomies are supposed to represent an option for [...] broadening access paths to the information resources via their collective indexing" [Peters, 2009] (page 287).

A study described in [Heymann, Koutrika et al., 2008] found out that popular query terms in a AOL query dataset and del.icio.us tags considerably overlap. Bischoff et al. in a study that used AOL query logs and del.icio.us tags concluded that "Regarding Del.icio.us, 71.22% of queries contain at least one Del.icio.us tag, while 30.61% of queries consist entirely of Del.icio.us tags. [...]due to the significant overlap Del.icio.us tags may help finding Web resources matching queries to tags" [Bischoff, Firan et al., 2008]. However, when they used images or music related tags, respectively from Flickr and Last.fm systems, the overlap was less significant (64.54% and 12.66% for Flickr, and 58.43% and 6% for Last.fm). Nevertheless, they emphasised the less reduced dataset used from these systems in comparison to the dataset from del.icio.us.

Currently there is no consensus regarding the retrieval effectiveness of systems that use folksonomies, with diverse researchers pointing other kind of problems, as low quality tags and tag spam, and the use of negative tags (tags that state what is not covered by a resource or a characteristic that it does not pertain) [Peters, 2009] (pages 332-336).

However, some strategies have emerged to overcome these problems. For instance, many anti-spam strategies have been proposed [Hotho, Jaschke et al., 2006; Xu, Fu et al., 2006; Koutrika, Effendi et al., 2007; Liu, Zhai et al., 2009], generally falling into one of the following categories [Heymann, Koutrika et al., 2007]: demotion, detection and prevention.

In sections 5.5.1 and 5.5.2 it is explained how to find similar resources and similar tags to one another using co-occurrence analysis (see section 5.4), respectively. This kind of analysis is not specific to tagging systems; it can be also carried out using the keywords associated to resources, for instance. However, due to the specificities of tagging systems, the effectiveness of other techniques alone, such as steaming and the use of dictionaries, to find similar tags, for example, might be compromised by the use of abbreviations, and a very informal writing style.

5.5.1 Resources related to other resources

In folksonomical systems it is possible to estimate which resources are syntagmatic related to another one, through the tags provided to them. Syntagmatic relations, as opposed to paradigmatic relations, where terms are connected through specific relations (commonly hierarchical, equivalence or association relations), are those emerged by co-occurrence of the terms. It should be useful or not, depending on the purposes, the consideration of all types of tags, or only those descriptive of the contents.

For example, if we consider the tags attributed to three resources R1, R2 and R3, which are T1, T2, T3, T4 and T5, and how many times they were used, as follows:

Table 18. Tags for each resource

	T1	T2	T3	T4	T5	Normalised value[36]
R1	5		1	3	3	6.633
R2	1	4	1	5		6.557
R3	3	2		2	3	5.099

[36] Normalised value means that the highest value of a set of values receives the value of 1 and the lowest 0.

Dividing the raw count by the normalised value (last column from Table 18), it is obtained the normalised vector for the resources (see Table 19).

Table 19. Normalised vector for the resources

	T1	T2	T3	T4	T5
R1	0.754	0.000	0.151	0.452	0.452
R2	0.152	0.610	0.152	0.762	0.000
R3	0.588	0.392	0.000	0.392	0.588

The dot product for each of the resource row in the previous table gives the results shown on Table 20. According to this, resource R3 is more similar to R1 than R2.

Table 20. Similarity matrix between the resources

	R1	R2	R3
R1	1.000	0.483	0.887
R2	0.483	1.000	0.628
R3	0.887	0.628	1.000

Considered a threshold, it is possible to state which resources are related to another one.

5.5.2 Tags related to other tags

Table 21 is a tag-resource matrix which contains the same data as Table 18. It was just rearranged to facilitate its analysis. Each row corresponds to the number of tags across resources, and a column represents a resource.

Table 21. Number of times a given tag was applied for each resource

	R1	R2	R3	Normalised value
T1	5	1	3	5.916
T2		4	2	4.472
T3	1	1		1.414
T4	3	5	2	6.164
T5	3		3	4.243

Repeating the processing explained in the previous section, Table 22 and Table 23 are obtained.

Table 22. Normalised values

	R1	R2	R3
T1	0.845	0.169	0.507
T2	0.000	0.894	0.447
T3	0.707	0.707	0.000
T4	0.487	0.811	0.324
T5	0.707	0.000	0.707

Table 23. Similarity matrix between the resources

	T1	T2	T3	T4	T5
T1	1.000	0.378	0.717	0.713	0.956
T2	0.378	1.000	0.632	0.871	0.316
T3	0.717	0.632	1.000	0.918	0.500
T4	0.713	0.871	0.918	1.000	0.574
T5	0.956	0.316	0.500	0.574	1.000

For instance, T1 and T5 are strongly correlated and the use of resources tagged with T5 can be considered for queries related to T1.

5.6 Conclusions

Diverse IR models and approaches were analysed in this chapter. One advantage of considering query expansion techniques is that it is not necessary to modify the internal functioning of the query processing, a point to consider when deciding to improve an IR system without altering its core. On the other hand, changing the whole intrinsic adopted model in an ended system is not a simple task.

In addition, the use of ontological query expansion methods has been reported with promising results. However, they are usually applied in a limited domain scope, as it is necessary to have ontologies to support the process – the principal drawback of these techniques.

In addition, the employment of folksonomies for knowledge representation in IR systems is in an early stage. The most important advantage and disadvantage of folksonomies are related to the liberal use of tags. However, they might be useful to attenuate the vocabulary problem in IR.

The IR evaluation has been focused on relevance. That concept has been dominant in the research in this area, despite the claims to consider also other aspects, such as user satisfaction.

THIRD PART: CONTRIBUTIONS

Chapter 6

LEARNING OBJECTS REPOSITORY - MODEL AND IMPLEMENTATION

In this chapter a model for repositories that considers the semantics and the pragmatics of learning objects is presented. In addition, the use cases for each actor are introduced. The learning technology specifications related to this model are also described.

In addition, this chapter discusses the chosen platform for the implementation of the model functionalities and the TREE application profile, which is based on the IEEE LOM standard.

6.1 Introduction

Learning objects repositories can be seen as storage and retrieval systems for learning resources. For instance, they usually make possible to find exercises related to a given topic. However, in fact, they can provide other functionalities to increase the findability of resources and to allow their characterisation in a richer way.

Repositories have a number of core functions. The basic functionalities to be provided are [The Consultative Committee for Space Data Systems, 2002]:

- **Administration** - It ensures the overall system operation;
- **Data ingest** – It is related to the addition of data to a repository;
- **Data management** – This functionality refers to the activities that make viable the use of data in the repository. It may include database updates, for instance, or generating reports;
- **Archival storage** – It is related to the technical systems and services that are necessary for the storage and retrieval of resources and metadata;
- **Preservation Planning** - It is necessary to guarantee that what is stored in a repository remains available for a long time, even if the original computing environment becomes outdated;
- **Providing access** – It refers to how contents are made available to end users.

These functional entities and the corresponding interfaces are visible in Figure 27.

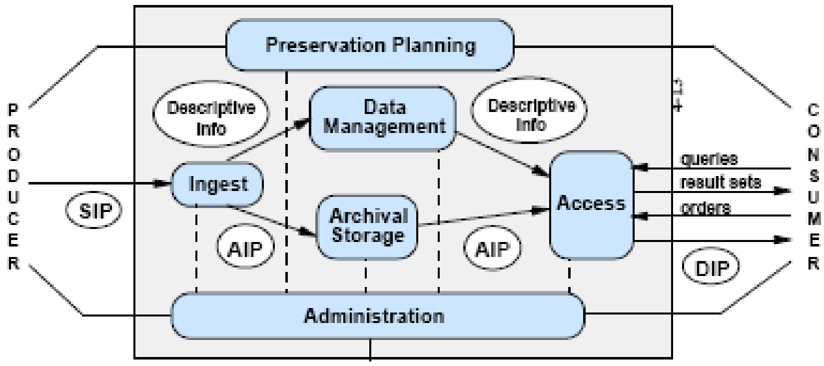


Figure 27. Repository core components
(Source: [The Consultative Committee for Space Data Systems, 2002])

The current practice in some repositories also provides some insights into the question of what are the commonly adopted features in repositories. A relatively recent study [Kallonis and Sampson, 2010] analysed 12 LORs, which are presented in Table 24 .The results are visible in Table 25, Table 26 and Table 27 (adapted from the mentioned article).

Table 24. The twelve repositories analysed in [Kallonis and Sampson, 2010]

LOR	**Web site**
ARIADNE	http://www.ariadne-eu.org
COSMOS	http://www.cosmosportal.eu
eAccess	http://www.eaccess2learn.eu
EdNA	http://www.edna.edu.au
FREE	http://free.ed.gov
LRE	http://lreforschools.eun.org
Jorum	http://open.jorum.ac.uk
Merlot	http://www.merlot.org
MIT OCW	http://ocw.mit.edu
Netlib	http://www.netlib.org
NLN Materials	http://www.nln.ac.uk
SMETE	http://www.smete.org

Some features offered by the 12 analysed repositories are presented in Table 25. In the sample, 67 percent of the repositories provide storage for LO, but FREE, LRE, Merlot and SMETE do not store LOs, but their links. Some of them also do not accept contribution from any users, and only selected users can upload their own digital educational resources.

Table 25. Features related to Los

Functionality	**Percentage**	**ARIADNE**	**COSMOS**	**eAccess**	**EdNA**	**FREE**	**LRE**	**Jorum**	**Merlot**	**MIT OCW**	**Netlib**	**NLN Materials**	**SMETE**
Storage	67	X	X	X	X			X		X	X	X	
Search/Browse	100	X	X	X	X	*X*	X	X	X	X	X	X	X
View	100	X	X	X	X	*X*	X	X	X	X	X	X	X
Download	75	X	X	X	X			X		X	X	X	X
Rate/Comment	50		X	X			X	X	X				X
Bookmark	33		X		X	*X*	X		X				
Wide submission	58		X	X	X	*X*	X	X	X				X

All the analysed repositories store metadata about resources (see Table 26). On the other side, only a few repositories allow to rate/comment LO metadata and contribute to LO metadata.

Table 26. Features related to LOs Metadata

Functionality	Percentage	ARIADNE	COSMOS	eAccess	EdNA	FREE	LRE	Jorum	Merlot	MIT OCW	Netlib	NLN Materials	SMETE
Storage	100	X	X	X	X	X	X	X	X	X	X	X	X
View	100	X	X	X	X	X	X	X	X	X	X	X	X
Download	17		X	X									
Rate/Comment	33	X		X			X		X				
Wide submission	50		X	X			X	X	X				X

From the features presented in Table 27, it seems that those related to users' personal accounts are the most commonly available.

Table 27. Other used features in repositories

Functionality	Percentage	ARIADNE	COSMOS	eAccess	EdNA	FREE	LRE	Jorum	Merlot	MIT OCW	Netlib	NLN Materials	SMETE
Creation and Management of Personal Accounts	75	X	X					X	X	X		X	X
Automatic Recommendations	8								X				
Supportive Tools	50	X	X	X	X	X		X					
Forums	50		X	X	X			X	X				X

From the data presented in the previous tables, it seems that 'Search/Browse' and 'View' of LO, and 'Storage' and 'View' of LO metadata are features undoubtedly available in repositories (100 percent in the considered sample). Another functionality that has that status is the submission of resources, even if not open to every user. Other common features (75 percent) are: 'Download' of LO and 'Creation and Management of Personal Accounts'. These features are related to some of the basic functionalities previously introduced ('Administration', 'Data ingest', 'Data management', 'Archival storage', 'Preservation Planning', and 'Providing access)'. The Preservation Plan, if used, is not visible to external users, and thus it is not possible to assess its adoption.

In this third part (Contributions) all the additional features, or those existing but in need of change, that should be considered in repositories to allow considering pragmatics and semantics of resources are presented.

This chapter introduces a model for repositories supported in methods of knowledge representation (see Chapter 4) and specifications from diverse areas (see Chapter 2 and Chapter 4), to characterise semantically and pragmatically learning objects. The implementation of the proposed model has the potential of offering other possibilities in IR (see Chapter 5). The topics of pragmatics and semantics were discussed in Chapter 3 and Chapter 4, and the next two chapters discuss their related components in the model.

In the development process a repository for a specific organisation, a higher education institute, was considered, the Instituto Superior de Engenharia do Porto (ISEP). The TREE (Teaching Resources for Engineering Education) group was constituted by initiative of the Scientific Council of ISEP in early 2008 and one of its objectives was to create a repository for storing / referencing educational resources (hereafter designated as the TREE repository), a common objective to the PhD work described in this document. Therefore, some of the decisions regarding the infrastructure preparation were based on the opinions expressed by the members of that group. However, in this chapter it was aimed to present the ideas and their implementation in a generic way, whenever possible.

The remaining of this chapter is organised as follows. The proposed model is introduced in section 6.2, but relating it to the next two chapters. Section 6.3 presents the actors and their use cases. In the process of identifying actors and use cases, situations are modelled as abstract enough to describe general requirements. Section 6.4 details the infrastructure preparation, with the discussion of the adoption of a repository platform and the metadata schema to be used. Finally, the last section provides some concluding remarks.

6.2 Proposal

In an early stage of this PhD work a preliminary model was defined. The feedback received from its public presentation in conferences made possible the refinement of some of the underlying ideas. The final version is shown in Figure 28.

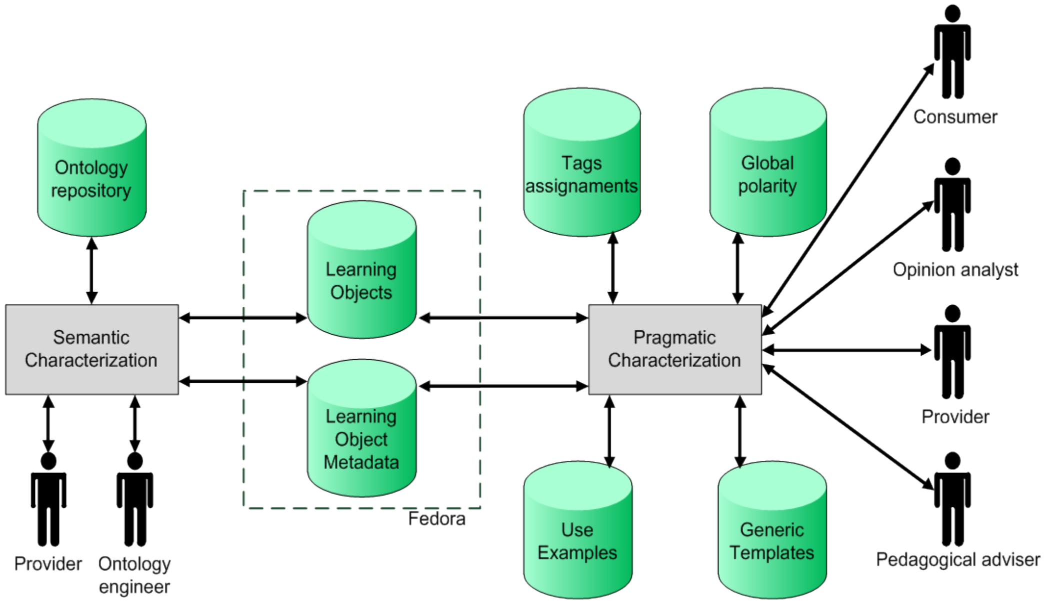


Figure 28. A proposed model for repositories
(Adapted from [Azevedo, Martins et al., 2006; Azevedo, Carvalho et al., 2008])

The proposed system, through a semantic and pragmatic characterisation of educational resources, intends to facilitate the use and discovery of resources of interest and to allow users to assign tags to resources, thereby contributing to a personal view about them. These tags, considered their local and global polarity, allow providers to have a personalised feedback.

On the other hand, upon submission of resources, one or more templates can be chosen. These templates illustrate a possible use form with the semi-automatic generation of a Unit of Learning that can be used in other compatible systems.

Another important point in the proposed model is the semantic characterisation of resources through the use of ontologies (and their acquisition).

Two important components of the model, 'Semantic characterisation' and 'Pragmatic characterisation', are fully explained in Chapter 7 and Chapter 8, respectively.

The components 'Learning Objects' and 'Learning Object Metadata' are part of the Fedora system, which was the choice for the basis of the development. Section 6.4 details that option.

The adopted learning technology specifications were:

- **IMS LD** – The component 'Use Examples' is compliant with IMS LD. It aims to exemplify how a learning object can be successfully used;
- **IMS CP** – In the component 'Use Examples' the learning object and the unit of learning are packed in a ZIP file compliant with the IMS CP specification. The IMS CP specification was considered important to allow the migration of the resources to other platforms;
- **IEEE LOM** – The metadata related to the learning objects are described in accordance with a LOM application profile, which is described later in this chapter.

6.3 Use cases

In this part only the additional or modified functionalities of a typical repository (the normal features of repositories were detailed in section 6.1) are regarded. However, having chosen a specific open-source platform for the system development, its particular functionalities were also considered in the list of use cases that are presented.

Thus, having this note in mind, the actors and their main tasks are as follows:

- **Administrator** – The Administrator(s) ensure(s) the proper functioning of the proposed system;
- **Consumer** – This actor represents a registered user that mainly uses the system to satisfy his interest in some topic(s), potentially for learning or teaching purposes. He can submit queries and analyse a set of resources that potentially suits his specific needs. He can also characterise the resources through tagging, providing his personal impressions and views about them. He can suggest additional templates to be widely considered in the repository;
- **Ontology engineer** – The Ontology engineer is responsible for the process of reusing ontologies, with the assistance of domain experts to have a list of core concepts of a certain domain and in the process of ontologies analysis and evaluation;
- **Pedagogical adviser** – This actor is concerned with the management of IMS LD templates and he has access to opinion statistics and can interact with providers as a consequence;
- **Provider** – This type of user submits resources to the repository and gets feedback about them. Institutions might consider as providers teachers, PhD students and/or researchers, among other alternatives;
- **Sentiment analyst** – This type of user has access to a list of words and their polarity opinion. They can change the assigned polarity to a tag regarding a particular resource or even its global polarity.

The additional functional requirements that must be considered in repositories of learning objects are described briefly in this section. It was chosen to briefly describe them aggregated by actors from subsections 6.3.1 to 6.3.6.

In Annex 3 the use cases are presented in detail, complementing their descriptions provided in this chapter.

6.3.1 Administrator's use cases

The identified use cases related to the actor 'Administrator' are represented in the diagram below (Figure 29)[37].

[37] In UML 2.0 an extend dependency is represented by means of <<extend>>, not <<extends>> as shown in the diagrams.

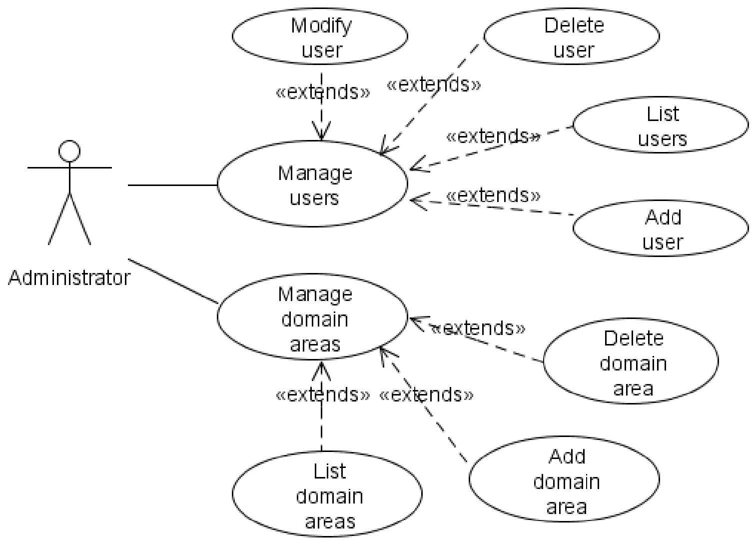


Figure 29. Administrator's use cases

Thus, besides the common tasks for a repository administrator, the use cases related to the administrator are as follows:

- **Manage users (UC01)** – The administrator is responsible for managing system' users;
  - **List users (UC02)** – The administrator can see information about all the registered users;
  - **Modify user (UC03)** – It allows the modification of the information about a specific user;
  - **Delete user (UC04)** – It allows the logical removal of a user;
  - **Add user (UC05)** – New users can be added by the administrator;
- **Manage domain areas (UC06)** – The administrator is responsible for managing the domain areas under usage in the repository. A domain area may have associated a domain ontology;
  - **Add domain area (UC07)** – The administrator can add a new domain area that is to be considered in the repository;
  - **List domain areas (UC08)** – It allows all domain areas to be listed;
  - **Delete domain area (UC09)** – If not anymore adequate to be used in the repository, a domain area can be deleted by the administrator.

The use cases related to management of users, besides representing a common feature in repositories, are presented because of the new types of users considered.

6.3.2 Consumer's use cases

For the consumer the generic use cases shown in Figure 30 were identified.

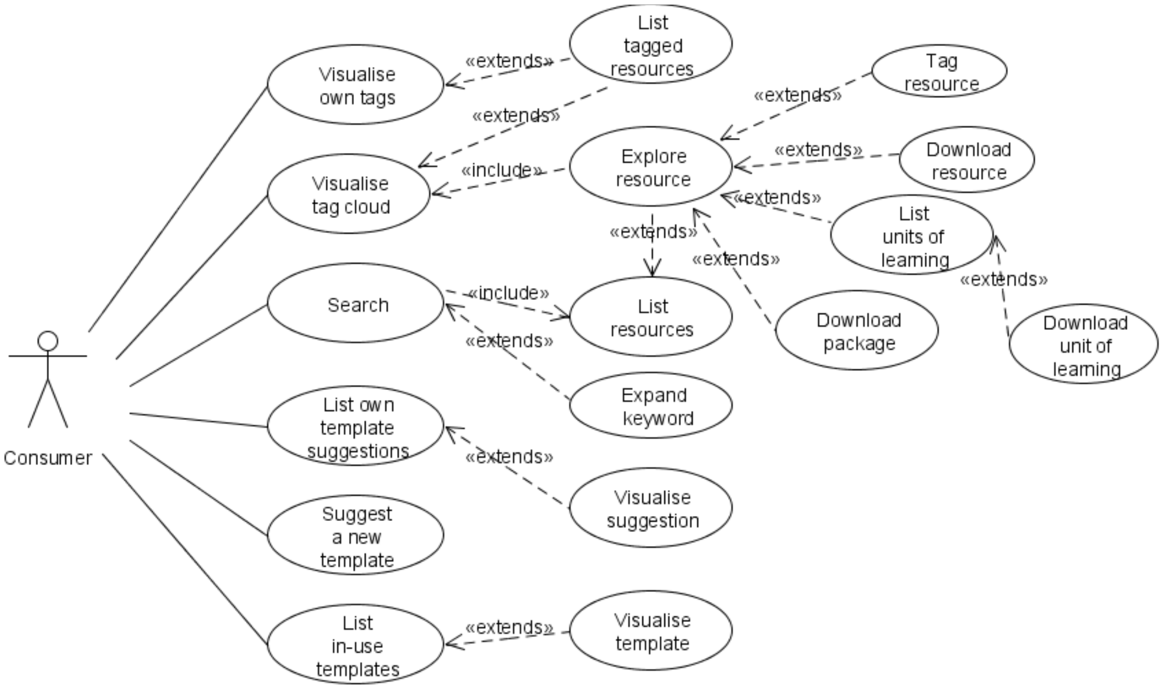


Figure 30. Consumer's use cases

Briefly described, the consumer's use cases are as follows:

- **Visualise own tags (UC10)** – The consumer can see the tags that he has previously applied;
    - **List tagged resources (UC11)** – Through a list of tags, the consumer has the choice to see all the resources to which a specific tag was assigned;
- **Visualise tag cloud (UC12)** – The consumer can see the tags used by other users in a way that puts in evidence their popularity;
- **Search (UC13)** – It is possible to submit queries to the system specifying one or more searchable fields;
    - **List resources (UC14)** – For a previously submitted query, the system provides a list of all resources that fulfil the query;
        - **Explore resource (UC15)** – All metadata about the resource is visible and also a list of tags that were assigned;
            - **Tag resource (UC16)** – The consumer can apply one or more tags to a resource;
            - **Download resource (UC17)** – If stored in the system, the consumer can download the resource. Otherwise, it is possible to visit the Web page associated to the resource;
            - **List units of learning (UC18)** – The consumer can visualise a list of the IMS LD files that exemplify how to use the associated resource;
                - **Download unit of learning (UC19)** – A file that exemplifies how to use the associated resource, which is in compliance with the appropriated specification, is available to be download and then used in any compatible system;

        - **Download package (UC20)** - A file compliant with the appropriated specifications, which encompasses metadata and a unit of learning (if it exists), is available to be downloaded and then used in any compatible system;
    - **Expand keyword (UC21)** – When specifying one or more keyword in the search form, it is possible to find other related keyword using the available domain ontologies;
- **List own template suggestions (UC22)** – A consumer can see his template suggestions and corresponding statuses;
    - **Visualise suggestion (UC23)** – A consumer can see detailed information about a previously suggested template;
- **Suggest a new template (UC24)** – A consumer can proposed a template to be considered in the system;
- **List in-use templates (UC25)** – It is possible to see all templates that are in use in the system;
    - **Visualise template (UC26)** – A consumer can see detailed information about an in-use template.

6.3.3 Ontology engineer's use cases

Domain characterisers help the ontology engineer in the specification of concepts relevant in the considered domain, as well as in the specification of competency questions to analyse the returned ontologies, but without direct intervention in the system. In a higher education institution, the users that assist the ontology engineer can be the responsible for some disciplines. However, the domain characterisers (or experts) do not directly interact with the system.

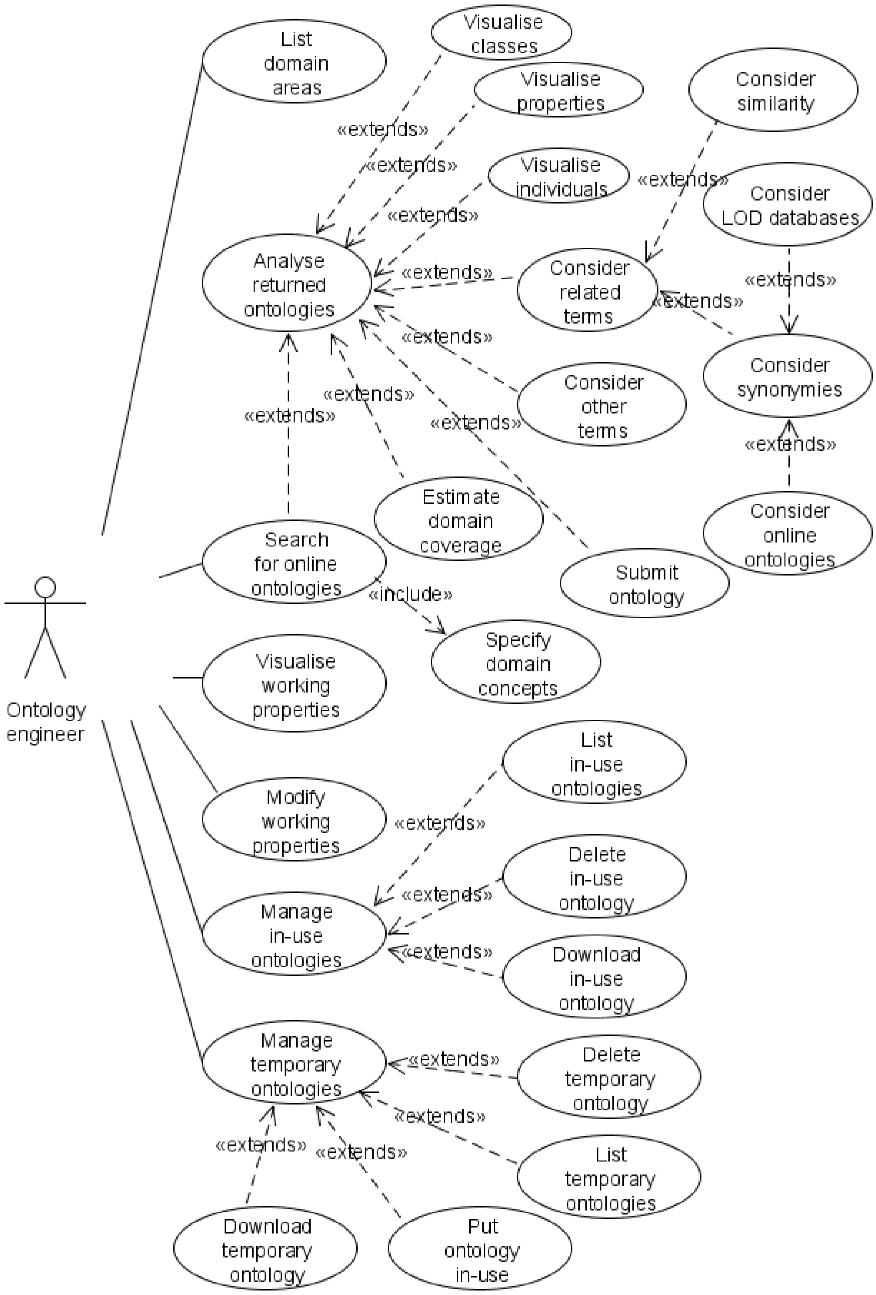


Figure 31. Ontology engineer's use cases

Figure 31 depicts the use cases identified for the ontology engineer, which are as follows[38]:

- **List domain areas (UC08)**;
- **Search for online ontologies (UC27)** – The ontology engineer characterises the concepts related to a specific domain and analyses the results;
    - **Specify domain concepts (UC28)** – The ontology engineer identifies the concepts related to a certain topic;
    - **Analyse returned ontologies (UC29)** – For each returned ontology some metadata fields are displayed. Also, although an ontology can be used and analysed in an external tool, some facilities are promptly available;

[38] Those without any description were already discussed in this chapter.

- **Visualise classes (UC30)** – The ontology engineer can visualise the classes of an ontology;
- **Visualise properties (UC31)** – It is possible to see the properties of an ontology;
- **Visualise instances (UC32)** – The individuals of an ontology can also be analysed;
- **Consider related terms (UC33)** – The ontology engineer has the option to consider other terms besides the specific ones that were provided;
    - **Consider similarity (UC34)** – Syntactic similarity can be used to catch very similar terms that are regarded to also have a quite analogous meaning, such as 'model' and 'modelling';
    - **Consider synonyms (UC35)** – For each domain term provided, a set of synonyms is obtained, but it is tried to estimate the synonyms most appropriated in the regarded area;
        - **Use Online ontologies (UC36)** – Online ontologies can be used to highlight the synonyms that make more sense in a given domain area;
        - **Consider LOD databases (UC37)** – Some Linked Open Data databases provide open data according to the W3C Linking Open Data project and they can be applied in order to determine the synonyms to use;
- **Consider other topics (UC38)** – Any domain term can be substituted and the whole process repeated;
- **Estimate domain coverage (UC39)** – It is possible to calculate the coverage degree of an ontology concerning the domain terms provided;
- **Submit ontology (UC40)** - It is possible to submit an ontology to the temporary ontologies repository in order to be then deeply analysed;

- **Visualise working properties (UC41)** – Various aspects are configurable in order to satisfy the current requirements;
- **Modify working properties (UC42)** – Various aspects are configurable in order to satisfy the current requirements;
- **Manage in-use ontologies (UC43)** – Regarding a domain area, it is possible to manage the ontologies in current usage:
    - **List in-use ontologies (UC44)** - The ontology engineer has access to a list of ontologies in current usage;
    - **Delete in-use ontology (UC45)** - The ontology engineer has the option to delete an ontology that no longer is to be used in the repository;
    - **Download in-use ontology (UC46)** - The ontology engineer has the option to download the ontology to use it in any ontology tool;
- **Manage temporary ontologies (UC47)** – There are some options available to the ontology engineer regarding the management of ontologies that were pre-selected to be deeply analysed later;
    - **List temporary ontologies (UC48)** - Regarding a domain area, it is possible to see the ontologies that were considered to be of possible interest to the repository;
    - **Delete temporary ontology (UC49)** – The ontology engineer can remove an ontology from the temporary repository;
    - **Put ontology in use (UC50)** – An ontology can be effectively made available to all related functions in the repository;

- **Download temporary ontology (UC51)** - The ontology engineer has the option to download the ontology to use it in any ontology tool.

6.3.4 Pedagogical adviser's use cases

For the pedagogical adviser the generic use cases shown in Figure 32 were identified.

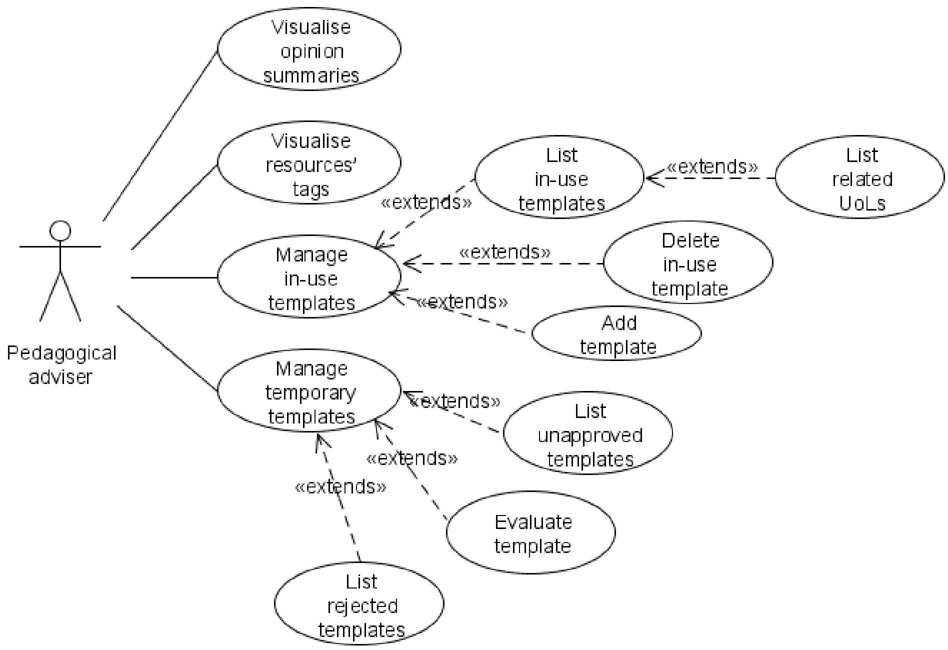


Figure 32. Pedagogical adviser's use cases

Summarising, the use cases recognised for the pedagogical adviser are as follows:

- **Visualise opinion summaries (UC52)** – The pedagogical adviser can see the results of the consideration of the tags regarding the opinions expressed by the users;
- **Visualise resources' tags (UC53)** – The pedagogical adviser can see the tags assigned to the resources in the repository;
- **Manage in-use templates (UC54)** –The IMS LD templates are supervised by the pedagogical adviser. They can see all the templates under usage, delete, modify or add templates;
    - **List in-use templates (UC55)** - It is possible to visualise the templates that were considered to be of possible interest to the repository;
        - **List related UoLs (UC56)** - It is possible to visualise the UoLs built on a given template;
    - **Delete in-use template (UC57)** - The pedagogical adviser has the option to delete a template that no longer is to be used in the repository;
    - **Add template (UC58)** – It is possible for the pedagogical adviser to add a new template to the repository;
- **Manage temporary templates (UC59) –** Temporary templates are those templates that were not evaluated or were rejected. The pedagogical adviser is responsible for supervising those templates in the repository:
    - **List unapproved templates (UC60)** – The pedagogical adviser can see a list of templates under approval consideration;

- **Evaluate template (UC61)** - It involves seeing the new IMS LD templates and approving or rejecting each of them. Approved templates become in-use in the repository;
- **List rejected templates (UC62)** – It is possible to visualise a list of the templates that were not considered to be used in the repository.

6.3.5 Provider's use cases

Figure 33 depicts the use cases identified for the provider.

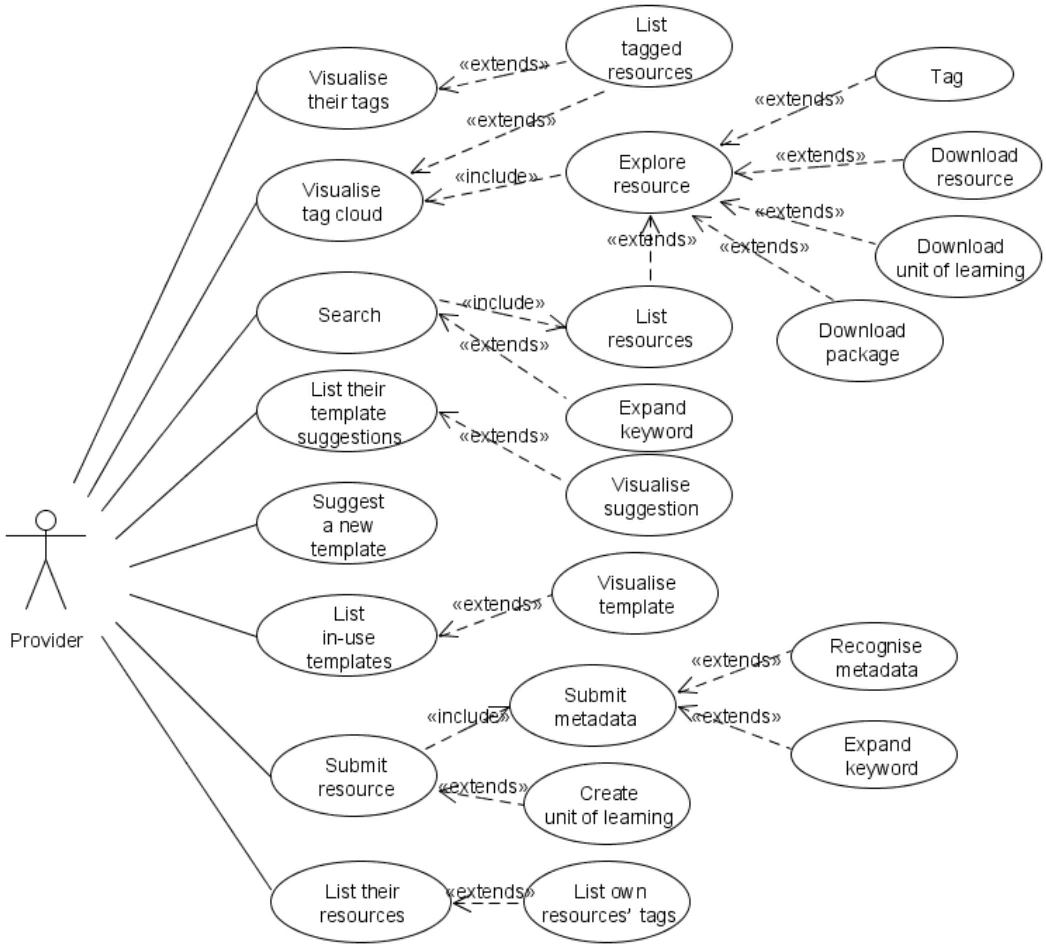


Figure 33. Provider's use cases

Briefly described, the provider's use cases are as follows[39]:

- **Visualise own tags (UC10)**;
  - **List tagged resources (UC11)**;
- **Visualise tag cloud (UC12)**;
- **Search (UC13)**;
  - **List resources (UC14)**;
    - **Explore resource (UC15)**;
      - **Tag resource (UC16)**;

[39] Those without any description were already discussed in this chapter.

        - **Download resource (UC17)**;
        - **List units of learning (UC18);**
            - **Download unit of learning (UC19)**;
        - **Download package (UC20)**;
    - **Expand keyword (UC21)**;
- **List their template suggestions (UC22)**;
    - **Visualise suggestion (UC23)**;
- **Suggest a new template (UC24)**;
- **List in-use templates (UC25);**
    - **Visualise template (UC26)**;
- **Submit a resource (UC63)** – The provider submits resources to the repository, supplying a file or an external link. Also, a number of metadata fields have to be provided;
    - **Submit metadata (UC64)** – In accordance with the metadata schema in force, some fields must be supplied to the system, while others are optional. The provider is responsible for their final provision;
        - **Recognise metadata (UC65)** – Some metadata fields can be automatically recognised by the system, if desired by the provider;
        - **Expand keywords (UC66)** – When submitting a resource, a provider can ask to see related keywords to those provided or automatically recognised. He can select some of them to be also considered;
    - **Create a unit of learning (UC67)** – Having in mind a way to use the resource, the provider can specify it;
- **List their resources (UC68)** – The provider has access to a list of the resources that he has previously submitted to the repository;
    - **List own resources' tags (UC69)** – A provider can see some statistic information regarding his submitted resources, and also a summary about the opinions expressed through tagging and all the tags assigned by the repository's users.

6.3.6 Opinion analyst's use cases

For the opinion analyst the generic use cases shown in Figure 34 were identified.

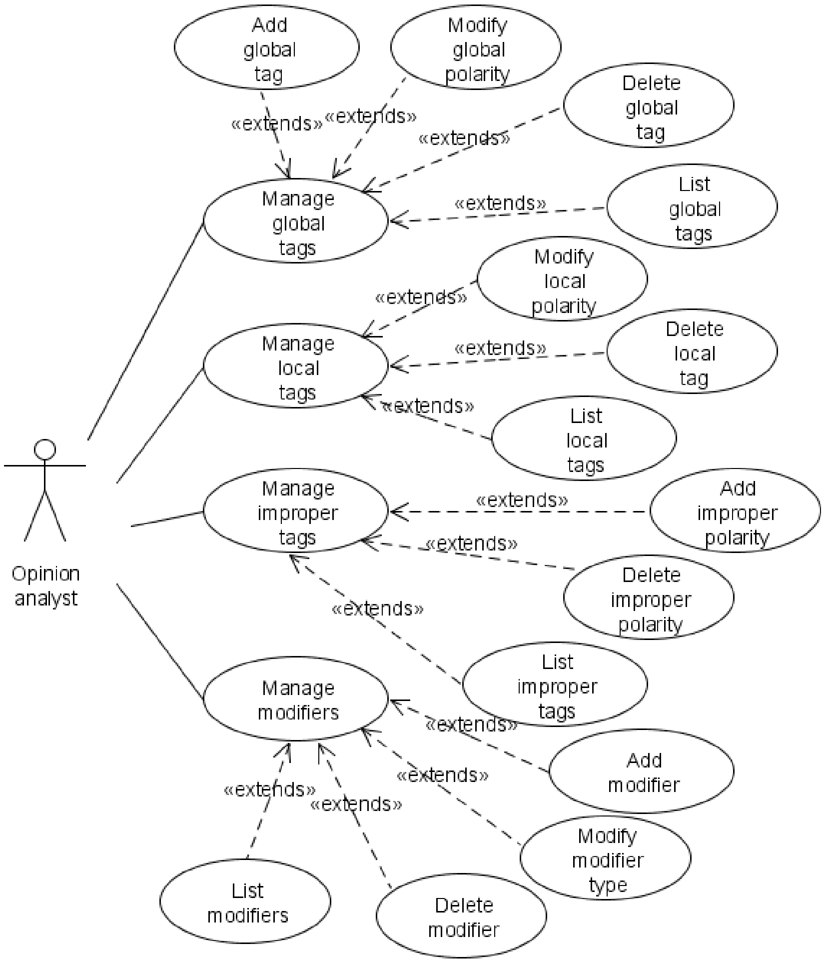


Figure 34. Opinion analyst's use cases

A total of eighteen use cases are associated with the opinion analyst, as follows:

- **Manage global tags (UC70) -** There are some options available to the opinion analyst regarding the management of global tags, which are tags for which a global polarity is predefined;
    - **Add global tag (UC71)** - The opinion analyst can add to the repository new tags and the respective global polarity;
    - **Modify global polarity (UC72)** – It is possible to modify a global tag polarity. That option is especially useful in the case of any error detection;
    - **Delete global tag (UC73)** – It is possible to remove a tag from the list of tags which have a global polarity assigned;
    - **List global tags (UC74)** - The opinion analyst can see the global list of tags and their polarity;
- **Manage local tags (UC75)** – Local tags are tags effectively assigned to a resource and can have a distinct polarity from the globally envisaged one due to some particularities of the resource under analysis. The opinion analyst has some options available to their management;
    - **Modify local polarity (UC76)** - The opinion analyst can modify a local tag polarity;

  - **Delete local tag (UC77) –** A local tag can be deleted (and become an improper one) if not appropriate to be used in a repository;
  - **List local tags (UC78)** – It is possible to visualise the list of local tags assigned to each resource in the repository;
- **Manage improper tags (UC79)** – Swear words or others considered to be offensive cannot be used in a repository. The opinion analyst is in charge of them, as he can add others, delete or visualise a list of them;
  - **Add improper tag (UC80)** – A tag if not appropriate to be used in a repository can be added to a list of forbidden tags;
  - **Delete improper tag (UC81)** – A tag can stop to be considered as inadequate to be used as tag, in consequence of a detected error or other reasons;
  - **List improper tags (UC82)** - The opinion analyst can visualise the tags which were previously classified as inadequate to be used in the system;
- **Manage modifiers (UC83) -** There are some options available to the opinion analyst regarding the management of modifiers, which are words that can restrict the meaning of other words in some way;
  - **Add modifier (UC84)** - The opinion analyst can add to the repository new modifiers;
  - **Modify modifier type (UC85)** – It is possible to change the type of modifier;
  - **Delete modifier (UC86)** – It is possible to remove a modifier;
  - **List modifiers (UC87)** - The opinion analyst can see the list of modifiers and their type (intensifier indicator or negation indicator).

6.4 Infrastructure preparation and implementation

Aiming to add new features to a common repository platform, it was important its selection and all the initial preparation work. Between a Learning Object Repository platform and a generic repository platform, the choice fell on the latter for its wide use [The JORUM Team, 2005] (page 44).

The availability of the source code of a repository platform was an important point in order to allow the customisation of the provided solution and the addition of new functionalities. Some of the best known and referred open source systems for repositories are ePrints, DSpace and Fedora. They represent a common denominator in a number of articles [Powel, 2005; The JORUM Team, 2005; Goh, Chua et al., 2006; Jeffery, 2006; Carr, 2007; Sreekumar, Sunitha et al., 2007; Amaral, 2008; Aschenbrenner, Blanke et al., 2008], and even in more recent ones [Marill and Luczak, 2009; Pyrounakis and Nikolaidou, 2009; Kökörčený and Bodnárová, 2010; Kökörčený, 2011]. All of them are initiatives with participation of universities, and many of them highlight the flexibility of the Fedora repository system. In fact, Fedora stands for "Flexible Extensible Digital Object Repository Architecture".

It was not important to choose the best open-source platform according to some criteria, but one often used and with some flexibility to be adapted to the specific work to be carried out. An aspect that was verified was the possibility to adapt or to change its default metadata schema.

Figure 35 depicts the components of the Fedora Digital Object Model, which are:

- **Persistent Identifier**, which is unique for each object;
- **Datastream(s)** hold(s) data or metadata and can be stored internally or remotely. Besides the default Dublin Core metadata datastream, other can also be used;
- **Disseminator(s)** state(s) an external service that can be used to view the object or its datastream content. More than one disseminator can be specified.

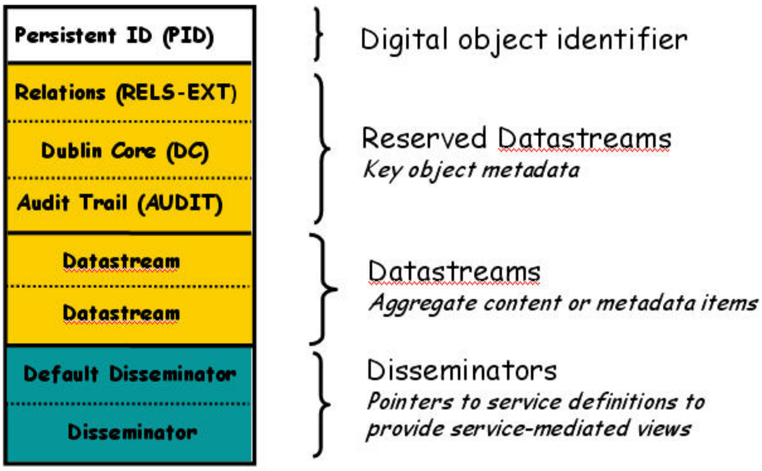


Figure 35. The Fedora digital object container model (Source: [The Fedora Development Team, 2007])

Thus, through the use of datastreams, different metadata schemas can be used in the Fedora repository system. The possibility offered for the construction of new metadata profiles different from the base version provided was important for the adoption of Fedora, as it enables the adoption of metadata specifications associated with the practice of teaching and learning, such as IEEE Learning Object Metadata, which considers not just metadata fields related to authorship, identification and brief descriptions of content.

Using the Fedora Generic Search Service (GSearch) it is possible to include fields from other metadata datastreams than DC.

Another important point was the flexibility of the system regarding the inclusion of new functionalities. As depicted in Figure 36, the interaction with the Fedora Repository System is made through web services on the frontend. On the backend, any needed data transformation is performed. REST requests and SOAP-based methods assure a web-based interface to Fedora digital objects, and they facilitate the integration of other functionalities.

Thus, Fedora, with a service-based architecture, was chosen since it provides extensibility capabilities and flexibility in the adoption of other metadata schema besides the default one.

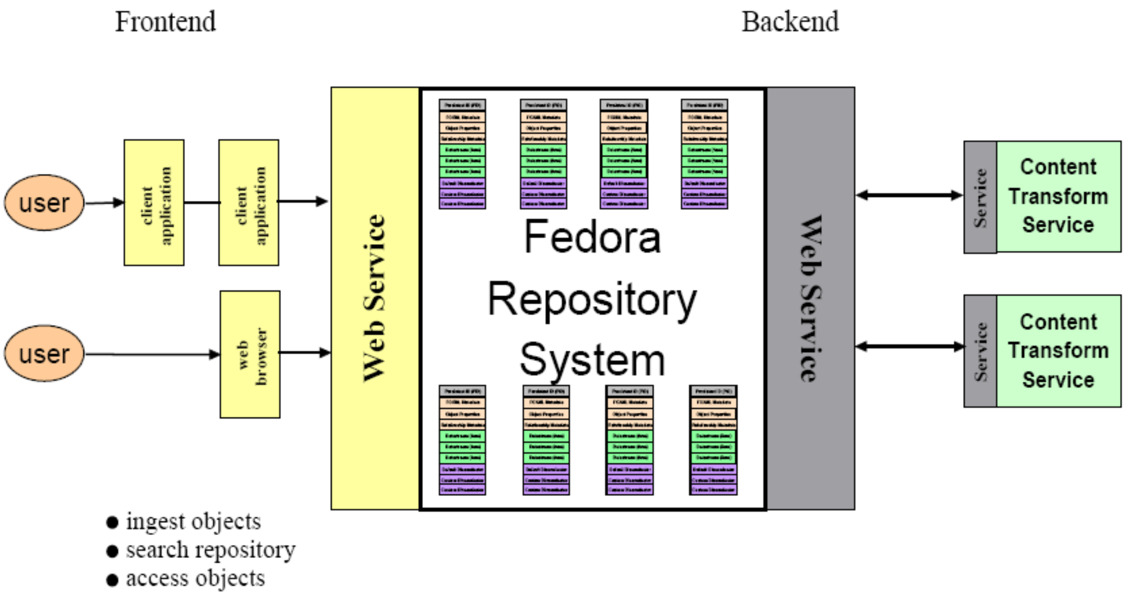


Figure 36. Interactions to the repository system (Source: [The Fedora Development Team, 2008])

Fedora installation requires the previous installation of a Java SE Development Kit (JDK), Apache Tomcat and a database management system, from a list of possibilities, and the choice for this latter requisite was PostgreSQL.

Fedora does not come with a front-end application, but there are available some third-party front end applications for Fedora, such as The Fascinator[40], Islandora[41], Blacklight[42], and Fez[43], among others[44]. At the time of the choice, the Fedora 2.x interfaces had already reached a mature stage, and among them all, Fez, a management and search front end, was adopted to be used since it satisfied the basic requirements.

Fez was developed by the University of Queensland Library as an open source project. It is available under a GNU General Public License. Before installing Fez, some packages must be installed, e.g., Apache Web Server, PHP, and MySQL.

Fez 2.1 Release Candidate 3 and Fedora 2.2.4 are the basis of the TREE repository, and their interconnection is shown in Figure 37.

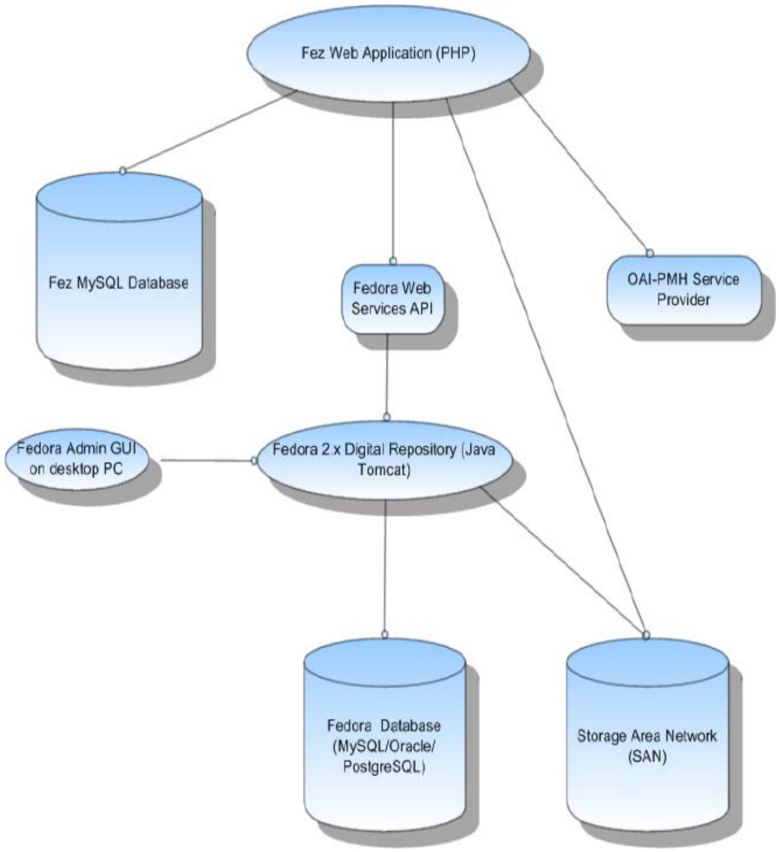


Figure 37. Fez and Fedora interconnection
(Source: [Webster, Kortekaas et al., 2007])

The installation process is detailed in [Seiça, Ortiz et al., 2009; Seiça, 2009].

Regarding the organisation of the resources, there are communities, one for each ISEP department:

- Chemical Engineering;
- Civil Engineering;

40 https://fascinator.usq.edu.au/trac
41 http://islandora.ca/
42 http://blacklight.rubyforge.org/
43 http://fez.library.uq.edu.au/wiki/Main_Page
44 A complete list is available at https://wiki.duraspace.org/display/DEV/Fedora+Tools.

- Electrical Engineering;
- Geotechnical Engineering;
- Informatics Engineering;
- Mechanical Engineering;
- Physics;
- Management;
- Mathematics.

These communities are visible in Figure 38. This organisation emerged from the opinions provided by the TREE group.

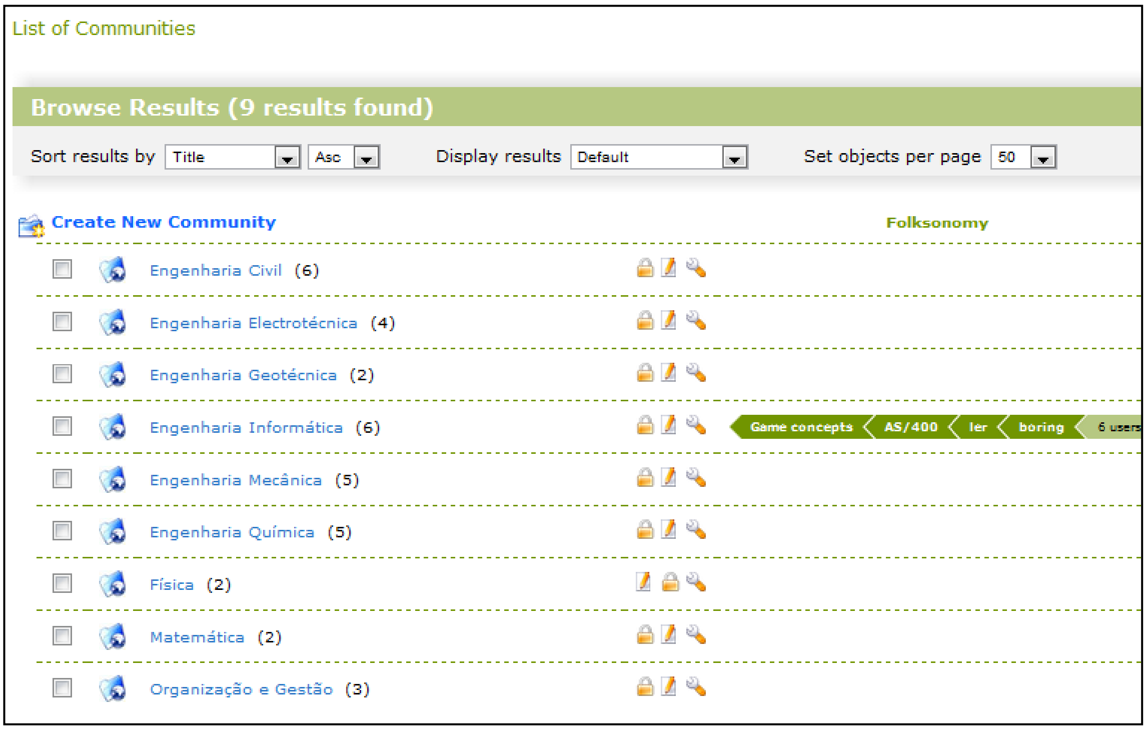


Figure 38. Communities in the TREE repository

Each community has collections, which in turn may have sub-collections. Learning objects are put in sub-collections, if they exist, or in collections. Figure 39 shows the collections in the Informatics Engineering community.

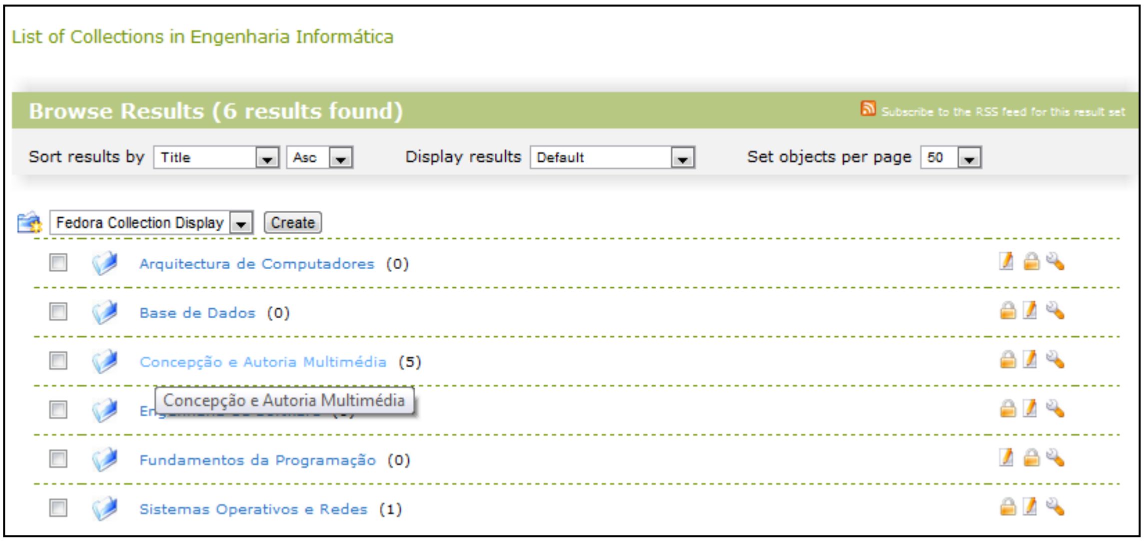


Figure 39. Collections in the TREE repository

Figure 40 shows the list of learning objects in a given collection, which does not have sub-collections.

Figure 40. Browsing learning objects in the TREE repository

Using Fedora/Fez it is possible to configure what fields are shown on the advanced search form and what search keys are used in basic searches. Fedora/fez return the results based on MySQL Boolean logic and they are ordered by search relevancy by default [Kortekaas, 2006].

6.4.1 TREE Application Profile

From the study documented in the second chapter, the necessity of compliance with the LOM standard was clear, although may not considering all fields. The TREE group was considered in the decision process of what fields to include in the TREE application profile. In September 2009 an early version of the repository was available for review by the elements of that group.

Considered the opinions of teachers from various departments of ISEP (TREE members) through semi-structured interviews carried out during three face meetings, and other application profiles, the adopted fields are listed from Table 28 to Table 36, grouped by categories. Some of the application profiles already presented in Chapter 2 are also detailed for comparison purposes. However, it should be noted that other application profiles were also analysed, such as:

- MACE application profile (versions 1.0[45] and 3.0[46]), which was developed under the MACE project (http://www.mace-project.eu);
- ELEONET application profile[47], developed under the ELEONET (European Learning Objects Network) project (http://www.eleonet.org/);
- MoodleCore[48], the LOM application profile for Moodle courses;
- The CELEBRATE Metadata application profile v1.1[49]. It was developed under the CELEBRATE (Context eLearning with Broadband Technologies) project (http://celebrate.eun.org).

However, it should be noted that the exact fields to be regarded in each institution need to be chosen carefully in order to meet its own current and envisioned necessities, whereas allowing a compromise between the deepest desired characterisation and its practical viability. For instance, regarding all fields to be mandatory might jeopardise the adoption of the repository, if the fields have to be manually provided.

Some fields were considered very important for the semantic and the characterisation of learning resources, as explained in the next two chapters, while others were signalled by the TREE members as essential for a learning object repository. Those fields are mandatory in the TREE application profile. For the optional elements the TREE members were also considered, as well as other application profiles analysed.

Table 28 presents the LOM elements of general category, which were included in the TREE application profile. Three active fields were considered to be mandatory (1.2.Title, 1.3.Language, and 1.5.Keyword).

Table 28. Use of general category elements in some application profiles, including TREE

IEEE LOM		ARIADNE	CanCore	HEAL	UK LOM Core	TREE
LOM number	**LOM name**					
1	General	M	O	M	M	M
1.1	Identifier		O	M	M	O
1.1.1	Catalog	M	O	M	M	O
1.1.2	Entry	M	O	M	M	O
1.2	Title	M	O	M	M	M
1.3	Language	M	O	O	M	M
1.4	Description	O	O	M	M	O
1.5	Keyword		O		O	M
1.6	Coverage			O	O	
1.7	Structure				O	O
1.8	Aggregation Level	O	O		O	O

[45] http://www.mace-project.eu/media/deliverables/JD/JD1/Annex%20-%20MACE%20application%20profile%20V1_0.pdf
[46] http://www.mace-project.eu/media/deliverables/WP7/D7_5/Annex%20A%20-%20MACE%20application%20profile%20V3_0.pdf
[47] http://www.eleonet.org/eleonet/eleonet_metadata_specifications.pdf
[48] http://docs.moodle.org/en/Metadata:MoodleCore
[49] http://celebrate.eun.org/docs/CELEB_AP_v1.1_2003-11-17.pdf

From the life cycle category, only the field that identifies the entities that contributed to the learning object (2.3.2. Entity) was considered to be mandatory, as shown in Table 29.

Table 29. Use of life cycle category elements in some application profiles, including TREE

IEEE LOM		ARIADNE	CanCore	HEAL	UK LOM Core	TREE
LOM number	**LOM name**					
2	Life Cycle	M	O	M	M	M
2.1	Version	O	O		O	O
2.2	Status				O	O
2.3	Contribute		O	M	M	O
2.3.1	Role	M	O	M	M	O
2.3.2	Entity	M	O	O	M	M
2.3.3	Date	O	O	O	M	O

All elements of the third category are optional in the TREE application profile (see Table 30).

Table 30. Use of meta-metadata category elements in some application profiles, including TREE

IEEE LOM		ARIADNE	CanCore	HEAL	UK LOM Core	TREE
LOM number	**LOM name**					
3	Meta-MetaData	M	O	M	M	O
3.1	Identifier		O	M	M	O
3.1.1	Catalog		O	M	M	O
3.1.2	Entry		O	M	M	O
3.2	Contribute		O	M	M	O
3.2.1	Role	M	O	M	M	O
3.2.2	Entity	M	O	O	M	O
3.2.3	Date	M	O	O	M	O
3.3	Metadata Scheme		O	O	M	O
3.4	Language	M	O	O	M	O

Few elements of the fourth category are included in the TREE application profile, even as optional ones, as detailed in Table 31.

Table 31. Use of technical category elements in some application profiles, including TREE

IEEE LOM		ARIADNE	CanCore	HEAL	UK LOM Core	TREE
LOM number	**LOM name**					
4	Technical	M	O	M	M	O
4.1	Format	M	O	M	O	O
4.2	Size	M	O	M[50]	O	O
4.3	Location	M	O	M	M	O
4.4	Requirement			O	O	

[50] Except for Web pages.

IEEE LOM		ARIADNE	CanCore	HEAL	UK LOM Core	TREE
LOM number	LOM name					
4.4.1	Or Composite				O	
4.4.1.1	Type	M		O	O	
4.4.1.2	Name	M		O	O	
4.4.1.3	Minimum Version	O			O	
4.4.1.4	Maximum Version				O	
4.5	Installation Remarks	O			O	
4.6	Other Platform Requirements	O	O		O	O
4.7	Duration		O	O	O	O

Table 32 presents the elements of the educational category that were included in the TREE application profile as optional or mandatory ones.

Table 32. Use of educational category elements in some application profiles, including TREE

IEEE LOM		ARIADNE	CanCore	HEAL	UK LOM Core	TREE
LOM number	LOM name					
5	Educational	M	O	O	M	M
5.1	Interactivity Type	M		O	O	O
5.2	Learning Resource Type	M	O	O	O	M
5.3	Interactivity Level	O	O	O	O	O
5.4	Semantic Density	O			O	
5.5	Intended End User Role	M	O		O	O
5.6	Context	O	O	O	O	O
5.7	Typical Age Range		O		O	O
5.8	Difficulty	O			O	O
5.9	Typical Learning Time	M	O	O	O	O
5.10	Description				O	O
5.11	Language		O		O	O

Most elements of the TREE application profile in the sixth category are mandatory, in line with the others application profiles included in Table 33.

Table 33. Use of rights category elements in some application profiles, including TREE

IEEE LOM		ARIADNE	CanCore	HEAL	UK LOM Core	TREE
LOM number	LOM name					
6	Rights	M	O	M	M	M
6.1	Cost	M	O		O	M
6.2	Copyright And Other Restrictions	M	O	M	M	M
6.3	Descriptions	M	O		M	O

All elements in the seventh and eighth categories of LOM standard are optional in the TREE application profile (see Table 34 and Table 35).

Table 34. Use of relation category elements in some application profiles, including TREE

IEEE LOM		ARIADNE	CanCore	HEAL	UK LOM Core	TREE
LOM number	**LOM name**					
7	Relation	O	O		O	O
7.1	Kind	O	O	O	O	O
7.2	Resource		O	O	O	O
7.2.1	Identifier		O	M	O	O
7.2.1.1	Catalog		O	M	O	O
7.2.1.2	Entry		O	M	O	O
7.2.2	Description	O		O	O	O

The eighth category allows the provision of comments (8.3 Description) and the statement of who (8.1 Entity) and when (8.2 Date) these comments were provided. Only the ARIADNE profile has mandatory fields in that category.

Table 35. Use of annotation category elements in some application profiles, including TREE

IEEE LOM		ARIADNE	CanCore	HEAL	UK LOM Core	TREE
LOM number	**LOM name**					
8	Annotation	O	O		O	O
8.1	Entity	M	O		O	O
8.2	Date	M	O		O	O
8.3	Description	M	O		O	O

The ninth category does not have any mandatory element in the TREE application profile, as shown in Table 36.

Table 36. Use of classification category elements in some application profiles, including TREE

IEEE LOM		ARIADNE	CanCore	HEAL	UK LOM Core	TREE
LOM number	**LOM name**					
9	Classification	M	O	O	O	O
9.1	Purpose	M	O		O	O
9.2	TaxonPath	M	O	O	O	O
9.2.1	Source	M	O	O	O	O
9.2.2	Taxon	M	O	O	O	O
9.2.2.1	ID		O	O	O	O
9.2.2.2	Entry	M	O	O	O	O
9.3	Description				O	O
9.4	Keyword		O	O	O	O

All mandatory fields in the TREE application profile are also mandatory in ARIADNE, except the field 1.5 Keyword. It was considered highly relevant for a higher education institution the existence of one or more keywords related to each resource. In fact, the tendency, considering the data provided by the ARIADNE foundation, is to increasingly use the keyword field. Figure 41 depicts the usage of some LOM fields[51] from the general category from 2005 to 2010 (9 September 2010). Keyword is the field with the most volume of data. However, it should be noticed that the keyword field may have many values, as some other fields.

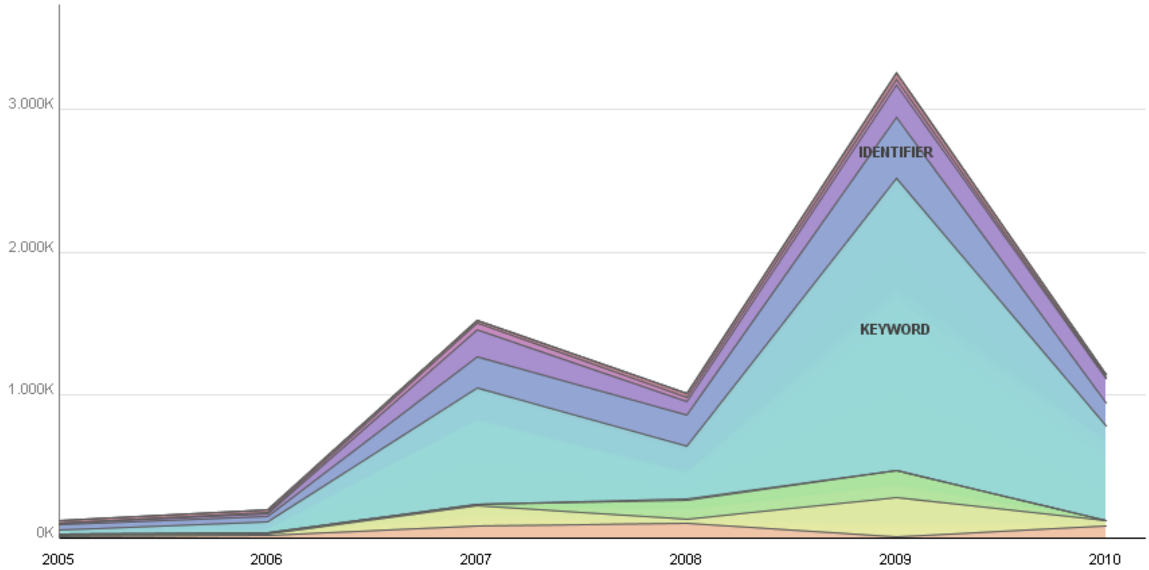


Figure 41. LOM general category – usage of its fields (Source: http://www-958.ibm.com/software/data/cognos/manyeyes/visualizations/lom-metadata-elements-through-time)

In addition, a number of vocabularies are suggested to be used with some field of the IEEE LOM standard. They were also considered for the TREE application profile. For example, the field 5.2 LearningResourceType may have the following value space: exercise, simulation, questionnaire, diagram, figure, graph, index, slide, table, narrative text, exam, experiment, problem statement, self assessment, and lecture, while the field 5.5 Intended End User Role may have the following values: teacher, author, learner, and manager. Thus, during the submission of a resource, the user has to choose from the presented list the most appropriate value, as depicted in Figure 42 and Figure 43.

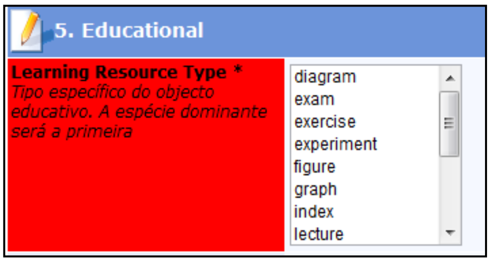


Figure 42. Vocabulary usage in the TREE repository for resource submission (Learning Resource Type)

[51] It was referred at http://www.teleurope.eu/pg/bookmarks/read/80556/many-eyes-lommetadataelement-sthroughtime

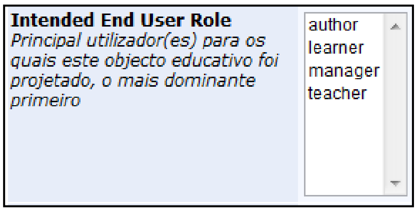


Figure 43. Vocabulary usage in the TREE repository for resource submission (Intended End User Role)

6.5 Conclusions

This chapter started the discussion of the contributions of the PhD work by presenting a model that allows the semantic and pragmatic characterisation of learning resources in a repository. That model was succinctly introduced, as the details were left to the next two chapters dedicated to the semantic and pragmatic components, respectively.

The model was supported by a series of use cases that were presented to each actor (Administrator, Consumer, Ontology engineer, Pedagogical adviser, Provider, and Sentiment analyst). Together they detail the behaviour of a system that follows the proposed model. The general descriptions were provided in this chapter, but Annex 3 complements their specification.

The steps described for the infrastructure preparation can help other people to set up a repository platform and to ensure its proper operation in the fulfilment of their specific requirements, but their main purpose was the statement of the basis for all the developments that followed that initial groundwork.

In addition, this chapter also presented the application profile developed for the TREE repository. Some of the mandatory fields are necessary for the characterisation of the resources as explained in Chapter 7.Semantic Component and Chapter 8.Pragmatic Component.

Chapter 7

SEMANTIC COMPONENT

This chapter presents the component that uses ontologies to encode knowledge to be used in the TREE repository and considers semantic characteristics of learning objects. It proposes a strategy to obtain knowledge to be used in the repository supporting that semantic characterisation and discusses how that knowledge can be applied with benefits to end users – providers and consumers.

7.1 Introduction

The barriers to sharing and reusing learning resources include "poor or inadequate search and discovery tools within the repository - if it cannot be found it cannot be reused or shared" [Philip and Cameron, 2003]. Different people apply different terms to describe the same concepts in a given area. Having a good knowledge source of common concepts and other related to them can help overcoming that problem, improving the resources' characterisation and their findability.

Thus, semantics applied to learning objects should improve their contents' description. The keyword field of the TREE application profile is one of the fields that can be used to characterise learning contents and it is often used in queries submitted by users.

Briefly stated, the two main questions addressed in this chapter are **(1)** how to obtain knowledge to be applied in a repository, and **(2)** how to use that knowledge. The former question is related to the 'Ontology repository' block (see Figure 44), while the latter uses the ontology repository but also information about learning resources.

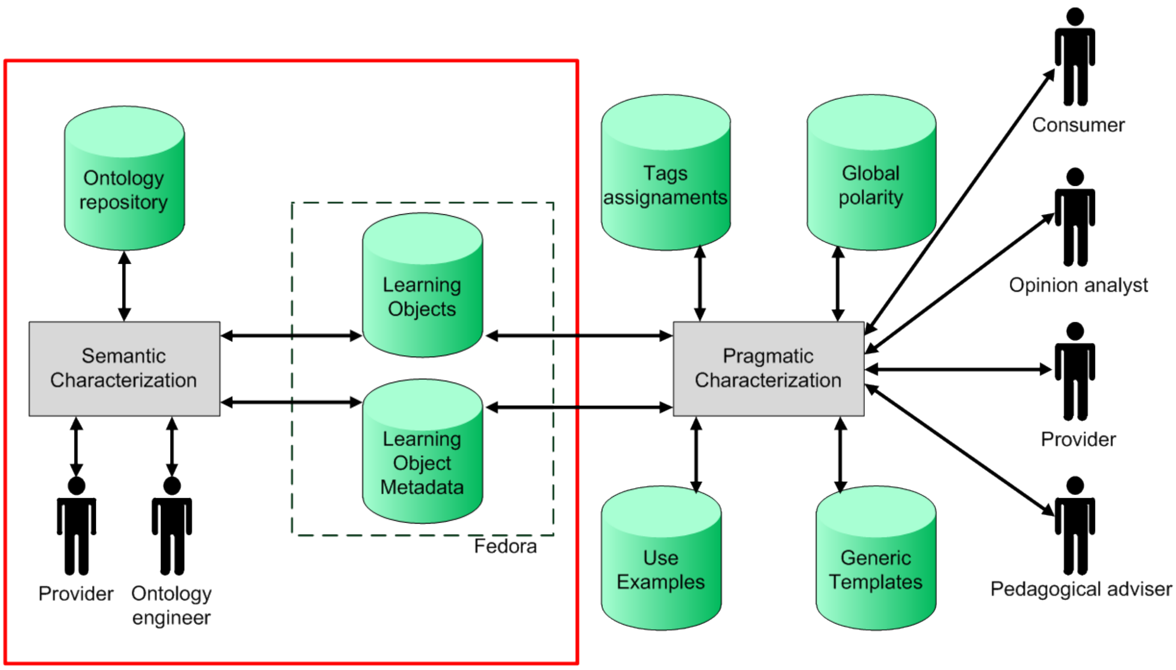


Figure 44. A proposed model for repositories – the semantic component

The use cases related to point **(1)** of the semantic component in the TREE repository (introduced in Chapter 6 and detailed in Annex 3) are as follows:

- List domain areas (UC27);
- Search for online ontologies (UC27);
    - Specify domain concepts (UC28);
    - Analyse returned ontologies (UC29);
        - Visualise classes (UC30);
        - Visualise properties (UC31);
        - Visualise instances (UC32);
        - Consider related terms (UC33);
            - Consider similarity (UC34);
            - Consider synonymies (UC35);

            - Use Online ontologies (UC36);
            - Consider LOD databases (UC37);
        - Consider other topics (UC38);
        - Estimate domain coverage (UC39);
        - Submit ontology (UC40);
- Visualise working properties (UC41);
- Modify working properties (UC42);
- Manage in-use ontologies (UC43);
    - List in-use ontologies (UC44);
    - Delete in-use ontology (UC45);
- Manage temporary ontologies (UC47);
    - List temporary ontologies (UC48);
    - Delete temporary ontology (UC49);
    - Put ontology in use (UC50).

The use cases related to point **(2)** of the semantic component in the TREE repository are as follows:

- Submit a resource (UC63);
    - Submit metadata (UC64),
        - Recognise metadata (UC65);
        - Expand keywords (UC66);
- Search (UC13);
    - Expand keyword (UC21).

Thus, the remaining of this chapter is organised as follows. Section 7.2 (Ontology reuse) deals with point **(1)**, which is materialised in section 7.3 (Ontology reuse in the TREE repository). This section includes a subsection (7.3.1.2) that is not strictly related to ontology reuse, but it is included in section 7.3 because it is used as a complement to the reuse approach. Then, section 7.4 (Ontologies application) addresses point **(2)**. Finally, the last section (7.5 Conclusions) summarises the most important parts of this chapter.

7.2 Ontology reuse

"Knowledge acquisition bottleneck" is a common expression often used to denote the difficulty to obtain knowledge to use in a given application (see, for example, [Cullen and Bryman, 1988; Sestito and Goss, 1994; Fernández-López, Gómez-Pérez et al., 1999; Suraweera, Mitrovic et al., 2005]). Knowledge reuse can be seen as a possibility to surmount that difficulty.

According to Alavi et al., knowledge reuse is the process by which an entity is capable of finding and applying shared knowledge [Alavi and Leidner, 2001]. The reuse of knowledge components probably started to be considered as an important subject when the Knowledge Sharing Effort, sponsored by some American organisations (the Air Force Office of Scientific Research, the Defense Advanced Research Projects Agency, the Corporation for National Research Initiatives, and the National Science Foundation), aiming to support the sharing of knowledge among different systems, suggested the connection of reusable units to build knowledge-based systems in 1991 [Nechs, Fikes et al., 1991].

Markus identified three main players in the process of reusing knowledge [Markus, 2001]:

- **Knowledge producer** - the knowledge creator, who registers explicit knowledge or transforms tacit knowledge into explicit,

- **Knowledge intermediary** - who adapts knowledge for reuse, with many roles, including sharing it,
- **Knowledge consumer** - who recovers the suitable knowledge piece and makes use of it.

The same questions that emerged from knowledge reuse, appeared, to some extent, in the attempts to reuse ontologies. Indeed, reusability is (or should be) an underlying property of ontologies. They are often seen as a way to allow "knowledge sharing and reuse" [Nechs, Fikes et al., 1991], as concepts represented in an ontology become explicit, reliable and convenient to share and, then, to reuse.

Many methodologies for ontology development consider ontology reuse, as described in section 4.5.1 of Chapter 4. Actually, as ontologies have been more and more used in many domains, several ontologies have been made available on the Web, with different scopes and quality. One of the motivations to make ontologies available is that an increase in their use and revision can boost their quality. Also, the applications that use them become (more) interoperable and are provided with a deeper, machine-processable understanding of the underlying domain.

Therefore, ontology reuse has started to be considered, following the reuse trend in knowledge, in general, and even before that, in the software engineering field. In fact, the reuse of ontologies may contribute to reducing costs and time in the same way that the reuse of software [Poulin, 1996] does, since, as stated before, the development of ontologies requires a great effort. However, the evaluation of software components can be done more easily than that of ontologies, in particular analysing the results obtained by their application [Guarino, 2004].

In addition, with the growing demand for the reuse of ontologies, there has been a need for criteria and standards to find out their quality. However, the quality of ontologies is not an easy topic, namely how to evaluate them. The measurements can be classified into two groups (absolute and relative), and can help to determine which ontologies are appropriate to be selected for reuse. Absolute measurements seek errors and conflicts, while the relative ones try to determine if a given ontology fits in a certain application.

Pinto recognises two purposes for reusing ontologies [Pinto and Martins, 2000]:

- **Ontology merging** - By merging different ontologies, another one is built with the combined concerned parts;
- **Ontology integration** - Different ontologies are used to build a new one, but with modifications and extensions.

In addition, Staab et al. detail how ontologies are reused: "'simple' reuse, or by aggregation, combination, condensation, abstraction, and by derivation of new knowledge from aggregations" [Staab, Schnurr et al., 2001].

The more the knowledge is available, the greater the likelihood of being reused. In fact, there are some tools that make easier finding ontologies suitable for some purpose. They play a role in ontology reuse situations, acting as knowledge intermediaries. These tools are:

- **Ontology registry** – It is an application used to register ontologies, maintaining some description fields, statistics about their contents, and a link to each registered ontology. A registry does not provide a central storage for ontologies, but a searchable list of them;
- **Ontology repository** – It maintains a local copy of ontologies, and their different versions, if they exist. For instance, the Protégé OWL library (http://protegewiki.stanford.edu/index.php/Protege_Ontology_Library) is maintained by the Protégé platform developers and BioPortal (http://bioportal.bioontology.org) is an example of an ontology repository for biomedical ontologies. This kind of tools usually only provide browse and download functionalities;
- **Ontology search engine** – It does not require an active action from ontology developers. This kind of tools automatically searches for and indexes the ontologies they discover. Some examples are Swoogle [Ding, Finin et al., 2004], Watson [d'Aquin, Sabou et al., 2007],

Sindice [Oren, R.Delbru et al., 2008] and Falcons[52] (http://iws.seu.edu.cn/services/falcons/objectsearch/index.jsp). They vary in the metadata provided for each ontology, as there is no standard for ontology metadata and exchange.

These kinds of tools facilitate the finding of suitable ontologies, but they differ in the way they describe ontologies, usually without any information about domain of interest, creation date or authorship, for instance.

In a near past the number of available ontologies was very limited [Guarino, 1998], but, currently, there are more than 10,000 online ontologies (http://swoogle.umbc.edu/), which does not mean that finding the suitable ontologies has become much easier. The problem has become to select the best ontology from the many available. Related to this topic is the ranking of ontologies. Search engines usually rank the found ontologies using one or more of the following approaches:

- **Popularity** – Ontologies are ranked in accordance with the number of times they are referred to in other ontologies [Ding, Finin et al., 2004]. That method is followed by Swoogle (http://swoogle.umbc.edu), with some drawbacks as the same ontology published in different sites or different versions of them can be differently scored, although consistent with the idea that the most used in other ontologies should have been considered suitable by many people;
- **Ontological structure** [Yu and Chen, 2009] – By applying metrics that estimate how elaborate is the knowledge structure of an ontology structure, the 'goodness' of an ontology can be estimated. Some of the approaches use the relation between the number of classes and properties [Buitelaar, Eigner et al., 2004], the centrality of a class in the whole hierarchy (Centrality measure [Alani and Brewster, 2006]) or estimate how rich is the definition of a concept (Density measure [Alani and Brewster, 2006]);
- **Concept coverage** – It is based on how well a concept is covered in an ontology. The matches between query terms and the labels in an ontology are regarded, and weighted in accordance with how perfect the matching is, usually just carried out at lexical level.

Moreover, a number of ontology development tools started to incorporate plug-ins that make easier the reuse of knowledge statements, like the Watson plug-in for the NeOn toolkit (http://kmi-web05.open.ac.uk/editor_plugins.html).

7.2.1 Search engines analysis

Ontology search engines accept queries in a format that varies from one tool to another. They usually provide results in an XML file. Their comprehensive designation is Semantic Web Search Engines (SWSE), as they provide Semantic Web documents.

Different from other types of platforms that can be used to find suitable ontologies, which usually only provide browse functionalities, ontology search engines permit another degree of automation.

The great amount of results provided by some SWSEs, which do not have concept or ontology search functionalities, disregard their consideration for an ontology reuse module. For instance, a query on Sindice for ontologies with the term 'Table'[53], returned more than 800,000 results, where near 300,000 of them were RDF files.

From the list previously mentioned, the more ontology-based search engines are Swoogle and Watson, which are described in sections 7.2.1.1 and 7.2.1.2, respectively. They have similarities, but also some differences, which are explored in those sections and also in section 7.2.1.3.

[52] It stands for stands for **F**inding, **A**ligning and **L**earning ontologies, ultimately for **C**apturing knowledge via **ON**tology-driven approaches.

[53] http://sindice.com/search?q=table

7.2.1.1 Swoogle

Swoogle was the first search engine dedicated to online semantic data. Its development was partially supported by DARPA and NSF (National Science Foundation).

The current version of Swoogle is 3.1, which has been available since January 2006. Its prior versions were[54]:

- Swoogle 2.2, launched in July 2005;
- Swoogle 2.1, which was available online in December 2004;
- Swoogle 2.0, launched in August 2004;
- Swoogle 1.0, which was released in May 2004 for demonstrations in international scientific meetings.

Some of Swoogle versions were demonstrated in reputed international conferences, such as version 1.0, presented in the Thirteenth International World Wide Web Conference (WWW 2004), version 2.1, demonstrated in the Third International Semantic Web Conference (ISWC 2004), and version 2.2, presented in the Twentieth National Conference on Artificial Intelligence (AAAI 2005).

Swoogle's architecture (see Figure 45) has four major components:

- **The Discovery component** – It is responsible for collecting candidate URLs. It caches Semantic Web Documents. Swooglebot is the Swoogle's Semantic Web Crawler that produces new candidates to be considered, but conventional search engines are also used for the same purpose. In addition, there is an option to submit sites and documents to be regarded;
- **The Indexing component** – It analyses the Semantic Web Documents (SWDs) found by the Discovery component and generates some metadata, which characterises the features associated with individual SWDs and Semantic Web Terms (SWTs), but also the relations among them. These metadata intend to improve searches;
- **The Analysis component** - It uses the metadata generated by the Indexing component to support ranking mechanisms;
- **The Search Services module** - It allows Swoogle to be used by agents and humans. It is mainly an interface component.

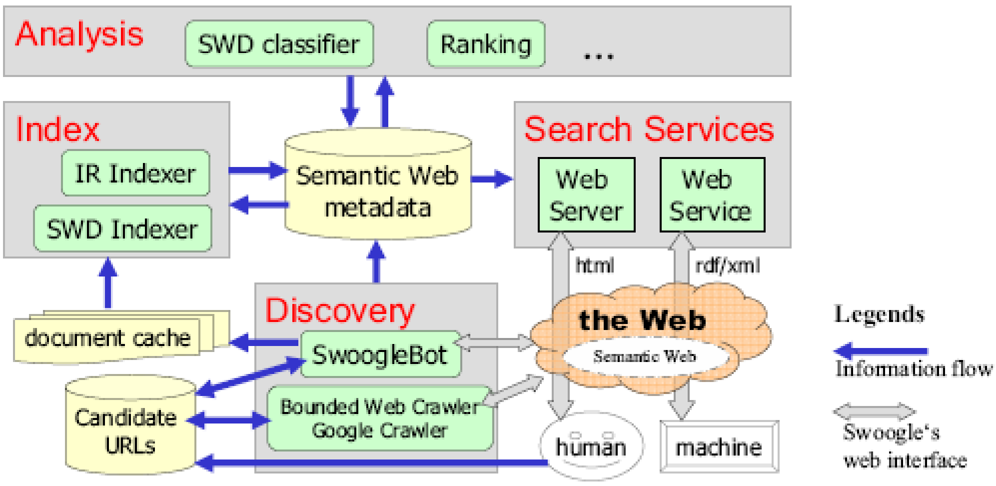


Figure 45. Swoogle 3.1 architecture
(From [Ding, Pan et al., 2005])

[54] http://130.85.88.69/index.php?option=com_swoogle_manual&manual=history

The Swoogle ranking method is based on the OntoRank algorithm, which is quite analogous to the PageRank algorithm (used by the Google search engine). Consider a page A which has n pages (T_1, T_2, ... T_n) with a link to it, as depicted in Figure 46.

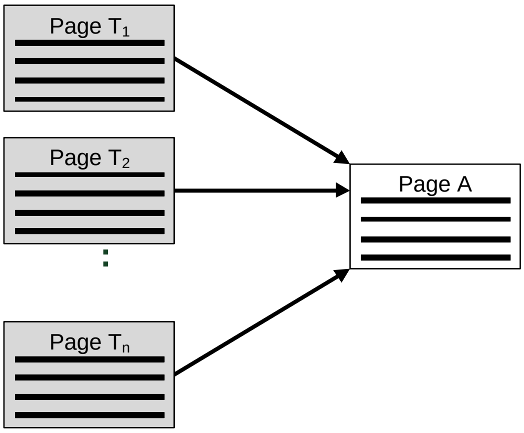


Figure 46. The idea behind PageRank algorithm (Adapted from [Page, Brin et al., 1999])

The PageRank of page A can be stated as follows [Page, Brin et al., 1999]:

$$PR(A) = (1 - d) + d\sum_{i=1}^{n} \frac{PR(T_i)}{C(T_i)}$$

In the above equation d is a normalising factor, whose value can vary from 0 to 1, $C(T_i)$ is defined as the number of links that T_i points to. The PageRank of A ($PR(A)$) considers the PageRank of each T_i ($PR(T_i)$). OntoRank adapts the PageRank approach "to expose more ontologies which are important to Semantic Web users" [Ding, Pan et al., 2005], using semantic relations between ontologies. Ding et al. detail the OntoRank method and compare it with the PageRank algorithm.

It is noteworthy that the keywords specified in a query do not influence the ranking process, just the inclusion of a given document in the results set.

7.2.1.2 Watson

Watson development was partially supported by the NeOn (http://www.neon-project.org) and the OpenKnowledge (http://www.openk.org) projects.

The functions (collecting, analysing and querying) of the core components of Watson (see Figure 47) do not differ significantly from those of Swoogle. These functions correspond to three different layers as follows:

- **The ontology crawling and discovery layer** is responsible for obtaining semantic data. Any document that cannot be parsed by Jena is disregarded as a way to guarantee that only documents that contain semantic data or ontologies are considered;
- **The validation and analysis layer** gathers metadata about the semantic data, which is also used for indexing purposes. In addition, semantic relations between ontologies are regarded for the retrieved ontologies (e.g. `owl:imports`, `rdfs:seeAlso`, namespaces, derefenceable URIs) in order to detect other sources of ontologies;
- **The query and navigation layer** is related to the available query interfaces that allow using the Watson functionalities.

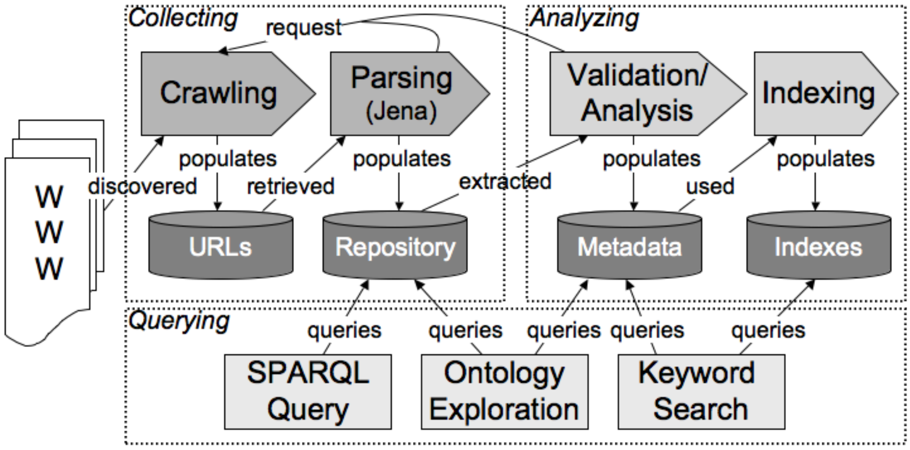


Figure 47. Watson architecture
(From [d'Aquin, Sabou et al., 2007])

Although it is declared in [d'Aquin, Sabou et al., 2007] that Watson uses a specialised crawler for Swoogle, it does not seem that it has been active (see section 7.2.1.3).

In [d'Aquin, Motta et al., 2008] it is mentioned that for ranking it is used "an initial set of measures that evaluate ontology complexity and richness". Also, in [d'Aquin, Sabou et al., 2007] the authors state that "the ranking mechanisms offered by Watson rely on a combination of simple, basic quality measures that are computed in the validation phase and stored along with the ontologies (i.e., structural measures, topic relevance, etc.)". However, the exact ranking method used by Watson is unknown.

7.2.1.3 Comparison and some remarks

The initial idea was to consider both Semantic Web Search Engines (SWSEs) in the ontology reuse module to develop, and that was done in the preliminary developments, but at some point it was impossible to follow through with that strategy due to the lack of uniformity in the search submissions and in the format of returned results, demanding for different strategies. There is not any standard related to queries submission to Semantic Web search engines, and therefore each one accepts queries with different options and syntax. Recently the Workshop on Semantic Repositories for Web (ISWC SERES 2010[55]) put this question on its agenda. This is a major problem that researchers will have to face in a near future. The workshop also discussed the need for interoperability between repositories and SWSE.

Table 37, Table 38 and Table 39 compare Swoogle and Watson under the number of results using query terms from different areas of Engineering.

The submitted queries used the default options, except the following ones:

- Classes and properties were considered;
- Just local names were regarded.

Table 37 shows the number of results for some search strings, comparing the results found by Swoogle and Watson, but also the number of available results considering only the best ten ranked documents in the results set. Swoogle and Watson do not cope with different writing styles. For instance, the results found for 'DataModel' do not include those returned for 'Data_model'.

[55] http://www.ontologydynamics.org/od/index.php/seres2010/

Table 37. Number of results for some Database concepts expressed in different ways

Search string	Swoogle		Watson	
	Number of results	Number of available results (Top Ten)	Number of results	Number of available results (Top Ten)
'Distributed_Database'	0	0	0	0
'DistributedDatabase'	2	2	0	0
'Distributed_Databases'	3	2	0	0
'DistributedDatabases'	0	0	0	0
'Data_model'	13	5	1	0
'DataModel'	11	7	1	1
'DataModels'	0	0	0	0
'Data_models'	3	2	1	1
'Table'	816	6	30	9
'Tables'	77	9	4	1

Table 38 provides the results obtained when some concepts from a Statistics course were considered.

Table 38. Number of results for some Statistics concepts expressed in different ways

Search string	Swoogle		Watson	
	Number of results	Number of available results (Top Ten)	Number of results	Number of available results (Top Ten)
'Sampling'	225	9	5	3
'Samplings'	0	0	0	0
'Probability'	232	6	10	6
'Probabilities'	2	0	0	0
'Linear_regression'	1	1	0	0
'LinearRegression'	10	2	2	0
'LinearRegressions'	0	0	0	0
'Linear_regressions'	0	0	0	0
'Probability_distribution'	1	1	0	0
'ProbabilityDistribution'	1	0	0	0
'ProbabilityDistributions'	0	0	0	0
'Probability_distributions'	0	0	0	0

Table 39 shows the results found using some concepts from a Chemical course.

Table 39. Number of results for some Chemical concepts expressed in different ways

Search string	Swoogle		Watson	
	Number of results	Number of available results (Top Ten)	Number of results	Number of available results (Top Ten)
'Periodic_table'	0	0	1	0
'PeriodicTable'	1	0	0	0
'PeriodicTables'	0	0	0	0
'Periodic_tables'	0	0	0	0
'solution'	400	6	13	9
solutions'	34	6	5	1
'Acid'	621	7	31	5
'Acids'	50	9	3	1
'Base'	1,625	4	27	6
'Bases'	48	5	3	1

From the experiments here documented and from others carried out, it was found that:

- Search strings that can be considered very generic, such as 'Base', 'Table' or 'Solution' return many results. However, a great number of those results are not really for the envisaged area. For example, the results returned for the search string 'Base', included ontologies with classes for baseball, database, and space subjects, among others;

- Although the common conventions of using the singular form in concept names and the CamelCase style to write compound words or phrases, followed by the W3C itself, these are not universally followed. The use of separator (underscore or no character in accordance with CamelCase naming convention) and singular or plural nominal word form in the submission of queries to SWSEs leads to different sets of results, which are not enclosed in the others;
- It was impossible to analyse all the returned results, but generally there is not an overlap in the top ten results provided by Watson and Swoogle. It can be a result from different ranking methods, but for some search strings, Watson provided no results, while Swoogle returned.

Although Watson has some version control mechanisms and it is "able to detect some form of duplication of ontologies" [d'Aquin, Baldassarre et al., 200], the same version of a given ontology can be returned by Watson, or even different versions of the same ontology. For instance, for the search string 'Base' the results returned by Watson are shown in Figure 48. The second and the third documents were analysed and they are exactly the same.

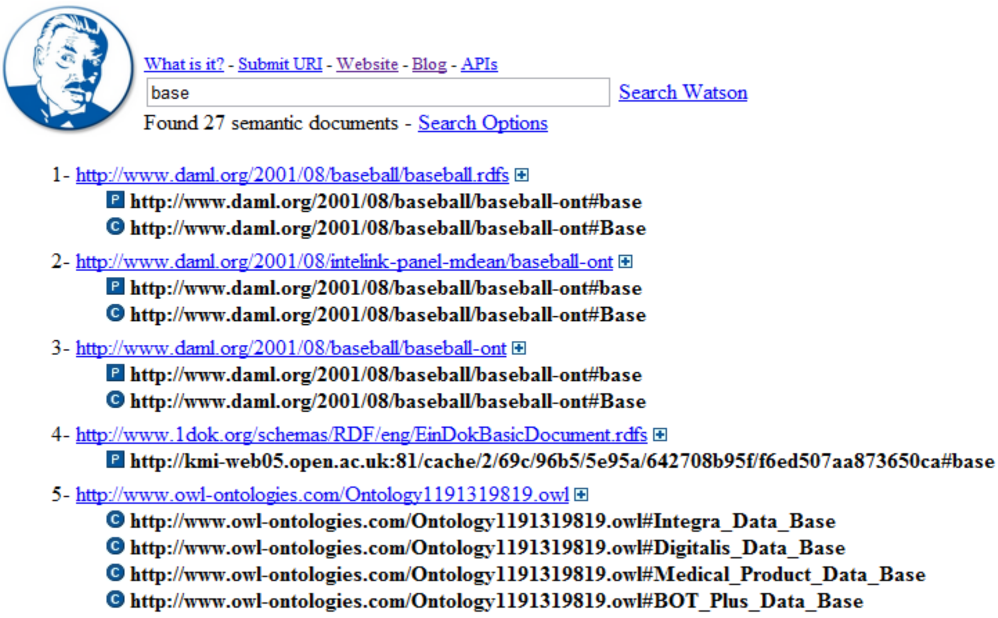


Figure 48. Top five results found by Watson using 'base' as search string

The number of results is not the only criterion to be considered, but it is important as it should be easier to find appropriate ontologies in a large set. However, the results were analysed to determine by sampling if the top ten results from Watson were more appropriated to be integrated in a repository than those provided by Swoogle. For instance, one of the results provided by Swoogle using 'Distributed_Databases' as search string is the computing ontology [Kamali, Cassel et al., 2004], which will be addressed later in this chapter. It is a reputed and documented ontology, which is returned by Swoogle but not by Watson. Regarding the domain of interest, generally it was found that the ontologies provided by Swoogle were more suitable for usage in the TREE repository.

Some other aspects that were studied were:

- The existence of a limit number of queries accepted;
- The existence of multiple options to sort the results;
- The metadata provided by each returned ontology;

- The possibility of specifying many terms, all to be considered (use of logical operator AND);
- The possibility of specifying many terms to be considered alternatively (use of logical operator OR);
- The ability to dynamically discover semantic data depends on available APIs to access the semantic resources collected by Semantic Web search engines.

These points are summarised in Table 40.

Table 40. Swoogle and Watson – a comparison

Search string	Swoogle	Watson
Unlimited number of queries	No	Yes
Multiple possibilities to sort the results	Yes	No
Provision of rich ontology metadata	Yes	No
Use of OR to specify possible terms	Yes	No
Use of AND to specify all terms	Yes	Yes
Available APIs	Yes	Yes

Some common problems were detected:

- In the results set there was a number of ontologies that were no longer available. For instance, one of the three returned ontologies for the query 'distributed_databases' has been unavailable (the one whose URI is http://what.csc.villanova.edu/twiki/pub/Main/OWLFileInformation/ComputingOntology-George2.rdf-xml.owl) for more than two years. In that case it is known that this ontology has a newer version (whose URI is http://what.csc.villanova.edu/twiki/pub/Main/OWLFileInformation/28Jul09.owl). Thus, it does not seem that Swoogle has any version control mechanism. However, as stated before, Watson suffers from the same problem;
- Other point to be improved in SWSEs is the use of wildcards, not their acceptance but how they are treated and the results of their usage. For instance, submitting a query string to Swoogle like "data*model" provides the same results as a query string like 'data_model', and they not include the results obtained with a query string like 'datamodel'. Watson has a similar problem. A query submitted to Watson specifying 'data*model' returns the same results returned by a query like "datamodel", not including the results returned by a query like "data_model"[56].

To sum up, the number of ontologies provided by Swoogle and their relatedness to the domain of interest, and the possibility to use the OR logical operator in the search string submitted for Swoogle were crucial aspects to the decision of using it in the ontology reuse module.

7.3 Ontology reuse in the TREE repository

Figure 49 generically delineates the architecture of the ontology reuse module that was developed. It takes two inputs: a list of domain concepts and the online ontologies found through semantic search engines using the given concepts.

[56] All Watson default options were kept for that query.

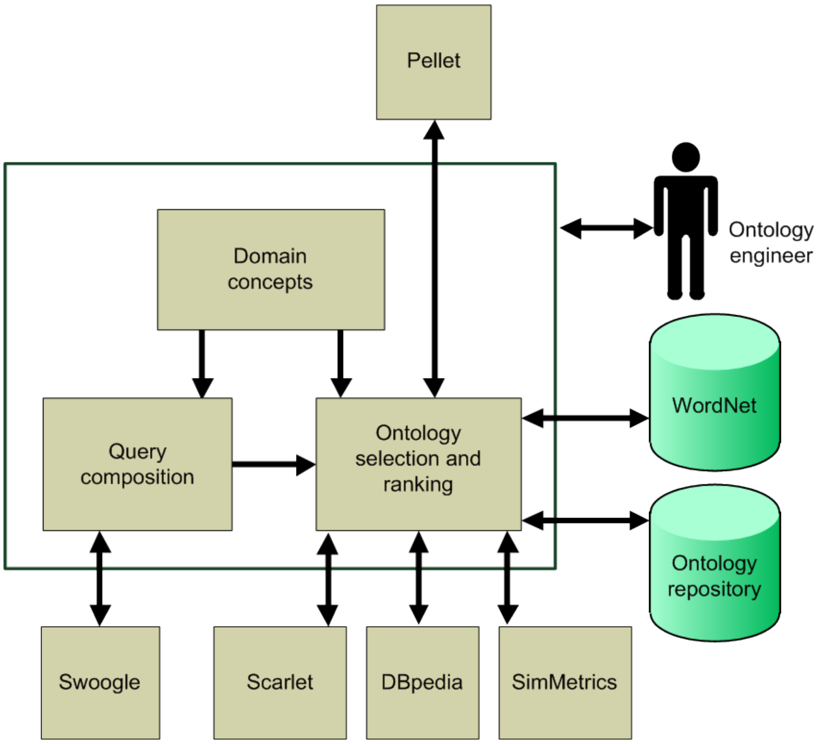


Figure 49. Architecture of the semi-automated ontology reuse module (Adapted from [Azevedo, Seiça et al., 2010])

Among many tools that facilitate the discovery of ontologies, Swoogle was selected, as explained in the previous section.

At ISEP each course has a description in English, which is mandatory, with some predefined fields to be supplied; one of them is related to the course contents (see Figure 50).

Course Contents

I - Probability
II - Probability distributions
III - Sampling
IV - Estimation of parameters
V - Testing hypotheses
VI - Linear regression

Figure 50.A course description extract

It is simple to obtain the relevant concepts for each course considering that specific field. Initially, the ontology engineer had the option to extract concepts from the course descriptions or even specify others and disregard or correct some of the obtained ones (see Figure 51). That enumeration of the important terms corresponds to one of the recommended steps to follow when developing an ontology [Noy and McGuinness, 2001].

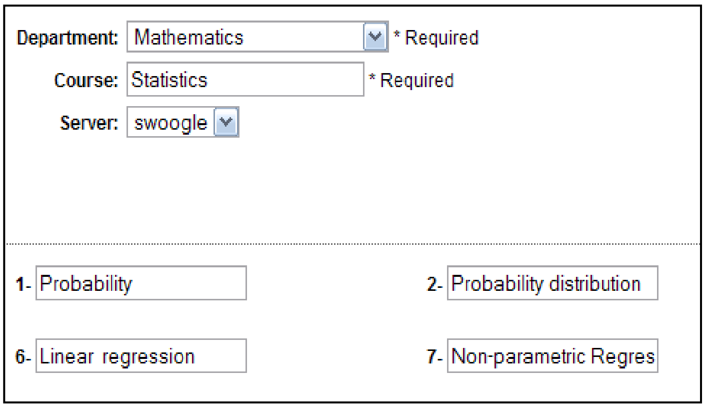


Figure 51. Specifying concepts to be regarded – an extract

However, for two different reasons that approach was disregarded:

- Firstly, the document format for course description was changed. It was not impossible to consider the new format, but some modifications would be necessary for that purpose;
- Secondly, it was realised that many times the course description was too general, but the concepts provided directly by the responsibles for the courses or by teachers from those courses were more precise. Mainly four courses were involved in that analysis: database, applications engineering, software engineering and laboratory, all from Informatics Engineering.

Thus, mainly for the second reason, it was decided to have the concepts provided by humans. Figure 52 shows how they can be introduced in the system.

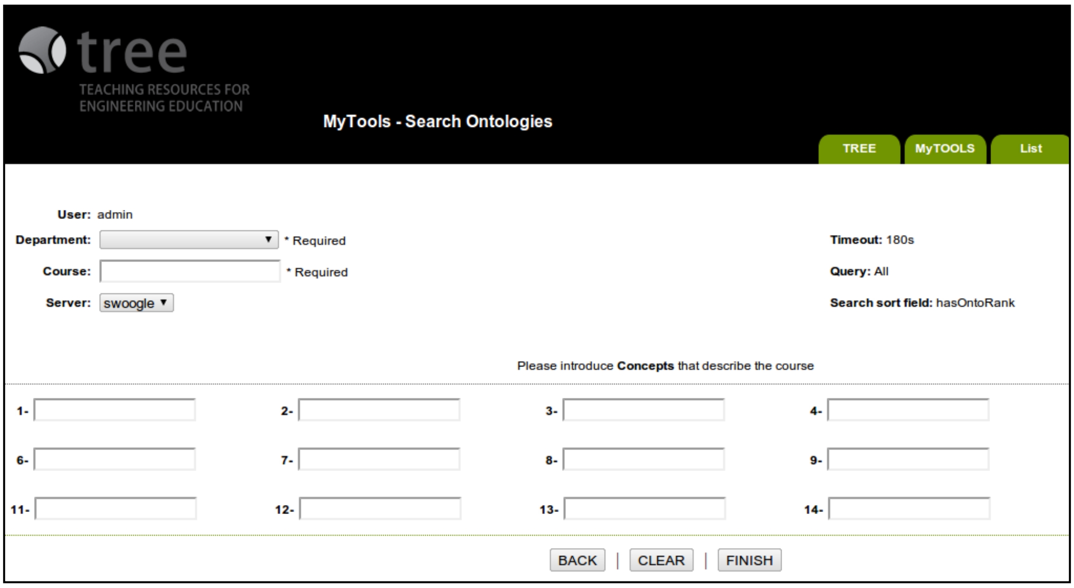


Figure 52. Searching for ontologies in the TREE repository

Having C_1, C_2, … C_N as concepts for a course, queries are tried with all the concepts, after disregarding those not covered by any online ontology. Thus, the initial query might have N terms, and then the system tries N-1 terms, and so on. Initially, having all concepts covered (by at least a predefined number of ontologies) or having tried each concept alone, the process was finished. However, later it was decided to continue the process until its end, as it is possible to have an ontology with concepts C_1 and C_2 that is more interesting than an ontology with concepts C_1, C_2 and C_3, for instance.

However, often ontologies with only one single concept are not related to the domain of interest. Also a query using only one search string might result in excessive results to be analysed. For instance, a query with 'table' as search string provides 816 results (as can be seen in Table 37).

The use of a separator (underscore separator or none in accordance with CamelCase naming convention) and singular or plural nominal word forms lead to different sets of results, which are not completely or partially enclosed in the others, as previously explained. Thus, for each concept it was decided to consider also some lexical variations. If C_i is represented by a set of terms Ti, the algorithm followed to submit queries to Swoogle can be expressed as indicated in Figure 53, considering four concepts C_1, C_2, C_3, and C_4, represented, respectively, by the sets of terms T1, T2, T3 and T4.

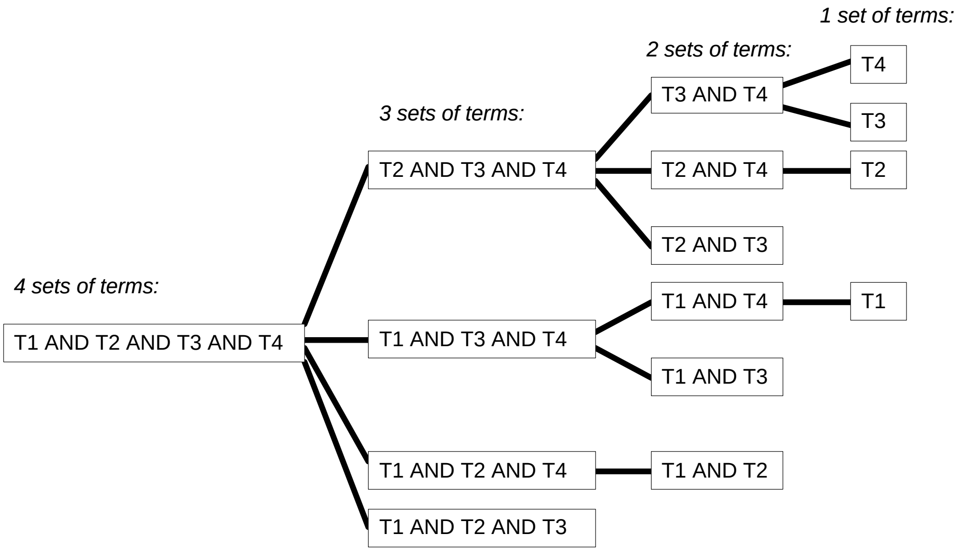


Figure 53. Algorithm used to submit queries to Swoogle

For instance, if C_1 is 'data model' then T1 is expressed by 'data_model OR datamodel OR data_models OR datamodels'. The use of upper case or not has no relevance for the queries in Swoogle.

In addition, using the same four concepts, C_1, C_2, C_3 and C_4, the top ontologies in the results set are those with the results considered all four sets of terms (T1 AND T2 AND T3 AND T4), and then those returned using three sets of terms, and so on. The order of each ontology obtained using a certain number of sets of terms is the same used by Swoogle, which in fact is controlled by a configuration variable.

There are some configuration values that affect the overall functioning of the system:

a) **The Swoogle key** – Initially the Swoogle key used was the one available to everyone (`demo`). However, that could lead to the limit of queries being reached before expected. Later a Swoogle key was asked for the Swoogle developers' team, which is the one presently in use. Although the ontology reuse module was developed to be used in the TREE repository, that variable can be useful if that approach is extended in the future. Also, changing this value, if the daily limit is reached, can allow other queries to be still accepted;

b) **Query definitions** – There are two ways to submit queries, using 'def', such as 'def:data_model', which led to the consideration of classes and properties, or everything otherwise. Sometimes it can be useful to extend the approach if the number of results is small;

c) **The sort field to be considered** – The default as explained before is OntoRank. But using Swoogle there are other possibilities to be used with Swoogle:

   - The last modification date;

- The document size;
- The number of triples;
- The percent of triples contributing to ontological definition (OntoRatio);
- The number of distinct terms;
- The number of distinct class-instance;

d) **Swoogle timeout** – The maximum number of seconds to be considered when waiting for Swoogle response. Currently that value is set to 180 seconds. When that time amount is reached, the processing is stopped;

e) **Pellet timeout** - The number of seconds to be considered when waiting for Pellet response. Presently that value is set to 600 seconds. When that time period is attained, the processing is stopped;

f) **Scarlet timeout** - The maximum number of seconds to be considered when waiting for Scarlet response. Currently that value is set to 600 seconds. Once that time amount is reached, the processing is stopped;

g) **Maximum number of ontologies to be automatically analysed** – That value controls the number of ontologies that are processed after they are returned by Swoogle;

h) **The ontological objects considered for domain coverage analysis**: class, individuals, and/or properties;

i) **Similarity threshold** – The similarity value (0-1) from which it is considered that a similar term found in an ontology has correspondence to a concept that was provided;

j) **Synonyms reduction** – When a synonym is considered to be used and it appears in some way in a given ontology, that ontology can be considered to totally cover that concept (no synonym reduction) or it can be specified a value in order to make that correspondence to be partially considered.

For the ontology retrieval, labels of instances, classes and properties in ontologies may be considered as lexically matching the query terms (configuration value (h)). As some other relevant concepts could remain unnoticed, some related projects try a more conceptual matching [Cantador, Fernández et al., 2007] [Brewster, Alani et al., 2004] and expand the initial concepts using WordNet [Miller, 1995], an electronic lexical database created by the Cognitive Science Laboratory of Princeton University. However, that approach was found more useful when there are already some ontologies to be analysed, as it will be later explained.

There are three possibilities when trying to find adequate ontologies:

- One or more ontologies are found with all the desired concepts;
- The system returns different ontologies with most of the concepts, but the set of ontologies covers all of terms;
- It is impossible to find an ontology with one or more of the concepts. When this happens, the Ontology engineer can repeat the process choosing other terms for the domain under analysis, but having the results of all steps: it is possible to choose synonyms provided by Wordnet, considering only those with strong relatedness to other concepts as indicated by Scarlet or DBpedia, as is later explained in this chapter. It never happened to have no ontologies found for the concepts introduced, except when Swoogle is offline (see, for instance, Figure 54 – the top two search results).

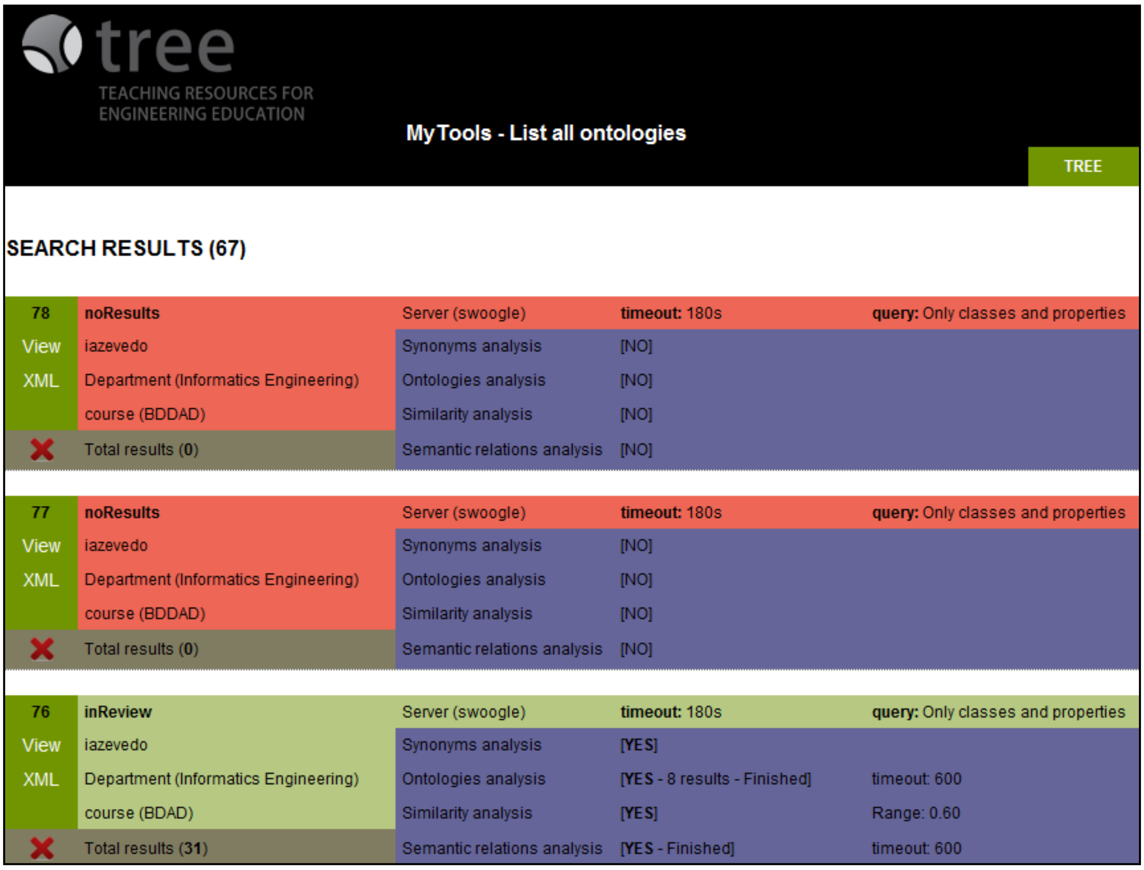


Figure 54. Unavailability of Swoogle – an example

Before carrying out the process of ontology selection, the ontologies whose URIs correspond to dead links are disregarded and the others are checked to verify if they are syntactically correct. Pellet 2.2.2 [Parsia and Sirin, 2004], an OWL DL Reasoner, is used for that purpose. Any ontology that cannot be basically processed by Pellet is disregarded.

7.3.1 Ontology ranking and selection

Once a system is reusing ontologies, their quality might affect the quality of the resulting ontologies and the desirable functioning of the system. Deficiencies in modelling may remain unknown for a long time until unpredictable or poor query results catch the attention of someone. With evaluation methods to estimate the quality of ontologies and rate them, the likelihood of that problem can be reduced, but not entirely eliminated.

However, evaluating a single ontology is different from evaluating a set of ontologies to decide on the one(s) to reuse, as an intensive evaluation of all candidates is not feasible.

Although more specific for the evaluation of a single ontology at a time, some of the more automatic approaches are:

- Use a certain ontology and realise its performance in a real environment with interest, allowing a functional evaluation. When evaluating a number of ontologies, the best is the one that provides the optimal fulfilment for an application. However, this technique is not often a feasible way to test and evaluate ontologies;
- Use a Gold Standard ontology. The most similar to a given ontology is considered the most suitable one and should be adopted. Some drawbacks of this approach are the lack of availability of the optimum ontology to compare with others, and the comparison itself, namely how to perform it;

- Realise the coverage of the domain concepts. The ontology that includes most of the concepts is the one to be used, but that analysis is usually based on matching at lexical level, as stated before.

The discovered ontologies are evaluated and ranked considering three different aspects, each one leading to a different number. These aspects are: their concepts coverage (considering the distance between the provided terms and those in the ontology), popularity (ontoRank value provided by Swoogle) and ontological structure (the fraction of classes and properties in a ontology, considered all triples – the ratio value visible in Figure 55).

During the concepts coverage analysis, some extraneous matching can occur between the terms provided and those in the ontology when carried out at lexical level, but applying a similarity metric that situation can be avoided to a certain extent, but not completely.

Levenshtein distance is the minimum number of character-level operations (insertion, deletion, or substitution) required to convert one string into the other [Levenshtein, 1965]. After computing the Levenshtein metric to estimate the distance between strings, only the terms that match some class, property or instance name (according to the configuration variable that defines what to consider) greater than a defined threshold, are regarded. For instance, that approach is able to keep the matching between 'relational_model' and 'relational_data_model' (similarity score near 0.76), and exclude the matching between 'Oracle' and 'Lawrence_Oracle_Ellison' (similarity score close to 0.26), depending on the similarity threshold.

SimMetrics-1.6.2 is used to compute the Levenshtein distance. SimMetrics (http://sourceforge.net/projects/simmetrics/) is an open source library of algorithms for calculating string measures of similarity between two given strings. It was developed by the UK Sheffield University.

Figure 55 shows the two ontologies that were returned as covering eight concepts (data*fragmentation, SQL, database, data*replication, distributed*database, table, normal*form, trigger). The second result will be used to illustrate the use of the Levenshtein metric (the first one was not available).

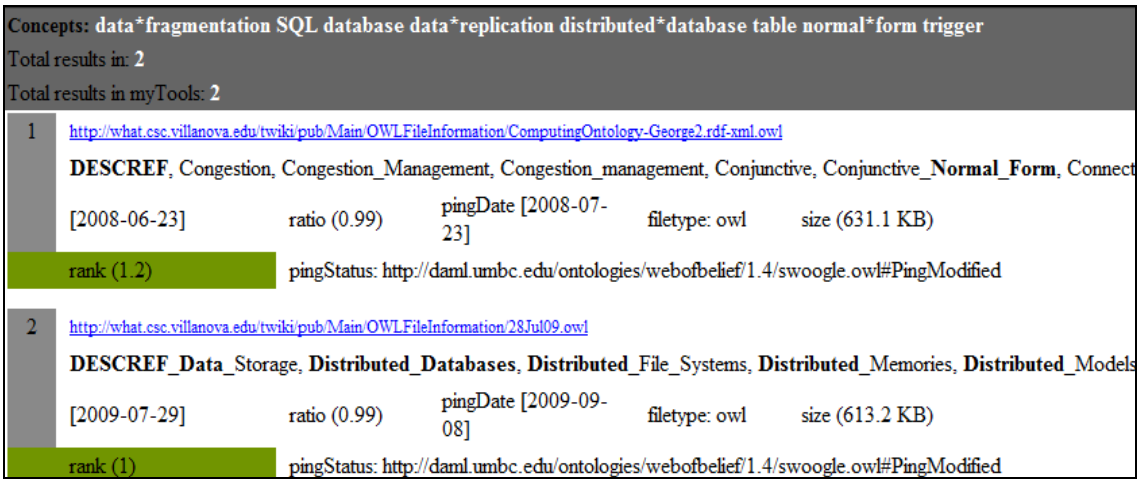


Figure 55. Ontologies found for 8 of 11 concepts specified

Thus, for the second ontology returned[57], the similarity values were computed. The column 'SiMM Concepts' in Figure 56 shows the values calculated when they are greater than the threshold value.

[57] http://what.csc.villanova.edu/twiki/pub/Main/OWLFileInformation/28Jul09.owl

	SiMM Concepts	Syn
data*fragmentation	1	
data*modelling	0.69	
SQL	1	
database	0.67	
data*replication	1	
distributed*database	0.95	
relational*model	0.79	
object*oriented*model	0.79	
table	[No SiMMClass]	board. tabular array. mesa. tabulate. tabularize. tabularise. put over. put off. defer. shelve. hold over. prorog back. postpone. remit.
normal*form	[No SiMMClass]	**normal** shape. **normal** manakin. **normal** mannequin. **normal** manikin. **normal** mannikin. **normal** cast. **normal** co contour. **normal** conformation. **normal** chassis. **normal** frame. **normal** figure. **normal** physique. **normal** build. **n** physical body. **normal** material body. **normal** soma. **normal** bod. **normal** flesh. **normal** anatomy. **normal** sort. kind. **normal** pattern. **normal** descriptor. **normal** word form. **normal** signifier. **normal** strain. **normal** variant. **no** course. **normal** grade. **normal** class. **normal** phase. **normal** forge. **normal** mold. **normal** mould. **normal** work. organize. **normal** organise. **normal** constitute. **normal** make. **normal** take form. **normal** take shape. **normal** spri
trigger	[No SiMMClass]	induction. initiation. gun trigger. activate. set off. actuate. spark off. trigger off. touch off. spark. trip.
	6.89	

Figure 56. Exemplifying the use of the Levenshtein metric – an extract

It is possible to analyse those values in more detail in an XML that logs the calculations (see Example 15).


```xml
<?xml version="1.0" encoding="iso-8859-1"?>
<simmetrics>
  <simm
search="http://what.csc.villanova.edu/twiki/pub/Main/OWLFileInformation/28Jul0
9.owl" id_onto="2" file_numb="2-2">
    <simm_word numb="1126">data_fragmentation - Data_Fragmentation -
1.0</simm_word>
    <simm_word numb="2841">data_fragmentation - Data_Translation -
0.64705884</simm_word>
    <simm_word numb="3539">data_modelling - Data_Cleaning -
0.6153846</simm_word>
    <simm_word numb="3576">data_modelling - Data_Model -
0.6923077</simm_word>
    <simm_word numb="3622">data_modelling - Data_Mining -
0.6153846</simm_word>
    <simm_word numb="4559">data_modelling - Domain_Modeling -
0.64285713</simm_word>
    <simm_word numb="5812">data_modelling - Modeling -  0.6153846</simm_word>
    <simm_word numb="5979">data_modelling - User_Modeling -
0.6153846</simm_word>
    <simm_word numb="7535">SQL - SQL -  1.0</simm_word>
    <simm_word numb="12182">database - XML_Databases -  0.6666666</simm_word>
    <simm_word numb="12287">data_replication - Data_Allocation -
0.73333335</simm_word>
    <simm_word numb="13378">data_replication - Data_Replication -
1.0</simm_word>
    <simm_word numb="14209">data_replication - Data_Visualization -
0.64705884</simm_word>
...
  </simm>
</simmetrics>
```


Example 15. An XML file with the similarity values

For any given ontology, its domain coverage is estimated with respect to the concepts initially provided, considering the sum of similarity values between ontology terms and those concepts that were specified. For the example represented in Figure 56 that value was computed as 6.89 (the sum of similarity values). When there is more than one correspondence for a certain concept, the highest similarity value is regarded.

The similarity processing is initiated by the ontology engineer for each ontology. Many times the simple metadata provided by Swoogle for a given ontology is able to discourage any further processing. For instance, ontology 21 (see Figure 57) has the term 'Table' and its description includes billiard, billiard-table, bingo, among other terms. It does not seem to be related to the database domain. Anyway, the ontology engineer can choose to proceed to the similarity analysis. A domain expert might be helpful in that phase.

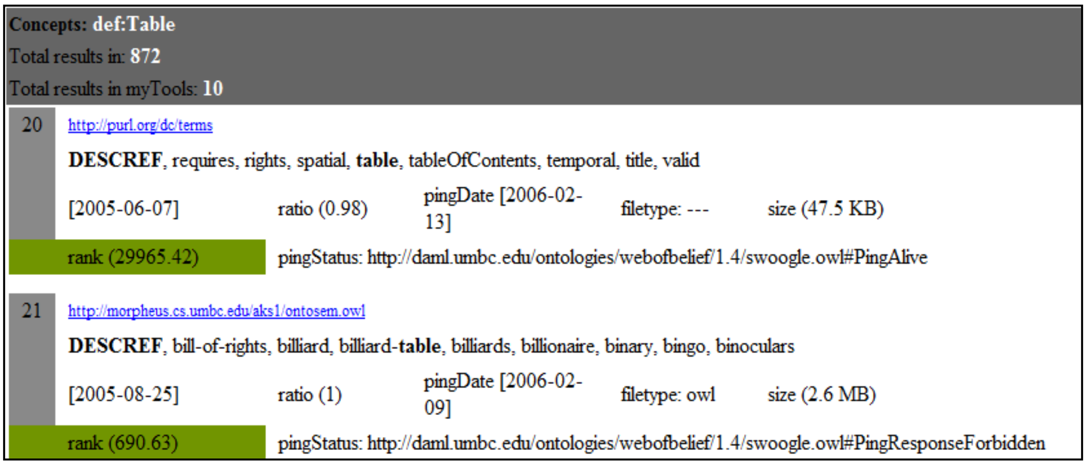


Figure 57. Partial view – ontology metadata

However, if a certain concept is not covered by a specific ontology, some additional steps are available, which are explained in section 7.3.1.1. These steps try other terms semantically related to the concepts initially provided, using synonyms provided by Wordnet (see column Syn in Figure 56), and for these alternative terms the ontology is reanalysed, and the similarity between them and the ontology terms are calculated.

Once a list of possible appropriate ontologies is ranked, an ontology engineer analyses it and manually selects the most suitable ones. In this phase some other criteria are also useful, such as:

- **The existence of available documentation** – This criterion serves useful purposes:
  - It can help understanding why the ontology was developed. An ontology developed as part of a course work may not be as useful as one developed under a project;
  - The documentation can include information about previous evaluations and their results;
- **The ontology developers** – An ontology developed by a reputed team may be of more interest than one developed by someone with no references in the area or not integrated in an organisation that has been working in the area of ontology engineering for some time.

Occasionally the assistance of an expert from the ontology area, namely when two or more ontologies have very similar ranking values, is also important. In that case, competency questions might be considered to decide on the most appropriate ontology. They have also proved valuable to detect missing parts of knowledge. For example, considering the database domain, an ontology might have to address the following competency questions:

- Which are the Normal Forms?
- Which are the subsets of SQL statements?

A final ontology can be obtained merging one or more ontologies (or extracts), or the ontologies of interest can be reused exactly as they are, and being then assigned to a community of the TREE repository, and uploaded to the ontology repository. The repository uses Sesame (http://www.openrdf.org/), an open source Java framework for querying and storing RDF data. Its development started under a European research project (On-To-Knowledge). The version in use is Sesame 2.3.2.

Alternatively, the ontology engineer can choose to eliminate what is not of interest in an ontology. The expression 'ontology winnowing' [Alani, Harris et al., 2005] is used to identify the process of removing what is not necessary and maintaining what is adequate. For instance, let us consider the ontology whose part is represented in Figure 58.

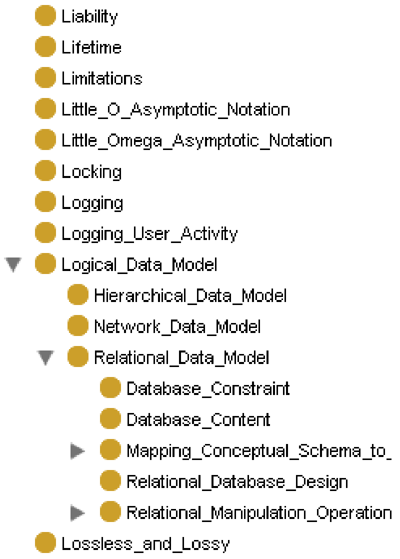


Figure 58. Ontology winnowing – the original ontology

If, for instance, the class Logical_Data_Model is the only desired, and also its super-class and the sub-classes up to 2 levels, the part visible in Figure 59 is then regarded.

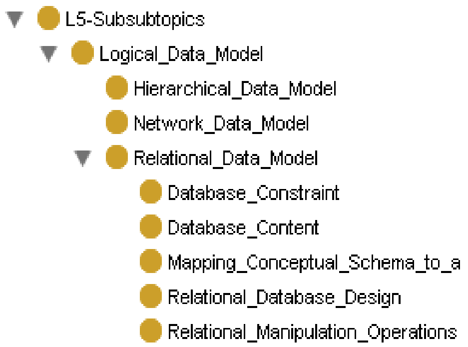


Figure 59. Ontology winnowing – the resulted ontology

7.3.1.1 Computation of semantic relatedness

Semantic similarity can be defined as a continuous variable that expresses the level of synonymy between two terms [Miller and Charles, 1991]. Later, a distinction between semantic similarity and semantic relatedness was introduced [Resnik, 1999]. Semantic relatedness provides a degree to which some terms are related, regarding any possible semantic relationship, and not only hyponymy/hypernymy relations.

In an ontology reuse module it does not make sense to use only the exact terms that are supplied as characterising the concepts of interest. Thus, for each concept it is necessary to consider other synonyms. However, not every synonym can be considered. For example, table in Information Systems is not a piece of furniture and in that field such a synonym must be disregarded.

WordNet is used to provide some synonyms. WordNet contains nouns, verbs, adjectives and adverbs. Words that are synonymous are grouped together in synsets - synonym sets. Each synset has an associated short definition named gloss. Polysemous words in WordNet are included in more than one synset, one for each meaning. For instance, the concept 'data modelling', which appears in the database course description, is not contained in any online ontology. Considering sister terms there are many possible substitutions for 'modelling' from different synsets, such as 'modeling', 'molding', 'moulding', 'model', 'pattern', and so on. Hypernyms (words from ancestor synsets) and hyponyms (words from descendant synsets) are not considered.

The computation of semantic relatedness between possible synonyms of the word table and others database concepts can provide a rather accurate indication of which synonym(s) to consider in a specific area, as some of them may not be appropriate.

Figure 60 exemplifies the approach. There are N terms (C_1, C_2, ..., C_N). Each of them has a number of synonyms. For instance, C1 has the following P synonyms: CS_{1_1}, CS_{1_2}, ..., CS_{1_P}. It is necessary to determine which of them are appropriate. This is achieved by determining if CS_{1_1} is related with C_2, C_3,..., C_N. The same is carried out for CS_{1_2},..., CS_{1_P}. The synonyms with more semantic relatedness discovered may be the more adequate to be used in the domain represented by C_1, C_2, ..., C_N. The same process is repeated for C_2, C_3,..., C_N.

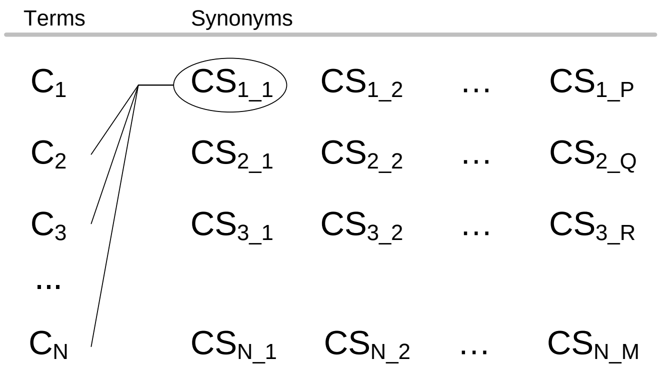


Figure 60. Semantic relatedness - comparing terms and synonyms

Two different strategies are used to find which synonyms provided by Wordnet can be considered as alternative to the initial provided concepts. The semantic relatedness discovered by using online ontologies is used to recheck the ontologies for which the initial concept was not found (considering the similarity threshold). The results encountered by the use of an RDF repository are used only to highlight possible synonyms to be considered if additional searches in the same domain are felt necessary. These strategies are described in the next subsections.

Use of online ontologies

The Scarlet (SemantiC relAtion discoveRy by harvesting onLinE onTologies) system was developed at the Open University. It is available at http://scarlet.open.ac.uk/, including a demo and documentation. It uses online ontologies as context for matching ontologies, its primary envisaged use. However, in a wider scope, Scarlet is "a technique for discovering relations between two concepts by harvesting the Semantic Web" [Sabou, d'Aquin et al., 2008], and can be used for different purposes.

Scarlet java API version 1.0 is used to verify if there are some relations between the probable substitutes supplied by WordNet and the other terms provided, choosing then the one with most relations found.

All types of named relations are regarded. Also, more than one ontology are considered in that approach and an inheritance depth equal to 5.

```
**************
Relation: database--R_FROM-http://www.w3.org/2000/10/swap/db#table-->table
In ontology: http://dev.w3.org/cvsweb/2000/10/swap/db.rdf?rev=1.2  because:
Entire path:
http://www.w3.org/2000/10/swap/db#Database--R_FROM-
http://www.w3.org/2000/10/swa
p/db#table-->http://www.w3.org/2000/10/swap/db#Table
Individual Relations in Path:
http://www.w3.org/2000/10/swap/db#Database--R_FROM-
http://www.w3.org/2000/10/swap/db#table--
>http://www.w3.org/2000/10/swap/db#Table

**************
Relation: database--R_FROM-http://www.w3.org/2000/10/swap/db#table-->table
In ontology: http://www.w3.org/2000/10/swap/db.rdf  because:
Entire path:
http://www.w3.org/2000/10/swap/db#Database--R_FROM-
http://www.w3.org/2000/10/swa
p/db#table-->http://www.w3.org/2000/10/swap/db#Table
Individual Relations in Path:
http://www.w3.org/2000/10/swap/db#Database--R_FROM-
http://www.w3.org/2000/10/swa
p/db#table-->http://www.w3.org/2000/10/swap/db#Table
```

Example 16. Scarlet output – relations between database and table

As can be observed, the same ontology or different versions of the same ontology can be regarded by Scarlet. Currently, it is only considered if a relation was discovered, and the number of ontologies is not used for any purpose.

However, using Scarlet an additional strategy was also considered necessary to compute semantic relatedness. As Scarlet considers upper level ontologies, it was necessary to disregard them in the process described in this section. Otherwise, the fact that two concepts are represented in the same ontology, for instance, could have little or no significance for the reason that led to the use of Scarlet.

Use of DBpedia predicates

The DBpedia project aims to extract structured data from Wikipedia and make them available in the DBpedia Knowledge Base [Bizer, Lehmann et al., 2009]. A DBpedia resource is described by a number of properties. The list of all properties is available at http://dbpedia.org/Datasets/Properties.

The use of some DBpedia properties can help disambiguating the best synonym(s) of a given word to use in a certain domain, through the estimation of the existence of semantic relatedness between terms. Previous works used Wikipedia as knowledge source for computing semantic relatedness, with interesting results reported [Strube and Ponzetto, 2006] [Gabrilovich and Markovitch, 2007]. The advantage of using DBpedia is that the information is already structured in a format easy to manipulate.

Due to its size, the text in Figure 61 is not readable, but its purpose is to show the current relevance of DBpedia, represented by the black circle in the centre. It is connected to many other repositories. The DBpedia data set currently describes 3.64 million 'things'[58].

[58] Information available at http://wiki.dbpedia.org/Datasets.

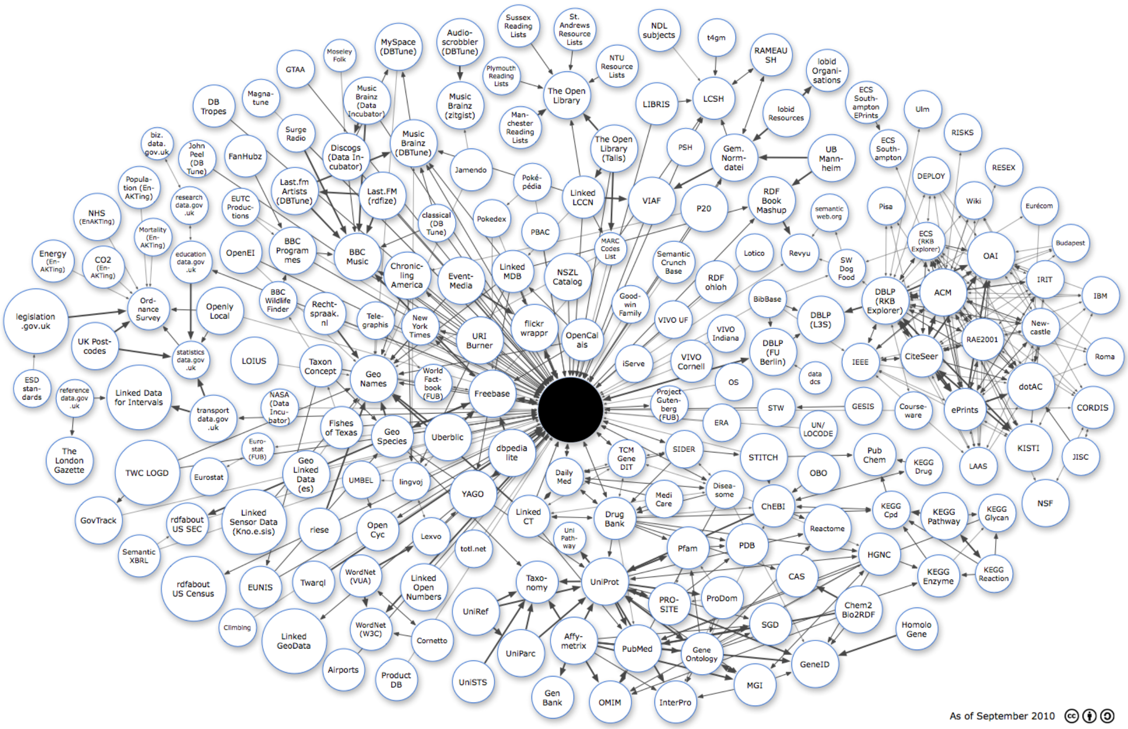


Figure 61. The magnitude of DBpedia
(Adapted from http://richard.cyganiak.de/2007/10/lod/lod-datasets_2010-09-22.png)

A Java module was developed to explore the DBpedia dataset by submitting SPARQL queries to DBpedia SPARQL endpoint. It uses ARQ (version 2.8.5), a query engine for Jena that supports the SPARQL RDF Query language.

Some research works have already begun using DBpedia and some tools have been developed, including RelFinder [Lehmann, Schüppel et al., 2007]. This tool finds the relationship between two IRI specified by the user, which was not the appropriate approach to this work. Its adaptation was initially considered, but then it was disregarded. For instance, it does not find any relationship between the concepts Database (represented by http://dbpedia.org/resource/Database) and Data_model (represented by http://dbpedia.org/resource/ Data_model), because it does not consider the property abstract.

Example 17 demonstrates how the property abstract is regarded in this work. For any DBpedia resource whose label property contains 'data_model', it is checked if its abstract property contains 'database'. Any two terms can be used, as they are supplied as variables in the developed module.

```
SELECT DISTINCT * FROM <http://dbpedia.org> WHERE {
                        ?s <http://www.w3.org/2000/01/rdf-schema#label> ?o.
                        FILTER (lang(?o) = 'en').
                        ?o <bif:contains> "data_model".
                        ?s <http://dbpedia.org/ontology/abstract> ?a.
                        FILTER (lang(?a) = 'en').
                        ?a <bif:contains> "database".
                       }
```

Example 17. DBpedia - Use of abstract property

This query returns 6 resources:

- http://dbpedia.org/resource/Data_model,
- http://dbpedia.org/resource/Standard_data_model,
- http://dbpedia.org/resource/Physical_data_model,

- http://dbpedia.org/resource/Many-to-many_%28data_model%29,
- http://dbpedia.org/resource/Data_model_%28ArcGIS%29,
- http://dbpedia.org/resource/Core_Architecture_Data_Model.

For conciseness, only two results are represented in Table 41. There are parts from the abstracts which were cut and these are represented through the use of '[…]".

Table 41. Results considering the abstract property

s (DBpedia resource)	o (Label)	a (Abstract)
http://dbpedia.org/resource/Standard_data_model	Standard data model@en	[...] Typically these use the popular relational model of **database** management, but some use the hierarchical model, especially those used in manufacturing or mandated by governments, e.g. , the DIN codes specified by Germany. Management consultant firms are often heavy users of standard models, providing cookie-cutter solutions" to many customers. These are however rarely as simple as off-the-shelf solutions, and may require customizations costing tens of millions of US$ and years to complete. The most complex data models known are in military use, and consortia such as NATO tend to require strict standards of their members' equipment and supply **databases**. [...]
http://dbpedia.org/resource/Physical_data_model	Physical data model@en	A physical data model is a representation of a data design which takes into account the facilities and constraints of a given **database** management system. In the lifecycle of a project it is typically derived from a logical data model, though it may be reverse-engineered from a given **database** implementation. A complete physical data model will include all the **database** artifacts required to create relationships between tables or achieve performance goals, such as indexes, constraint definitions, linking tables, partitioned tables or clusters. The physical data model can usually be used to calculate storage estimates and may include specific storage allocation details for a given **database** system. At present, there are six main **databases** in the business market; Oracle, Postgres, SQL Server, Sybase, DB2 and MySQL. There are a great many other RDBMS systems out there, but these tend either to be legacy **databases** or used within academia such as universities or further education colleges. A physical data model on each implementation would be significantly different, not least due to the underlying OS requirements that sit underneath them. Examples would be SQL Server which only run on Microsoft Windows operating systems, while Oracle and MySQL can run on Solaris, Linux and other UNIX-based operating systems as well as Windows. This means that the disk requirements, security requirements and many other aspects of a physical data model will be influenced entirely by the RDBMS that a **database** administrator (or their organization) chooses to use. [...]

The property wikiPageRedirects is used to automatically redirect some pages to others. Using a browser this process is unnoticeable by the user. Really, it allows dealing with syntactic variations of some designations provided to DBpedia resources. For instance, using a browser to access the Web page http://dbpedia.org/resource/Chomsky_Normal_Form, a user will reach the page http://dbpedia.org/resource/Chomsky_normal_form.

Example 18 demonstrates how the property wikiPageRedirects is regarded in this work, using the term 'normal_form'.

```
SELECT DISTINCT * FROM <http://dbpedia.org> WHERE
                    {
                     ?s <http://dbpedia.org/ontology/wikiPageRedirects> ?o.
                     ?o <http://www.w3.org/2000/01/rdf-schema#label> ?a.
                     FILTER (lang(?a) = 'en').
                     ?a <bif:contains> "normal_form".
                    }
```

Example 18. DBpedia - Use of property wikiPageRedirects

Some of the results are represented in Table 43, purposely chosen were some that are not simple writing variations.

Table 42. Results considering the property owl:sameAs

s	o	a
http://dbpedia.org/resource/Full_disjunctive_normal_form	http://dbpedia.org/resource/Disjunctive_normal_form	Disjunctive normal form@en
http://dbpedia.org/resource/Sum-of-product	http://dbpedia.org/resource/Disjunctive_normal_form	Disjunctive normal form@en
http://dbpedia.org/resource/3-CNF	http://dbpedia.org/resource/Conjunctive_normal_form	Conjunctive normal form@en
http://dbpedia.org/resource/3NF	http://dbpedia.org/resource/Third_normal_form	Third normal form@en
http://dbpedia.org/resource/2NF	http://dbpedia.org/resource/Second_normal_form	Second normal form@en
http://dbpedia.org/resource/4NF	http://dbpedia.org/resource/Fourth_normal_form	Fourth normal form@en
http://dbpedia.org/resource/1NF	http://dbpedia.org/resource/First_normal_form	First normal form@en
http://dbpedia.org/resource/Normalized_form	http://dbpedia.org/resource/Normal_form	Normal form@en
http://dbpedia.org/resource/Jordan_canonical_form	http://dbpedia.org/resource/Jordan_normal_form	Jordan normal form@en
http://dbpedia.org/resource/Classical_canonical_form	http://dbpedia.org/resource/Jordan_normal_form	Jordan normal form@en
http://dbpedia.org/resource/Boyce-Codd	http://dbpedia.org/resource/Boyce-Codd_normal_form	Boyce-Codd normal form@en
http://dbpedia.org/resource/BCNF	http://dbpedia.org/resource/Boyce-Codd_normal_form	Boyce-Codd normal form@en

Another property that was considered as relevant to compute semantic relatedness between two terms was the property owl:sameAs. The way that the property owl:sameAs is used in DBpedia has been analysed in some works, such as [Halpin, Hayes et al., 2010] and [Ding, Shinavier et al., 2010].

Example 19 shows a query that use the property in an attempt to discovers terms related to 'SQL'.

```
SELECT DISTINCT * FROM <http://dbpedia.org> WHERE
                    {
                     ?s  <http://www.w3.org/2002/07/owl#sameAs>  ?o.
                     ?o <http://www.w3.org/2000/01/rdf-schema#label> ?a.
                     ?a <bif:contains> "SQL".
                    }
```

Example 19. DBpedia - Use of property owl:sameAs

This property is mainly used to connect DBpedia resources to resources from other repositories, being not relevant to the purposes of this work. The top 10 results are represented in Table 43.

Table 43. Results considering the property owl:sameAs

s	o	a
http://mpii.de/yago/resource/Microsoft_SQL_Server	http://dbpedia.org/resource/Microsoft_SQL_Server	Microsoft SQL Server@en
http://mpii.de/yago/resource/SQL_Server	http://dbpedia.org/resource/SQL_Server	SQL Server@en
http://mpii.de/yago/resource/SQL	http://dbpedia.org/resource/SQL	SQL@en
http://mpii.de/yago/resource/SQL-Ledger	http://dbpedia.org/resource/SQL-Ledger	SQL-Ledger@en
http://mpii.de/yago/resource/SQL	http://dbpedia.org/resource/SQL	SQL@de
http://mpii.de/yago/resource/SQL_Slammer	http://dbpedia.org/resource/SQL_Slammer	SQL Slammer@en
http://mpii.de/yago/resource/Mimer_SQL	http://dbpedia.org/resource/Mimer_SQL	Mimer SQL@en
http://mpii.de/yago/resource/NonStop_SQL	http://dbpedia.org/resource/NonStop_SQL	NonStop SQL@en
http://mpii.de/yago/resource/SQL_injection	http://dbpedia.org/resource/SQL_injection	SQL injection@en
http://mpii.de/yago/resource/SQL	http://dbpedia.org/resource/SQL	SQL@es

Another property that started to be considered was rdfs:seeAlso; however it seems that its usage has been discontinued in DBpedia along the time, and currently it is not in use in the developed module.

7.3.1.2 A final strategy

When no ontology was discovered considering the provided concepts, and it is impossible to provide others, regarding information supplied by Wordnet, DBpedia and Scarlet, as described later in this chapter, another solution can be followed. However, that last solution must be considered when all the others have failed.

Thus, when all other attempts to find ontologies related to a concept have proved unsuccessful, a set of documents containing that concept can be provided, which is used for ontology extraction, with some parts subsequently incorporated into the ontology already obtained earlier, after their revision.

Text2Onto (http://code.google.com/p/text2onto/) [Cimiano and Volker, 2005] is the tool used to support the semi-automatic ontology extraction from textual corpus. It was developed by the AIFB institute in Universität Karlsruhe. It accepts as input a collection of text, HTML or PDF documents. It applies machine learning and linguistic processing algorithms to extract concepts, instances and relations.

The ability to choose different algorithms was important for its choice. It relies on the GATE framework for some of the linguistic processing, which also offers some flexibility in the algorithms to use. This tool was seen as adequate for the current necessities but also able to support future developments in the area.

This last approach to obtain an ontology with one or more concepts is significantly more time-spending than the others. Also, ontology learning algorithms should not be considered as possible of providing optimal results currently nor in a near future. However, it has been more and more regarded, with incorporation of some ontology learning tools in ontology development applications[59], and ontology learning is being considered in ontology engineering methodologies, including On-To-Knowledge (see section 4.5.1.2).

It was not really integrated in the ontology reuse module. A simple interface to the application was developed (see Figure 62), which accepts up to three files to be analysed and be considered in the

[59] For instance, OntoLT in Protégé and Onto2Text in Neon toolkit.

generation of an OWL file. The Text2Onto and GATE API versions used for building that interface are the following ones: Text2Onto 180607 and GATE 4.0 beta 2693.

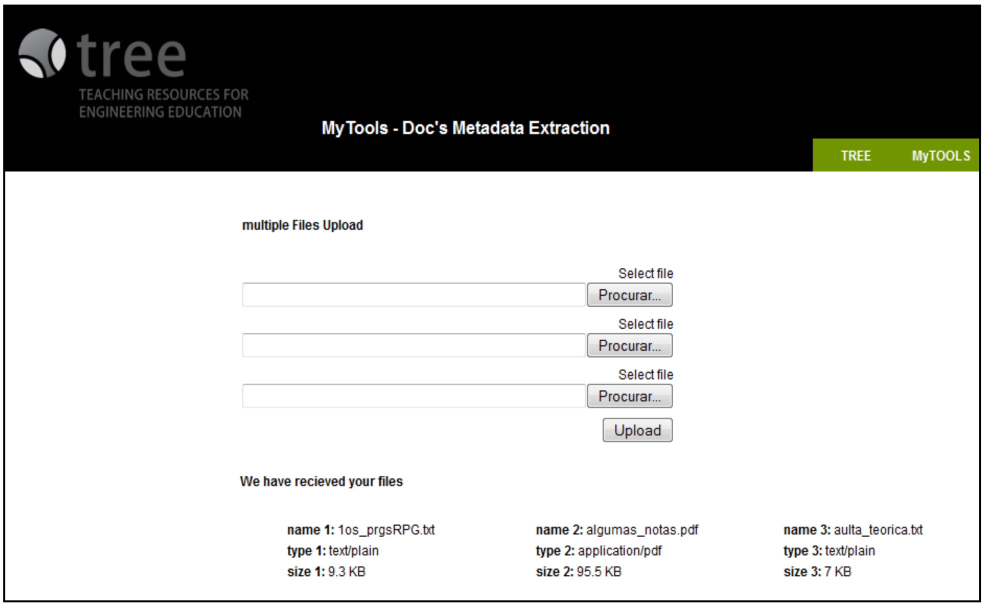


Figure 62. The ontology learning component

This simple interface is mainly useful for people without knowledge in the area. However, the resulting ontology has to be analysed by an ontology engineer.

Besides Text2Onto, there are also other tools that can be used for ontology learning purposes, such as:

- OntoLT – It is a Protégé plug-in (http://olp.dfki.de/OntoLT/OntoLT.htm). It can be used to extract concepts and relations from text collections [Buitelaar, Olejnik et al., 2004];
- DODDLE-OWL - Domain Ontology rapiD DeveLopment Environment-OWL extension (DODDLE-OWL) extends previous tools, DODDLE and DODDLE II. It is available at http://doddle-owl.sourceforge.net/en/;
- OntoBuilder - It extracts ontological structures based on heuristic methods learned from a training set of HTML documents [Gal and Modica, 2004]. It is available at http://ontobuilder.bitbucket.org/.

A recent study compared these tools and also Text2Onto [Park, Cho et al., 2010]. A number of different aspects were considered, including the quality of the constructed ontology. Text2Onto received the best global score, but not in all criteria. For instance, OntoLT received a better score in relations extraction.

7.3.2 Related work

SENSUS methodology was committed to the reuse of large ontologies. The underlying idea is that top-level ontologies can be reused in the construction of domain ontologies.

In the example used to illustrate the METHONTOLOGY approach in [Gómez-Pérez, Fernández-López et al., 2004] (pages 127 and 128), the activities involved in the reuse of two ontologies are high-level ones, but, as described in the methodology, they are not detailed. On-to-knowledge also considers the reuse of top-level ontologies [Sure, Staab et al., 2003], but suffers from the same problem.

The most related approach is described in [Cantador, Fernández et al., 2007]. The tool (WebCORE) is not available. The similarities are as follows:

- The concepts that characterise a domain are provided by the users;
- WebCORE uses WordNet synonyms for the concepts provided by the users.

The main differences are the following:

- The approach uses a local ontology repository and the problem should be stated to help the users in the selection of the most appropriate ones from that local repository regarding the specified concepts;
- The synonyms obtained from Wordnet are put under user consideration in WebCORE, without any advice regarding the most adequate ones. Also, that process is carried out before any attempt to find ontologies of interest;
- The approach used in the TREE repository deals with writing variations of the provided concepts;
- If N concepts are specified and there is no ontology with all of them, the strategy presented in this chapter tries N-1 concepts, and so on.

There have been some works related to ontology reuse, but not works trying to automate the process as much as feasible, considering Semantic Web Search Engines. The proper notion of feasibility is subjective, but the amount of the necessary human intervention, being not negligible, is not mandatory in all steps.

7.3.3 Some remarks

The approach followed in the ontology reuse module, in its initial steps, can be seen as an adaptation of the keyword-based search method used in IR, where the resources of interest are ontologies.

Section 5.4 provided some insights into query expansion methods. It described the possibility of using dictionaries, as well as stemming and conflations methods. They were followed, a variation for the latter, in the ontology reuse module. However, the lexical variations are considered before the synonyms obtained from Wordnet.

Other word association technique presented in section 5.4, the appliance of co-occurrence statistics, was not followed, but future developments might consider it using the keywords supplied (LOM field 1.5) when documents are submitted or co-occurrences in tags provided by users.

The final version had more than 70 tests in order to check its maturity and feasibly. The tests were conducted using concepts from a database course, a statistics course, an engineering applications course, a chemical engineering course, a software engineering course, and an English course. However, the most detailed tests were carried out in the database domain, because of the background of the researcher.

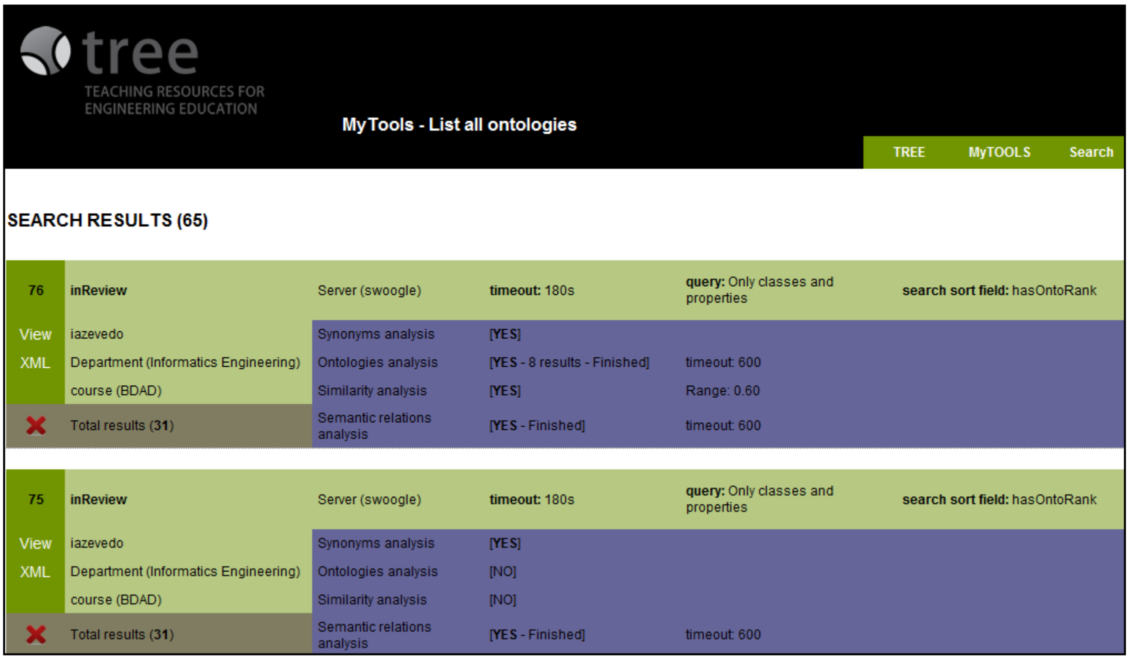


Figure 63. List of ontology search results

It was possible to find very interesting ontologies, regarding their domain coverage, but also the supporting documentation available. The most relevant discovered ontology was the Computing Ontology [Kamali, Cassel et al., 2004], already mentioned in this chapter. It has 2,970 named classes. The DL expressivity of this ontology is $\mathcal{SI}(\mathcal{D})$. It has been developed by a reputed team, and some additional information is available at http://what.csc.villanova.edu/twiki.

The research community concerned with the ontology reuse problem is recent and not very-well implanted. Therefore, due to the lack of availability of knowledge reuse tools, and thus performance reports, the delineated method was just tested for its effectiveness in discovering ontologies and comparative tests were impossible to be carried out. However, the results are dependent on many aspects, such as the keywords provided, but also on the available ontologies in a given area, which is unknown.

The lack of specifications related to the functionalities and usage of Semantic Web search engines puts also another problem, which might affect the developed ontology reuse module. One Semantic Web search engine was used, with a particular format of queries and results. A change from the Swoogle side can compromise the functionality of the developed module. During the presentation of [Azevedo, Seiça et al., 2010] the same image represented in Figure 64 was used to discusses the problem. The level of standardisation in Semantic Web technologies is far from achieved in the learning technology area, with specifications regarding query submissions and results provision by different repositories (see Chapter 2).

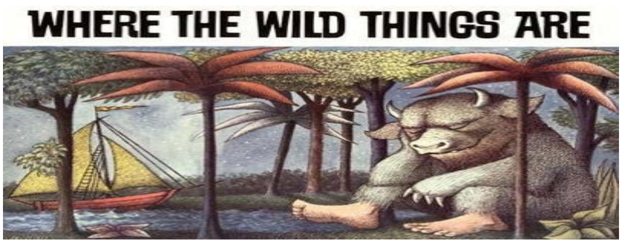


Figure 64. The “wild things” in ontology reuse

At the moment of this writing some SWSEs have been inactive, such as Falcons[60], unavailable for months, and SWSE[61], unavailable for weeks. That area lacks maturity and will certainly be advanced in upcoming years. Also, the Swoogle REST API[62] has been also inaccessible for weeks.

The implemented solution was integrated in a repository of learning resources, available only in an intranet, and it was developed under that premise from the beginning. However, the whole scientific community can benefit from a robust system for ontology reuse. That can represent a future research area, and may imply the recognition of more general requisites and approaches.

A final note about the strategy that was adopted: the possibility to consider the synonyms (the most appropriated, once computed the semantic relatedness) when submitting queries to Swoogle was weighted. The specification of the HTTP protocol does not impose any maximum length to a URL, but in practice server software might enforce some limits[63]. For instance, the following REST request submitted to Swoogle, using 3,741 characters (here abbreviated), is handled by Swoogle.

```
http://logos.cs.umbc.edu:8080/swoogle31/q?queryType=search_swd_ontology&search
String=def:programming+OR+def:programming1+OR+def:programming2+OR+def:programm
ing3+OR+def:programming4+OR+def:programming5+OR+def:programming6+OR+def:progra
mming7
...
+OR+def:programming170&key=demo
```

Example 20. A long REST request

However, using for each concept, others related to it, and considering different writing styles and plural/singular forms, in a query, can reach an excessive number of characters. Swoogle documentation does not provide any information about that aspect. Thus, that hypothesis was disregarded.

7.4 Ontologies application

The reasons for the adoption of domain ontologies in the TREE repository are twofold:

- Allow a detailed description of the resources;
- Attend to users' keywords included in a query, suggesting additional terms to be considered by applying a query expansion method.

Section 7.4.1 discusses how some metadata fields are automatically recognised in the TREE repository, while section 7.4.2 presents the query expansion module that is used when users search for resources.

7.4.1 Metadata specification

Before the recognition of some metadata fields, several initial steps were found necessary, which is explained in subsection 7.4.1.1. Then, the next two subsections (7.4.1.2 and 7.4.1.3) explain how the language and the keyword metadata fields are extracted from documents being submitted to the repository. Subsection 7.4.1.4 summarises the main points covered in this section.

7.4.1.1 General preparatory processing

The users can be very concise when providing metadata about the documents being submitted, even for the keywords that characterise them, with consequences in the retrieval of these resources

[60] http://ws.nju.edu.cn/falcons/
[61] http://swse.deri.org/
[62] An example, using demo key:
http://logos.cs.umbc.edu:8080/swoogle31/q?queryType=search_swd_ontology&searchString=def%3Acom puter+def%3Aprogramming&key=demo
[63] See http://www.danrigsby.com/blog/index.php/2008/06/17/rest-and-max-url-size/.

in keyword-based searches. We realised that it was necessary to automate as much as possible this process.

When a document is submitted to the TREE repository, some metadata fields are automatically or semi-automatically filled, like the keyword metadata field (a subelement of the General element in the IEEE LOM standard), which can have multiple values.

In the TREE repository the following file types were considered for the automatic metadata extraction: txt, doc and pdf. However, for the metadata specification, most of the applications used for automatic metadata extraction required a txt file as input. Thus, the following tools were used to convert the original file to a txt one:

- **ExtractText** (http://pdfbox.apache.org/commandlineutilities/ExtractText.html) – It is a Java command line application that extracts all text from the given PDF document. It is part of the Apache PDFBox (http://pdfbox.apache.org/index.html), which is an open source Java PDF library for working with PDF documents. The Apache PDFBox has many features, including the PDF to text extraction. The version in use in the repository is PDFBox 1.4.0;
- **Apache POI**[64] (http://poi.apache.org/)– It includes a number of APIs for manipulating various file formats based upon Microsoft's OLE 2 Compound Document format using Java. One of the POI subcomponents handles Word97 format, HWPF[65], which was used in the TREE repository to convert that type of files to text (specially the class org.apache.poi.hwpf.extractor.WordExtractor).

Other formats can also be dealt with using the same approach, converting the files to text files. For instance, currently the HWPF subcomponent of Apache POI has a partner for the Word 2007 docx format, the subcomponent XWPF, and both provide similar features, but the latter was not used in the repository.

Another API that was extensively used was the Java client API that employs the Google Translate service (http://code.google.com/p/google-api-translate-java/), as many processing steps mentioned in this chapter required document in English, and not in any language. Some changes were carried out in order to accept an input file to be translated, the original language and the target language (always English, in the currently utilisation in the TREE repository). Also it was necessary to modify the source to overcome the limitation of 5,120 characters. The text file to be translated is subdivided into sets of near 5,120 characters, and it was ensured that each segment contains complete sentences to be translated (sentences ending with "\ n" or "."). When necessary, incomplete sentences are added to the following text segment. The language identifications are represented in ISO 639-1 format, and thus the language is represented by two characters, for instance, 'en' is used for English language. Figure 65 illustrates that approach.

[64] POI was originally a humorous acronym for 'Poor Obfuscation Implementation' (see http://www.javaworld.com/javaworld/jw-03-2004/jw-0322-poi.html).

[65] See http://poi.apache.org/hwpf/index.html. HWPF was a joking acronym for 'Horrible Word Processor Format'.

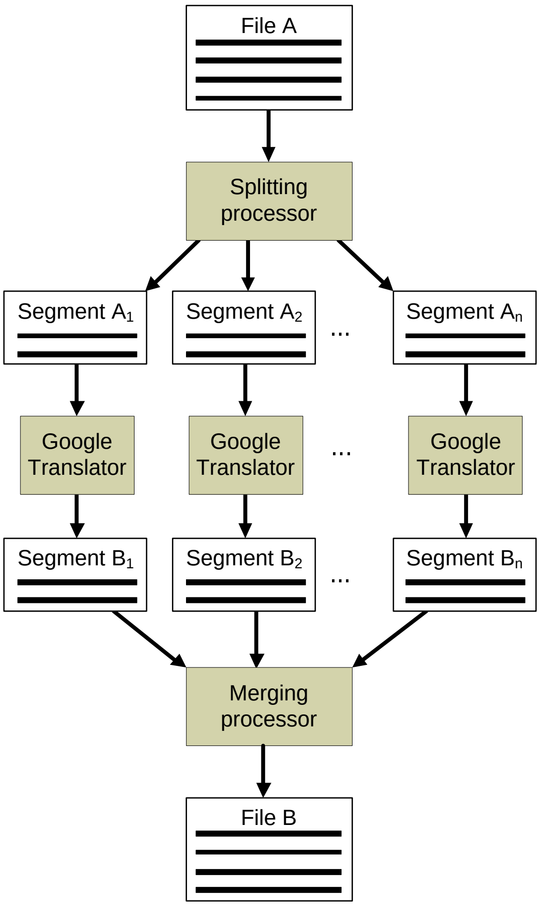


Figure 65. Translation strategy

It was tested mainly with resources written in Portuguese and Spanish, with good results, but any of the languages supported by the Google Translate API can be used (more than 50). Google Translate received the best rank in a recent study, among 10 free machine translators, in the translations of sentences in a formal style [Hampshire and Salvia, 2010]. The Google Translate 0.7 API was used.

Although it was not a problem during our experiments, the main drawback is that the Google Translation service may not be always accessible within an adequate timeframe, or it can be simply unavailable due to maintenance operations. Thus, the time amount necessary for the translation processing was a major concern. Using this document in text format when it had 268 pages, some experiments were carried out (see Table 44).

Table 44. Amount of time necessary for translation processing

Experiment number	Time amount
1	10s94ms
2	7s08ms
3	10s77 ms
4	6s08 ms
5	3s87 ms
6	10s13 ms
7	3s05 ms
8	6s26 ms
9	3s33 ms
10	7s12 ms

As shown in Table 44, there are fluctuations in the time amounts, but they never exceeded eleven seconds. Network congestions may cause problems, but, as explained, they have not been realised yet.

Before integrating the translation component in the TREE repository, it was extensively tested. It is still available in isolation in the TREE repository allowing its usage simply to translate documents, as shown in Figure 66.

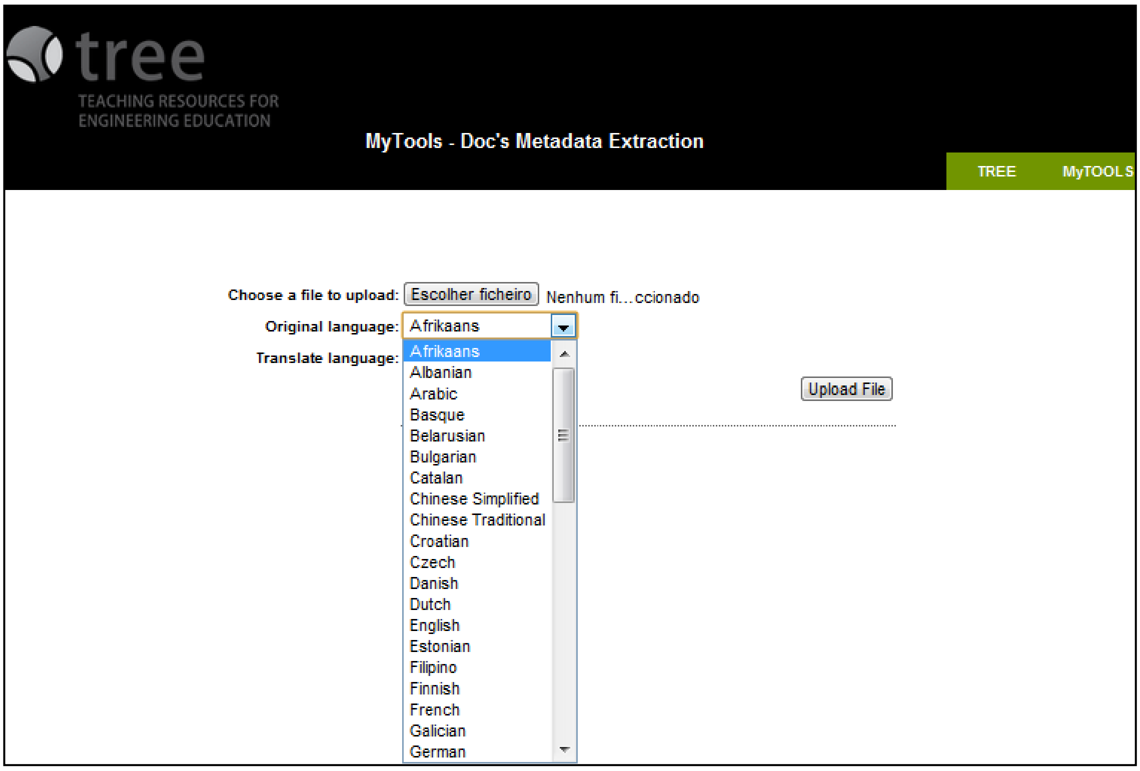


Figure 66. Translation service in the TREE repository

In this component users choose the file and state the original and the target language. After the translation the resulting file is available (see Figure 67).

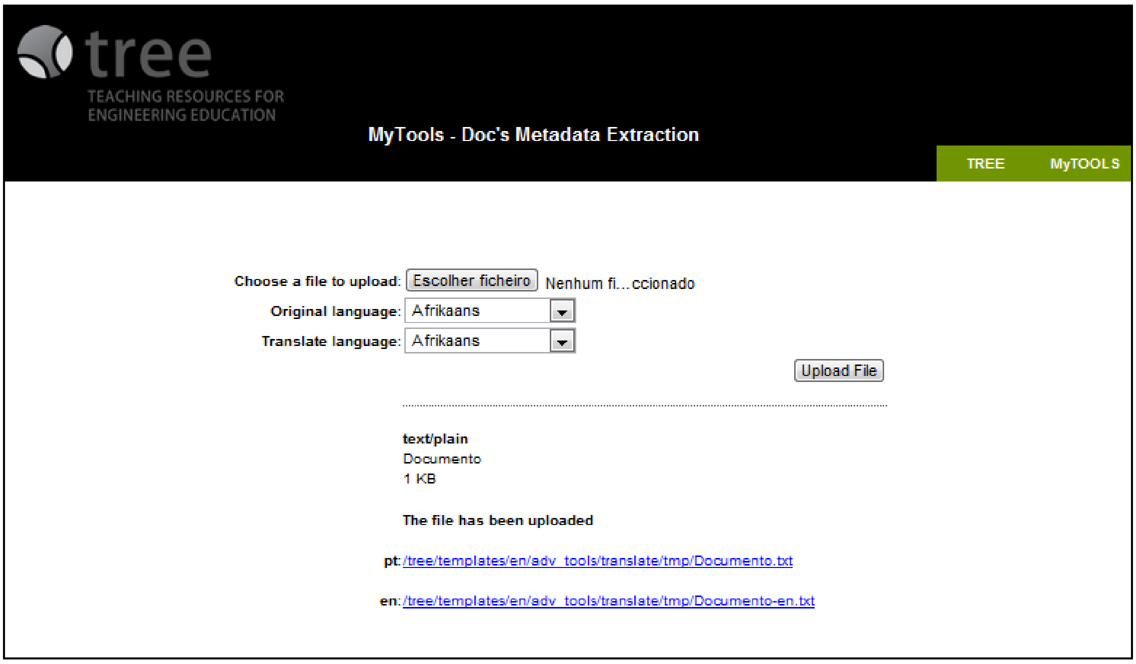


Figure 67. Translation service in the TREE repository - results

Example 21 presents an extract from the document available at http://www.dei.isep.ipp.pt/~iazevedo/algumas_notas.pdf, which was submitted to the TREE repository.

```
Ao longo dos anos a IBM tem convertido o AS/400 de um sistema "centrado em
servidor", com terminais ligados ao servidor, que tem todos os dados,
programas e capacidade de execução, para um sistema "centrado em
aplicações", que já permite que dados e programas possam estar em sistemas
separados, e a aplicação (que é composta por dados e programa) pode executar
noutro sistema.
```

Example 21. Original Portuguese text

Example 22 shows the translated text in English, using the Google Translate service, as previously explained. This translation and others obtained in other experiments are not perfect but satisfactory for the purpose to extract its main keywords.

```
Over the years IBM has converted the AS/400 system "centered on server" with
terminals connected to the server that has all data, programs and
implementation capacity, for a system "focused on applications", which
already allows which data and programs may be on separate systems, and
application (which consists of data and programs) can run on another system.
```

Example 22. Translated English text

7.4.1.2 Language metadata field

Prior to the translation of a given document, it is necessary to specify the original language, as it is not automatic recognised by the Google Translate service. On the other side, that processing was also found necessary for the automatic provision of the LOM metadata field Language (General category).

Textcat (http://www.let.rug.nl/~vannoord/TextCat/) was the first available implementation of the text categorisation algorithm described in [Cavnar and Trenkle, 1994]. A java version of Textcat (http://sourceforge.net/projects/textcat/), version 1.0.1, was applied for language guessing[66].

The application splits the text to be analysed into pieces called N-Grams. An N-gram is an N-character part of a longer string, considering contiguous characters only. Summing up, the document is broken into N-grams that are compared to some 'language models' (lm). In Textcat these are text files containing frequency data for N-grams in various languages, and have a file extension of 'lm'. The language with the highest ranking for a given document is identified as its language.

The data in the language models are ordered by frequency of use (see Table 45). The underlying idea is based on the Zipf's law of word occurrence [Zipf, 1949], which basically states that there is a set of words with high frequency usage in any language, but it was adapted to N-grams.

Table 45. Part of the Portuguese and English language models

Beginning of the Portuguese model		Beginning of the English model	
N-gram	**Frequency**	**N-gram**	**Frequency**
_ (space)	35328	_ (space)	20326
a	10423	e	6617
e	10132	t	4843
o	8919	o	3834
s	6795	n	3653
r	6033	i	3602
i	5443	a	3433
n	4588	s	2945
d	4531	r	2921
t	4217	h	2507
m	3476	e_	2000
u	3404	d	1816
o_	3240	_t	1785
a_	3029	c	1639
e_	2879	l	1635

In [Cavnar and Trenkle, 1994] the authors provided some evaluation data using articles from newsgroups in various languages. For example, for articles from Portugal the percentage of correct classification is 100 percent for six of the eight length profiles considered in their study. In the experimentations carried out in the TREE repository, incorrect language detection only occurred when using documents with purposely a very small number of characters (less than 100 characters).

Figure 68 shows how the language recognition is carried out when a resource is being uploaded to the repository.

[66] The complete list of supported languages is available at http://www.let.rug.nl/vannoord/TextCat/list.html.

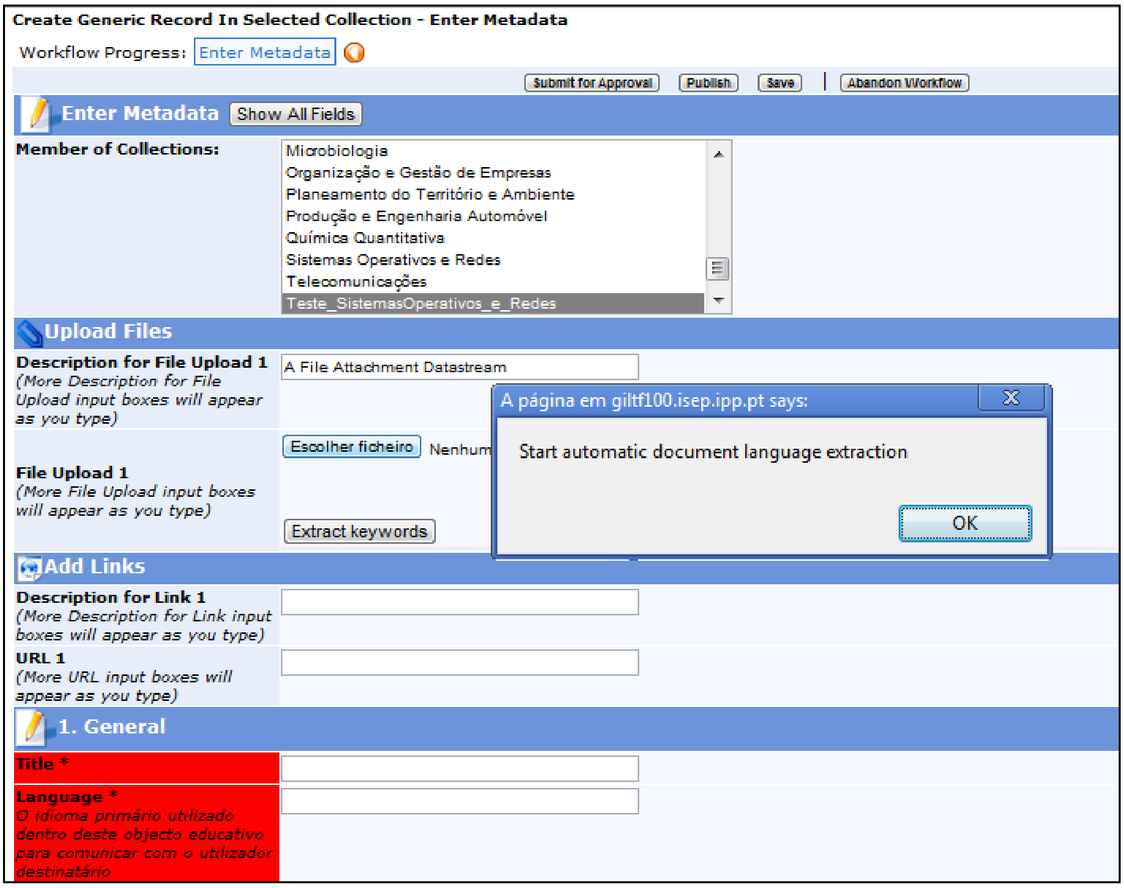


Figure 68. Automatic language recognition

Figure 69 depicts the language field usage considering a dataset from ARIADNE from 2005 to 2010. A huge number of resources do not have a language associated. The approach followed for language recognition can be useful for other repositories as well. However, these data must be read with care as it is unknown if 'x-none' value was used for images, for instance.

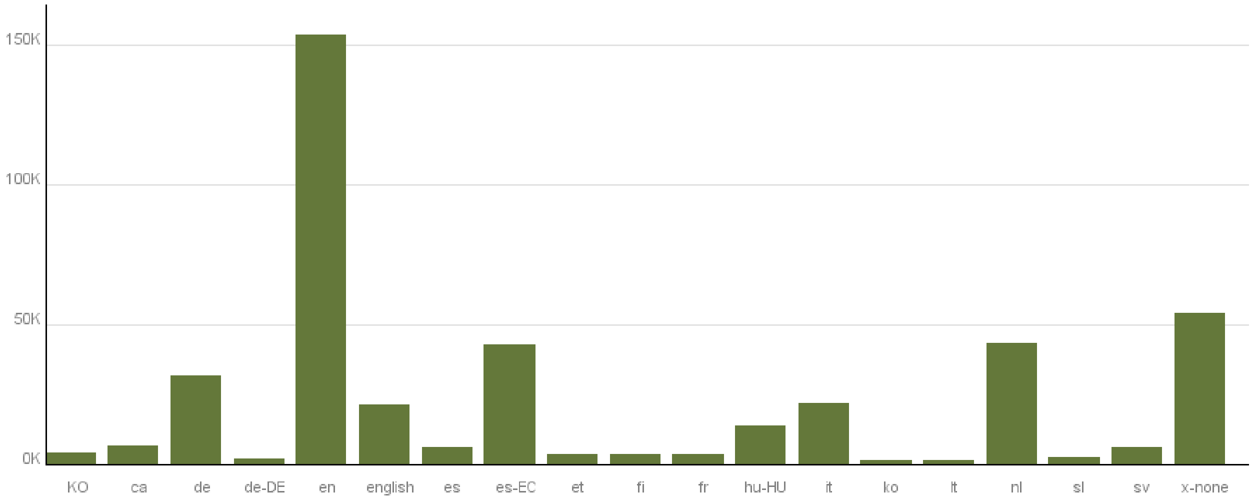


Figure 69. Language field usage – a dataset form ARIADNE
(From http://www-958.ibm.com/software/data/cognos/manyeyes/visualizations/language-usage)

7.4.1.3 Keyword metadata field

Another concern was the automatic recognition of keywords or keyphrases which are representative of a given resource when it is being submitted.

Phrase is a group of words which have a particular meaning when used together. For instance, 'data model' is composed of two words, each with their particular significance, while together they have a precise meaning in database domain. Keyword is usually applied to refer to a word that semantically characterise a resource, technically being only one word. Keyphrase is used to denote a group of words with a particular meaning that characterises a resource. The keyphrases associated to a document can be 'data model' or 'distributed database', for example. A keyphrase can have one or more words – that is the meaning used in this document. In practice, the two expressions 'keyword' and 'keyphrase' are often applied interchangeably[67], and that is the case in this document too.

To improve the keyword (or more precisely, keyphrase) representation of the resources uploaded to the TREE repository, two complementary steps were considered. First, the relevant keywords are extracted from textual English resources (pdf, doc or txt files) using XtraK4Me[68] keyphrase extractor [Schutz, 2008a; Schutz, 2008b], which makes use of GATE (General Architecture for Text Engineering) [Bontcheva, Tablan et al., 2004] components.

The following GATE components are used by XtraK4Me:

- Tokeniser;
- Sentence splitter;
- Part-of-Speech Tagger;
- Morphological Analyser.

XtraK4Me provides a ranked list of keyphrase candidates, after considering a number of steps [Schutz, 2008b]. A statistical measure is computed over basic forms of words (lemmas)[69] of lexical items. The top 25% relevant lemmas of the document are used to form clusters using the Monge-Elkan string-similarity metric [Monge and Elkan, 1996]. For each cluster, one element is chosen to represent it through the utilisation of an intra-cluster similarity. These elements are keyphrase candidates. Depending on their scope over the document, they are included in the output list provided by XtraK4Me.

[67] See, for instance, http://www.eflaunt.com/seo-glossary/keyphrase.htm.
[68] http://smile.deri.ie/projects/keyphrase-extraction.
[69] For instance, the infinitive form of a verb can be chosen for lemma.

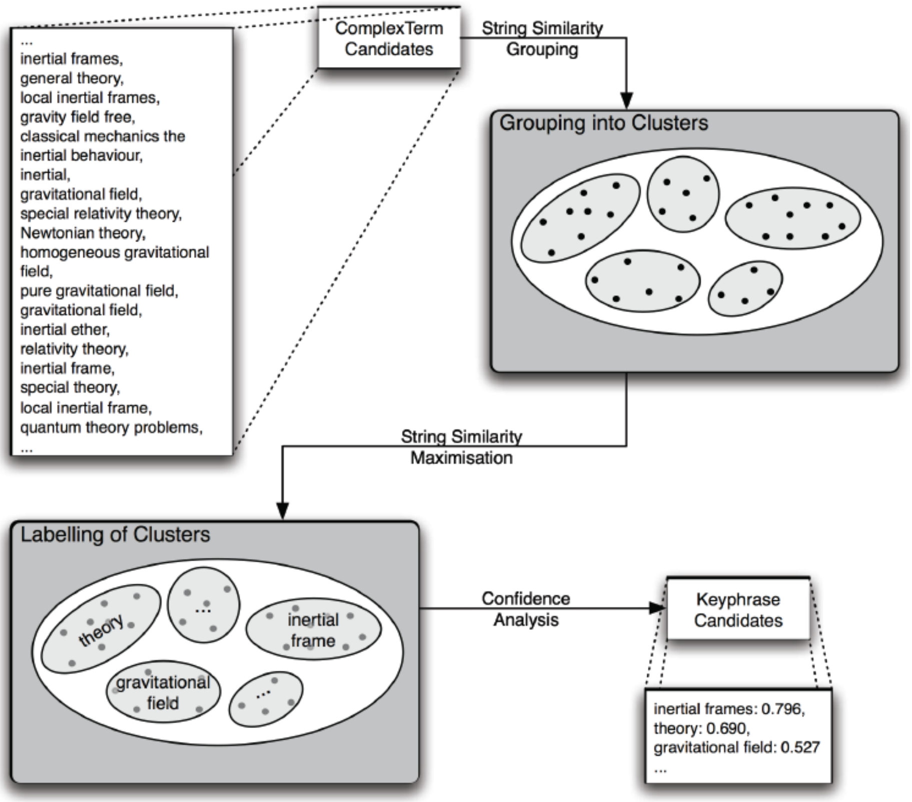


Figure 70. Overview of part of the strategy followed by XtraK4Me keyphrase extractor (Source: [Schutz, 2008b] (page 55))

The main drawback of the approach, which are reported in [Schutz, 2008b], is that it might not provide optimal keyphrases identification with documents with less than 500 words. The selection of XtraK4Me as the keyphrase extractor to be used in the TREE repository was dictated by the following aspects:

- It does not require a training step using some documents, prior to its application, such as KEA[70], for instance. This tool is definitely one of the most cited keyphrase extractors. However, before its usage, Kea demands the creation of a model that learns the extraction policy from manually (high quality) indexed documents. The model uses an input directory where there are files with the extension '.key' with the same names that the related document. This file includes assigned keyphrases, one per line. That requirement poses a major problem for its adoption in the TREE repository, but can be useful for repositories with manually indexed resources that desire to start applying a more automatic approach. There is no predefined minimum number of files that have to be in the input directory, but a very small number of files affects the model construction;
- XtraK4Me makes use of GATE, a very well known and popular tool which has been used in a number of projects, for some of their processing steps. The fact of lying in a very mature platform make it a robust choice to consider;
- Many projects have reported the satisfactory use of XtraK4Me, such as those reported in [Thai and Handschuh, 2009; Abela and Cortis, 2010; Farrugia, 2010; Barker, 2011; Gütl, Lankmayr et al., 2011], among others.

Then, using those keywords, SPARQL queries are submitted to the ontologies repository where the selected domain ontologies, associated to the TREE communities (civil engineering, informatics, mathematics, etc.), are stored.

[70] http://www.nzdl.org/Kea/description.html

In the ontologies there are sets of concepts and relations between them, which are used to obtain additional possible keywords. All these keywords are put under user consideration and he can modify them or add new ones.

For the document available at http://www.dei.isep.ipp.pt/~iazevedo/algumas_notas.pdf the keyphrases that were recognised are shown in Figure 71 (top part). In that document the phrase 'library list' and some variations of it are often used. The way that XtraK4Me chooses a representative for each cluster made 'Libraries libraries' to be considered a viable candidate.

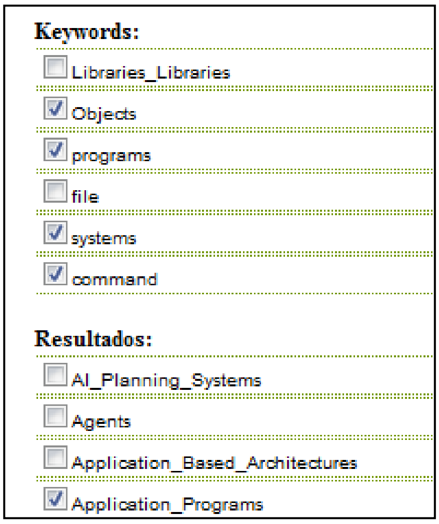


Figure 71. Selecting keywords

After the selection of the keyphrases of interest, these appear in the metadata provision page (see Figure 72). In that page it is possible to modify these keywords, add new ones, or delete one or more keywords. Thus, 'Library List' can be specified in that page, or 'Libraries libraries' could had been previously selected to be then modified.

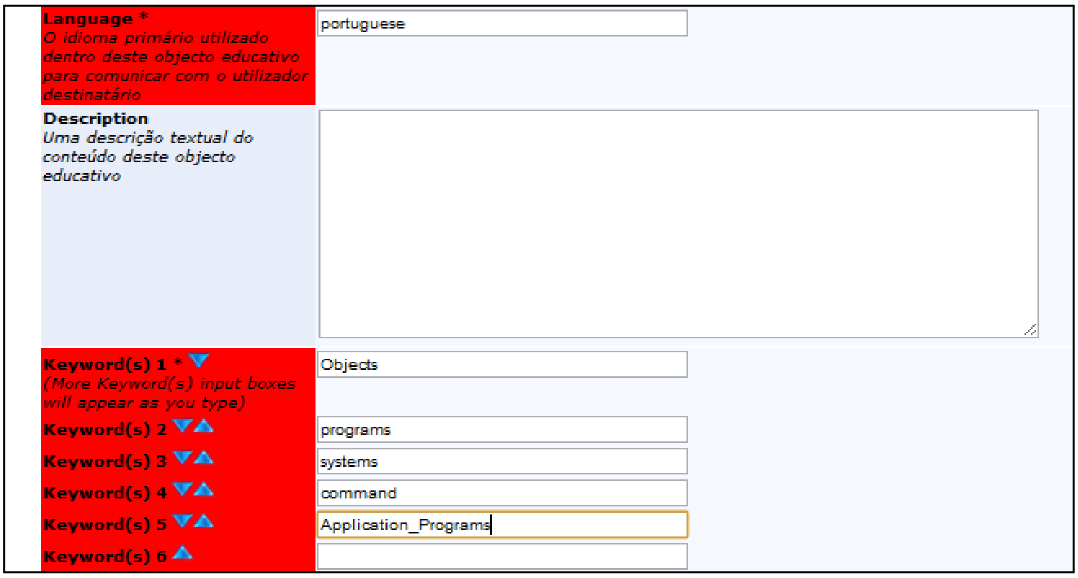


Figure 72. Language and keyword fields being automatically filled

SPARQL language is used to perform semantic querying and to derive new knowledge related to the extracted keywords.

The domain ontologies related to the community in which the resource is being added is queried in order to find relevant related terms to those mined keywords, but also ontologies associated to other collections, as it is not uncommon to have many different subjects being addressed in the same learning resource, especially the ones related to disciplines taught during the first years of degree or master's degree courses.

To expand the original keywords, subsumption or supersumption relations (one level only), meronymy and equivalence relations and instance data are considered. Thus, 'Physical data model' can be expanded to 'Data Model' (see Figure 73), for example, but not to 'Logical data model' as sibling relations are not considered.

When the same keyword is found in two or more ontologies additional keywords can be used to determine the most appropriated one, disambiguating the meaning, and avoiding the problems of polysemy. However, that step was not implemented in the prototype.

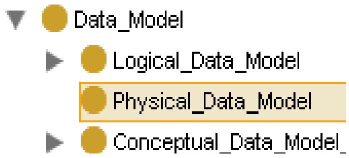


Figure 73. Partial view of an ontology

The extracted keywords and the related ones are put under user consideration and he can accept or not the suggestions, or supply others.

7.4.1.4 Summary

With the approach described in section 7.4.1 a richer characterisation of the resources can be achieved, with positive effects on recall rate.

Besides the TREE group, three people actively tested the approach described in this section, in conjunction with that described in the next section (7.4.2). They appreciated the new functionalities. The members of the TREE group did not report any problem.

Figure 74 summarises the main steps used in the TREE repository for automatic metadata extraction. The 'metadata extractor' block basically controls all the process, calling the various subcomponents and dealing with the results. File size and file type are recognised by Fedora during resource submission.

As explained before, some of the tools that were integrated in the TREE repository needed to be customised for the specific requirements of the followed approaches.

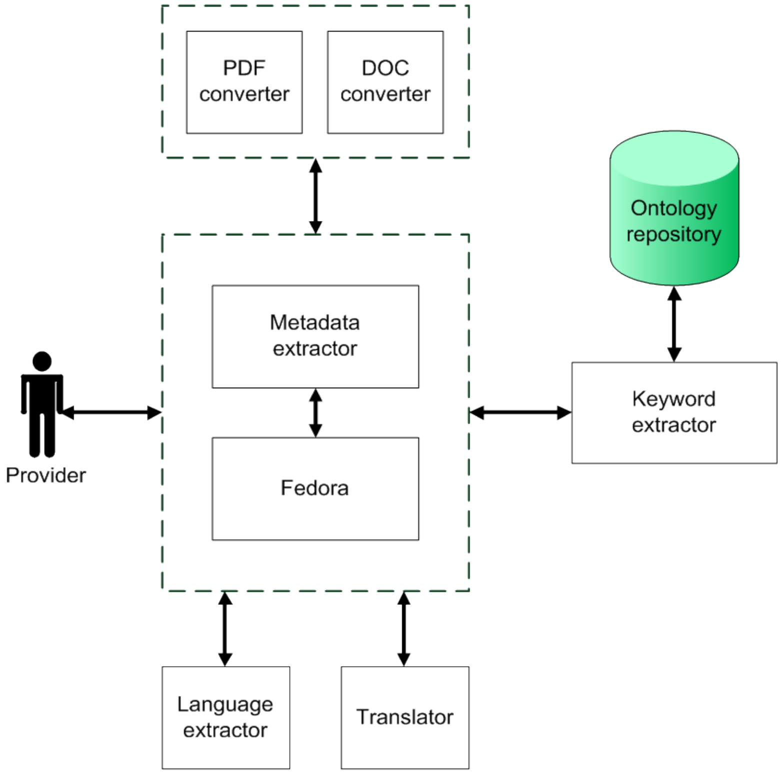


Figure 74. Interaction of the various subcomponents used for metadata extraction

To sum up, the tools used by the subcomponents in Figure 74, which were discussed in this section, are:

- PDF converter – ExtractText tool of PDFBox 1.4.0;
- DOC converter – Apache POI (the subcomponent HWPF);
- Language extractor – Textcat version1.0.1;
- Translator – Java client API for using Google Translate service;
- Keyword extractor – XtraK4Me.

7.4.2 Query expansion

Nowadays users are familiarised with keywords usage in search engines, such as Google and Yahoo, which return the resources that include character strings identical to those specified keywords. Despite their high popularity, the use of semantic technologies has been envisaged as a way to improve the provided results. For instance, Google announced in May 2009 the intention to adopt a technology that can "better understand associations and concepts related" searches[71].

The ability to find the right resource in the shortest time is pursued by information retrieval systems and by their users. Many techniques have been employed to improve and expand the initial query. Query Expansion (QE) is a technique that aims to enhance the number of appropriate documents retrieved by using further terms to the search query.

Figure 75 shows a partial view of the search form in the TREE repository. The case sensitive property was found helpful when dealing with acronyms.

71 http://googleblog.blogspot.com/2009/03/two-new-improvements-to-google-results.html

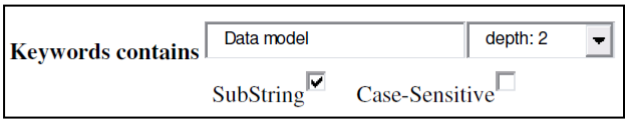


Figure 75. Partial view of the search form

Considering an inheritance level of 2, partial matchings ('relational_data_model' will be regarded if searching for 'data_model') and case-insensitive, as represented in Figure 75, part of the SPARQL query that is used to query the local ontologies repository is shown in Example 23.

```
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl:http://www.w3.org/2002/07/owl#

SELECT DISTINCT * WHERE
    {
        {
            {
               ?x rdfs:subClassOf ?y.
               FILTER (isIRI(?y)).
               FILTER (regex(str(?x)," '.$var_query.'","i")).
               FILTER (isIRI(?y)).
               OPTIONAL {?y rdfs:subClassOf ?z}.
               FILTER (isIRI(?z)).
            }
            UNION
               ...
        }
    }
```

Example 23. Part of a SPARQL query used for query expansion

Each time a query is submitted, every query term is expanded to related ones. The expansion uses ontologies previously stored in the ontology repository, considering subsumption or supersumption relations (up to two levels), meronymy (`partOf` properties) and equivalence relations (`owl:equivalentClass` and `owl:sameAs` properties) and instance data. The procedure is very similar to the expansion explained in section 7.4.1.3. Figure 76 shows some of the keywords related to 'Data model' that resulted from a query expansion – see also Figure 73.

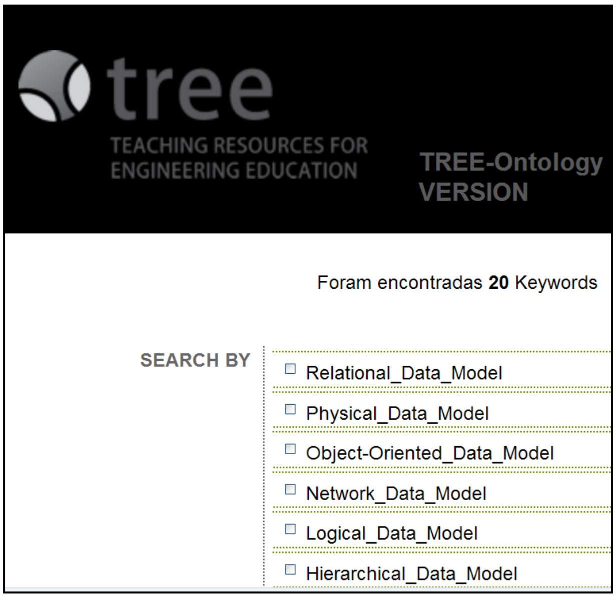


Figure 76. Some of the results of an query expansion

Differently from the processing carried out during the discovering of domain ontologies for a specific area, the query expansion have to be done in short time, as it is an interactive process.

A module was developed to expand the query terms provided by the TREE users, but allowing the users to agree or not with the use of the additional terms (see Figure 77).

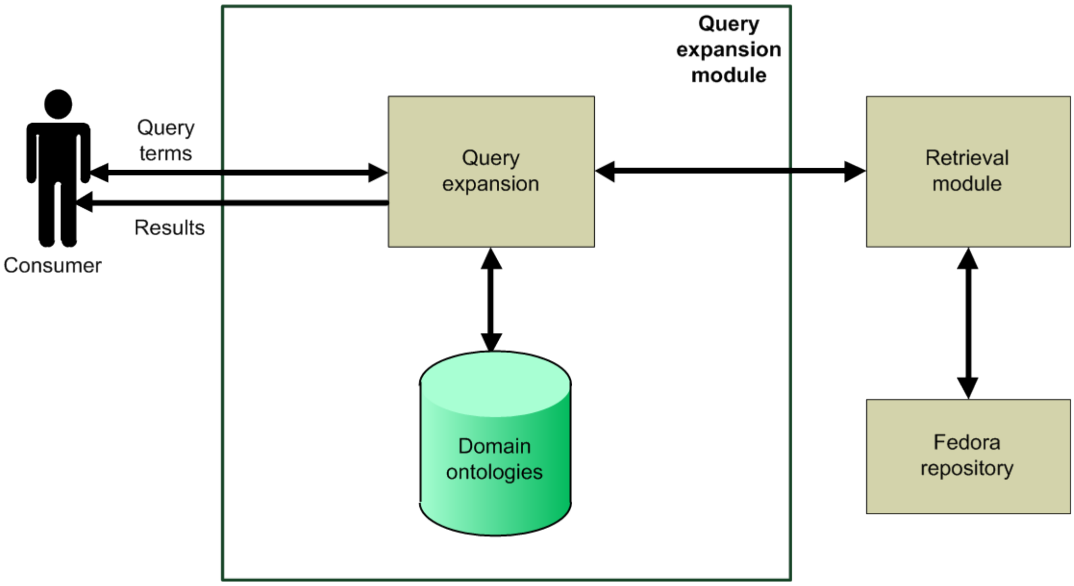


Figure 77. Architecture of the query expansion module

The architecture of the query expansion module allows the interaction with different components, and it is detailed in Figure 77. The document corpus is located in the Fedora repository and the fundamental element of the architecture is the domain ontologies repository, in which all relevant knowledge is stored.

The results are displayed using the Fedora/Fez sort option selected by the user, as shown in Figure 78.

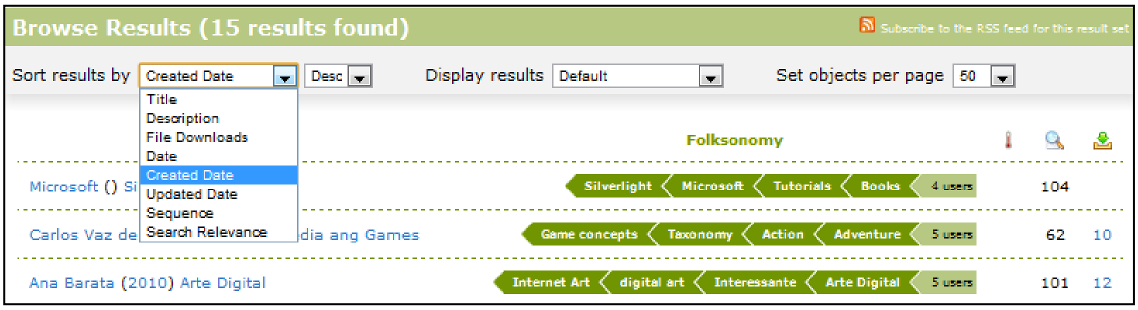


Figure 78. Sort options

Some ontologies were selected to be reused and one of those was developed to represent the computing and information related disciplines together [Kamali, Cassel et al., 2004], and part of it[72] is shown in Figure 79.

[72] From the 28 July 2009 version - http://what.csc.villanova.edu/twiki/bin/view/Main/OWLFileInformation.

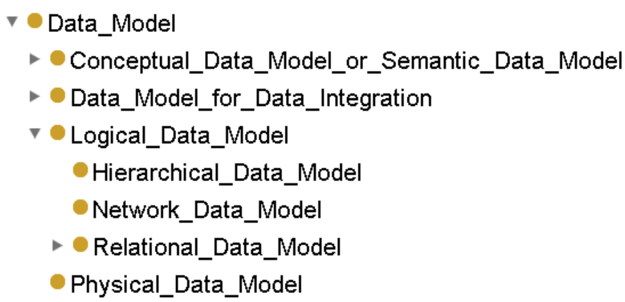


Figure 79. An extract from a reused ontology (ComputingOntology.owl)

Using that ontology a query that uses the term 'Logical Data Model' can be expanded to also consider the terms 'Data Model', 'Hierarchical Data Model', 'Network Data Model', and so on.

Different from other approaches [Ali and Khan, 2008] [Huang and Hsu, 2008] [Lee, Tsai et al., 2008], this one does not rely on a single ontology specifically developed for its application in the repository or a well-validated by the scientific community in the area.

Currently, although aware of its usefulness, the query terms used and if it was possible to achieve an expansion or not are not being logged to later analysis in order to identify gaps in the knowledge sources. However, some impossibilities of expansion may be due to misspellings.

Some approaches were not implemented in the prototype, such as:

- The rdfs:label property can be used to conform with different languages requisites;
- Others query keywords supplied can be used to disambiguate the meaning. For example, the word Law can be related to Boolean Laws (Mathematics) or Cybercrime Law (Law).

But they might be useful. In a well populated and actively used repository, those strategies should be studied.

7.5 Conclusions

In the development related to the semantic component of the TREE repository, many different tools were integrated, for translation, but also for generic information extraction, with the use of texts as input to produce structured data as output. Special attention was given to the keyword metadata field. This field was considered a mandatory one in the TREE application profile, but an effort was made to facilitate its provision, as keywords are often used in IR as a way to express user needs and they are used in retrieving the most probable proper resources. With the keyword recognition processing it is possible to obtain a faithful representation of the resources' contents.

Domain ontologies are used for the purpose of having the core concepts of a domain well-related in order to help users in the specification of keywords of interest, possibly improving the performance of information retrieval in the TREE repository, by expanding the initial query terms provided. The information in the ontologies should answer what concepts are related to another given one, which is also useful when users are characterising resources. In both situations, the user is responsible for agreeing or not with the expanded terms and, thus, with their usage.

With all domain areas of interest covered by a number of ontologies, a pure metadata annotation strategy for the keyword field might be achieved by attaching semantic descriptions to resources, linking them to items defined in ontologies.

It is noteworthy that the Fedora retrieval system was extended. Basically, an outer layer was developed and, through that layer, the system started to 'interpret' the information needs. Some 'interpretation' was also added to the resources' representation.

Differently from other approaches for ontology reuse that mainly acts at lexical level, in this chapter the strategy adopted in the TREE repository was presented, which also considers semantic relatedness between concepts. A number of tools were used for that purpose. The achieved solution cannot be used by everyone. It is envisaged to be used by an ontology engineer, who should be able to understand and follow all the processing steps.

Most of the attempts to reuse ontologies cannot determine that two terms T1 and T2 are semantically very close, and if they are not lexically similar. Thus, when searching for ontologies with T1, ontologies will be disregarded if they do not contain T1, but T2. A step in that direction was achieved in this work.

In summary, this chapter described how some metadata fields are obtained and how ontologies are used when resources are submitted to the TREE repository and users submit queries that represent their information needs. Moreover, it discussed how ontologies can be reused in a repository, detailing the whole process in a method that can be followed in other repositories as well.

Chapter 8

PRAGMATIC COMPONENT

To pragmatically characterise learning resources, two complementary approaches were chosen: the use of folksonomies, which allows collecting and using users' perspectives and opinions on various resources, and the utilisation of the IMS LD specification to illustrate possible ways to use resources.

This chapter describes the pragmatic component of the TREE repository and justifies the adopted options.

8.1 Introduction

The word pragmatic is used in this document, and particularly in this chapter, with a dual meaning: the approach is feasible (one of the ordinary significances of the term), but a context, which is an important point in pragmatics, is also considered when deciding opinion polarity and submitting resources (a use context).

A learning objects repository typically stores instructional materials and therefore it has mainly two types of users: the people that submit resources (providers) and the people that use them for teaching or learning purposes (consumers). Ideally the former provide a context in which the resource makes sense, while the latter provide feedback about the resources.

Those people that have probably used a learning resource over and over, the providers, should have the possibility to state how it was used in practical terms, which can be seen as a pedagogical advice that can potentially facilitate its reuse.

Pragmatics is also concerned with what is understood in a given context. In a LOR it is important to know how resources are regarded by who use them. The existing approaches to supplying feedback to providers are not satisfactory, even in many well-established repositories. Tags provided by users in repositories are a valuable source of opinions about the resources, but currently not used for this purpose. This chapter also discusses a pragmatic approach to a learning objects repository that uses tags for pragmatic characterisation of resources, including sentiment analysis to capture users' opinions expressed through tags assignments.

The model already presented in Chapter 6, Learning Objects Repository - Model and Implementation, is repeated here but highlighting what is covered in this chapter (see Figure 80).

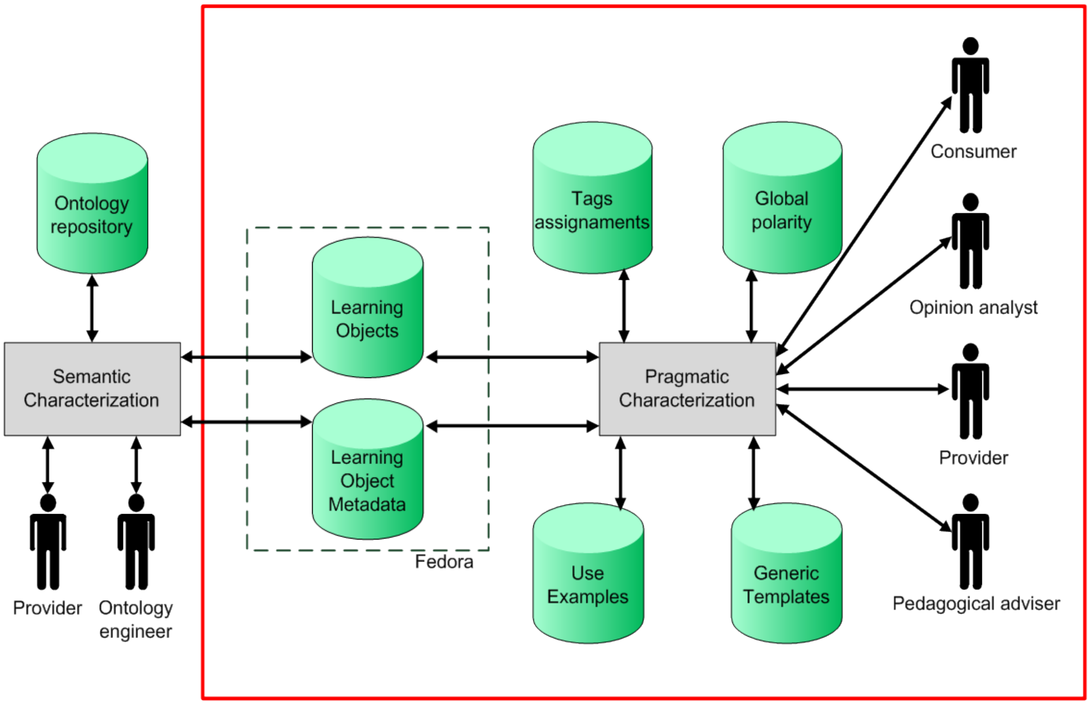


Figure 80. A proposed model for repositories – the pragmatic component

This chapter is structured as follows. Section 8.2 provides the rationale for adopting and using learning design templates in the TREE repository. It also elucidates how they are considered. Tagging as a communicative function is introduced in section 8.3 and also its application to supplying an opinion summary to providers. Finally, section 8.4 summarises the main findings of this chapter and some concluding remarks.

8.2 Use of learning design templates

The term 'pedagogical content knowledge' was coined by Lee Shulman to describe the required knowledge to teach successfully in a discipline [Shulman, 1986]. It interconnects two dimensions: pedagogical knowledge and content knowledge (see Figure 81), and it comprises knowledge about how to teach a particular content, i.e. "the most useful forms of representation of those ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations—in a word, the ways of representing and formulating the subject that make it comprehensible to others".

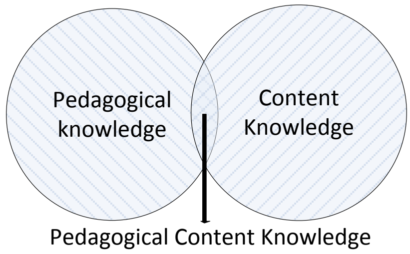


Figure 81. Pedagogical content knowledge overview (Adapted from [Herr, 2008])

In a learning objects repository, the resources should be connected to a description of how to apply them, supporting some pedagogical knowledge. Thus, it is advisable to state how to use a learning object, and what learning outcomes to expect.

These concerns might help making learning resources widely used by other individuals than those people that have developed them and know exactly how to successfully apply them. To design an effective learning process it is necessary not only to consider the learning resources, but also the activities and interactions that will support the knowledge transfer, and the whole learning situation.

Moreover, the use of learning technologies and specifications should be considered for that wide contextualisation of the resources. The IMS Learning Design specification can be used to sustain this approach and state how to use a Learning Object.

In fact, several learning design repositories have appeared, based (or not) on the IMS LD specification. There is a Learning Design Repository available at http://www.idld.org, result of Implementation and Deployment of Learning Design Specification (IDLD) project, which is a Canadian project committed to the diffusion of "basic educational modelling concepts, like learning designs patterns and examples, in order to produce a repository of IMS LD compliant units of learning". The IDLD repository encloses a number of learning scenarios.

The European Network for Lifelong Competence Development, TENCompetence (http://www.tencompetence.org), was a four-year project supported by the European Commission through the IST Programme running from December 2005 to November 2009. It used IMS LD in the context of Life Long Competency Development. It developed systems using IMS LD, in the context of one of its work packages that focussed specifically on Learning Activities.

Also, OpenDock (http://www.opendockproject.org/), a Leonardo programme project, developed an IMS LD aware repository called OpenDocument.net [Elferink, Griffiths et al., 2006]. It was available in an early phase of this work (see Figure 82), but it has been unavailable since 2008.

Figure 82. The OpenDocument.net repository

Another remarkable project from the pre-IMS LD era was the Learning Designs Project [Australian Universities Teaching Committee, 2003]. This project started in 2000 and generated "generic/reusable learning design resources". It resulted in 32 narrative learning designs divided into five focus groups:

- Collaborative,
- Concept/Procedure Development,
- Problem Based Learning,
- Project/Case Study,
- Role-play.

It is important to note that past projects were very important for the attention gradually given to learning designs, for the effort to launch a specification and to make it reach a more mature usage (see Annex 1 to an overview of projects and tools). CooperCore is an engine (see Annex 1 – section A1.2.1) that was developed by the Open University of Netherlands under the ALFANET project. For a long time it was the only one developed. It was incorporated in many players. Also, many projects websites maintain available detailed information on their use of the IMS LD specification.

The LAMS (Learning Activity Management System) community of practice (available at http://www.lamscommunity.org/lamscentral/) maintains a repository of learning scenarios in LAMS sequence format (.las files). Example 24 shows an extract from a LAMS file.

```
<org.lamsfoundation.lams.learningdesign.dto.LearningDesignDTO>
  <learningDesignID>8</learningDesignID>
  <title>Learning to Learn</title>
  <firstActivityID>60</firstActivityID>
  <firstActivityUIID>1</firstActivityUIID>
  <maxID>17</maxID>
  ...
<org.lamsfoundation.lams.learningdesign.dto.AuthoringActivityDTO>
  <activityID>60</activityID>
  ...
```

Example 24. LAMS sequence format

It is expected that Learning Design Repositories will become increasingly popular, as the possibility to search on learning design properties, for example, is a very interesting functionality to provide to end users.

Besides those learning design repositories, there have been some approaches proposed for systematising activities through the use of the IMS LD specification in an easier way [Currier, Campbell et al., 2005]. Some initiatives that exploit informal descriptions of learning designs are:

- **LearningMapR** – It aimed to facilitate the use of the IMS LD specification through templates that can be personalised to a specific situation [Buzza, Richards et al., 2005]. It has two main components. One of them uses a simplified version of Bloom's taxonomy with six types of learning objectives while the other component aims at helping instructors in the identification of the level of difficulty that learners face with a particular subject;
- **8LEM** – The 8 Learning Events Model (8LEM) [Verpoorten, Poumay et al., 2007] is a pedagogical framework developed under the European project iClass (http://www.iclass.info). It considers teachers, students and their interrelations. The eight events are: *receive*, *imitate*, *explore*, *practice*, *debate*, *reflect*, *create* and *experiment*, in the student perspective;
- **DialogPlus** – It was a five-year project that ended on 31 January 2008. Under this project a toolkit was developed, which is available at http://www.nettle.soton.ac.uk/toolkit/. It uses 'nuggets', which are activities/elements that guide the creation, alteration and sharing of learning activities and resources [Conole and Fill, 2005]. The types of the DialogPlus taxonomy of learning activities fall into the following categories: Adaptive, Assimilative, Communicative, Experiential, Information Handling, and Productive.

It is noteworthy that these approaches do not represent totally distinct ideas but they have similarities and a common objective of facilitating the production of a Unit of Learning. For instance, in [Conole, Littlejohn et al., 2005] the authors map DialogPlus taxonomy onto 8LEM learning events. The mapping suggestions for Assimilative, Communicative, and Productive categories are represented in Table 46. Only one DialogPlus category does not have a direct mapping, the Experiential one. This has a correspondence to practice, imitate and explore of the 8LEM model. See Appendix Six of [Conole, Littlejohn et al., 2005] for a complete list.

Table 46. A partial correspondence between 8LEM and DialogPlus

DialogPlus task type	8LEM event
reading	receive
viewing	receive
listening	receive
discussing	debate
presenting	debate
debating	debate
creating	create
produce	create
writing	create
drawing	create
composing	create
critiquing	create
synthesising	create
re-mixing	create

The 8LEM model was chosen for the initial creation of templates because of its simple and very usable approach in modelling common situations in learning processes. However, the DialogPlus taxonomy can also be useful for detailing the process.

Each 8LEM event encompasses learner and teacher actions, which are not necessarily carried out by them. For instance, sometimes a student may be assigned traditional teacher tasks.

Some 8LEM events were selected for the development of the initial templates to be used in the system prototype. The approaches followed for the creation of the templates are depicted from Figure 83 to Figure 85 in a graphical representation based on that followed in [Leclercq and Poumay, 2005].

A template based on the Exercise event that can be used for quiz tests or questionnaires is shown in Figure 83. It is a single-activity template.

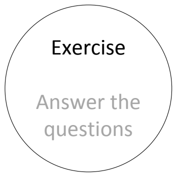


Figure 83. A exercise template

The template shown in Figure 84 uses two activities, but they can also be considered alone in separated templates as the activity 'Summarise the main ideas' can be seen as encompassing the activity 'Read' or any other that let a user be aware of the ideas presented in a given resource.

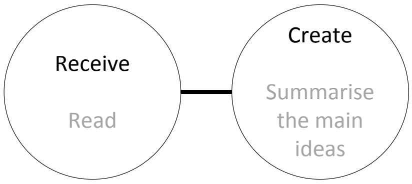


Figure 84. A receive + create template

A template based on three events is shown in Figure 85. It can be used for lab works or tutorial documents, for instance. Again a remark similar to that provided for the previous template: the 'receive' event can be disregarded or used for a 'Listen' activity, among others.

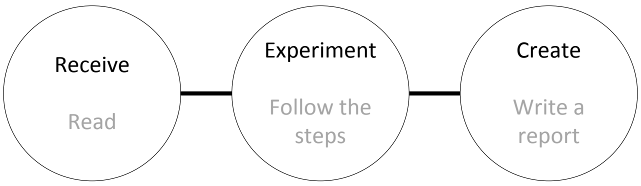


Figure 85. A receive + experiment + create template

The exact templates to be considered have to be carefully planned, as the usefulness of all other steps is clearly connected to them. However, the developed prototype has a number of functions related to them (see section 6.3.4. Pedagogical adviser's use cases). Thus, it is possible to add new templates (Use case Add template (UC58)), which are subject to evaluation by the Pedagogical Adviser, remove templates (Use case Delete in-use template (UC57)), or visualise active templates (Use Case List in-use templates (UC55)) and the associated UoLs (List related UoLs (UC56)), among others.

Templates are defined to support different pedagogical approaches and to provide a practical context for learning objects. They represent learning strategies that can be used for a number of resources, as shown in Figure 86.

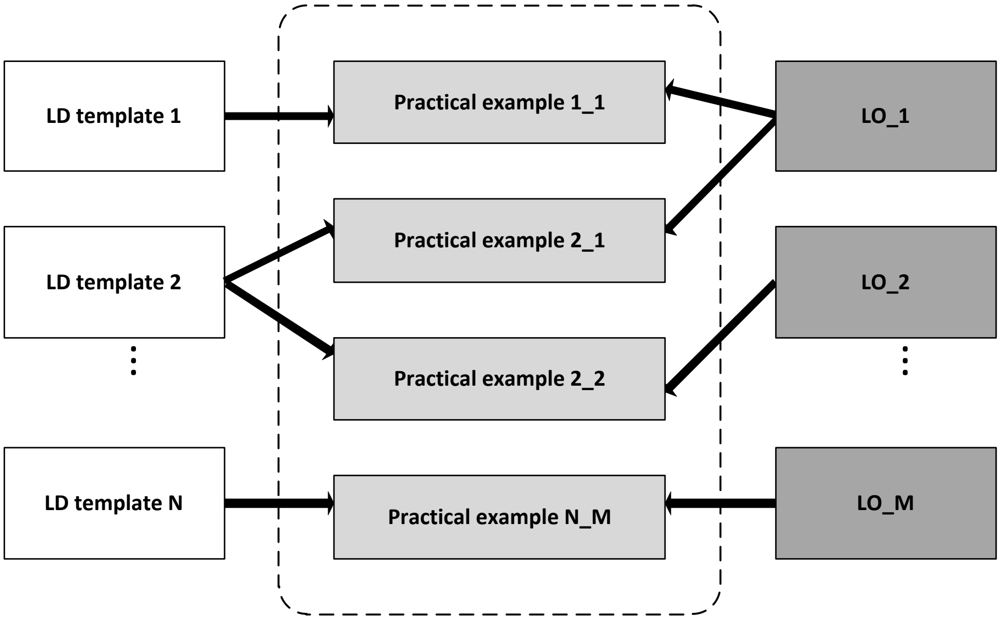


Figure 86. Use of IMS LD templates
(Adapted from [Azevedo, Carrapatoso et al., 2008a])

Heathcore distinguishes between small learning activities templates that encompass a single activity and templates that encapsulate more activities [Heathcote, 2006]. In this document the expression 'learning design template' refers to both approaches, and thus one or more activities can be used in a learning design template, as evidenced in the templates presented.

Also, it is noteworthy that the level of detail on the definition of learning activities was purposely low in order to allow them fit different concrete situations.

The use cases related to the use of learning design templates in the TREE repository are as follows:

- Download unit of learning (UC19);
- Download package (UC20);
- List own template suggestions (UC22);
    - Visualise suggestion (UC23);
- Suggest a new template (UC24);
- List in-use templates (UC25);
    - Visualise template (UC26);.
- Manage in-use templates (UC54);
    - List in-use templates (UC55);
        - List related UoLs (UC56)
    - Delete in-use template (UC57);
    - Add template (UC58);
- Manage temporary templates (UC59);
    - List unapproved templates (UC60);
    - Evaluate template (UC61);
    - List rejected templates (UC62).

These use cases were introduced in Chapter 6 and they are detailed in Annex 3. They are related to the parts of the pragmatic component shown in Figure 87.

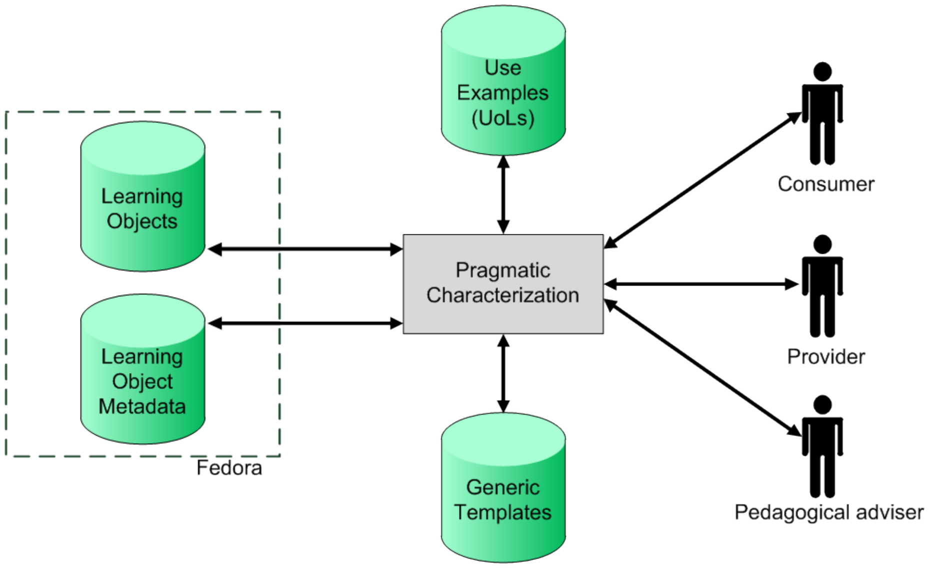


Figure 87. Pragmatic component – use of templates

8.2.1 Generation of Units of Learning

The main focus of the work described in this section is to create a close connection between a resource and its use, advancing the approach of allowing users to add pedagogical comments to submitted resources [Dalziel, 2008].

Figure 88 depicts some information about a MERLOT resource. The submitter also provided some pedagogical advice for its application. Although undoubtedly useful, the comments in a text format might vary in the details provided and they might not facilitate the reuse of the resources.

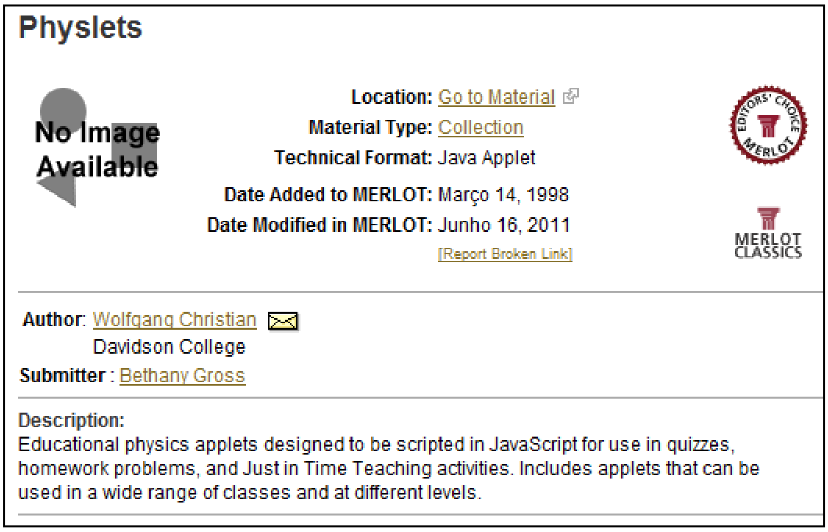


Figure 88. Example of the use of pedagogical comments in the MERLOT repository (From http://www.merlot.org/merlot/viewMaterial.htm;jsessionid=4A745CC43B373FAF680EA9B75F1106AF?id=75000)

In the TREE repository it is also possible to provide pedagogical comment through the LOM field 8.3 Description, as a complement of the use of learning design templates.

Figure 89 highlights how a UoL is generated in the system. There are resources and the corresponding metadata fields, which are important for the UoL. The Provider can initiate the process, not only for resources he previously submitted, but for any other that he had used before or have an idea of how to apply it successfully. The option to generate a UoL is immediately shown to the provider, when a resource is submitted. But it is also available after the submission.

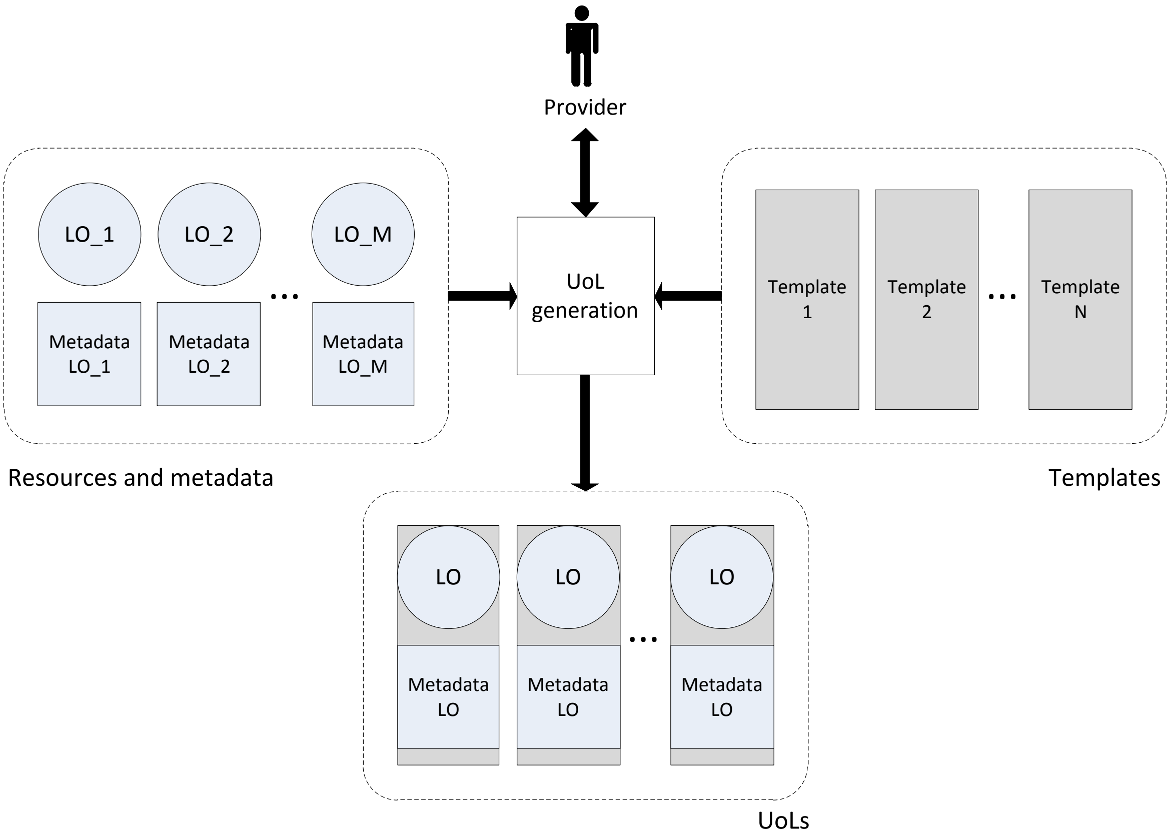


Figure 89. Overview of the UoLs generation
(Adapted from [Azevedo, Carrapatoso et al., 2008b])

A learning design template has associated a number of activities and, for each one, the generic user (a student, a monitor, a teacher) implicated is specified, and how the task is conducted. The individual activities are in a sequence (a "method" in the IMS Learning Design nomenclature). A UoL may also have one learning objective for the set of specified activities. In fact, the specification allows each activity to have its own learning objective, but in the prototype only a global learning objective for all activities is considered.

For a template creation, it is basically necessary to state the number of activities to be considered (see Figure 90). Only then is the information related to each activity provided. The data model for the templates was created in accordance with the IMS LD XML schema. It contains level A elements.

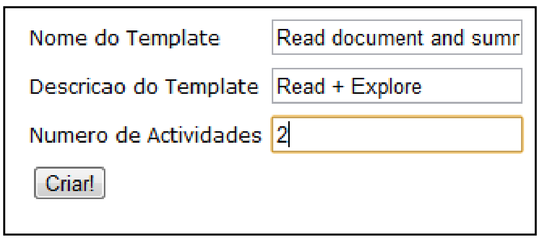
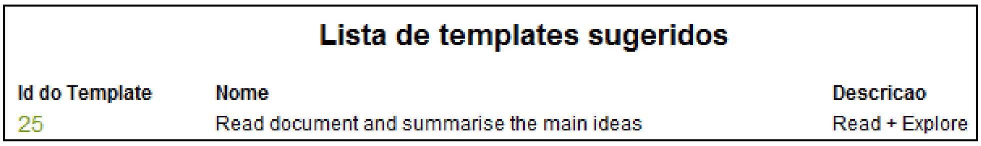


Figure 90. Creating a template

Once a template is created, it is then available for approval. The Pedagogical Adviser has access to a list of suggested templates (see Figure 91), and can see the details of each one.



Lista de templates sugeridos

Id do Template	Nome	Descricao
25	Read document and summarise the main ideas	Read + Explore

Figure 91. List of suggested templates

For a resource, the user decides on the template to be used (see Figure 92).

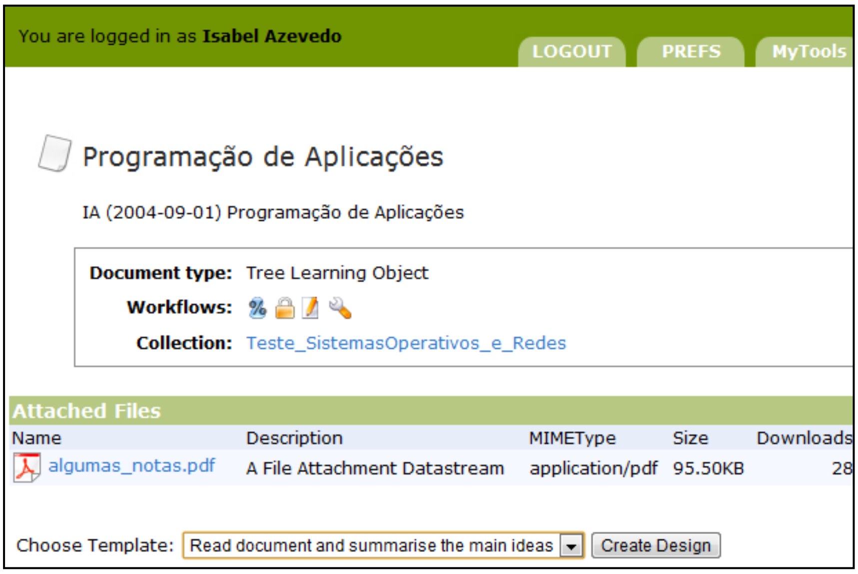


Figure 92. Choosing a template

A form is dynamically generated to instantiate the chosen template with the resource with some fields to be filled in. Then the system generates a UoL, an IMS CP package, which is a zip file containing the imsmanifest.xml file and the chosen resource.

The next example shows part of the xml document generated for the template represented in Figure 84. It has two activities: 'Read the document' and 'Summarise the main ideas'.

```
<imsld:activities>
  <imsld:learning-activity identifier="Learning1">
    <imsld:title>Read the document.</imsld:title>
             …
    <imsld:activity-description>
      <imsld:item identifierref="R1" identifier="I1">
        …
      </imsld:item>
    </imsld:activity-description>
    …
  </imsld:learning-activity>
  <imsld:learning-activity identifier="LA_RPI">
    <imsld:title>Summarise the main ideas.</imsld:title>
             …
    <imsld:activity-description>
      <imsld:item identifierref="R1" identifier="I2"/>
        …
      </imsld:item>
    </imsld:activity-description>
  </imsld:learning-activity>
  …
</imsld:activities>
```

Example 25. A UoL extract

The next example shows some elements of the LOM part (general category) in a generated manifest for an IMS Content Package.

```
           <imsld:metadata>
                  <imsmd:lom>
                    <imsmd:general>
                      <imsmd:title>
                        <imsmd:langstring xml:lang="pt">
                           Programação de Aplicações
                        </imsmd:langstring>
                      </imsmd:title>
                      <imsmd:catalogentry>
                        <imsmd:catalog>CASPOE</imsmd:catalog>
                        <imsmd:entry>
                          <imsmd:langstring xml:lang="pt">
                              treepid:77
                        </imsmd:langstring>
                        </imsmd:entry>
                      </imsmd:catalogentry>
                      <imsmd:language>pt</imsmd:language>
                      <imsmd:keyword>
                        <imsmd:langstring
xml:lang="pt">iSeries</imsmd:langstring>
                        </imsmd:keyword>
                      <imsmd:keyword>
                        <imsmd:langstring
xml:lang="pt">AS/400</imsmd:langstring>
                        </imsmd:keyword>
                      <imsmd:aggregationlevel>
                        <imsmd:source>
                          <imsmd:langstring
xml:lang="pt">LOMv1.0</imsmd:langstring>
                        </imsmd:source>
                        <imsmd:value>
                          <imsmd:langstring xml:lang="pt">2</imsmd:langstring>
                        </imsmd:value>
                      </imsmd:aggregationlevel>
                    </imsmd:general>
    …
```

Example 26. A UoL extract – LOM metadata

After the creation of the XML file, the system validates it using the IMS LD XSD, informing the user of its successful or unsuccessful completion. The generated IMS CP package, if valid, is then made available in the system and ready to be used in any tool that supports the IMS LD specification, for example, the .LRN platform [Cid, Valentín et al., 2007].

Another available option is the creation of an IMS CP package, but only containing the resource and its metadata.

The system produces UoLs using level A of the LD specification. The data model considers even the elements not currently in use in the prototype. For instance, although not regarded a learning objective for an activity alone, the related table has that field in order to support any further modifications (see Figure 93).

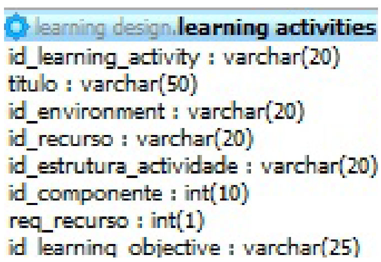


Figure 93. Partial view of the learning activities table

In the developed prototype only the possibility of a unique resource to be related to a template was considered. For instance, it is not possible to state that a document must be read (one activity) and a questionnaire (another resource) has to be then answered (a second activity). Two different UoL have to be designed for that purpose.

The approach was tested and the generated UoLs were successfully used in the players described in Annex 1. To be used in an LMS, if that possibility becomes a common practice, it is necessary to consider the specific users of interest, possibly the students that are following a module, for instance, and other detailed information. As stated in [Tattersall, Vogten et al., 2005] "the design time concept (the Unit of Learning) must be augmented with additional, runtime concepts". A UoL may have many associated runs, as "a run adds runtime information to a Unit of Learning by defining a start and end date and binding specific individuals into the roles modelled in the learning design part of the Unit of Learning".

8.2.2 Some remarks about the use of learning design templates

A potential advantage of the use of learning design templates is the reinforcement of good design practices. However, the user responsible for the associate functionalities, the Pedagogical Adviser, needs to have adequate credentials and experience. Also, it can be seen as a source of inspiration to less experienced teachers, or even to those with significant experience.

The presented strategy permits considering learning not only in a perspective of contents, but it is also regarded the activities that can be carried out using the resources.

Yule provided a possible definition of pragmatics as "the study of contextual meaning" [Yule, 1996], entailing the interpretation of what people mean in a certain context. Allowing users to specify a context for using learning objects contributes to their pragmatic characterisation.

The evaluation of this pragmatic approach was carried out by a group of 5 people, all familiarised with learning object repositories. Their opinions were collect in interviews. The functionalities regarding that approach were considered easy to use. In addition they found useful the approach to state how to apply a learning resource, but useless on the other side when noticed the impossibility to use the generated Units of Learning in an LMS, such as Moodle, presently in use at ISEP. Currently, that is the main drawback of the approach. The limited support to the IMS LD specification, especially by renewed learning platforms, does not facilitate its adoption. For instance, in the past Moodle announced for future versions some functions related to the specification, but they were not provided yet.

Finally, the current approach is not only restricted to the engineering field, used to validate it, but can be applied to others as well.

8.3 Use of tags

A social tendency has definitely reached learning environments. Researchers mention a new generation of students, the "Digital Natives" [Small and Vorgan, 2008], who would not know the world without the Internet. Frequently, their motivation for learning is social, e.g., they aim to find interesting information to make an impact on peers or to help with a task in a group work.

Teachers, students and staff are used to social networking functionalities. An innovative project in the e-learning field, called Metadata Ecology for Learning and Teaching (MELT) (http://info.melt-project.eu/ww/en/pub/melt_project/welcome.htm), ran from 1/10/2006 to 31/03/2009 and studied the impact of tagging by teachers, with very promising results. Although the tags were then reviewed by expert indexers, it was found that in 25% of the cases, the tags contained additional value for the repository, while in other cases many of the tags were already in some way represented in descriptors metadata. In its final report it is said that "at the end of MELT, we suspect that we are just beginning to see the tip of the social tagging iceberg as far as its potential added value is concerned" [Assche, Ayre et al., 2009].

Additionally, the EU-cofunded MACE project on Metadata for Architectural Contents in Europe (http://ec.europa.eu/information_society/activities/econtentplus/projects/edu/mace/index_en.htm), ran from 01/09/2006 to 31/08/2009, and developed a Web site for social search, available at http://portal.mace-project.eu/Tag#.

There is a DCMI Social Tagging Community (http://dublincore.org/groups/social-tagging/), which discuss "how tagging relates to the activities and practices of the Dublin Core general community, or how tags relate to other metadata".

The use cases related to the use of tags in the TREE repository are as follows:

- Visualise own tags (UC10);
    - List tagged resources (UC11);
- Visualise tag cloud (UC12);
- Tag resource (UC16);
- Visualise opinion summaries (UC52);
- Visualise resources' tags (UC53);
- Manage global tags (UC70);
    - Add global tag (UC71);
    - Modify global polarity (UC72);
    - Delete global tag (UC73);
    - List global tags (UC74);
- Manage local tags (UC75);
    - Modify local polarity (UC76);
    - Delete local tag (UC77);
    - List local tags (UC78);
- Manage improper tags (UC79);
    - Add improper tag (UC80);
    - Delete improper tag (UC81);
    - List improper tags (UC82).

These use cases were presented in Chapter 6 and they are detailed in Annex 3. They are related to the parts of the pragmatic component shown in Figure 94.

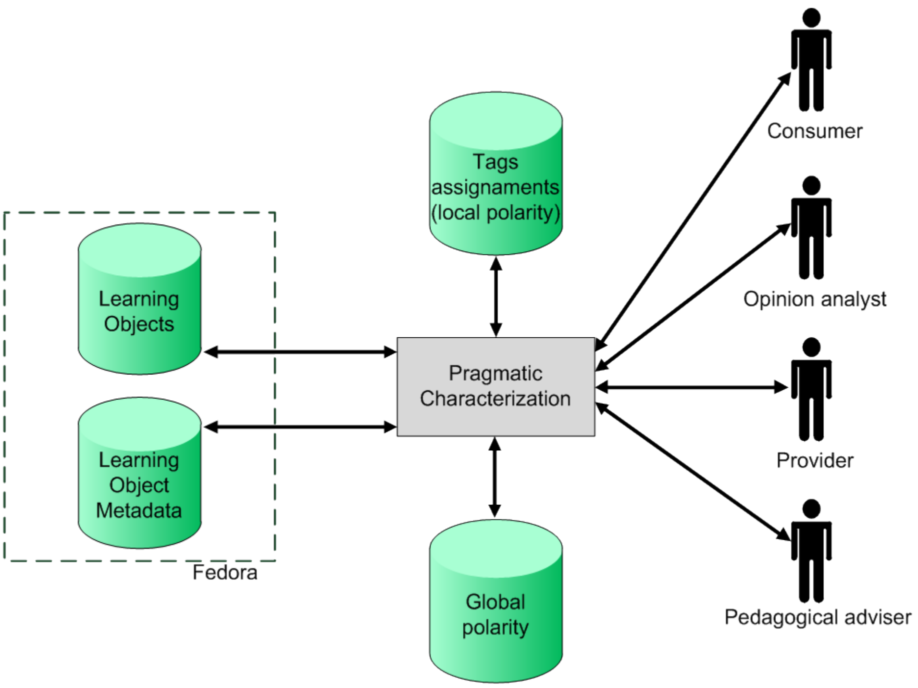


Figure 94. Pragmatic component – use of tags

8.3.1 The more traditional use of tags in the TREE repository

A user can assign a tag to a resource (see Figure 95). If that tag is chosen to be private ('privado' check box in the image), that tag is not visible to consumers and providers.

Figure 95. Adding tags – partial view

Although many institutions have a code of conduct that highlights the importance of using proper vocabulary and language, as tags assigned by users are visible to everyone, if they are public, there is a list of tags that cannot be applied in the TREE repository. If a user assigns an improper tag to a resource, that it not considered by the system, even if it is a private tag.

Under Advance Search menu it is possible to specify a number of criteria for the desired resources. The options can be used in conjunction and one of them can be used to specify the tags that a resource must have in order to be considered (see Figure 96).

Figure 96. Searching by tag

Many approaches for tags visualisation have emerged in the last years. Del.icio.us soup[73] and del.icio.us graph[74] can be used to show tag similarities. Cloudalicious uses a timeline to represent the chronological development of tags [Russell, 2006], while taglines[75] show "tags with floating thumbnails along a primary axis of time". More advanced tags visualisation approaches have also come out, like visualisation schemas that incorporate emoticons or have a 'topigraphical' style [Fujimura, Fujimura et al., 2008].

Tag cloud is one of the most notorious visualisation methods for folksonomies. The representation of tags is dependent on their usage frequency, with use of colours or a greater font size to signal popular tags. Tags in a tag cloud are hyperlinks to the resources that had that tag assigned. Because of its simplicity, that approach was adopted. The list of the most commonly used tags in the TREE repository is available in a tag cloud (see Figure 97). Its design was subject of a separate study [Silva, Azevedo et al., 2010].

Figure 97 also shows some statistics about tag usage. When that image was obtained, 91 tags were assigned ('91 Tags'), but only 80 represented different tags ('80 Words') and a total of 6 users have tagged 6 resources ('6 Tagged Objects').

Figure 97. Tag cloud in the TREE repository

In addition, users can see the tags that they used before in a tag cloud (see Figure 98). The tags in that tag cloud are hyperlinks but not to all resources with that tag, but to all resources that have that tag assigned by the user. In the image the same statistical information shown in the global folksonomy is available again. It can be seen that 4 tags were used, but only three different tags (designated as 'Words' in the figure). In the image the user who is visualising the tags previously used only tagged two resources ('2 Tagged objects').

[73] http://www.zitvogel.com/delicioussoup/screenshots.html
[74] http://www.cheesy.at/download/Folksonomy.pdf
[75] http://tagsonomy.com/index.php/taglines-tag-visualization-from-yahoo-research/

Figure 98. User's tags

That possibility can be used to support social bookmarking facilities. Social bookmarking systems provide links management functions, they are "a class of collaborative applications that allow users to save, access, share and describe shortcuts to web resources" [Braly and Froh, 2006]. Spiteri refers some problems to list favourites kept by browsers or other approaches that are desktop-dependent: "Our bookmarks or favourite lists are mushrooming out of control. Many of us have folders within folders within folders. We find ourselves bookmarking the same time a dozen times because we can't remember where we filed it. Alternatively, we simply 'Google' it to save time. [...] The problem is exacerbated when people use different computers (e.g., one at work, one at home, a laptop, etc.); they do not keep the same information across the different computers they use" [Spiteri, 2006].

The view provided in My Tags tab (see Figure 98) highlights the most used tags, but other possibilities could also be considered, as the most recently tags, for instance.

The MyDocs tab that is visible in Figure 97 and Figure 98 is used to provide some feedback to providers, concerning the resources that they had submitted, which is explained in the next sections.

8.3.2 A general proposal for considering users' opinions

Differently from ontologies, folksonomies are often used as a mean to express feelings or how good something is perceived. In a repository of learning resources, the latter should be more likely to occur than the former and could be used as a subjective indexing. People might want to read a resource that many others have considered an interesting one. Thus, folksonomies allow systems applying them to recommend (or not) tagged resources, implicitly or explicitly, which was highlighted by Golder and Huberman in a study focused on Del.ici.ous [Golder and Huberman, 2006b].

In regard to non-aboutness tags, "it is safe to say that tags are an important communication mechanism in addition to a categorisation technique" [Dennis, 2006].

Tags expressing a sentiment towards a resource are called 'emotional tags' or 'affective tags' [Kipp, 2007]. Yanbe and colleagues found that content tags are on average more common than sentiment tags (in a ratio of 10:1), after having analysed the use of tags in the Japanese social bookmarking service 'Halena Bookmarks' [Yanbe, Jatowt et al., 2007]. Figure 99 shows the top 54 'emotional tags' collected during their study.

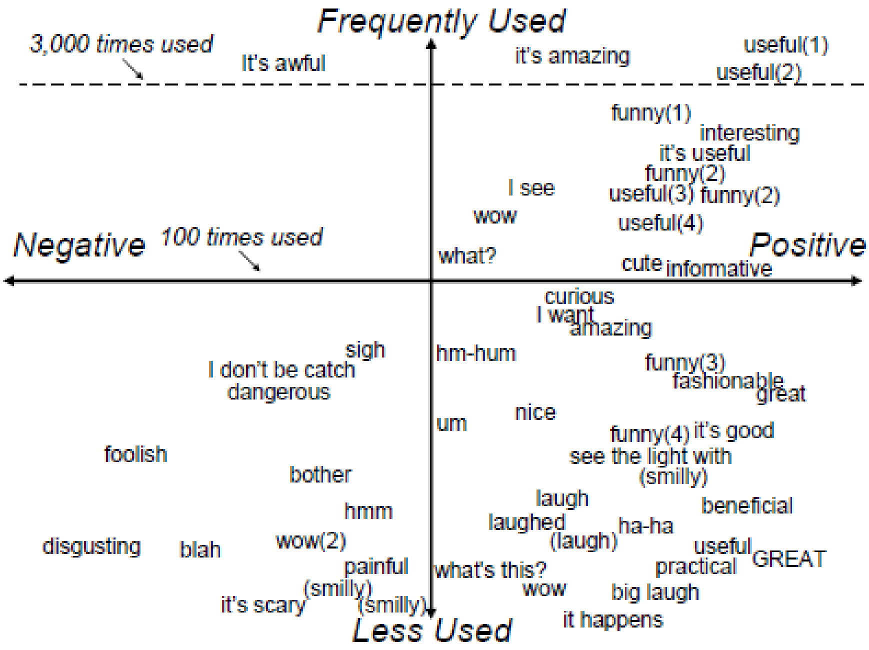


Figure 99. Top 54 emotional tags used in a Japanese Social Bookmarking Service (Source: [Yanbe, Jatowt et al., 2007])

Due to the existence of more positive sentiment tags than negative ones, the authors concluded that users usually do not tag resources to which they have negative feelings. However, in their study a Japanese service is used and the predominance of positive sentiment tags may be caused by cultural characteristics of the Japanese people (see section 3.5).

The quality of the resources in a repository can be highly variable. In this section a general approach that uses opinions expressed through tagging is presented.

At least two tables, TG and TB, have to be considered, each one enclosing tags that could be used to rate a resource as good (in TG), meaning that the resource was appreciated, or bad (in TB), meaning that the resource was disliked. Some of the tags that can appear in TG are: good, toread, ok, cool, fine, funny, fun, great, fantastic, among many others. In TB there might be tags like bad, uninteresting, dump, notinteresting, boring, geek, nerd, for example.

The tables TB and TG need to be populated. But a tag management module must ensure the possibility to add or remove tags.

TB and TG can be constructed by inspecting tag assignments in a similar system[76] or the system in consideration if tagging is not a new functionality. Some sentiment lexicons can also be regarded for that purpose.

A possible ranking of the resources is made by the subtraction of the number of 'bad' tags attributed from the number of 'good' ones. If that number is positive, the resource is well-regarded, otherwise, it is not.

The hypothesis to have a scale assigned to each tag in order to have different degrees of good or bad can be considered. For example, 'excellent' should have a stronger positive polarity than 'good'.

That approach can be used to have a rating of the resources. By explicitly asking users to rate a resource, using a number in a predefined interval or choosing the number of stars (typically from one to five stars), it is also possible to achieve similar results, but it should be difficult to decide if a resource deserves one or two stars, for example, but easy to say if it was boring or wonderful.

[76] For instance, tags posted in a Social Bookmarking System (http://delicious.com) from October to December 2003 can be obtained from a large dataset (http://www.dai-labor.de/index.php?id=1726&L=1) [Wetzker, Zimmermann et al., 2008].

Some of the main advantages of the suggested approach are as follows:

- It is possible to provide a summary information about a given resource,
- In a repository, the providers can see what was liked and disliked and consider the given opinions when submitting a new version.

Its main drawback is that it is impossible to have a list of all possible tags for TG and TB tables. But it is possible to consider the known tags in each category and analyse the tags provided to collect news terms for each table.

8.3.3 Tag Polarity Assessment in the TREE repository

To fully benefit from the use of Learning Objects it is necessary that they are stored in repositories with some essential functionality of storage, tracking, maintenance and retrieval.

We believe that the ability of users to assign tags to resources in a repository can contribute to enrich their characterisation in a more comprehensive way, which was the reason to add the tagging functionality to the TREE repository. Some educational tagging systems are the Calibrate portal, currently known as the Learning Resource Exchange (LRE) portal[77], the LeMill Web community for finding, authoring and sharing open educational resources[78], and the Open Educational Resources (OER) Commons platform[79]. A recent study reported that the three tags most often used in LRE portal were *good foreign resource*, *average foreign resource* and *weak foreign resource* (their equivalent words in Hungarian were applied – here translated) [Vuorikari, Poldoja et al., 2010]. These tags clearly reflect an opinion about the tagged resources, although none of the mentioned repositories uses these opinions, at least in a more automatic way.

Opinion mining in LOR can be useful to highlight resources that need to be improved, helping to achieve a high quality repository. The rationale for the use of opinion tags for this purpose is provided in section 8.3.3.1, while subsection 8.3.3.2 provides a detailed description of which kind of tags were considered as reflecting an opinion about the resources, and how they are processed. Some recognised improvements that can be considered and some issues that need to be further investigated are discussed in subsection 8.3.3.3.

8.3.3.1 Motivation

Over the years many learning objects repositories were developed. The Multimedia Educational Resource for Learning and Online Teaching (MERLOT) repository, available at http://www.merlot.org/merlot/index.htm, is one of the most often cited Learning Objects Repositories (LORs). At the time of this writing it encompasses 28.052 materials, while in 2005 it included about 10,730 learning objects [MacLeod, 2005].

MERLOT has been developed and provided by the California State University Center for Distributed Learning, and it has been available since 1997. It can be seen as an example of *architectures of participation*[80], such as Wikipedia[81], as any user can register and then contribute with resources to the repository. However, it is interesting to note that although MERLOT is often described as a "peer reviewed repository" (see [Northrup, 2007] (page xi), for instance), in 2005 near only 14% of the materials submitted were actually reviewed [Hanley, 2005], a value that decreased to 13% in 2009 - guez, Dodero et al., 2010], being now 11% (3,167 of 28,052 resources). It seems that as the number of items augments, it has been difficult to maintain the

[77] http://lreforschools.eun.org
[78] http://lemill.net/
[79] http://www.oercommons.org/
[80] Tim O'Reilly first used the expression when discussing open-source software development (in T. O'Reilly. (2003). "The Architecture of Participation" O'Reilly Developer Weblogs. Available at http://www.oreillynet.com/pub/wlg/3017.), but it was later extended to Web 2.0 tools, or platforms where the user has a leading role in the sharing of resources.
[81] http://en.wikipedia.org/wiki/Main_Page

peer-review policy, since the review of resources demands considerable investment of time and effort: a study from 2002 estimated the average word count of Amazon reviews as 181.5 words [Ketzan, 2002].

In additional, another important point is how many users contribute to the reviews in order to make the process more reliable. We carried out some queries in order to have an idea of its current review policy and the feedback that those who submit resources are receiving (see Table 47). For instance, searching for *database* in *science and technology* category, returned 170 results, but none of them were reviewed by more than one person. In fact, the entirety of the reviewed resources for each search term was reviewed by only one person.

Table 47. Data on searches carried out in the MERLOT repository – the review practice

Search term	Category	Results	Resources reviewed	Resources reviewed by only one person
database	science and technology	170	29	29
programming	all	204	10	10
global warming	all	40	1	1
evaluation	education	61	8	8
matrix	mathematics and statistics	35	11	11
Europe	all	64	8	8

From the data in Table 47, it does not seem that the people who upload resources to the MERLOT repository are getting sufficient feedback from other users, through any repository facility.

Other repositories seem to face the same problem. Although Intute[82], a free online service provided by a consortium of seven UK universities, is described as having "reviewed and evaluated thousands of resources" to help users choose key websites in their subject, it seems that the review process was done for selection of which resources to include in the repository. Another repository, the Digital Library for Earth System Education[83] (DLESE), which comprises 13,380 resources about the Earth system at all levels, has a specific collection with 1,059 resources (near 8% of the resources) that were reviewed and found to meet seven specific review criteria.

It appears that the review of resources is not a process massively carried out due to the expenses involved. However, that activity is important to assess the quality of the available resources and to have those who submit resources knowing aspects that can be improved.

MERLOT has also a rating mechanism. Table 48 has essentially the same data as the previous one (the category column is not shown for conciseness), but shows the number of rated resources. Although having the contribution of more users, still, most of the results are rated by only one user. In addition, the resources' providers only have a final number, without any insight into what might need to be improved.

Table 48. Data on searches carried out in the MERLOT repository – the rating practice

Search term	Results	Resources reviewed	Rated resources	Resources rated by only one user
database	170	29	26	18
programming	204	10	48	30
global warming	40	1	5	4
evaluation	61	8	5	2
matrix	35	11	6	4
Europe	64	8	8	6

[82] http://www.intute.ac.uk/

[83] http://www.dlese.org/library/index.jsp

Thus, having realised the problems of the current approaches adopted in many repositories to estimate how resources are regarded by users, other possible alternatives started to be considered in the TREE repository.

8.3.3.2 Tag polarity classifier

Adjectives are often applied in evaluations [Boiy, Hens et al., 2007], and previous research works have demonstrated the existence of a strong positive correlation between the use of adjectives and the presence of opinion [Hatzivassiloglou and Wiebe, 2000]. A category of tags previously discussed is the one that aggregates mostly adjectives provided by users, expressing their opinion about the resources.

It is interesting to note that tags like *to_read*, *todo*, *towatch*, among others, describe a planned action in regard to the tagged resource, embodying a performative act (see section Speech acts). The use of any Illocutionary Force Indicating Devices (see section 2.1.1) is not common and, therefore, it might be not clear if the user is doing a promise (to read something), a suggestion (to others read something), or stating a plan (to read something), for instance. Some researchers refer to this kind of tags as *functional tags* [Kroski, 2005], *signalling tags* [Dennis, 2006], and *task related tags* [Kipp, 2007]. The tagged resources are likely to have been regarded as useful as these kinds of tags “express a response from the user rather than a statement of the aboutness of the document; […] they suggest an active engagement with the text, in which the user is linking the perceived subject matter with a specific task or a specific set of interests” [Kipp and Campbell, 2006].

In order to provide some feedback about the resources to the users who submitted them, we considered the opinions provided through tags assignment. To this purpose the tags under the categories *functional tags* and *subjective tags* were considered in the system as both of them are used to signal directly or indirectly an opinion about the resources. Recently, in a small sample, considering the positive and negative opinion tags, it was found that the percentages of positive (Spearman correlation: Rs =0.856, p = 0) and negative tags (Spearman correlation: rs =-0.739, p = .001) assigned to resources at the online library Amazon (http://www.amazom.com) were related to the average score given directly by users in a five-star rating system [Morim and Azevedo, 2011]. In this study, also subjective and functional tags were considered.

Among other opinion lexicons, the subjectivity lexicon was chosen [Wilson, Wiebe et al., 2005] for its availability under a GNU General Public License, simplicity and its successful usage in a number of research projects, such as [Breck, Choi et al., 2007; Gyamfi, Wiebe et al., 2009; Klenner, Petrakis et al., 2009], among many others.

The selected lexicon has 8,220 subjectivity single-word clues in English, which have subjective usages. Wilson et al. studied, as part of their research, if words appear to induce something positive or negative out of any context. The clues in the subjectivity lexicon are classified as positive, negative, both or neutral. 92.8% of them are marked as having a prior polarity, positive (33.1%) or negative (59.7%). For instance, *interesting* and *boring* are words classified as having a positive and a negative polarity, respectively.

However, the subjectivity lexicon also considers some others parts of speech, besides adjectives. It can also provides a list of verbs (*refuse*, for instance), adverbs (e.g., *successfully*) and nouns (e.g., *garbage*), and their polarity.

In the TREE repository users utilise mainly two languages: Portuguese and English. Thus, the Google Translate 0.7 API[84] was used to collect possible translations in Portuguese for the subjectivity clues, which were translated in isolation.

The clues from the subjectivity lexicon in English and in Portuguese were the basis of the *Multilanguage Opinion polarity lexicon* (see Figure 101), after their translation revision. In that revision, three people analysed the translated words in order to detect linguistic peculiarities potentially affecting their polarity, but for some words it was chosen an alternative synonymy in common use

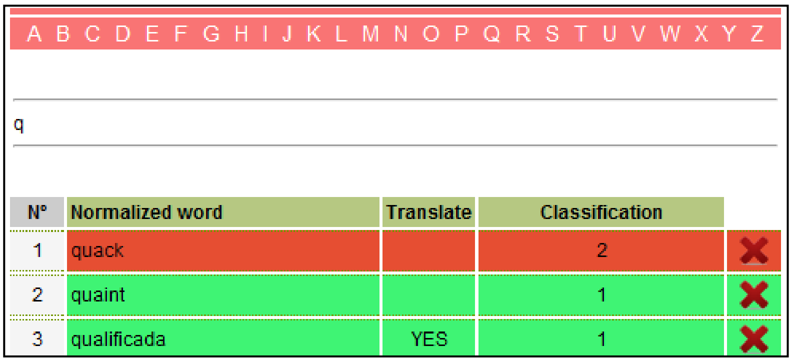
[84] http://code.google.com/p/google-api-translate-java

in Portugal, instead of the suggested translation closer to the Brazilian Portuguese. For instance, for the English word *mess* the initial suggestion was *bagunça*, but the word *confusão* was used instead.

Also, many English words have only one form for masculine and feminine, which is not the case in Portuguese language. Thus, it was necessary to consider both forms in Portuguese. For instance, for the Portuguese word *qualificada* (see Figure 100), it was not necessary to considerer its masculine equivalent (*qualificado*). It was not detected any other kind of problems, and even those did not affect the assigned polarity, but might cause some words to be disregarded by the automatic classifier. However, this fact does not mean that the revision is not necessary if that approach is used to have polarity lexicon in other languages as well.

In [Banea, Mihalcea et al., 2008] the subjectivity lexicon used in this work was also translated for similar purposes. The authors concluded that "machine translation offers a viable alternative in the construction of resources and tools for subjectivity classification". In their experiments the new target languages that were analysed were Romanian and Spanish.

Figure 100 shows some of the words that begin with the letter Q. Those words that were obtained by translation are signalled (column 'Translate'). Their polarity is visible in column 'Classification' (1-Positive, 2-Negative and 3-Neutral).



Nº	Normalized word	Translate	Classification
1	quack		2
2	quaint		1
3	qualificada	YES	1

Figure 100. List of tags

Their polarity was considered, but not if they are strongly or weakly subjective (the type value in the lexicon considered). However, some others tags were regarded, in addition to those words from the subjective lexicon. In a previous work [Silva, Azevedo et al., 2010] we had access to a dataset from Delicious (http://www.delicious.com/), a social bookmarking service, and many tags were selected from it. However, it was impossible to understand the meaning of some tags, such as *peep*, *suli*, *jegyek*, *kiina*, *kulttuuri, kiina*, *kulttuuri*, *sanasto*, *lex*, *kanada*, *sieci*. They could correspond to words in other languages than English, acronyms, abbreviations or even to very impolite words.

The literature review [Zollers, 2007] [Kipp and Campbell, 2006] [Kipp, 2007] also provided some additional tags to consider and insert in the opinion indicators table, such as *trash, ripoff,* and also tags with two or more words, e.g. *waste of time*, among others, with their recognised polarity.

In Figure 101 the architecture of the opinion classifier module is visible. For each tag assigned to a resource, the system verifies if the tag is part of a metadata field (keyword or title) associated to the resource (*Contextual Information* in Figure 101). In that case it always considers the tag to have a neutral sense. Otherwise it is considered the opinion, the negation and the intensifier indicators. For instance, the tag *garbage* assigned to a resource from the community *Computer Science* whose title is *Garbage collection* is always considered to be neutral, despite having a negative global polarity. That way we are regarding the context in which the tag was applied. In addition, tags such as *sleep* or *tosleep* applied to resource whose title is *Understanding Sleep and Dreaming* is considered to be neutral, but assigned to a resource which introduces some computer programming concepts, it is regarded as negative, which is the global polarity they have in the Multilanguage Opinion Polarity Lexicon.

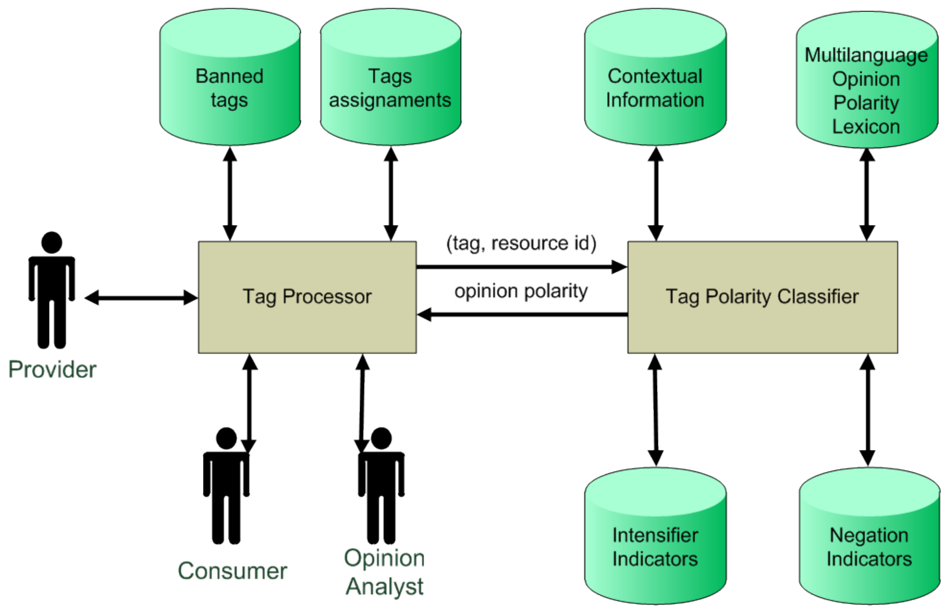


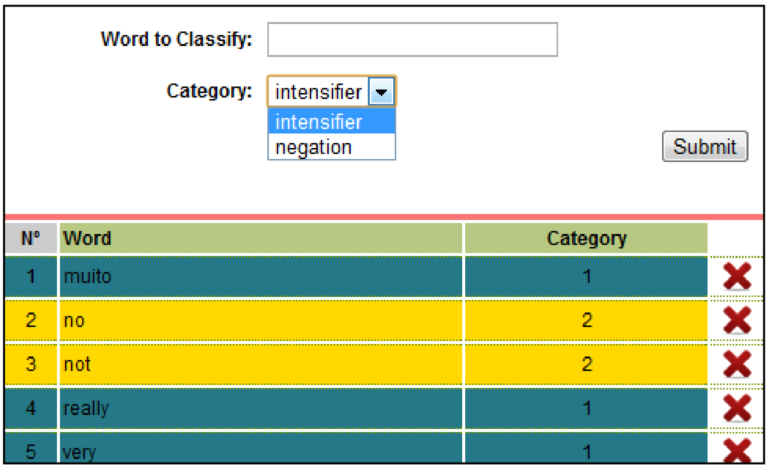
Figure 101. Architecture of the automatic opinion classifier

Moreover, in the polarity analysis there are two groups of words that receive a great attention: intensifier and negation words, represented in Figure 101 respectively as *intensifier indicators* and *negation indicators*.

The use of negation words (e.g., *not* in *not good* or *no* in *no good*) changes the polarity to the opposite of the other words in isolation. Thus, having the word *good* a positive polarity, the tag *not good* is considered to have a negative polarity.

The use of intensifier words, such as *pure* in *pure garbage* or *very* in *very good*, also *so*, *oh*, among others, do not change the polarity, but makes the opinion stronger.

Besides the negation and intensifier words in use in the TREE repository, the opinion analyst can add others as well, or remove (see Figure 102: 1 is used for intensifier words and 2 for negation ones), which can be particular useful in case of a previous misunderstanding.



Nº	Word	Category
1	muito	1
2	no	2
3	not	2
4	really	1
5	very	1

Figure 102. Management of intensifier and negation words in the TREE repository

The tags provided by the users in the TREE repository are visible under different tags clouds, mainly in the repository homepage, where users can see ‘their’ tags clouds (accessible in the tab *MyTags* – see Figure 103), with the tags they had previously provided. Under the tab *MyDocs* (see Figure 103) each user has access to a summary of the tags assigned to the resources he/she had previously submitted to the repository. For each resource the number of positive, negative and neutral tags is visible. Also, there is a hyperlink to a web page where the tags can be inspected in detail.

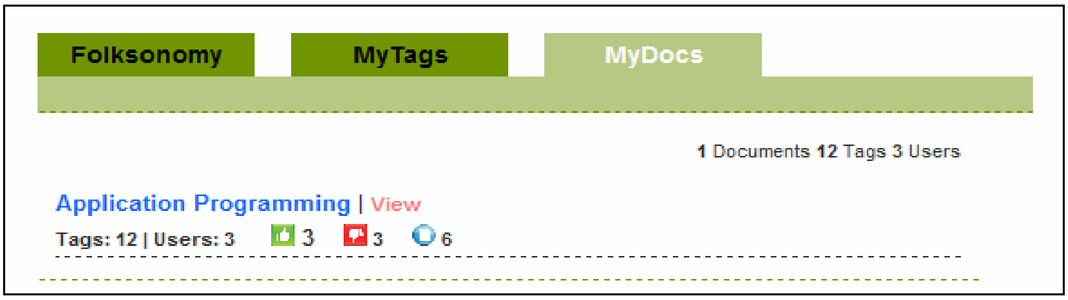


Figure 103. Visualising an opinion summary for a resource – a partial view

Thus, besides the summary about the opinions provided by other users with the number of positive, negative and neutral tags assigned, a more detailed information with the tags supplied and how their polarity are regarded by the system (through the hyperlink 'View' shown in Figure 103) without the taggers' identification. A short extract from the tags provided to a resource whose title is 'Programação de Aplicações' is shown Figure 104. The neutral tags are in blue, the positive ones in green and the negative tags are shown in red.

Programação de Aplicações

Original word	Date
IBM	28/05/10 06:36 pm
AS/400	28/05/10 06:36 pm
AS/400	04/06/10 05:56 pm
PDM	28/05/10 06:36 pm
toRead	28/05/10 06:36 pm
ler	28/05/10 06:37 pm
ler	07/07/10 03:45 pm
confuso	28/05/10 06:37 pm
boring	28/05/10 08:06 pm
boring	05/11/10 06:49 pm

Figure 104. Part of the tags assigned to a specific resource, as visible to the provider

However, although considered to the accounting of positive and negative tags assigned to a resource, if a tag is private it is not shown to the provider. In that case, the opinion summary can show a number of tags that is greater than the quantity that appear in the list accessible to providers.

Once a day the automatic classifier runs and checks the new tags to decide about their polarity. In any case it is possible to manually change the assigned polarity, but that functionality is only available to the Opinion Analyst (see Figure 105). That option is especially useful when an error in the assigned polarity is encountered, probably by the resource's provider.

User ID	Original word	Classify Original	Classify	Private	Date
adela	Eye movement	neutral	neutral	NO	09/07/10 12:23 pm
1090016	Eye Tracking	neutral	neutral	NO	08/07/10 04:30 pm
1090016	Concentration	neutral	neutral	NO	08/07/10 04:30 pm
1090016	Comprehension	neutral	neutral	NO	08/07/10 04:31 pm
1090016	Heat Maps	neutral	neutral	NO	08/07/10 04:32 pm
1090016	Session Maps	neutral	neutral	NO	08/07/10 04:32 pm
1030483	publicity	neutral	neutral	NO	09/07/10 11:21 am
1030483	website design	neutral	neutral	NO	09/07/10 11:51 am
adela	looking	neutral	neutral	NO	09/07/10 12:21 pm
adela	eye tracker	neutral	neutral	NO	09/07/10 12:23 pm
adela	vision	neutral	neutral	NO	09/07/10 12:25 pm

0 0 11 0 Submit

Figure 105. List of tags assigned to a certain resource – opinion analyst view

8.3.3.3 Possible enhancements

Recognising that *excellent* might comprise a more positive opinion than *good*, an improvement to this work can be the consideration of different degrees of positive or negative in the opinion summariser. SentiWordNet 3.0 is a lexical resource for sentiment classification and opinion mining applications [Baccianella, Esuli et al., 2010] that can be applied for that future improvement.

In SentiWordNet 3.0 each synset of WordNet is associated to three numerical scores Pos, Neg, and Obj, respectively related to the notions of *positivity*, *negativity*, and *neutrality*. These scores vary from 0 to 1, their sum being always 1. For instance, for the word *excellent* the only synset that comprises it has the following scores: Pos=1, Neg=0 and Obj=0, while the best positive score for a synset that include the word good is 0.875 (Neg=0, Obj= 0.125). The same is valid to the use of intensifier words: they are recognised but currently they do not affect the opinion strength, although *very good* might express a more positive opinion than *good*.

Yet related to the problem of determining opinion strength, currently we do not address what Yus called "textual deformation", since it was not found necessary.

In addition, the use of deictic expressions in tagging systems needs to be analysed in future research works. Deixis was introduced in section 3.2.1. Some tags use deictic expressions[85] (*this*, *that*, *now*, *there*, among others). It is not clear if users apply them to mark resources as adequate to specific activities to be performed at some time or place, and thus indirectly expressing an opinion about the tagged resource.

8.3.4 Some remarks on the use of tags

For the consideration of users opinions provided through tagging, cultural differences have to be regarded if the results in one repository are to be used in comparison to those from another repository, especially if their users are from distinct cultures. But these cultural differences may also affect the observation of the results of polarity analysis using resources from two different courses if one of them has a great number of students from other countries. Namely, politeness and irony are phenomena subject to cultural variations that can affect how resources are tagged. For instance,

[85] See, for instance, some resources tagged with this at the del.icio.us system (http://www.delicious.com/tag/this).

using a tag such as 'very good' to mean the opposite (irony) might affect the polarity analysis, especially if this tagging behaviour is frequent.

However, in face-to-face communications there are generally clues that suggest the use of irony. Tags are generally short expressions composed by few words, and the contextual information is limited. Thus, in this work it was supposed that the set of assumptions guiding a conversation suggested by Grice might also apply to tagging to some extent (see section 3.2.2), namely the following rules:

- Do not say what you believe to be false,
- Do not say that for which you lack adequate evidence,
- Make your contribution to the conversation as informative as necessary,
- Make your contributions relevant,
- Avoid obscurity,
- Avoid ambiguity,
- Be brief,
- Be orderly.

If irony can be difficult to detect in tagging, it is not expect to occur frequently, which might not be the case of the use of politeness and the cultural differences in its usage. In fact, disparity in the use of politeness could affect the reading of the results if they are to be used to in comparisons involves users from very different geographic places. For instance, German speakers seem to be significantly less polite in complaints than English speakers [Levinson, 1983] (page 376).

Sentiment analysis deals with the identification of positive and negative opinions, emotions, and evaluations. A number of research works dedicated to sentiment analysis at sentence and document levels has been carried out, but in this chapter a novel approach which considers the tags provided by users to have an opinion summary in repositories was proposed.

To highlight functional problems many tests were carried out during and after the development. The detected errors were corrected.

In an evaluation exercise, 5 people browsed the resources in the TREE repository for 30 minutes. They received the indications to analyse 5 resources under Multimedia category and explore any desired functionality. They did not receive any other instruction. Three of them naturally assigned tags to some of the resources, but they only provided neutral tags. The dimension of the group involved in the exercise was too small to reach definitive conclusions, and also the time available for it. But the fact that the users provided tags to some resources may mean a desire to collaborate, when that option is available, potentially beneficiating the whole community.

Table 49 summarises the results. For all the resources, more than 80% of tags were not used by the provider as keywords. Only two tags were provided as private.

Table 49. Exercise results – a summary

Resource	Number of tags	Number of tags that were not provided as keyword	Number of relevant tags that were not provided as keyword	Number of distinct tags not enclosed in the keyword metadata field
1	17	14	11	13
2	11	10	3	10
3	19	17	9	14
4	18	16	12	13
5	18	15	4	14

However, the providers of these 5 resources (fourth column) did not consider all the tags as relevant. More than 70% of the tags used for two resources (resources 1 and 4) were considered appropriate, but for resource 2, for instance, only 30% received the same classification.

The will to collaborate and the possibility to assign their own tags to resources (bookmarking them) were their main motivations for tagging, according to what the users said in semi-structured interviews. In the controlled experience all the tags provided were neutral.

Further research is necessary to assess if tags can enrich the traditional metadata provided when resources are uploaded, but they can certainly be used in different ways, such as a source of opinions about resources, complementing or not other adopted strategies.

However, the appropriateness of the proposed approach needs to be validated thoroughly, not in a single study involving only one repository, but with a broader scope. It is expected that other works can also provide additional perspectives on the approach proposed in this chapter.

Finally, this work can be extended and applied to other tagging systems, not only for learning objects repositories, but the contextual information to consider need to be well thought out.

8.4 Conclusions

The interest in pragmatics has grown over the years and this study area came to other fields beyond linguistics and philosophy of language. This chapter stated how some concepts related to pragmatics can be applied in a learning objects repository, considering a context and the two sides of a communication process – what was intended (by providers) and what was perceived (by consumers). Also, as covered in Chapter 3, communication does not only comprise words, but also emotions, which may contradict or reinforce the message being transmitted, even in computer mediated communications.

This chapter discussed the pragmatic approach to a learning object repository through two different strategies:

- The use of learning design templates to provide a pedagogical and practical context to resources, considering a number of specifications;
- The use of tags to uphold a more complete understanding of how resources are regarded by users.

They can be adopted in isolation, as they are not interdependent. However, the former allows integrating the provider's view, while the latter permits the consideration of the consumer's view. Moreover, their use in conjunction allows a characterisation of the resources in a wide perspective.

The point of using learning design templates supporting a pedagogical context and their utilisation for generating UoL is not to support instant reuse, but to sustain reuse with conscience. The final users are still responsible for instructional decisions and they have to carefully think about them. Any necessary edition can be done in simple text editors, but demanding a great knowledge of the IMS LD specification, or IMS LD editors (see Annex 1), that vary in the degree of ease of use. The use of some fields of the TREE application profile complement the approach, namely the fields 1.4.Description and the fields under the Educational category (see section 6.4.1).

Through the use of tags, consumers can provide their opinion about different characteristics: contents, usefulness, ease of understanding, among others. That diversity allows capturing various aspects of the tagged resources, which can have different applications. In the TREE repository it was used to support bookmarking functionalities, new retrieval options (based on tags) and to give feedback to providers. The approach can also be used to introduce user satisfaction in conjunction with other variables in the ranking of results when users search for resources.

As far as the author of this document knows, this is the first approach to suggest using opinions expressed through tags assignments in a learning objects repository. With opinion summarisation through tags analysis, a user can easily see how the others feel about a resource, and can get the reasons why they like it or not, providing to the person who submitted a resource some insights about what can be enhanced. In a repository that summary can also be useful to highlight resources that need to be deeply revised in less automatic ways, helping to improve the quality of the resources.

Chapter 9

CONCLUSIONS

This chapter summarises the work developed in this thesis, referring to its contributions and the achieved objectives. Its main limitations are also presented, and some aspects that might be the basis for future work in the area are described. This chapter finalises with some concluding remarks.

9.1 Introduction

The research questions (see Chapter 1) were answered along this document. To overcome some limitations of learning objects repositories, this thesis proposed a model for repositories that incorporates semantic and pragmatic concerns in the representation of learning resources.

After a brief introduction that mainly discussed the research questions and objectives, in the second part of this document different research fields were examined. With the exception of the second chapter, which was dedicated to Learning Technology Standards, Specifications and Models, the other chapters were not directly related to repositories or learning resources, but they provided insights into some fields (Pragmatics, Knowledge representation and Engineering, and Information Retrieval) in areas where these fields have been comprehensively studied.

In the third part of this document the literature review was more directed to the problems being addressed since that part was concerned with the presentation of possible answers to the more specific research questions that were previously introduced.

Chapter 6, besides presenting the proposal for a repository model, also served as an introduction to Chapter 7 and Chapter 8, respectively dedicated to the semantic and pragmatic components of the model. The presentation of these components in distinct chapters was useful mainly for the document organisation. However, for instance, tags also provide an understanding of contents of resources, and they can also be regarded as highlighting their semantics. Also, the use of context in some of its sub-components was also necessary for the semantic component. In fact, as it was previously mentioned, the boundary between semantics and pragmatics is subtle. Salmon classifies the classical distinction based on context as being “more of a slogan than a clarification or explanation” [Shalmon, 2005].

Lastly, this chapter concludes the main body of this document. It is structured as follows. Section 9.2 describes the main contributions of the thesis. Some limitations of the work are detailed in section 9.3. The most challenging open research issues are described in section 9.4. Finally, section 9.5 provides some ultimate remarks about the work, including a more personal view.

9.2 Main contributions

This thesis presented a research on the characterisation of learning objects in repositories. In particular, the concepts of semantics and pragmatics applied to learning objects were investigated, as well as learning technology specifications and their suitability for the description of learning resources.

How to consider knowledge in repositories and to make possible the use of semantics in the description of learning resources and in their search and retrieval was also studied. In addition, and complementing that previous approach, a method to obtain knowledge lying in ontology reuse was proposed and a tool was developed for that purpose. The proposed method considers online ontologies and in one of its steps considers semantic relatedness to assess the most appropriate synonyms.

Not only was the use of ontologies considered, but also the use of folksonomy to permit a wide description of resources. The tags provided by users allow considering diverse aspects of learning resources, which can be applied to provide feedback to those users that submit resources, or even to highlight resources that may need an in-depth revision.

In addition, a learning design specification was used for the purpose of stating how to use learning resources.

A prototype was developed, which integrated the approaches for semantic and pragmatic characterisations of learning resources, assessing their viability.

To sum up, the main contributions of this thesis are the following:

- A comprehensive review of previous studies related to the subject of this research was provided (objectives 1, 2, 3 and 4);

- The application of knowledge in repositories was proposed and implemented with a dual objective of facilitating the description of resources in accordance with the IEEE LOM standard, and also their search and retrieval (objective 2);
- A method for reusing ontologies was proposed, and a tool was developed (objective 3);
- The notion of pragmatics in repositories was put forward, considering the provider and the consumer. A novel approach to capture users' opinions on resources was suggested, regarding tag assignments. A learning technology specification was used to state the providers' opinions of how to use the resources (objective 4);
- A model for repositories was proposed to put pragmatic and semantic concerns in the representation of learning resources. A prototype that considers the proposed model and all the previous points was developed and validated (objective 5).

The objectives mentioned in the previous list are those presented in section 1.2.

The main goal of this thesis was achieved and a general conclusion can be drawn from this work: it is possible to consider semantics and pragmatics in learning objects repositories to support the broad description of learning contents and additional services, benefiting their users.

9.3 Limitations

Usability, "the effectiveness, efficiency and satisfaction with which specified users can achieve specified goals in particular environments" [ISO/IEC, 1998], was not directly related to the research questions of this work and, thus, it was not a concern during the work carried out. However, usability needs to be considered in repository platforms if they are to be adopted and widely used.

The next two subsections address more specific limitations of the work.

9.3.1 Semantic component

The ideas underlying the method for ontology reuse can be applied in other areas and it can be expanded to consider other Semantic Web Search Engines. However, the developed tool is integrated in a LOR and lies in a single Semantic Web Search Engine (Swoogle), as explained in Chapter 7. These options were satisfactory to the thesis objectives, but not to a more mature solution to be widely available to other research projects, or to be used in other fields. Moreover, the unavailability of Swoogle or changes in its query processing can compromise the operation of the ontology reuse module, but it can be adapted to other SWSEs as well, which can alleviate the problem.

The semantic relatedness component of the ontology reuse tool considers direct relationship between concepts through the use of DBpedia resources. However, if a concept C1 is related to a concept C2 that mentions concept C3 in its abstract, the approach does not find a relationship between concepts C1 and C3. The approach followed can be extended to consider indirect relationships up to some level also providing a value for the semantic relatedness between concepts, after studied its viability.

9.3.2 Pragmatic component

As explained in Chapter 8, cultural differences may also affect the comparison of the results of the polarity analysis using resources from two different courses if one of them has a large number of students from diverse geographic regions.

In addition, the use of irony through tagging is not expected to occur frequently, assuming that the set of assumptions guiding a conversation suggested by Grice might also apply to tagging (see section 3.2.2).

The use of the IMS LD specification itself might be seen as a limitation, not properly on the thesis itself, but on its immediate application. The specification compliance does not reach the desired level and the most used learning platforms do not support it. However, it is expected that the situation will change in the future (see Annex 1).

9.4 Future work

The limitations discussed in the preceding section can also be seen as research opportunities.

Moreover, this thesis raised many questions that were considered out of its scope. Some of the most challenging ones are summarised as follows:

- In this research work, folksonomies and ontologies were applied but in isolation. An integrated approach can be used to explore their conjugated application, where one can fully benefit from the other;
- Not only the types of tags but the behaviour of users in regard to tagging can be useful in repositories. The logging of tagging details can be useful to assess the kind of expansion more useful when submitting resources or queries to a LOR. In addition, how tagging information should be incorporated in learning technology specifications will be an important point to consider in the future;
- The use of semantic relatedness can be expanded to consider other terms that are not synonyms of those originally provided, but are semantically very close to them. This approach can be useful in queries submitted to a LOR, but also in the discovery of ontologies of interest;
- Ontology reuse approaches can benefit from a wide consideration of semantic relatedness between terms, but also lexical variations. Two sides can be regarded: an effort can be made in a consumer application side (the one that is followed in this research work but that can be extended as described in the previous section), but even SWSEs can be improved to deal with a number of situations that currently are not addressed (see, for instance, section 7.2.1).

To deal with these points two research projects were submitted in 2011 to the Portuguese Foundation for Science and Technology: FLOR - Folksonomies for Learning Objects Repositories (PTDC/EIA-EIA/117071/2010) and ARO - Automating the Reuse of Ontologies (PTDC/EIA-EIA/122179/2010). The researcher actively worked in their scientific descriptions and substantiations.

9.5 Final remarks

The approval of a research project with the preliminary ideas of this thesis in 2007 (PTDC/EIA/65387/2006 'CASPOE – CAracterização Semântica e Pragmática de Objectos Educativos') is probably an indicator of the thesis relevance.

A recent issue of the IEEE Learning Technology Newsletter (July 2011) had the special theme of "Adopting Standards and Specifications for Educational Content" and eight articles were published under this particular topic[86]. The previous special theme (April 2011) had six articles[87]. Beforehand (January 2011) there was an issue dedicated to "Semantic Web Technologies for Technology Enhanced Learning" with only three papers[88], one of them related to the work presented in this document. At least since 2008, a special theme had never had so few articles. However, far from meaning that the Semantic Web topic has no interest for the e-learning field and its community, it is conviction of the author of this document that there have been difficulties in transposing some theoretical ideas into practice in the e-learning area, and specifically in LORs.

[86] http://lttf.ieee.org/issues/july2011/IEEE-LT-Jul11.htm
[87] http://www.ieeetclt.org/issues/april2011/IEEE-LT-Apr11.html
[88] http://lttf.ieee.org/issues/january2011/IEEE-LT-Jan11.htm

In fact, a number of previous works related to the usage of ontologies in e-learning were cited along this thesis, but they usually show the possible advantages that can be achieved from their consideration, or use a small number of ontologies, typically one, in their experiments. This work can represent a step forward in the change of the current situation.

On the other side, the field of Pragmatics does not have a tradition in Informatics subjects, and this work might also be seen as a contribution to the awareness and diffusion of that field. However, as linguistic perspectives are more and more considered in many areas, also the pragmatic concerns will be increasingly regarded. In fact, the Pragmatic Web [Schoop, Moor et al., 2006] and the Language Action Perspective [Schoop, 2001] can be seen as examples of that trend.

The work developed under this thesis has received some attention by the scientific community, in particular the articles published in an early phase. Some of the papers that cited those articles are [Arapi, Moumoutzis et al., 2007; Kloos, Pardo et al., 2007; Lucia, Francese et al., 2008; Farhat, 2010; Valentín, 2011], which include two PhD dissertations, conference papers and a journal article.

Finally, some personal comments about the work carried out are also provided. Many people worked in the implementation of the prototype under my direct supervision. The ability to work with people with different backgrounds, academic formations, and personalities was definitely a useful and rewarding experience.

Also, many subjects were studied within this thesis for some years. During this period the author of this document faced many research and implementation problems, which were satisfactorily overcome. Some ideas were abandoned, while others were improved. It is my conviction that this maturing process was not only useful for the work presented in this document, but also for all research projects that will be embraced in the future.

FOURTH PART: REFERENCES AND APPENDICES

BIBLIOGRAPHY

C. Abela and K. Cortis (2010). SemChat: Extracting Personal Information from Chat Conversations. Workshop on Personal Semantic Data, EKAW 2010, Lisbon, Portugal, October 2010.

ADL. (2000). "Sharable Content Object Reference Model (SCORM) ".Advanced Distributed Learning (ADL) Available at http://www.adlnet.gov/scorm/.

W. Adrion (1993). "Research methodology in software engineering (Summary of the Dagstuhl workshop on future directions in software engineering: February 17–21, 1992, Schloß Dagstuhl)". ACM SIGSOFT Software Engineering Notes. **18**(1): 36-37.

A. Al-Ajlan and H. Zedan (2008). Why Moodle. 12th IEEE International Workshop on Future Trends of Distributed Computing Systems, Kunming, China, October 21-October 23**:** 58-64.

H. Alani and C. Brewster (2006). Metrics for Ranking Ontologies. Fourth International Evaluation of Ontologies for the Web Workshop held as part of the 15th International World Wide Web Conference, Edinburgh, United Kingdom, May 22nd, 2006.

H. Alani, S. Harris and B. O'Neil (2005). Ontology Winnowing: A Case Study on the AKT Reference Ontology. International Conference on Intelligent Agents, Web Technology and Internet Commerce, Vienna, Austria, 28-30 November 2005.

M. Alavi and D. E. Leidner (2001). "Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues". MIS Quarterly. **25**(1): 107-136.

W. Ali and S. Khan (2008). Ontology driven query expansion in data integration. The 2008 Fourth International Conference on Semantics, Knowledge and Grid (SKG‘ 08),IEEE Computer Society**:** 57–63.

M. Amaral (2008). "Institutional Repositories, Open Source Options, and Libraries". Open and Libraries Class Journal. **1**(1).

J. Angele, H. Boley, J. d. Bruijn, D. Fensel, P. Hitzler, M. Kifer, R. Krummenacher, H. Lausen, A. Polleres and R. Studer. (2005). "Web Rule Language (WRL)". Available at http://www.w3.org/Submission/WRL/.

L. E. Anido-Rifón, J. M. Santos-Gago, J. S. Ródríguez-Estévez, M. Caeiro-Rodríguez, M. J. Fernández-Iglesias and M. Llamas-Nistal (2002). A Step ahead in E-learning Standardization: Building Learning Systems from Reusable and Interoperable Software Components. Proceedings of the Eleventh International World Wide Web Conference (WWW2002), Honolulu, Hawaii, USA, 7-11 May 2002.

L. E. Anido, M. J. Fernandez, M. Caeiro, J. M. Santos, J. Rodriguez and M. Llamas (2002). "Educational metadata and brokerage for learning resources". Computers & Education. **38**: 351-374.

P. Arapi, N. Moumoutzis, M. Mylonakis, G. Theodorakis and G. Stylianakis (2007). Supporting Personalized Learning Experiences within the LOGOS Cross-Media Learning Platform. Workshop on Cross-Media and Personalized Learning Applications on top of Digital Libraries (LADL2007) in conj. with ECDL2007 Conference, Budapest, Hungary, September 2007.

Ariadne. Available at http://www.ariadne-eu.org.

J. C. Arpírez, O. Corcho, M. Fernández-López and A. Gómez-Pérez (2001). WebODE: A Scalable Workbench for Ontological Engineering. First International Conference on Knowledge Capture (KCAP01), Victoria, British Columbia, Canada, October 22-23, 2001.

A. Aschenbrenner, T. Blanke, D. Flanders, M. Hedges and B. O'Steen (2008). "The Future of Repositories? Patterns for (Cross-)Repository Architectures". D-Lib Magazine. **14**(11/12).

J. L. Austin (1962). How to Do Things with Words. Oxford University Press, New York.

J. L. Austin (1970). Philosophical Papers. Oxford University Press, Oxford.

Australian Universities Teaching Committee. (2003). "Learning Designs Project". Available at http://www.learningdesigns.uow.edu.au/project/index.htm.

D. Avison, F. Lau, M. Myers and P. A. Nielsen (1999). "Action Research". Communications of the ACM. **42**(1): 94-97.

I. Azevedo, E. Carrapatoso and C. V. Carvalho (2008a). A Framework to Scaffold the Reuse of Learning Objects. Eighth IEEE International Conference on Advanced Learning Technologies, Santander, Cantabria, 1-5 July 2008.

I. Azevedo, E. Carrapatoso and C. V. Carvalho (2008b). The use of templates to support learning design. The 2008 Frontiers in Education conference (ASEE/IEEE), Saratoga Springs, NY, October 22-25, 2008.

I. Azevedo, C. Carvalho and E. Carrapatoso (2008). "Effective Characterization of Learning Objects". International Journal of Advanced Media and Communication. **2**(3): 236 - 246.

I. Azevedo, C. Martins, E. Carrapatoso and C. V. Carvalho (2006). A model for the effective reuse of learning objects in repositories. The 6th IEEE International Conference on Advanced Learning Technologies (ICALT 2006), Kerkrade, Holanda, 5 - 7 July 2006.

I. Azevedo, R. Seiça, A. Ortiz, E. Carrapatoso and C. V. d. Carvalho (2010). A Semantic Approach for Learning Objects Repositories with Knowledge Reuse. EKAW 2010, Lecture Notes in Artificial Intelligence, Lisbon, Portugal,Springer-Verlag Berlin Heidelberg**:** 580–589.

F. Baader and W. Nutt (2003). Basic Description Logics. The description logic handbook: theory, implementation, and applications. Chapter 2. F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi and P. F. Patel-Schneider, Cambridge University Press**:** 43 - 95.

S. Baccianella, A. Esuli and F. Sebastiani (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. Proceedings of the Seventh Conference on International Language Resources and Evaluation,European Language Resources Association.

R. Baeza-Yates and B. Ribeiro-Neto (1999). Modern Information Retrieval. Addison Wesley.

S. M. Bagheri, B. Poizat and M. Pourmahdianc (2009). "An arithmetical view to first-order logic". Annals of Pure and Applied Logic.

C. Banea, R. Mihalcea, J. Wiebe and S. Hassan (2008). Multilingual Subjectivity Analysis Using Machine Translation. Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, Honolulu, October 2008**:** 127–135.

Y. Bar-Hillel (1971). "Out of the Pragmatic Wastebasket". Linguistic Inquiry. **2**: 401-407.

T. Barker (2011). "An Automated Individual Feedback and Marking System: An Empirical Study". Electronic Journal of e-Learning. **9**(1).

M. J. Bates (1989). "The Design of Browsing and Berrypicking Techniques for the Online Search Interface". Online Review. **13**(5): 407-424.

S. Battle, A. Bernstein, H. Boley, B. Grosof, M. Gruninger, R. Hull, M. Kifer, D. Martin, S. McIlraith, D. McGuinness, J. Su and S. Tabet. (2005). "Semantic Web Services Language (SWSL)". Available at http://www.w3.org/Submission/SWSF-SWSL/.

S. Bechhofer, F. v. Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider and L. A. Stein. (2004). "OWL Web Ontology Language Reference". Available at http://www.w3.org/TR/owl-ref/.

G. Begelman, P. Keller and F. Smadja (2006). Automated Tag Clustering: Improving search and exploration in the tag space. The 15th World Wide Web Conference, Edinburgh, UK, May 22–26, 2006.

S. Bennett, D. Parrish, G. Lefoe, M. O'Reilly, M. Keppell and R. Philip (2008). Chapter 18: A Needs Analysis Framework for the Design of Digital Repositories in Higher Education. Handbook of Research on Learning Design and Learning Objects: Issues, Applications, and Technologies. L. Lockyer, S. Bennett, S. Agostinho and B. Harper, IGI Global.

K. Bischoff, C. S. Firan, W. Nejdl and R. Paiu (2008). Can All Tags be Used for Search? The 17th ACM conference on Information and knowledge management, Napa Valley, CA, USA, October 26 - 30, 2008.

C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak and S. Hellmann (2009). "DBpedia - A crystallization point for the Web of Data". Journal of Web Semantics. Web Semantics: Science, Services and Agents on the WorldWideWeb. **7**(3): 154–165.

S. Bloom (1956). Taxonomy of educational objectives, the classification of educational goals, Handbook I: Cognitive Domain. McKay, New York.

M. Boekaerts (2003). "Towards a model that integrates motivation, affect and learning". British Journal of Educational Psychology Monograph, Series II, 2, Development and Motivation. **1**(1): 173-189.

S. E. Böer and W. G. Lycan (1976). The myth of semantic presupposition. Indiana University Linguistics Club, Bloomington, IN.

E. Boiy, P. Hens, K. Deschacht and M. F. Moens (2007). Automatic Sentiment Analysis in On-line Text. The 11th International Conference on Electronic Publishing, Vienna, Austria, June 2007.

E. P. Bontas and M. Mochol. (2005). "A Cost Model for Ontology Engineering. Technical Report TR-B-05-03". FU Berlin. Available at ftp://ftp.inf.fu-berlin.de/pub/reports/tr-b-05-03.pdf.

K. Bontcheva, V. Tablan, D. Maynard and H. Cunningham (2004). "Evolving GATE to meet new challenges in language engineering". Natural Language Engineering. **10**(3–4): 349–373.

M. D. Braly and G. B. Froh (2006). Social Bookmarking in the Enterprise. Proceeding of the 17th Annual ASIS&T SIG/CR Classification Research Workshop, Austin, Texas, USA.

E. Breck, Y. Choi and C. Cardie (2007). Identifying expressions of opinion in context. Proceedings of the Twentieth International Joint Conference on Artificial Intelligence.

C. Brewster, H. Alani, S. Dasmahapatra and Y. Wilks (2004). Data-driven Ontology Evaluation. Proceedings of the 4th International Conference on Language Resources and Evaluation, Lisbon, Portugal, 26-28 May 2004.

D. Brickey and R. V. Guha. (2004). "RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation".W3C. Available at http://www.w3.org/TR/PR-rdf-schema.

D. Brickley. (1996). "Towards an open question-interchange framework". Available at http://www.ilrt.bris.ac.uk/netquest/about/lang/motivation.html.

R. B. Briner and T. Kiefer (2007). Psychological Research into the Experience of Emotion at Work: Definitely Older, But are we Any Wiser? Research on Emotion in Organizations. N. M. Ashkanasy, W. J. Zerbe and C. E. J. Härtel, JAI Press. **1:** 281-304.

B. C. Brookes (1980a). "The foundations of information science: Part I. Philosophical aspects". Journal of Information Science. **2**(3–4): 125–133.

B. C. Brookes (1980b). "The foundations of information science: Part II. Quantitative aspects: Classes of things and the challenge of human individuality". Journal of Information Science. **2**(5): 209–221.

B. C. Brookes (1980c). "The foundations of information science: Part III. Quantitative aspects: Objective maps and subjective landscapes". Journal of Information Science. **2**(6): 269–275.

B. C. Brookes (1981). "The foundations of information science: Part IV. Information science: The changing paradigm". Journal of Information Science. **3**(1): 3–12.

C. H. Brooks and N. Montanez (2005). An analysis of the effectiveness of tagging in blogs. AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs.

C. Buckley and E. M. Voorhees (2000). Evaluating Evaluation Measure Stability. The 23rd annual international ACM SIGIR conference on Research and development in information retrieval, Athens, Greece, July 24 - 28, 2000.

K. Bühler (1934). Sprachtheorie. Fisher, Jena.

P. Buitelaar, T. Eigner and T. Declerck (2004). OntoSelect: A Dynamic Ontology Library with Support for Ontology Selection. The International Semantic Web Conference (Demo Session), Hiroshima, Japan, 7-11 November 2004.

P. Buitelaar, D. Olejnik and M. Sintek (2004). A protégé plug-in for ontology extraction from text based on linguistic analysis. Proceedings of the 1st European Semantic Web Symposium (ESWS), Heraklion, Greece.

D. Buzza, L. Richards, D. Bean, K. Harrigan and T. Carey (2005). "LearningMapR: A Prototype Tool for Creating IMS-LD Compliant Units of Learning". Journal of Interactive Media in Education. **17**.

M. Caeiro-Rodríguez, L. Anido-Rifón and M. Llamas-Nistal (2010). "Challenges in Educational Modelling: Expressiveness of IMS Learning Design." Educational Technology & Society. **13**(4): 215–226.

M. Caeiro-Rodríguez, M. Llamas-Nistal and L. Anido-Rifón (2005). "Towards a benchmark for the evaluation of LD expressiveness and suitability". Journal of Interactive Media in Education. **2005** (04).

C. Caffi (1994). Metapragmatics. Encyclopedia of languages and linguistics. R. Asher. Oxford, Pergamon and Aberdeen University Press**:** 2461–2466.

L. M. d. Campos, J. M. Fernández-Luna and J. F. Huete (2004). "Bayesian networks and information retrieval: an introduction to the special issue". Information Processing and Management. Special issue: Bayesian networks and information retrieval **40**(5).

Cancore.Canadian Core Learning Resource Metadata. Available at http://www.cancore.org.

I. Cantador, M. Fernández and P. Castells (2007). Improving Ontology Recommendation and Reuse in WebCORE by Collaborative Assessments. Proceedings of the 1st International Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007), at the 16th International World Wide Web Conference (WWW 2007), Banff, Canada.

L. Carr (2007). Open access to science. VI Workshop CALSI | Valencia, Spain, 14-16 November 2007.

W. B. Cavnar and J. M. Trenkle (1994). N-Gram-Based Text Categorization. Proceedings of Third Annual Symposium on Document Analysis and Information Retrieval, Las Vegas, NV, 11-13 April 1994**:** 161-175.

CEN-WS LT. (2010). "CEN-WS LT Learning Technology Standards Observatory". Available at http://www.cen-ltso.net/.

CEN. (2009). "The European Committee for Standardization". Available at http://www.cen.eu/cen/AboutUs/Pages/default.aspx.

CEN ISSS (2005). A Simple Query Interface Specification for Learning Repositories. CEN Workshop Agreement 15454. European Committee for Standardization. B. Simon, E. Duval, D. Massart, F. V. Assche and S. Ternier.

CEN WS-LT. (2010). "IMS Content Packaging - Overview". Available at http://www.cen-ltso.net/main.aspx?put=219.

CEN WS_LT. (2011a). "ISO/IEC JTC1 SC36 WG4 (MD) - Current Version". Available at http://www.cen-ltso.net/main.aspx?put=993.

CEN WS_LT. (2011b). "ISO/IEC JTC1 SC36 WG4 (MD) - Summary". Available at http://www.cen-ltso.net/main.aspx?put=992.

CEN/ISSS (2005). A Simple Query Interface Specification for Learning Repositories. CEN Workshop Agreement 15454. European Committee for Standardization. B. Simon, E. Duval, D. Massart, F. V. Assche and S. Ternier.

N. Chomsky (1957). Syntactic Structures. Mouton de Gruyter. 2nd edition (December 31, 2002).

N. Chomsky (1965). Aspects of the Theory of Syntax. The MIT Press (March 15, 1969).

J. P. E. d. Cid, L. d. l. F. Valentín, S. Gutiérrez, A. Pardo and C. D. Kloos (2007). "Implementation of a Learning Design Run-Time Environment for the .LRN Learning Management System". Journal of Interactive Media in Education. **2007**(7).

CiscoSystems (1999). Reusable Information Object Strategy. Definition, Creation Overview, and Guidelines.

C. W. Cleverdon (1991). The significance of the Cranfield tests on index languages. Proceedings of the 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Chicago, Illinois, USA, October 13-16, 1991**:** 3-12.

A. Cockburn (2000). Writing effective use cases. Addison-Wesley.

G. Conole and K. Fill (2005). "A learning design toolkit to create pedagogically effective learning activities". Journal of Interactive Media in Education(8).

G. Conole, A. Littlejohn, I. Falconer and A. Jeffery. (2005). "Pedagogical review of learning activities and use cases". Bristol. Available at http://www.elframework.org/refmodels/ladie/ouputs/LADIE lit review v15.doc.

O. Corcho, M. Fernández, A. Gómez-Pérez and A. López-Cima (2005). Building legal ontologies with methontology and webode. Law and the Semantic Web, LNAI. R. Benjamins, P. Casanovas, J. Breuker and A. Gangemi, Springer-Verlag. **3369:** 142–157.

M. Crampes and S. Ranwez (2000). Ontology-Supported and Ontology-Driven Conceptual Navigation on the World Wide Web. 11th ACM Hypertext Conference, San Antonio, TX, USA, May 30 - June 03, 2000**:** 191-199.

N. Craswell and D. Hawking (2004). Overview of the TREC-2004 Web Track. TREC-2004, Gaithersburg, Maryland, USA, November 2004.

N. Craswell, D. Hawking and S. Robertson (2001). Effective Site Finding using Link Anchor Information. The 24th annual international ACM SIGIR conference on Research and development in information retrieval New Orleans, LA, USA, September 09 - 12, 2001.

J. Cullen and A. Bryman (1988). "The Knowledge Acquisition Bottleneck: Time for Reassessment?" Expert Systems. **5**(3): 216–225.

J. Culler (1981). The pursuit of signs: semiotics, literature, deconstruction. Cornell University Press.

S. Currier and L. M. Campbell (2005). Vocabulary Management Technologies Review, Final report of JISC Pedagogical Vocabularies Project, JISC.

S. Currier, L. M. Campbell and H. Beetham. (2005). "Pedagogical Vocabularies Review. JISC Pedagogical Vocabularies Project". Available at http://www.jisc.ac.uk/uploaded_documents/PedVocab_VocabsReport_v0p11.doc.

M. d'Aquin, C. Baldassarre, L. Gridinoc, M. Sabou, S. Angeletou and E. Motta (200). Watson: supporting next generation semantic web applications. WWW/Internet 2007 Conference, Villareal, Spain, 8-9 Oct 2007.

M. d'Aquin, E. Motta, M. Sabou, S. Angeletou, L. Gridinoc, V. Lopez and D. Guidi (2008). "Toward a New Generation of Semantic Web Applications". IEEE Intelligent Systems. **23**(3): 20 - 28.

M. d'Aquin, M. Sabou, M. Dzbor, C. Baldassarre, L. Gridinoc, S. Angeletou and E. Motta (2007). Watson: A Gateway for the Semantic Web. European Semantic Web Conference (ESWC 2007), Innsbruck, Austria, 3-7th, June 2007.

J. Dalziel (2008). Learning design: Sharing pedagogical know-how. Opening up education: The collective advancement of education through open technology, open content, and open knowledge. T. Liyoshi and M.S.V.Kumar. Cambridge, MA, MIT Press: 375–387.

T. H. Davenport, S. L. Jarvenpaa and M. C. Beers (1996). " Improving knowledge work processes". Sloan Management Review. **37**(4): 53-65.

R. Davis, H. Shrobe and P. Szolovits (1993). "What is a Knowledge Representation?" AI Magazine. **14**(1): 17-33.

M. Day and P. Cliff. (2009). "RDN cataloguing guidelines (Version 1.1)". Available at http://www.rdn.ac.uk/publications/cat-guide/.

DCMI. (1999). "Dublin Core Metadata Element Set, Version 1.1".Dublin Core Metadata Initiative. Available at http://dublincore.org/documents/dces/.

DCMI (2008). Dublin Core Metadata Element Set, Version 1.1, Dublin Core Metadata Initiative.

R. de Hoog (1998). Methodologies for Building Knowledge Based Systems: Achievements and Prospects. Handbook of Expert Systems. J. Liebowitz. Boca Raton, Florida, CRC Press: 1.1-1.13.

S. Decker, S. Melnik, F. van Harmelen, D. Fensel, M. Klein, J. Broekstra, M. Erdmann and I. Horrocks (2000). "The Semantic Web: The Roles of XML and RDF". IEEE Internet Computing. **4**(5): 63-74.

S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer and R. Harshman (1990). "Indexing by latent semantic analysis". Journal of the American Society for Information Science. **41**(6): 391-407.

B. Dennis (2006). Foragr: Collaboratively Tagged Photographs and Social Information Visualization. Collaborative Web Tagging Workshop, WWW2006, Edinburgh.

D. Derks, A. H. Fischer and A. E. R. Bos (2008). "The role of emotion in computer-mediated communication: A review". Computers in Human Behavior. **24**(3): 766-785.

V. Devedzic, J. Jovanovic and D. Gasevic (2007). "The Pragmatics of Current E-Learning Standards". IEEE Internet Computing: 16-24.

A. K. Dey (2001). "Understanding and Using Context". Personal and Ubiquitous Computing. **5**(1).

L. Ding, T. Finin, A. Joshi, R. Pan, R. S. Cost, Y. Peng, P. Reddivari, V. Doshi and J. Sachs (2004). Swoogle: A Search and Metadata Engine for the Semantic Web. Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management, November 2004.

L. Ding, R. Pan, T. Finin, A. Joshi, Y. Peng and P. Kolari (2005). Finding and Ranking Knowledge on the Semantic Web. The 4th International Semantic Web Conference, Galway, Ireland, November 2005.

L. Ding, J. Shinavier, T. Finin and D. L. McGuinness (2010). owl:sameAs and Linked Data: An Empirical Study. Proceedings of the WebSci10: Extending the Frontiers of Society On-Line, Raleigh, NC: US, April 26-27th, 2010.

B. L. Doan, Y. Bourda and N. Bennacer (2004). Building a Pedagogical Ontology to Improve Information Retrieval. Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004, Chesapeake, VA,AACE**:** 5193-5197.

A. Doms and M. Schroeder (2005). "GOPubMed: exploring PubMed with the Gene Ontology". Nucleic Acid Research. **33**: W783-786.

E. Duval (2004). Learning Technology Standardization: Making Sense of it All. ComSIS. **1:** 33-43.

E. Duval, E. Forte, K. Cardinaels, B. Verhoeven, R. Van Durm, K. Hendrikx, M. Wentland-Forte, N. Ebel, M. Macowicz, K. Warkentyne and F. Haenni (2001). "The ARIADNE Knowledge Pool System." Communications of the ACM. **44**(5): 72-78.

P. Ekman (1989). The argument and evidence about universals in facial expressions of emotions. Handbook of Social Psychophysiology. H. Wagner and A. Manstead. New York, Wiley**:** 143-64.

P. Ekman (1992). "An Argument for Basic Emotions". Cognition and emotion. **6**(3/4): 169-200.

P. Ekman (1999). Basic emotions. The handbook of cognition and emotion. T. Dalgleish and T. Power. New York, John Wiley & Sons**:** 45-60.

P. Ekman and R. J. Davidson, Eds. (1994). The Nature of Emotion: Fundamental Questions, Oxford University Press.

P. Ekman, W. V. Friesen and P. Ellsworth (1982). What emotion categories or dimensions can observers judge from facial behavior? Emotion in the human face. P. Ekman. New York, Cambridge University Press**:** 39-55.

R. Elferink, D. Griffiths and E. Veenendaal (2006). The OpenDock project: putting in place the infrastructure for sharing learning activities. Proceedings of International Workshop in Learning Networks for Lifelong Competence Development, TENCompetence Conference, Sofia, Bulgaria, September 12th, 2006,TENCompetence.

R. Farhat (2010). Approche d'assistance aux auteurs pour la réutilisation d'objets d'apprentissage, Thèse de doctorat de Telecom & Management SudParis dans le cadre de l'école doctorale SITEVRY en co-accréditation avec l'Université d'Evry Val d'Essonne.

S. Farrugia. (2010). "TaskConsole: Managing Information on the Semantic Desktop". Available at http://staff.um.edu.mt/cabe2/supervising/undergraduate/overview/stefania_farrugia.pdf.

M. Fernández-López, A. Gómez-Pérez, A. Pazos and J. Pazos (1999). "Building a Chemical Ontology Using Methontology and the Ontology Design Environment". IEEE Intelligent Systems & their applications. **4**(1): 37-46.

M. Fernández, A. Gómez-Pérez and N. Juristo (1997). METHONTOLOGY: from Ontological Art towards Ontological Engineering. Proceedings of Workshop on Ontological Engineering: AAAI-97. Stanford, CA, USA, Spring Symposium Series**:** 33-40.

S. Fisher, L. Tozer, N. Friesen and A. Roberts. (2002). "CANCORE INITIATIVE. CanCore Learning Object Metadata". Available at http://cancore.athabascau.ca/guidelines/CanCore%20Guidelines%20version%201.1.doc.

G. Forger (2003). Hidden Treasures on the Internet: The Gateway to Educational Materials (GEM) Project. Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2003, Chesapeake, VA,AACE.

M. Foucault. (1969). Archaeology of Knowledge. Harper, New York.

E. A. Fox and S. Sharat (1986). A Comparison of Two Methods for Soft Boolean Interpretation in Information Retrieval. Technical Report TR-86-1, Virginia Tech, Department of Computer Science.

F. L. G. Frege (1952). On Sense and Reference. Translations from the Philosophical Writings of Gottlob Frege. P. Geach and M. Black. Oxford, Blackwell**:** 56-78.

N. Friesen. (2002). "E-learning Standardisation: An Overview". Available at http://www.cancore.ca/pdfs/e-learning_standardization_overview.pdf.

N. Friesen (2004). Learning Objects and Standards:Pedagogical Neutrality and Engagement. Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT'04), Joensuu, Finland, 30 Aug.-1 Sept. 2004.

N. Friesen (2005). CanCore: In your Neighborhood and around the World. CanCore Report.Canadian Core Metadata Intitiave.

N. Friesen and R. McGreal (2002). "International E-learning Specifications". International Review of Research in Open and Distance Learning. **3**(2).

N. H. Frijda (1986). The emotions. Cambridge University Press, New York.

N. H. Frijda (1993). Moods, emotion episodes, and emotions. Handbook of emotions. M. Lewis and J. M. Haviland. New York, The Guilford Press**:** 381-403.

K. Fujimura, S. Fujimura, T. Matsubayashi, T. Yamada and H. Okuda (2008). Topigraphy: visualization for large-scale tag clouds. Proceeding of the 17th international conference on World Wide Web (WWW 2008), Beijing, China, April 21-25, 2008**:** 1087-1088.

G. W. Furnas, T. K. Landauer, L. M. Gomez and S. T. Dumais (1987). "The vocabulary problem in human-system communication". Communications of the ACM. **30**(11): 964 - 971.

E. Gabrilovich and S. Markovitch (2007). Computing semantic relatedness using wikipedia-based explicit semantic analysis. Proceedings of The Twentieth International Joint Conference for Artificial Intelligence, Hyderabad, India.

A. Gal and G. Modica (2004). OntoBuilder: fully automatic extraction and consolidation of ontologies from Web sources. Proceedings of the 20th International Conference on Data Engineering (ICDE).

D. Gasevic (2006). Model Driven Engineering and Ontology Development. Springer.

E. Goffman (1974). Frame analysis: An essay on the organization of experience. Northeastern University Press, Boston, MA.

D. H.-L. Goh, A. Chua, D. A. Khoo, E. B.-H. Khoo, E. B.-T. Mak and M. W.-M. Ng (2006). "A checklist for evaluating open source digital library software". Online Information Review. **30**(4): 360-379.

C. Golbreich and E. K. Wallace. (2008). "OWL 2 Web Ontology Language: New Features and Rationale". Available at http://www.w3.org/TR/2008/WD-owl2-new-features-20081202/.

S. A. Golder and B. A. Huberman (2006a). "Usage patterns of collaborative tagging systems". Journal of Information Science. **32**(2): 198–208.

S. A. Golder and B. A. Huberman (2006b). "Usage Patterns of Collaborative Tagging Systems". Journal of Information Science. **32**(2): 198-208.

A. Gómez-Pérez, M. Fernández-López and O. Corcho (2004). Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. Springer-Verlag.

D. Good (1996). "Pragmatics and Presence". AI & Society. **10** (3-4): 309 - 314.

C. Goodwin and A. Duranti (1992). Rethinking Context: an Introduction. Rethinking Context: Language as an Interactive Phenomenon. A. Duranti and C. Goodwin, Cambridge University Press**:** 1-42.

H. P. Grice (1975). Logic and conversation. Syntax and semantics. P. Cole and J. L. Morgan. New York, Academic Press. **3:** 41-58.

T. R. Gruber (1993). "A translation approach to portable ontology specifications". Knowledge Acquisition. **5**(2): 199-220.

M. Grüninger and M. Fox (1995). Methodology for the Design and Evaluation of Ontologies. Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95, Montreal.

N. Guarino (1997). Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology. M. T. Pazienza, Springer Verlag**:** 139-170.

N. Guarino (1998). Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998,IOS Press**:** 3-15.

N. Guarino (2004). "Toward a formal evaluation of ontology quality. Why Evaluate ontology technologies? Because it works!" IEEE Intelligent Systems.

N. Guarino and C. Welty (2002). "Evaluating ontological decisions with OntoClean". Communications of the ACM. **45**(2): 61-65.

C. Gütl, K. Lankmayr, J. Weinhofer and M. Höfler (2011). "Enhanced Automatic Question Creator – EAQC: Concept, Development and Evaluation of an Automatic Test Item Creation Tool to Foster Modern e-Education". Electronic Journal of e-Learning. **9**(1).

Y. Gyamfi, J. Wiebe, R. Mihalcea and C. Akkaya (2009). Integrating Knowledge for Subjectivity Sense Labeling. Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics**:** 10-18.

H. Haberland and J. L. Mey (1977). "Editorial: Pragmatics and linguistics". Journal of Pragnatics. **1**(1): 1-16.

H. Halpin, P. J. Hayes, J. P. McCusker, D. L. McGuinness and H. S. Thompson (2010). "When owl:sameAs isn't the Same: An Analysis of Identity in Linked Data". The Semantic Web – ISWC 2010. Lecture Notes in Computer Science. **6496/2010**: 305-320.

S. Hampshire and C. P. Salvia (2010). "Translation and the Internet: Evaluating the Quality of Free Online Machine Translators". Quaderns. Rev. trad. . **17**(2010).

G. Hanley (2005). Enabling Open Education with MERLOT. Proceeding of the OpenEd Conference at Utah State University, September 2005.

D. Harley (2008). Why Understanding the Use and Users of Open Education Matters. Opening Up Education. The Collective Advancement of Education through Open Technology, Open Content, and Open Knowledge. T. Iiyoshi and M. S. V. Kumar, The MIT Press.

V. Hatzivassiloglou and J. M. Wiebe (2000). Effects of adjective orientation and gradability on sentence subjectivity. Proceedings of the 18th conference on Computational Linguistics (CONLING).

R. Hausser (2001). Foundations of Computational Linguistics: Human-Computer Communication in Natural Language. Springer-Verlag. Second edition (October 16, 2001).

Heal.Heath Education Assets Library. Available at http://www.healcentral.org.

L. Heathcote (2006). Learning Design templates - a pedagogical just-in time support tool. JISC Innovating e-Learning 2006: Transforming Learning Experiences Online Conference.

N. Herr. (2008). "Pedagogical Content Knowledge in Science Teaching". Available at http://www.csun.edu/science/ref/pedagogy/pck/index.html.

P. Heymann, G. Koutrika and H. Garcia-Molina (2007). "Fighting spam on social web sites: A survey of approaches and future challenges". IEEE Internet Computing. **11**(6): 36–45.

P. Heymann, G. Koutrika and H. Garcia-Molina (2008). Can Social Bookmarking Improve Web Search? The international conference on Web search and Web data mining, Stanford, CA, USA, February 11-12, 2008.

P. E. Higgs, S. Meredith and T. Hand. (2002). "Technology for sharing - a research project to inform VET Australia about learning objects and digital rights management including systems and metadata to support them: flexible learning leader, 2002 report".Australian Flexible Learning Framework (AFLF). Available at http://leaders.flexiblelearning.net.au/fl_leaders/fll02/finalreport/final_hand_higgs_meredith.pdf.

I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof and M. Dean. (2004). "SWRL: A Semantic Web Rule Language Combining OWL and RuleML". Available at http://www.w3.org/Submission/SWRL/.

W. Horton and K. Horton (2003). E-Learning Tools and Technologies. Wiley Publishing, Indianapolis.

A. Hotho, R. Jaschke, C. Schmitz and G. Stumme (2006). "Information retrieval in folksonomies: Search and ranking". The Semantic Web: Research and Applications In The Semantic Web: Research and Applications. **4011**: 411-426.

Y.-F. Huang and C.-H. Hsu (2008). "PubMed smarter: query expansion with implicit words based on gene ontology". Knowledge-based Systems. **21**(2008): 927-933.

L. Huberman (1979). História da Riqueza do Homem. Zahar Editors, Brazil, page 194.

IEEE. (1995). "Glossary of General SE Terms. IEEE Standard 730.1-1995. " Available at http://www.computer.org/portal/site/seportal/menuitem.092750c1f9509be2169d1205682079ec/index.jsp?&pName=se_level1&path=seportal/subpages&file=glossarym.xml&xsl=generic.xsl&;jsessionid=JnjGGg6nm1yJ0gqlYmj9W4WqXnPNMf2my2s9hMdFBccy99gJsMGV!-1582182879.

IEEE. (2002). "IEEE Standard 1484.12.1-2002, Learning Object Metadata (LOM)". Available at http://ltsc.ieee.org/wg12.

IETF. (2007). "The Dublin Core Metadata Element Set". Available at http://www.ietf.org/rfc/rfc5013.txt.

M. H. Immordino-Yang and A. Damasio (2007). "We Feel, Therefore We Learn: The Relevance of Affective and Social Neuroscience to Education". Mind, Brain, and Education. **1**(1): 3–10.

M. H. Immordino-Yang and M. Faeth (2009). The Role of Emotion and Skilled Intuition in Learning. Mind, Brain, and Education. D. A. Sousa. Bloomington, Solution Tree Press: 66-81.

IMS. (1999a). "IMS Enterprise Specification ".IMS Global Learning Consortium. Available at http://www.imsglobal.org/enterprise/index.html.

IMS. (1999b). "IMS Learning Resource Meta-Data Specification".IMS Global Learning Consortium. Available at http://www.imsglobal.org/metadata/index.html.

IMS. (2000a). "IMS Content Packaging Specification".IMS Global Learning Consortium. Available at http://www.imsglobal.org/content/packaging/.

IMS. (2000b). "IMS Question and Test Interoperability Specification".IMS Global Learning Consortium. Available at http://www.imsglobal.org/question/.

IMS. (2001). "IMS Learner Information Package Specification ".IMS Global Learning Consortium. Available at http://www.imsglobal.org/profiles/index.html.

IMS. (2002). "IMS Reusable Definition of Competency or Educational Objective ".IMS Global Learning Consortium. Available at http://www.imsglobal.org/competencies/index.html.

IMS (2003a). IMS Digital Repositories Interoperability - Core Functions Information Model. K. Riley and M. McKell, IMS Global Learning Consortium.

IMS. (2003b). "IMS Learning Design Best Practice and Implementation Guide. Version 1.0 Final Specification".IMS Global Learning Consortium. Available at http://www.imsglobal.org/learningdesign/ldv1p0/imsld_infov1p0.html.

IMS. (2003c). "IMS Learning Design Information Model. Version 1.0 Final Specification ".IMS Global Learning Consortium. Available at http://www.imsglobal.org/learningdesign/ldv1p0/imsld_infov1p0.html.

IMS. (2003d). "IMS Simple Sequencing Specification ".IMS Global Learning Consortium. Available at http://www.imsglobal.org/simplesequencing/index.html.

IMS (2010). IMS GLC Learning Object Discovery and Exchange (LODE). Base Document Version 1.0 D. Massart, N. Nicholas and N. Ward, IMS Global Learning Consortium.

ISO. (2003). "ISO Standard 15836-2003". Available at http://www.iso.org/iso/search.htm?qt=15836&searchSubmit=Search&sort=rel&type=simple&published=on.

ISO/IEC (1998). ISO-9241-11. Ergonomic requirements for office work with visual display terminals (VTDs) - Part 11. Guidance on usability.

ISO/IEC. (2009). "Text of ISO/IEC FCD 19788-1 – ITLET – Metadata for Learning Resources – Part 1: Framework". Available at http://isotc.iso.org/livelink/livelink/fetch/2000/2122/327993/806742/4920436/4920313/WG4_N0373_Text_of_FCD_for_ISO_IEC__19788-1_--_ITLET_--_Metadata_for_Learning_Resources_--_Part_1_Framework.pdf?nodeid=8501949&vernum=0.

C. E. Izard (1971). The face of emotion. Appleton-Century-Crofts, New York.

R. Jakobson (1960). Closing statement: linguistics and poetics. Style in language. T. A. Sebeok. Cambridge, MA, MIT Press: 350-377.

K. Jarvelin and J. Kekalainen (2002). "Cumulated gain-based evaluation of ir techniques". ACM Transactions on Information Systems. **20**(4): 422–446.

K. G. Jeffery (2006). "Open Access: an introduction". European Research Consortium for Informatics and Mathematics News. **64**: 16-17.

E. Jokisalo. (2010). "LRE metadata and thesaurus". Available at http://insight.eun.org/ww/en/pub/insight/interoperability/learning_resource_exchange/metadata.htm.

K. S. Jones (1974). "Automatic indexing". Journal of Documentation. **30**: 393–432.

K. S. Jones and C. J. van Rijsbergen (1976). "Information retrieval test collections". Journal of Documentation. **32**(1): 59–75.

S. Jones, S. W. Karen and S. E. Robertson (2000). "A probabilistic model of information retrieval: Development and comparative experiments". Information Processing and Management. **36**(6): 779–808, 809–840.

KACTUS. (1996). "The KACTUS Booklet version 1.0. Esprit Project 8145 KACTUS". Available at http://www.swi.psy.uva.nl/projects/NewKACTUS/Reports.html.

Y. Kagolovsky and J. R. Möhr (2001). "A new approach to the concept of "relevance" in information retrieval (IR)". Studies In Health Technology and Informatics. **84**(Pt 1): 348-352.

P. Kallonis and D. G. Sampson (2010). Examining Learning Object Repositories from a Knowledge Management perspective. 10th IEEE International Conference on Advanced Learning Technologies, Sousse, 5-7 July 2010**:** 289 - 293.

R. Kamali, L. Cassel and R. LeBlanc (2004). Keeping family of computing related disciplines together. The 5th conference on Information Technology Education, Salt Lake City, UT, USA, October 28 - 30, 2004**:** 241 - 243.

L. Karttunen (1974). "Presupposition and Linguistic Context". Theoretical Linguistics. **1**: 3-44.

R. M. Kempson (1975). Presupposition and the Delimitation of Semantics. Cambridge University Press, Cambridge.

S. Kendal and M. Creen (2006). An Introduction to Knowledge Engineering. Springer.

A. Kent, M. Berry, F. U. Leuhrs and J. W. Perry (1955). "Machine literature searching VIII: Operational criteria for designing information retrieval systems". American Documentation,. **6**(2): 93–101.

E. Ketzan. (2002). "Amazon.com and the new democracy of opinion". Available at http://www.themodernword.com/pynchon/papers_ketzan1.html.

M. E. I. Kipp (2007). @toread and Cool : Tagging for Time, Task and Emotion. 8th Information Architecture Summit Las Vegas, Nevada (US), 22-26 March 2007.

M. E. I. Kipp and D. Campbell (2006). Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices. The 17th Annual General Meeting of the American Society for Information Science and Technology, Austin, Texas, USA.

M. Klenner, S. Petrakis and A. Fahrni (2009). Robust Compositional Polarity Classification. International Conference Recent Advances in Natural Language Processing 2009 Borovets, Bulgaria, September 14-16, 2009**:** 180–184.

C. D. Kloos, A. Pardo, M. M. Organero and L. D. L. F. Valentin (2007). "E-LANE: an e-learning initiative based on open source as a basis for sustainability". International Journal of Continuing Engineering Education and Life Long Learning. **17**(1): 57 – 66.

M. Knapp and J. Hall (1997). Nonverbal Communication in Human Interaction. Harcourt Brace College Publishers. 4th edition.

M. Kobayashi and K. Takeda (2000). "Information Retrieval on the Web". ACM Computing Surveys. **32**(2): 144-173.

M. Kökörčený (2011). "Architecture and flexibility of digital libraries systems". International Journal of Systems Applications, Engineering and Development. **5**(2): 237-244.

M. Kökörčený and A. Bodnárová (2010). Comparison of digital libraries systems. Proceedings of the 9th WSEAS international conference on Data networks, communications, computers, Faro, Portugal, , November 3-5, 2010**:** 97-100.

R. Koper. (2001). "Modeling units of study from a pedagogical perspective the pedagogical meta-model behind EML". Available at http://dspace.learningnetworks.org/bitstream/1820/36/1/Pedagogical%20metamodel%20behind%20EMLv2.pdf.

R. Koper and B. Olivier (2004). Representing the learning design of units of learning. Education, technology and society. **7:** 97-111.

R. Koper and C. Tattersall (2005). Learning Design: A Handbook on Modelling and Delivering Networked Education and Training. Berlin Heidelberg, Springer.

R. Korfhage (1997). Information storage and retrieval. Wiley.

C. Kortekaas (2006). Don't keep it under your hat: A Fez design description and case study. Fedora Users Conference 2006, June 19-20, 2006**:** 1-21.

G. Koutrika, F. A. Effendi, Z. Gyongyi, P. Heymann and H. Garcia-Molina (2007). Combating spam in tagging systems. The 3rd international workshop on Adversarial information retrieval on the web (AIRWeb), Banff, Alberta, Canada, May 8, 2007.

E. Kroski. (2005). "The Hive Mind: Folksonomies and User-Based Tagging". Available at http://infotangle.blogsome.com/2005/12/07/the-hive-mind-folksonomies-anduser-based-tagging/.

D. T. Langendoen and H. Savin (1971). The Projection Problem for Presuppositions. Filmore and Langendoen.

O. Lassila and D. McGuinness (2001). The Role of Frame-Based Representation on the Semantic Web, Technical Report KSL-01-02, Knowledge Systems Laboratory, Stanford University.

O. Lassila and R. Swick. (1999). "Resource Description Framework (RDF): Model and Syntax Specification. W3C Recommendation". Available at http://www.w3.org/TR/REC-rdf-syntax/.

D. Leclercq and M. Poumay. (2005). "The 8 learning Events Model. Release 2005.1. LabSET".University of Liège. Available at http://www.labset.net/media/prod/8LEM.pdf.

M.-C. Lee, K. H. Tsai and T. I. Wang (2008). "A practical ontology query expansion algorithm for semantic-aware learning objects retrieval". Computers and Education. **50**(2008): 1240–1257.

G. N. Leech (1983). Principles of Pragmatics. Longman.

J. Lehmann, J. Schüppel and S. Auer (2007). Discovering Unknown Connections - the DBpedia Relationship Finder. Proceedings of the 1st SABRE Conference on Social Semantic Web (2007)**:** 99-110.

D. B. Lenat and R. V. Guha (1990). Building Large knowledge-Based systems. Addison-Wesley.

D. H. Leo, J. I. A. Perez and Y. Dimitriadis (2004). IMS Learning Design Support for the Formalization of Collaborative Learning Patterns. The IEEE International Conference on Advanced Learning Technologies, Joensuu, Finlandia, August 2004**:** 350–354.

V. Levenshtein (1965). "Binary Codes for Correcting Deletions, Insertions, and Reversals". Doklady Akademii Nauk SSSR. **163**(4): 845–848.

S. C. Levinson (1983). Pragmatics, Cambridge Textbooks in Linguistics. Cambridge University Press (July 29, 1983).

J. Lindholm, E. Axdorph, K. E. J. Gilbert, J. Hylén, H. Jonsson, L. Olsson and e. al. (2009). "Project "Open Educational Resources in institutional repositories" - executive summary". Available at http://www.kb.se/Docs/about/projects/openaccess/OERIIR_executive_summary.pdf.

B. Liu, E. Zhai, H. Sun, Y. Chen and Z. Chen (2009). Filtering Spam in Social Tagging System with Dynamic Behavior Analysis. The 2009 International Conference on Advances in Social Network Analysis and Mining Athens, Greece, July 20–22, 2009.

A. D. Lucia, R. Francese, I. Passero and G. Tortora (2008). The CD-Lomas project: an RCP-based Collaborative Distributed Learning Object Management System. The 3rd Italian Workshop on Eclipse Technologies, Bari, Italy, November 17-18, 2008.

D. MacLeod. (2005). "Learning Object Repositories: Deployment and Diffusion. CANARIE Discussion Paper." Available at http://www.canarie.ca/funding/elearning/2005_LOR_final_report.pdf.

C. D. Manning, P. Raghavan and H. Schütze (2008). An Introduction to Information Retrieval. Cambridge University Press.

F. Manola, E. Miller and B. McBride. (2004). "RDF primer. W3C recommendation".W3C. Available at http://www.w3.org/TR/2004/REC-rdf-primer-20040210/.

R. Manor (2001). "On The Overlap Of Pragmatics And Semantics". Synthese. **128**(1-2): 63–73.

G. Marchionini (2004). From Information Retrieval to Information Interaction. Proceeding of the 26th European Conference on Information Retrieval Research, Sunderland, UK, April 2004,Springer-Verlag**:** 1-11.

J. L. Marill and E. C. Luczak (2009). "Evaluation of Digital Repository Software at the National Library of Medicine ". D-Lib Magazine. **15**(5/6).

M. L. Markus (2001). "Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success". Journal of Management Information Systems. **18**: 57-93.

C. Marlow, M. Naaman, D. Boyd and M. Davis (2006). HT06, tagging paper, taxonomy, Flickr, academic article, to read. Proceedings of the seventeenth conference on Hypertext and hypermedia. SESSION: Social networks, networking & virtual communities, Odense, Denmark, August 22 - 25, 2006, ACM.

M. E. Maron and J. Kuhns (1960). "On relevance, probabilistic indexing and information retrieval". Journal of the Association for Computing Machinery. **7**(3): 216–244.

M. E. Maron, J. L. Kuhns and L. Ray (1958). Some experiments with probabilistic indexing (45pp.). Ramo-Wooldridge.

M. E. Maron, J. L. Kuhns and L. C. Ray (1959). Probabilistic indexing: A statistical technique for document identification and retrieval. Technical Memorandum No. 3. Data Systems Project Office, 91pp, Ramo-Wooldridge.

J. L. McClelland and D. E. Rumelhart (1986). Parallel distributed processing: Explorations in the microstructure of cognition, Vol. II. MIT Press, Cambridge, MA.

C. McNaught (2003). Chapter 16: Identifying the complexity of factors in the sharing and reuse of resources. Reusing Online Resources: A Sustainable Approach to E-Learning. Kogan Page Ltd**:** 199–210.

C. T. Meadow, B. R. Boyce, D. H. Kraft and C. L. Barry (2007). Text information retrieval systems. Academic Press. Third edition.

J. Mey, H. Haberland and K. Fischer (2010). "Editorial". Pragmatics and Society. **1**(1): 1-8.

J. L. Mey (1993). Pragmatics: An Introduction. Wiley-Blackwell. 2nd edition (February 15, 2001).

D. K. Meyer and J. C. Turner (2002). "Discovering emotion in classroom motivation research". Educational Psychologist. **37**: 107-114.

Y. Miao, K. Hoeksema, H. U. Hoppe and A. Harrer (2005). CSCL Scripts: Modelling Features and Potential Use. Proceedings of the 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!, Taipei, Taiwan.

Y. Miao, H. Vogten, H. Martens and R. Koper. (2007). "The Complementary Roles of IMS LD and IMS QTI in Supporting Effective Web-based Formative Assessment". Available at http://dspace.ou.nl/handle/1820/951.

J. Millea. (2003). "The EDNA Metadata Standard". Available at http://www.educationau.edu.au/jahia/webdav/site/myjahiasite/shared/papers/edna_metadata.pdf.

G. A. Miller (1995). "WordNet: A Lexical Database for English. New Horizons in Commercial and Industrial Artificial Intelligence". Communications of the ACM. **38**(11): 39-41.

G. A. Miller and W. G. Charles (1991). "Contextual correlates of semantic similarity. " Language and Cognitive Processes. **6**(1): 1-28.

R. Mizoguchi (2003). "Tutorial on Ontological Engineering: Part 1: Introduction to Ontological Engineering". New Generation Computing. **21**(4): 365-384.

V. Monfort, M. Khemaja and S. Hammoudi (2010). Extending the IMS LD standard with Adaptability. The 10th IEEE International Conference on Advanced Learning Technologies, Sousse, Tunisia, July 05-July 07.

A. Monge and C. Elkan (1996). The Field Matching Problem: Algorithms and Applications. Proceedings of Knowledge Discovery and Data Mining Conference.

C. N. Mooers (1950). The Theory of Digital Handling of Non-Numerical Information and its Implications to Machine Economics. Proceedings of the meeting of the Association for Computing Machinery at Rutgers University, New Brunswick, N. J., March 1950.

J. L. Morgan (1977). "Linguistics: The Relation of Pragmatics to Semantics and Syntax". Annual Review of Anthropology. **6**: 57-67.

D. Morim and I. Azevedo (2011). Utilizing Opinion Tags - a case study (Utilização de Tags de Opinião – um estudo de caso). 6ª Conferência Ibérica de Sistemas e Tecnologias de Informação, Chaves, Portugal, 15-18 July 2011.

C. W. Morris (1938). Foundations of the theory of signs. The University of Chicago Press, Chicago.

P. Morville (2005). Ambient Findability: What We Find Changes Who We Become. O'Reilly Media. 1st edition (September 26, 2005).

B. Motik, B. C. Grau, I. Horrocks, Z. Wu, A. Fokoue and C. Lutz. (2009). "OWL 2 Web Ontology Language Profiles. W3C Recommendation 27 October 2009". Available at http://www.w3.org/TR/owl2-profiles.

P. J. Muñoz-Merino, C. D. Kloos and J. F. Naranjo (2009). "Enabling interoperability for LMS educational services". Computer Standards & Interfaces. **31**(2009): 484–498.

J. Najjar, S. Ternier and E. Duval (2003). The actual use of metadata in ARIADNE: An empirical analysis. Proceedings of the 3rd Annual ARIADNE Conference.

R. Nechs, R. Fikes, T. Finin, T. Gruber, R. Patil, T. Senator and W. Swartout (1991). "Enabling Technology for Knowledge Sharing". AI Magazine. **12**(3): 36-56.

S. Neumann, M. Klebl, D. Griffiths, D. Hernández-Leo, L. Valentín, H. Hummel, F. Brouns, M. Derntl and P. Oberhuemer (2010). "Report of the Results of an IMS Learning Design Expert Workshop. " International Journal Of Emerging Technologies In Learning (IJET). **5**(1): 58-72.

H. Neuroth and T. Koch (2001). Metadata mapping and application profiles: approaches to providing the cross-searching of heterogeneous resources in the EU project Renardus. International Conference on Dublin Core and Metadata Applications 2001, Tokyo, Japan, 24-26 October 2001.

F. Neven and E. Duval (2002). Reusable learning objects: a survey of LOM-based repositories. Proceedings of the tenth ACM international conference on Multimedia, ACM Press.

M. Nilsson, M. Palmér and J. Brase (2003). The LOM RDF binding - principles and implementation. Proceedings of the Third Annual ARIADNE conference.

NISO. (2007). "ANSI/NISO Z39.85 - The Dublin Core Metadata Element Set ". Available at http://www.niso.org/kst/reports/standards?step=2&gid=&project_key=9b7bffcd2daeca6198b4ee5a848f9beec2f600e5.

H. Nissenbaum (1985). Emotions and focus. CSLI, Stanford, CA.

P. T. Northrup (2007). Learning Objects for Instruction: Design and Evaluation. Information Science Publishing.

J. Novak (1995). "Concept mapping to facilitate teaching and learning". Prospects. **5**(1): 79-86.

N. Noy and D. L. McGuinness (2001). Ontology development 101: A guide to creating your first ontology. Technical Report KSL-01-05. Stanford, Stanford Medical Informatics.

T. O'Reilly. (2003). "The Architecture of Participation" O'Reilly Developer Weblogs. Available at http://www.oreillynet.com/pub/wlg/3017.

X. Ochoa (2008). Learnometrics: Metrics for Learning Objects. Leuven, KATHOLIEKE UNIVERSITEIT LEUVEN. **PhD thesis**.

E. Ochs and B. Schieffelin (1989). "Language has a heart". Text 9. **1**: 7-25.

C. K. Ogden and I. A. Richards (1923). The Meaning of Meaning: A Study of the Influence of Language Upon Thought and of the Science of Symbolism. Routledge & Kegan Paul, London.

E. Oren, R.Delbru, M. Catasta, R. Cyganiak, H. Stenzhorn and G. Tummarello (2008). "Sindice.com: A document-oriented lookup index for open linked data". International Journal of Metadata, Semantics and Ontologies. **3**(1).

A. Ortiz, I. Azevedo, R. Seiça, E. Carrapatoso and C. V. Carvalho (2009). Benchmarking of concept maps tools for learning objects. FINTDI (Promotion and Innovation with New Technologies in the Teaching of Engineering) Congress, Vigo, Spain, 14-15 December 2009.

A. Ortony, G. L. Clore and M. A. Foss (1987). "Cognitive Science". The referential structure of the affective lexicon. **11**: 341-364.

A. Ortony and T. J. Turner (1990). "315-331". Psychological Review. **97**(3): 315-331.

J.-O. Ostman (1988). Adaptation, variability and effect: comments on IPrA Working Documents 1 and 2. Working Document #3. Antwerp, International Pragmatics Association: 5-39.

L. Page, S. Brin, R. Motwani and T. Winograd (1999). The pagerank citation ranking: Bringing order to the web. Stanford Digital Library Technologies Project, Tech. Rep. SIDL-WP-1999-0120.

C. P. Paice (1984). "Soft Evaluation of Boolean Search Queries in Information Retrieval Systems". Information Technology, Res. Dev. Applications. **3**(1): 33-42.

G. Paquette and O. Marino (2006). Learning Objects, Collaborative Learning Designs and Knowledge Representation. Technology, Instruction, Cognition and Learning. **3:** 33-42.

J. Park, W. Cho and S. Rho (2010). "Evaluating ontology extraction tools using a comprehensive evaluation framework". Data & Knowledge Engineering. **69**: 1043–1061.

T. K. Park (1994). "Towards a theory of user-based relevance: a call for a new paradigm of inquiry". Journal of the American Society for Information Science. **45**: 135 - 141.

W. G. Parrott, Ed. (2001). Emotions in Social Psychology. Key Readings in Social Psychology, Psychology Press.

B. Parsia and E. Sirin (2004). Pellet: An owl dl reasoner. Proceedings of the International Workshop on Description Logics. International Semantic Web Conference 2004.

A. Paschke, D. Hirtle, A. Ginsberg, P.-L. Patranjan and F. McCabe. (2008). "RIF Use Cases and Requirements. W3C Working Draft". Available at http://www.w3.org/TR/rif-ucr/.

C. S. Peirce (1955). Logic as Semiotic: the Theory of Signs. Philosophical Writings of Peirce. J. Bucher. New York.

R. Pekrun (2005). "Progress and open problems in educational emotion research". Learning and Instruction. **15**: 497-506.

R. Pekrun, T. Goetz, W. Titz and R. P. Perry (2002). "Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research". Educational Psychologist. **37**: 91-105.

J. Peregrin (1996). What does one need when she needs 'higher-order' logic? LOGICA'96, Prague**:** 75-92.

I. Peters (2009). Folksonomies. Indexing and Retrieval in Web 2.0, Knowledge & Information: Studies in Information Science. De Gruyter.

R. Philip and L. Cameron (2003). Sharing and Reusing Learning Designs: Contextualising Enablers and Barriers. World Conference on Educational Multimedia, Hypermedia and Telecommunications (EDMEDIA) 2008, Vienna, Austria, June 30, 2008.

H. S. Pinto and J. P. Martins (2000). Reusing Ontologies. AAAI 2000 Spring Symposium on Bringing Knowledge to Business Processes, Karlsruhe, Germany AAAI Press**:** 77-84.

H. S. Pinto and J. P. Martins (2004). "Ontologies: How can They be Built?" Knowledge Information Systems. **6**(4).

K. R. Popper (1949). "Towards a Rational Theory of Tradition". The Rationalist Annual.

J. S. Poulin (1996). Measuring Software Reuse: Principles, Practices, and Economic Models. Addison-Wesley Professional.

A. Powel. (2005). "Notes about possible technical criteria for evaluating institutional repository (IR) software".University of Bath. Bath. Available at http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/ir-software.pdf.

G. Pyrounakis and M. Nikolaidou (2009). Comparing Open Source Digital Library Software. Handbook on Digital Libraries,. Idea Group Inc: 51-60.

D. R. Rehak, P. Dodds and L. Lannom (2005). A Model and Infrastructure for Federated Learning Content Repositories. The 14th International World Wide Web Conference (WWW 2005), Chiba, Japan, May 10--14, 2005.

P. Resnik (1999). "Semantic Similarity in a Taxonomy: An Information-based Measure and its Application to Problems of Ambiguity in Natural Language". Journal of Artificial Intelligence: 95-130.

B. A. N. Ribeiro and R. Muntz (1996). A belief network model for IR. Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval, Zurich, Switzerland.

G. E. Rice (1980). "On cultural schemata". American Ethnologist. **7**(1): 152–171.

À. Rius, M.-A. Sicilia and E. García-Barriocanal (2008). "An Ontology to Automate Learning Scenarios? An Approach to its Knowledge Domain". Interdisciplinary Journal of E-Learning and Learning Objects. **4**: 151-165.

S. Robertson (1997). The probability ranking principle in IR. Readings in information retrieval San Francisco, CA, USA, Morgan Kaufmann Publishers Inc.

S. Robertson and H. Zaragoza (2009). "The Probabilistic Relevance Framework: BM25 and Beyond". Foundations and Trends in Information Retrieval. **3**(4): 333–389.

D. E. Rumelhart and J. L. McClelland (1986). Parallel distributed processing: Explorations in the microstructure of cognition, Vol. I. MIT Press, Cambridge, MA.

T. Russell (2006). Cloudalicious: folksonomy over time. Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries, Chapel Hill, NC, USA, June 11 - 15, 2006.

J. Sabini and M. Silver (2005a). "Emotion and Experience". Psychological Inquiry. **16**(1): 49 – 56.

J. Sabini and M. Silver (2005b). "Why Emotion Names and Experiences Don't Neatly Pair". Psychological Inquiry. **16**(1): 1 – 10.

G. Salton, E. A. Fox and H. Wu (1983). "Extended Boolean Information Retrieval". Communications of the ACM. **26**(12).

G. Salton and M. E. Lesk (1983). Introduction to Modern Information Retrieval. McGraw-Hill.

- nchez-Alonso (2010). "Ranking Learning Objects through Integration of Different Quality Indicators". IEEE Transactions on Learning Technologies. **3**(4): 358 - 363.

M. Schoop (2001). "An introduction to the language-action perspective". Newsletter ACM SIGGROUP Bulletin **22**(2): 3-8.

M. Schoop, A. Moor and J. L. G. Dietz (2006). "The pragmatic web: a manifesto". Communications of the ACM. **49**(5): 75 - 76.

G. Schreiber, H. Akkermans, A. Anjewierden, R. Hoog, N. Shadbolt, W. V. d. Velde and B. Wielinga (1999). Knowledge Engineering and Management: The CommonKADS Methodology. MIT Press, Cambridge, Mass.

A. Schutz (1970). Reflections on the Problem of Relevance. Yale University Press, New Haven.

A. Schutz. (2008a). "XtraK4Me: Extraction of keyphrases for metadata creation".Semantic Information Systems and Language Engineering Group (SmILE), DERI. GATE plug-in software., Galway, Ireland. Available at http://smile.deri.ie/projects/keyphrase-extraction.

A. T. Schutz (2008b). Keyphrase Extraction from Single Documents in the Open Domain Exploiting Linguistic and Statistical Methods. Galway, National University of Ireland. **Masters of Applied Science**.

P. A. Schutz and S. L. Lanehart (2002). "Introduction: Emotions in education". Educational Psychologist. **37**: 67-68.

J. Searle, F. Kiefer and M. Bierwisch (1980). Introduction. Speech Acts Theory and Pragmatics J. Searle, F. Kiefer and M. Bierwisch. Dordrecht, Reidel**:** vii-ix.

J. R. Searle (1976). A Classification of Illocutionary Acts. Language in Society. **5:** 1-23.

R. Seiça, A. Ortiz, I. Azevedo, C. V. Carvalho and E. Carrapatoso (2009). Tree - A Repository of Learning Objects for Engineering based on open source tools. FINTDI Conference, Vigo, Spain, 4 - 15 December 2009.

R. A. R. Seiça (2009). TREE – Um Repositório de Objectos Educativos de Software Aberto com Perfil de Metadados, Faculdade de Engenharia da Universidade do Porto. **Master Degree**.

S. Sen, S. Lam, A. M. Rashid, D. Cosley, D. Frankowski, J. Osterhouse, F. M. Harper and J. Riedl (2006). Tagging, communities, vocabulary, evolution. Conference on Computer Supported Cooperative Work Banff, Alberta, Canada**:** 181-190.

S. Sestito and S. Goss (1994). Automated knowledge acquisition for diagnosis. Proceedings of the 1994 Second Australian and New Zealand Conference on Intelligent Information Systems**:** 427-431.

N. Shalmon (2005). Two conceptions of semantics. Semantics versus pragmatics. Z. G. Szabó, Oxford University Press.

A. Sheth, C. Ramakrishnan and C. Thomas (2005). "Semantics for the Semantic Web: The Implicit, the Formal and the Powerful". International Journal on Semantic Web & Information Systems. **1**(1): 1-18.

L. S. Shulman (1986). "Those who understand: Knowledge growth in teaching". Educational Researcher. **15**(4): 4–14.

J. Silva, I. Azevedo and C. V. Carvalho (2010). Visualização com nuvens de tags (Visualizing with cloud tags). I Encontro Internacional TIC e Educação (http://ticeduca.ie.ul.pt), Lisbon, Portugal, 19 -20 November 2010.

K. Siorpaes and E. Simperl (2010). "Human Intelligence in the Process of Semantic Content Creation". World Wide Web. **13**: 33–59.

O. Sitthisak and L. Gilbert (2009). Improving the pedagogical expressiveness of IMS LD. International conference on Technology Enhanced Learning Conference, Taipei.

R. Slavin (1995). Cooperative learning: Theory, research, and practice. Allyn & Bacon, Boston.

B. Smith and C. Welty (2001). FOIS introduction: Ontology - towards a new synthesis. The _International Conference on Formal Ontology in Information Systems, Ogunquit, Maine, USA.

H. V. d. Sompel, M. L. Nelson, C. Lagoze and S. Warner (2004). "Resource Harvesting within the OAI-PMH Framework". D-Lib Magazine. **10**(12).

J. F. Sowa (1999). Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks/Cole, Pacific Grove, CA.

J. F. Sowa. (2009). "Building, Sharing, and Merging Ontologies". Available at http://www.jfsowa.com/ontology/ontoshar.htm.

K. Sparck Jones and C. van Rijsbergen (1976). " Information retrieval test collections". Journal of Documentation. **32**(1): 59-75.

D. Sperber and D. Wilson (1986). Relevance: Communication and Cognition. Blackwell, Cambridge MA.

L. F. Spiteri (2006). "The use of folksonomies in public library catalogues". The Serials Librarian,. **51**(2): 75-89.

M. G. Sreekumar, T. Sunitha, P. Sreejaya, R. Biju, J. Kuriakose and P. Kalathil (2007). Institutional Repositories for Knowledge Management in Academic and Research Institutions. International Conference on Semantic Web and Digital Libraries (ICSD-2007)**:** 260-273.

S. Staab, H.-P. Schnurr, R. Studer and Y. Sure (2001). "Knowledge Processes and Ontologies". IEEE Intelligent Systems. **16**(1): 26–34.

R. Stalnaker (1972). Pragmatics. Semantics of Natural Language. D. Davidson and G. Harman. Dordrecht, Reidel**:** 380-97.

R. Stalnaker (1974). Pragmatic Presuppositions. Semantics and Philosophy. M. Munitz and P. Unger. New York, New York University Press**:** 197-213.

L. Stojanovic, S. Staab and R. Studer (2001). eLearning based on the Semantic Web. World Conference on the WWW and Internet - WebNet.

P. F. Strawson (1952). Introduction To Logical Theory. Methuen, London.

M. Strube and S. P. Ponzetto (2006). Wikirelate! computing semantic relatedness using Wikipedia. Proceedings of the 21st national conference on Artificial intelligence, Boston, Massachusetts,AAAI Press.

R. Studer, V. R. Benjamins and D. Fensel (1998). "Knowledge engineering: principles and methods". Data and Knowledge Engineering. **25**: 161–197.

D. Subramanian, R. Greiner and J. Pearl (1997). "The relevance of relevance". Artificial Intelligence (Special issue on relevance). **97**(1-2): 1–5.

P. Suraweera, A. Mitrovic and B. Martin (2005). A Knowledge Acquisition System for Constraint-based Intelligent Tutoring Systems. Proceeding of the 2005 conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology, Amsterdam, The Netherlands, July 18-22, 2005,IOS Press Amsterdam**:** 638-645.

Y. Sure (2003). Methodology, tools and case studies for ontology based knowledge management, Universit¨at Karlsruhe. **PhD**.

Y. Sure, M. Erdmann, J. Angele, S. Staab, R. Studer and D. Wenke (2002). OntoEdit: Collaborative Ontology Engineering for the Semantic Web. First International Semantic Web Conference (ISWC'02) (Lecture Notes in Computer Science LNCS 2342), Sardinia, Italy,Springer-Verlag**:** 221–235.

Y. Sure, S. Staab and R. Studer (2003). On-To-Knowledge Methodology. Handbook on Ontologies. S. Staab and R. Studer. Seiten, Springer**:** 117-132.

B. Swartout, P. Ramesh, K. Knight and T. Russ (1997). Toward Distributed Use of Large-Scale Ontologies. AAAI'97 Spring Symposium on Ontological Engineering, Stanford University, California**:** 138–148.

J. A. Swets (1969). "Effectiveness of information retrieval methods". American Documentation. **20**(1): 72-89.

J. Tague-Sutcliffe and J. Blustein (1995). A statistical analysis of the TREC-3 data. the Third Text Retrieval Conference (TREC-3)**:** 385–398.

C. Tattersall, D. Burgos, H. Vogten, H. Martens and R. Koper (2006). How to use IMS Learning Design and SCORM 2004 together. Proceedings of the SCORM 2006 conference, Taipei, Taiwan, January 17-19, 2006.

C. Tattersall, H. Vogten, F. Brouns, R. Koper, P. Van Rosmalen, P. Sloep and J. Van Bruggen (2005). "How to create flexible runtime delivery of distance learning courses". Journal of Educational Technology and Society. **8**(3): 226.

S. Ternier, D. Massart, A. Campi, S. Guinea, S. Ceri and E. Duval (2008). "Interoperability for Searching Learning Object Repositories. The ProLearn Query Language ". D-Lib Magazine. **14**(1/2).

V. Thai and S. Handschuh (2009). IVEA: toward a personalized visual interface for exploring text collections. Proceedings of the 14th international conference on Intelligent user interfaces (IUI '09), Sanibel Island, USA, Feb. 8-11, 2009.

The Consultative Committee for Space Data Systems. (2002). "Reference Model for an Open Archival Information Systems". Available at http://public.ccsds.org/publications/archive/650x0b1.pdf.

The Fedora Development Team. (2007). "Overview: The Fedora Digital Object Model. Fedora Release 2.2.1". Available at http://fedora-commons.org/download/2.2.1/userdocs/digitalobjects/objectModel.html.

The Fedora Development Team. (2008). "Fedora Tutorial #1: Introduction to Fedora. Fedora 3.0". Available at https://wiki.duraspace.org/download/attachments/4718930/tutorial1.pdf?version=1&modificationDate=1218459761506.

The JORUM Team. (2005). "Report on Open Source Learning Object Repository System". Available at http://www.jorum.ac.uk/docs/pdf/JORUM_osswatch_final.pdf.

S. S. Tomkins (1984). Affect theory. Approaches to emotion. K. R. Scherer and P. Ekman. Hillsdale, NJ, Erlbaum**:** 163-195.

J. Tuominen, T. Kauppinen, K. Viljanen and E. Hyvnen (2009). Ontology-based query expansion widget for information retrieval. The 5th Workshop on Scripting and Development for the Semantic Web (SFSW 2009), 6th European Semantic Web Conference (ESWC 2009), May 31 - June 4 2009.

A. Turani and R. Calvo (2006). Sharing Synchronous Collaborative Learning Structures using IMS Learning Design. The 7th International Conference on Information Technology Based Higher Education and Training, July 2006**:** 25–34.

H. Turtle and W. B. Croft (1991). "Evaluation of an inference network-based retrieval model." ACM Transactions on Information Systems. **9**(3): 187–222.

UKLOM. "UK Learning Object Metadata Framework". Available at http://www.cetis.ac.uk/profiles/uklomcore/.

UnfoldProject. "Learning design tools currently available or under development". Available at http://www.unfoldproject.net:8085/UNFOLD/general_resources_folder/tools/currenttools.

M. Uschold and M. King (1995). Towards a methodology for building ontologies. Workshop Held in Conjunction with IJCAI on Basic Ontological Issues in Knowledge Sharing**:** 19.1-19.12.

L. Valentín (2011). Orchestration of learning activities through the integration of third-party services in IMS learning design, UNIVERSIDAD CARLOS III DE MADRID. **PhD Thesis**.

C. J. van Rijsbergen (1979). Information Retrieval. Butterworths, London. 2nd edition.

V. Varma (2007). Use of Ontologies for Organizational Knowledge Management and Knowledge Management Systems. Ontologies. A Handbook of Principles, Concepts and Applications in Information Systems. Raj Sharman, Rajiv Kishore and R. Ramesh. **14:** 21-47.

A. Vazsonyi (1974). "Pragmatics: Rational Approach to Irrationality". Interfaces. **4**(3): 40-45.

K. Verbert and E. Duval (2004). Towards a global architecture for learning objects: a comparative analysis of learning object content models. The World Conference on Educational Multimedia, Hypermedia and Telecommunications (ED-MEDIA), Lugano, Switzerland, June 21-26, 2004.

D. Verpoorten, M. Poumay and D. Leclercq (2007). "The eight learning events model: A pedagogic conceptual tool supporting diversification of learning methods". Interactive Learning Environments. **15**(2): 151-160.

J. Verschueren (1987). Pragmatics as a theory of linguistic adaptation. Working document Nr. 1. Antwerp, International Pragmatics Association.

L. Vignollet, M. L. Bote-Lorenzo, J. I. Asensio-Pérez and Y. A. Dimitriadis (2009). A generic specification to the data-flow issue in the learning design field. Proceedings of the 6th International Conference on Advanced Learning Technologies, Riga, Latvia**:** 56—58.

E. Voorhees. (2000). "Presentation of the Text REtrieval Conference (TREC-9)".NIST. Gaithersburg, MD, USA. Available at http://trec.nist.gov/presentations/TREC9/intro/.

E. M. Voorhees and D. Harman (1999). Overview of the Seventh Text Retrieval Conference (TREC-7). The Seventh Text Retrieval Conference, August 1999**:** 1–23.

R. Vuorikari, H. Poldoja and R. Koper (2010). "Comparison of educational tagging systems – any chances of interplay?" International Journal of Technology Enhanced Learning. **2**(1-2): 111-131.

D. Watson, L. A. Clark and AukeTellegen (1988). "Development and validation of brief measures of positive and negative affect: The PANAS scales". Journal of Personality and Social Psychology. **54**(6): 1063–1070.

K. Webster, C. Kortekaas and A. Bennett (2007). Making Fedora easier to implement with Fez: A free open source content model and workflow management front-end to Fedora. Open Repositories 2007, San Antonio, Texas USA 23-26 January, 2007.

H. M. Weiss and R. Cropanzano (1996). An affective events approach to job satisfaction. Research in Organizational Behavior. **18:** 1-74.

T. Wharton (2009). Pragmatics and Non-Verbal Communication. Cambridge University Press.

D. Wiley (2002). Learning objects. Educational Technology: An Encyclopedia. Santa Barbara: ABC-CLIO. A. Kovolchick & K. Dawson (Eds.).

D. Wiley. (2006). "RIP-ping on Learning Objects". Available at http://opencontent.org/blog/archives/230.

R. Wilkinson and P. Hingston (1991). Using the cosine measure in a neural network for document retrieval. The 14th annual international ACM SIGIR conference on Research and development in information retrieval, Chicago, IL, USA, October 13 - 16, 1991.

T. Wilson, J. Wiebe and P. Hoffmann (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. Human Language Technology Conference/Conference on Empirical Methods in Natural Language Processing, Vancouver, Canada, October 6-8, 2005.

M. Winget (2006). User-defined classification on the online photo sharing site Flickr … Or, how I learned to stop worrying and love the million typing monkeys. Advances in Classification Research. 17th ASIS&T SIG/CR Classification Research Workshop, Philadelphia, PA, November 2006.

J. Xu and W. B. Croft (2000). "Improving the effectiveness of information retrieval with local context analysis". ACM Transactions on Information Systems. **18**(1): 79–112.

X. Xu, W. Zhu, X. Zhang, X. Hu and I.-Y. Song (2006). A Comparison of Local Analysis, Global Analysis and Ontology-based Query Expansion Strategies for Bio-medical Literature Search. IEEE International Conference on Systems, Man and Cybernetics (SMC '06), Taipei, 8-11 Oct. 2006**:** 3441 - 3446.

Z. Xu, Y. Fu, J. Mao and D. Su (2006). Towards the semantic web: Collaborative tag suggestions. WWW2006: Proceedings of the Collaborative Web Tagging Workshop, Edinburgh, Scotland, 23-26th May.

Y. Yanbe, A. Jatowt, S. Nakamura and K. Tanaka (2007). Can Social Bookmarking Enhance Search in the Web? Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries, Vancouver, BC, Canada, June 18 - 23, 2007.

W. Yu and J. Chen (2009). Ranking Ontology Based on Structure Analysis. Second International Symposium on Knowledge Acquisition and Modeling, China, Novermber 30-December1, 2009.

G. Yule (1996). Pragmatics, Oxford Introduction to Language Study Series. Oxford University Press.

F. Yus (2001). Ciberpragmática. El uso del lenguaje en Internet. Ariel, Barcelona,Spain.

F. Yus (2005). "The linguistic-cognitive essence of virtual community". IBÉRICA 9: 79-102.

F. Yus (2008). "Alterations of Relevance in Cyber-media". Universitas Psychologica. **7**(3): 629-642.

G. K. Zipf (1949). Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology. Addison-Wesley, Cambridge, MA.

A. Zollers (2007). Emerging Motivations for Tagging: Expression, Performance, and Activism. WWW 2007, Banff, Canada, May 8--12 2007.

Annex 1.IMS LD PROJECTS AND TOOLS

European Commission, JISC and some other institutions have been funding and carrying out work on IMS LD specification, under which some tools have also been developed.

The most prominent programmes that have supported the research about the employment of the IMS LD specification are:

- JISC Design for Learning Programme[89] had a two years' duration (till May 2008) and worked around Learning Design and it was not only focused on IMS LD specification. It funded 14 projects. It is part of a broad programme: JISC eLearning Programme[90], which will end in 2012;
- European Commission eContent Plus[91] was a four-year programme that finished in 2008 and funded projects that support the use of open standards, not only in the e-learning field, but in all the following areas: geographic content, educational content, cultural, scientific and scholarly content, addressing the accessibility and usability enhancement of digital resources in a multilingual environment;
- Framework Programme 6[92], FP6 for short, and Framework Programme 7 [93](FP7) funded some projects about Technology Enhanced Learning that used the IMS LD specification. The former ran from 2002 to 2006 and the latter will end in 2013. The Information Society Technologies (IST) and the Information and Communication Technologies (ICT) priorities in FP6 and FP7, respectively, are the most related to Technology Enhanced Learning;
- Leonardo da Vinci programme[94] finished in 2006 its second phase that started in 2000. It supported cooperation projects in the field of vocational training involving many Member States of the European Commission. One of the supported projects (OpenDock) implemented an IMS LD aware repository;
- The Lifelong Learning Programme[95] is a European Commission funded programme dedicated to the education and training area, covering learning opportunities from childhood to old age, during the period 2007-2013. It is the descendant of the Socrates, Leonardo da Vinci and eLearning programmes.

---

[89] http://www.jisc.ac.uk/whatwedo/programmes/elearning_pedagogy/elp_designlearn.aspx
[90] http://www.jisc.ac.uk/whatwedo/programmes/elearning.aspx
[91] http://europa.eu.int/information_society/activities/econtentplus/index_en.htm
[92] http://ec.europa.eu/research/fp6/index_en.cfm?p=0
[93] http://ec.europa.eu/research/fp7/index_en.cfm
[94] http://ec.europa.eu/education/programmes/leonardo/leonardo_en.html
[95] http://ec.europa.eu/education/programmes/newprog/index_en.html

In section A1.1 some projects that use IMS LD specification are presented, while IMS LD tools, many of them implemented under these discussed projects, are further characterised in section A1.2. Last section summarises this annex.

A1.1. IMS LD projects

In this section some projects are cited and discussed; not all the projects objectives and developments are mentioned but only the strands related to the IMS LD specification. Some early projects are discussed in section A1.1.1. Those projects were the basis for more advanced developments that later were carried out by more recent projects that are described in section A1.1.2.

A1.1.1. Early research projects

Some early projects were very important for the attention gradually given to the specification and to make it reach its current use (in research projects and in the academic world - not really a generalised employment). CooperCore is an engine that was developed by the Open University of the Netherlands under ALFANET project. It was the first one to be developed and it has been incorporated in many players. The UNFOLD project has many available information on IMS LD. Although the project has already finished, the repository continues up to date.

ACETS (http://www.acets.ac.uk/, United Kingdom)

Assemble, Catalogue, Exemplify, Test & Share (ACETS) was a JISC funded project dedicated to research of pedagogical use of reusable learning objects in a broad range of healthcare educational settings. It ran from September 2002 to November 2005. Among others recommendations, the project suggests that reusable learning objects can be readily accessed and aggregated into learning activities, making easier the use of RLO as they are strong related to practical use.

Semi-structured interviews based on the IMS LD specification were conducted in order to create a formal way to documenting current practice, which can be used to recognise teaching scenarios to be studied and reused. The ACETS exemplars are available at http://www.acets.ac.uk/exempgrid.asp.

ALFANET (http://alfanet.ia.uned.es/alfanet/, Europe)

Active Learning for Adaptive Internet (ALFANET) project ran from May 2002 to April 2005 and provided "an eLearning adaptive platform that allows individuals to have an interactive, adaptive and personalised learning through the internet". It implemented CopperCore, the first IMS LD engine that covered A, B and C levels, later disseminated by the UNFOLD project. Copper Core was released under GNU General Public License.

E-LANE project (http://e-lane.org, Europe and Latin America)

European and Latin American New Education (E-LANE) was a one-year project that started on 30 March 2006. It was funded by the @LIS, a programme of the European Commission. As a result of this project, .LRN platform provided support for IMS LD.

Mod4L (http://mod4l.com, Canada)

It was a JISC funded project under Design for Learning programme that ran from 2002 to 2004 involving some Canadian universities mainly dedicated to "extending and enhancing the contributions and efforts in the areas of IMS Learning Design". One of its activities was the exploration and evaluation of the LAMS tool in 2004 concluding that "the LAMS system represents a solid first step towards an IMS Learning Design environment that focuses on context rather than content".

OpenDock project (http://www.opendockproject.org, Europe)

The OpenDock project, funded by the Leonardo Programme, sought to "stimulate the sharing of eLearning activities and resources in Vocational Education and Training (VET) by demonstrating how Units of Learning (UoLs) defined in IMS Learning Design can reuse learning resources and be shared between different institutions" [Elferink, Griffiths et al., 2006]. A new Open Source distributed file repository called OpenDocument.net was developed with built in support for Open Document Standards, including IMS LD.

R2R (http://tlc.ucalgary.ca/projects/r2r, Canada)

Repository to Reality (R2R) project ran from 2002 to 2004 involving some Canadian universities mainly dedicated to "extending and enhancing the contributions and efforts in the areas of IMS Learning Design". One of its activities was the exploration and evaluation of the LAMS tool on 2004 concluding that "the LAMS system represents a solid first step towards an IMS Learning Design environment that focuses on context rather than content".

Reload (http://www.reload.ac.uk/, United Kingdom)

Reusable eLearning Object Authoring & Delivery (RELOAD) was a JISC (http://www.jisc.ac.uk/) funded project (X4L strand B) for "developing tools to facilitate the use of emerging Learning Technology Interoperability specifications such as those produced by ADL and IMS". It started on 1 October 2002 and ended on 31 July 2005. Under this project, among others developments, an IMS LD player and an editor were released as open source software.

SCOPE (http://www.tecn.upf.es/scope/showcase, Europe)

Structuring Content for Online Publishing Environments (SCOPE) was a European Commission project (eContent programme) that started on 1st January 2002 and ended on 30th June 2003. Its objective was "to demonstrate how the scientific intellectual property held by many universities and public bodies can be reused and transformed into a valuable resource and high value products". A commercial service has been launched, G&H Continuada (http://www.ghcontinuada.com), providing academic information about Gastroenterology and Hepatology for medical professionals that belong to the Spanish speaking community. Also a workflow solution was proposed to assist coordination and collaboration in the editorial process of content creation of a medical journal. In collaboration with the RELOAD project (see below) classification tools were developed for course authors to assist them in the task of using IMS LD to describe content and organise it into courses.

TELCERT (http://www.opengroup.org/telcert/, Europe)

Technology Enhanced Learning Certification - European Requirements & Testing (TELCERT) was an EU-funded project that started on 1 January 2004 and ended on 30 June 2006. It was based on RELOAD tools and has developed tools and resources to achieve better interoperability in the e-learning field. It worked on conformance tests for IMS LD specification.

UNFOLD (http://www.unfold-project.net/, Europe)

Understanding New Frameworks of Learning Design (UNFOLD) was a project that was supported by the UE Sixth Program Framework during 2004 and 2005. The UNFOLD project "was conceived of as a measure to promote and coordinate the adoption, implementation and use of IMS Learning Design and related specifications". Although it finished, there are many resources related to IMS LD specification available in its Web site. Also the UNFOLD forums and activities were moved to the Learning Networks for Learning Design Web site (http://imsld.learningnetworks.org/login/index.php).

A1.1.2. Recent research projects

In this section some projects that recently used IMS LD specification for some purpose are briefly discussed.

Calibrate (http://calibrate.eun.org, Europe)

It was a STREP (Specific Targeted Research Projects) project that aimed at preparing a federation of repositories and it also planned to provide tools for teachers based on IMS LD to "support the collaborative use and exchange of learning resources in schools". It started on October 2005 and finished on March 2008.

Cooper (http://www.cooper-project.org, Europe)

Collaborative Open Environment for Project-Centred Learning or Cooper, for short, was a project funded by the IST programme of the European Commission. It was a two-year project that started on 1 December 2005. Under this project, two pedagogical scenarios were considered: Virtual Company (VC) and Virtual Project (VP). This project used IMS LD to support cooperation among group of students, distance located from each other.

D4LD (http://sled.open.ac.uk/web/project/d4ld.jsp, Europe)

It was a JISC funded project initiated on 1 May 2006 and ended on 31 October 2007. D4LD aimed to produce a stable LD player. One of the project deliverables was an enhanced open source release of the SLeD LD player and the CopperCore LD engine. It also carried on some research about the re-use of IMS LD units of learning.

Fle (http://mlab.uiah.fi/fle/, Finland)

Future Learning Environment (FLE) project was supported by the Finnish Technology Development Centre, industrial partners and the Finnish Ministry of Education. It started in June, 1997 and ended in 1999. It later was continued by the Fle2 project funded by Nordunet2, a research programme financed by the Nordic Council of Ministers and by the Nordic Governments. The most recent generation of Fle Software is Future Learning Environment 3 (FLE3) (http://fle3.uiah.fi/). Fle3 research and development was supported by the European Commission in the IST framework. Fle3 software uses IMS LD as its file format to specify its courses and the latest version (1.5), which is an open-source and free software released under the GNU General Public Licence (GPL).

iClass (http://www.iclass.info, Europe)

It was a project funded by European Commission IST Framework Programme 6. It started on 1st January 2004 and it ended on 31st July 2008. It aimed to "to develop an intelligent cognitive-based open learning system and environment, adapted to individual learners' needs and ensuring their take-up in the education sector at a European level". Under this project a graphical authoring tool for Learning Design, ASK-LDT, was developed.

IDLD (http://www.idld.org, Canada)

Implementation and Deployment of Learning Design Specification was a Canadian project committed to the diffusion of "basic educational modelling concepts, like learning designs patterns and examples, in order to produce a repository of IMS LD compliant units of learning". The main result of the project was a portal available at http:// www.idld.org. Under this project a learning design classification based on LOM was carried, useful for learning object repositories with learning design items.

All the resources developed under this project are in the public domain using eCommons licenses. A Canadian University (Télé-université) has been maintaining the IDLD portal.

LD4P (http://www.hope.ac.uk/ld4p/, England)

IMS Learning Design for Practitioners (LD4P) started on June 2006 and ended on November 2007 and as the project title suggests it was centred on the IMS LD specification. The project worked with "FE/HE teaching practitioners to evaluate the RELOAD IMS LD editor in preparing learning designs (specifically IMS LD) to support delivery of their modules". It aimed at defining a set of requirements for a user interface to an IMS LD editor, focussing on usability aspects.

The project ran a number of workshops for practitioners related to the use of the RELOAD IMS LD editor.

PROLEARN (http://www.prolearn-project.org/, Europe)

PROLEARN project was a Network of Excellence in Professional Learning financed by the IST programme of the European commission. It ran from the years 2004 till 2008. PROLEARN aimed to close "the gap between learning and academic". In some PROLEARN events the IMS LD specification was discussed and also the modelling of learning activities using the specification.

Prolix (http://www.prolixproject.org/index.html, Europe)

Prolix was a 4-years integrated project supported by the European Commission under the Sixth Framework Programme. It started on December 2005. Prolix aimed to "align learning with business processes in order to enable organisations to faster improve the competencies of their employees according to continuous changes of business requirements". It used IMS LD to prepare the learning activities needed when there was a change in a business process. Under this project an IMS Learning Design editor was developed, Prolix GLM (Graphical Learning Modeller).

TENCompetence (http://www.tencompetence.org, Europe)

The European Network for Lifelong Competence Development (TENCompetence) was a 4 years project supported by the European Commission through the IST Programme. It ran from December 2005 to November 2009. It used IMS LD in the context of Life Long Competency Development and developed systems that employs IMS LD, in the context of one of its work packages, WP6 (Learning Activities & Units of Learning), that focussed specifically on Learning Activities employing IMS LD. Also one of its main tasks was the development of an open source Learning Design editor (Recourse).

Sharing the Load (http://www.ucel.ac.uk/load/)

Sharing the LOAD: Learning Objectives, Activities and Designs (or Sharing the LOAD, for short) was an 18-month project that started on 1 May 2006. It was supported by JISC and explored the reusability of learning designs. It conducted a series of workshops where participants created learning designs. They worked on the creation of a learning design taxonomy with 12 key attributes, supported by their own collection. They matched LO attributes to LD, as can be seen is Figure 106. The Sharing the LOAD learning object pedagogical attributes (in yellow) were mapped to the environment, roles and activities of the IMS Learning Design Framework (in blue).

IMS Learning Design Framework

Environment	Roles	Activity
Text	Recipient of information	MCQ
Audio	Navigator	Answer Selection
Images	Active participant	Drag and drop
Animation	Contributor	Text entering
Video	Self-assessor	Image selection
Interactive elements	Problem solver	Image manipulation

Inter-activity	Objective	Integration	Context	Richness	Pre-requisites	Support	Feedback	Self- direction	Navigation	Assess-ment	Alignment
None	Not stated	None	Total	Very Low	Very Low	None	None	Very Low	None	None	None
Low	Minimal match	Low	High	Low	Low	Low	Low	Low	Low	Low	Low
Medium	Medium match	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium
High	High match	High	Low	High	High	High	High	High	High	High	High
Very high	Complete match	Very high	Neutral	Very high	Very high	Very high	Very high	Very high	Very high	Very high	Very high

Learning Object Pedagogical Attributes

Figure 106. Matching learning object pedagogical attributes to IMS LD (From http://www.ucel.ac.uk/load/docs/IMSLD_LOattributes_map.pdf)

A1.2. IMS LD Tools

IMS LD is used to formally describe educational processes, which are denominated Unit of Learning (UoL) by the specification. These educational processes are encoded in XML. Though IMS Learning Design is considered a complex specification, many tools have being developed making possible the work with the IMS LD at various levels.

Some platforms are much closer to the specification, and some, although very intuitive, are not. This section will discuss some of these tools and their features.

The LD tools can be categorised as:

- Higher-level tools or
- Lower-level tools.

The distinction between them is in the knowledge level about LD indispensable to successful use the tool [Koper and Tattersall, 2005]. Higher-level tools are usually considered appropriate to be used by teachers. CoSMoS (Collaboration Script Modelling System) is not a higher-level authoring tool, and ASK-LDT and LAMS are.

Another possible classification of Learning Design tools is:

- Bottom-up tool,
- Top-down tool.

This classification is related to how a learning design emerges: from the lower level details to the learning design or from a top view that later is refined to reach the resources, activities, roles, and so on.

The LD tools are stand-alone tools, although the authoring of UoL usually is not an individual task. Integrated tools should appear in the future, making possible multiple users with different roles to work around a learning design.

There are many IMS LD related tools currently available:

- engines,
- editors,
- players.

Each of these different kinds of tools is explained in the following sections, but for a complete list of the tools available or under development, see [UnfoldProject].

A1.2.1. IMS LD engines

LD is a declarative language, without any implementation information or details. An IMS LD engine is a software application that interprets the XML notation, and usually appears incorporated in an LD player.

CopperCore (http://coppercore.sourceforge.net/) is a J2EE runtime IMS LD engine. It can be incorporated in other tools, so it is very interesting for system developers, especially to player developers. When version 2.2 was released on December 2002, it was the first open source engine to provide support for all 3 IMS LD levels. The last version (3.1) was released at the end of March 2007 and was funded by D4LD Developing for Learning Design project.

CopperCore engine supplies APIs to build user interface, but it does not supply interfaces ready to use.

The Learning Management System .LRN has support for IMS LD and has its own engine [Cid, Valentín et al., 2007].

A1.2.2. IMS LD editors

LD Editors are Learning Design authoring tools built around the concepts of "activity" and "learning flow", usually providing support to guarantee the validity of the generated document.

Reload, Ask-LDT, CopperAuthor and CoSMoS are LD editors (see Table 50). We can say that Reload is the reference editor for IMS LD. Most available LD authoring tools only allow the editing of UOLs at Level A, but Reload and CoSMoS provide full support to edit learning design at levels A, B, and C. Levels B and C require dealing with some technical vocabulary that does not facilitate its operation in a friendly way.

LD Editors can be classified as:

- Tree-based Authoring Tools,
- Diagram-based Authoring Tools.

A tree-based authoring tool presents the elements of LD aggregated in more and more specialised levels (like a tree), with a well defined hierarchy. Users can navigate through the structure and enter values for the shown elements. Some LD authoring tools that adopted a tree-based approach are Cosmos, CopperAuthor, and RELOAD.

A diagram-based authoring tool represents IMS LD elements as nodes and their relations as arrows in a graphical representation. Examples of diagram-based authoring tools are MOT+ (http://www.licef.teluq.uquebec.ca/gp/eng/productions/mot.htm) and ASK-LDT. LAMS (http://lamsfoundation.org/) is a diagram-based authoring tool. Not all diagram-based tools can be considered higher-level tools: MOT+ is not one.

In a tree-based authoring tool it is easy to understand the major components of a learning design, but it is easier to understand the relations between the different elements in a diagram-based tool. But with a great number of elements it can be difficult to put all of them in a diagram; for this reason MOT+ has two layers. Also, as recognised by Miao, in a diagram-based tool it is impossible or very difficult to use the graphical information created in different tools.

Only one note: both MOT+ and LAMS are not very close to the IMS LD specification, and thus they are considered LD tools, but learning design tools with LD export capabilities.

Table 50. Learning Design editors

Tool Name	Web site	Author	Status	Levels
ASK LDT	www.ask4research.info/	University of Piraeus (iClass project)	-	A, B
Collage	http://gsic.tel.uva.es/collage	GSIC/EMIC group of the University of Valladolid, Spain	Integrated in Reload	A
CopperAuthor	www.copperauthor.org	OUNL	Available as open source	A
Cosmos	www.unfold-project.net:8085/UNFOLD/general_resources_folder/cosmos_tool.zip	University of Duisburg	Available as open source	A, B, C
EduCreator Editor	-	Chronotech	Proprietary, under development	A
elive LD Suite	http://www.elive-ld.com/content/index_ger.html	Elive Learning Design – a German company	Proprietary, under development	LD level A import – export
HyCo-ALD Editor	-	University of Salamanca	-	A
Prolix GLM	http://sourceforge.net/projects/prolix-glm/.	Prolix project	Available as open source	A, B
Recourse	http://tencompetence-project.bolton.ac.uk/ldauthor/index.html	TENCompetence project	Available as open source	A, B
Reload LD Editor	www.reload.ac.uk/ldeditor.html	University of Bolton	Available as open source	A, B, C

A1.2.3. IMS LD players

An IMS LD player is an application that interprets an XML file and, considering all the roles of the unit of learning, presents the activities in a sequenced order, based on the specified properties and conditions. The most popular player is SLED player, but others are available or being developed (see Table 51).

Table 51. Learning Design players

Tool Name	Web site	Author	Status	Levels
EduBox	-	OUNL and Blackboard	Proprietary	A, B, C
SLED	http://sled.open.ac.uk/web/	OU UK and OUNL	Demo available (it is necessary to have the Coppercore engine installed)	A,B,C
RELOAD				
IMS LD Package for the .LRN platform	http://strauss.gast.it.uc3m.es/	University Carlos III of Madrid	Under development	A

Another LD player currently available is CopperCore Player, but it is more a working prototype to show how UOLs run and it is not a very functional tool as it doesn't have a user-friendly interface.

A1.1. Summary

Many projects that used the IMS LD specification were presented in this annex. These projects are spread almost all over the world, but especially over Europe, Canada and Australia.

Also many tools have been developed. However, one of the limitations of the specification is the number of e-learning platforms that support the specification, but the large number of projects and tools developed or under development, also in the commercial sector, can be seen as a promising indicator that it is a matter of time to have it largely adopted. Furthermore, the .LRN platform has LD support, and the possibility of the LD integration into Moodle was previously announced[96].

[96] See, for instance, http://docs.moodle.org/dev/Portfolio_API.

Annex 2. EMOTION

Emotions have a close connection to language and are related to one segment of pragmatics as humans beings are emotional creatures. This is an aspect highlighted in the following extract from a paper entitled 'Pragmatics: Rational Approach to Irrationality': "The user of an information system, the receiver of the communication, is a component of the information system. Therefore, we must know something about the user, and must make assumptions about him. We want him to react to the information received in accordance with our intentions. If he reacts contrary to our intentions, our information system is worthless. One of the most absurd assumptions we make about users of our information systems is that they will react rationally" [Vazsonyi, 1974].

It is interesting to note that the number of words in a given language dedicated to emotions varies. Also, not all emotions have a word associated to them in every language. Even the word 'emotion' did not always existed in English. Its first usages date from mid 16th century, but denoting a public disturbance. Its actual meaning dates from early 19th century, according to the Oxford Advanced Learner's Dictionary.

In this annex some definitions are presented related to the theme of emotion.

A2.1. Emotions

In an etymological perspective, the word 'emotion' is made from two Latin words: 'ex' (out) and 'motion' (movement). Actually, emotions, or their manifestation, are related to movements, and not only facial ones.

It is easy to think about emotions, talk about sadness or happiness, usually seen as emotions. However, it is much more difficult to define them, even for psychologists, a point that was highlighted by Parrot in his book 'Emotions in Social Psychology'. Notwithstanding that assumed difficulty, he advances a possible definition for emotions: "ongoing states of mind that are marked by mental, bodily, or behavioural symptoms" [Parrott, 2001] (page 3). He distinguishes mood and emotion, using what philosophers have called 'object directedness' [Nissenbaum, 1985]. Thus, **emotions** are directed to something, e.g. person, object or event. On the other side, **moods** have no object associated, or might have a very general one, lacking the very specific orientation to something.

An additional difference between mood and emotion is provided in [Briner and Kiefer, 2007]. According to the authors, an emotion causes short-lived, rapidly changing, and strong in intensity reactions, whereas mood has a lower intensity and is more diffuse.

Another concept related to emotion is **affect,** which can be defined as "an umbrella for a set of more specific mental processes including emotions, moods and possible attitudes" [Bagozzi, Gopinath et al., 1999]. Therefore, affect is a more general notion than emotion.

The study of emotion and its influence has gathered attention from many researchers from different fields, such as the study of emotion at workplace. For instance, Weiss and Cropanzano argue that events at work that trigger positive affective reactions over time influence workplace attitudes, like job satisfaction, organisational trust and commitment, while mitigating behaviours such as absenteeism, lateness and low productivity, among others [Weiss and Cropanzano, 1996].

It is common to classify affect into positive or negative, according to the general pleasantness or unpleasantness of an eliciting event [Frijda, 1993]. **Positive Affect** (PA) and **Negative Affect** (NA) are two relatively independent affective state dimensions. Watson et al. characterise PA as "the extent to which a person feels enthusiastic, active, and alert", with a high PA being "a state of high energy, full concentration, and pleasurable engagement", whereas NA is considered a "general dimension of subjective distress and unpleasurable engagement that subsumes a variety of aversive mood states, including anger, contempt, disgust, guilt, fear, and nervousness, with low NA being a state of calmness and serenity" [Watson, Clark et al., 1988].

Numerous Positive Affect (PA) and Negative Affect (NA) scales have been proposed along the years. For instance, the Positive Affect and Negative Affect Scale (PANAS) comprises two 10-item mood scales [Watson, Clark et al., 1988].

Another hard task is the definition of what is an emotion or not, as researchers have adopted classification schemes that differ one from the others. For instance, for Ortony et al. surprise is regarded a cognitive state [Ortony, Clore et al., 1987], as the affective state of a surprised person can be positive, negative, or even neutral, and they deny the possibility of an emotion being affectively neutral. Others regard surprise as an emotion, and a basic one [Ekman, Friesen et al., 1982] [Frijda, 1986] [Izard, 1971] [Tomkins, 1984].

Ortony lists more than thirteen reasons that different researchers have considered to classify an emotion as a fundamental one or not [Ortony and Turner, 1990]. Not surprising, there is not a universal agreement about which emotions are basic. Ekman initially considered the following emotions as basic: *happiness*, *sadness*, *fear*, *surprise*, *anger*, and *disgust*, but later he expanded his initial list to comprise eleven more: *amusement*, *contempt*, *contentment*, *embarrassment*, *excitement*, *guilt*, *pride in achievement*, *relief*, *satisfaction*, *sensory pleasure*, and *shame* [Ekman, 1999]. Those lists were based on notion of universal facial expressions.

A2.1.1. The role of emotions in learning processes

The importance of emotions in education was almost entirely neglected until 2002 [Schutz and Lanehart, 2002]. In that year there was a special issue of Educational Psychologist wholly devoted to this theme. Many studies about the subject appeared since then, such as [Pekrun, 2005] [Boekaerts, 2003] [Pekrun, Goetz et al., 2002], among others.

Also, although there has been a tendency to separately study cognition, motivation, and emotion, there is recent indication from educational research that they are interlinked and emotions represent an integral part of instructional interactions [Meyer and Turner, 2002].

Recent research in neuroscience has pointed to the existence of a relationship between emotion, social functioning, and decision making. Mary Helen Immordino-Yang and Antonio Damasio highlight in the paper 'We Feel, Therefore We Learn: The Relevance of Affective and Social Neuroscience to Education' that these studies can potentiality change the "understanding of the role of affect in education" and recommend that educators should consider the importance of students' emotions, otherwise they will fail to take into consideration a significant force in learning processes [Immordino-Yang and Damasio, 2007].

Later, Mary Helen Immordino-Yang and Matthias Faeth added that "the message from social and affective neuroscience is clear: no longer can we think of learning as separate from or disrupted by emotion, and no longer can we focus solely at the level of the individual student in analyzing effective strategies for classroom instruction" [Immordino-Yang and Faeth, 2009].

Annex 3. USE CASES

The purpose of this appendix is to provide a formal description of the functionalities associated with the model proposed for learning objects repositories. Use cases are requirements, mainly functional or behavioural, that specify what the system does. This appendix describes in detail the use cases previously introduced in Chapter 6.

The description of the different use cases is done by presenting a series of tables in a common format. Table 52 presents the use case template that is followed for all use cases in this annex. This template represents a compromise between the casual and the fully dressed use case format [Cockburn, 2000] (pages 120-122). The first column presents a template main section, while the second one gives a brief explanation about its usage.

Table 52. Use case template

Section	Description
Abbreviated designation	UCXY, where X and Y are digits and they (together) are unique for each use case.
Extended designation	It is a use case designation that summarises an action related to the use case.
Primary actors	The principal actors that demands system services to fulfil a goal.
Stakeholders and interests	This part states who cares about the use case, and what they want.
Preconditions	This part states what must be true on start.
Postconditions	This part states what must be true on successful completion.
Main success scenario	A typical path scenario of success.

A3.1. List of use cases

Table 53 describes Use case UC01.

Table 53. Use case UC01

Abbreviated designation	UC01	
Extended designation	Manage users	
Primary actors	Administrator	
Stakeholders and interests	Administrator	Wants accurate, easy and fast management of users.
	Institution manager	Wants a repository that allows the management of users, ensuring their satisfaction, but also an accurate process.
	Consumer/ Ontology engineer/ Opinion analyst/ Provider/Pedagogical adviser	Wants a system able to easily and accurately process his information with regard to his functions in the repository.
Preconditions	The Administrator is identified and authenticated.	
Postconditions	All the data pertinent to the performed operations are saved.	
Main success scenario	1-The Administrator chooses the option to manage users; 2-The system shows a number of options available to the Administrator, which includes the options to manage users (see UC02, UC03, UC04, UC05), but also to exit from the user management section; 3-The Administrator select the desired option and the system proceeds in accordance with his selection; 4-Steps 2 and 3 might be repeated until the Administrator chooses the exit option.	

Table 54 provides some information about Use case UC02.

Table 54. Use case UC02

Abbreviated designation	UC02	
Extended designation	List users	
Primary actors	Administrator	
Stakeholders and interests	Administrator	Wants accurate, easy and fast provision of users list.
	Institution manager	Wants a repository that allows listing the registered users and corresponding registered data, which could serve different purposes.
Preconditions	The Administrator is identified and authenticated and the option to list users was selected (UC01).	
Postconditions	-	
Main success scenario	1-The system shows the sort options available to the Administrator; 2-The Administrator selects the desired sort option; 3-The system proceeds in accordance with the selected option, providing the users list.	

Table 55 describes Use case UC03.

Table 55. Use case UC03

Abbreviated designation	UC03	
Extended designation	Modify user	
Primary actors	Administrator	
Stakeholders and interests	Administrator	Wants easy, accurate and fast option to modify user data.
	Institution manager	Wants a repository that allows the management of users with regard to the modification of some of his data, ensuring an accurate processing of these modifications.
	Consumer/ Ontology engineer/ Opinion analyst/ Provider/Pedagogical adviser	Wants a system able to process with modifications in his data whenever necessary.
Preconditions	The Administrator is identified and authenticated and the option to modify user was selected (UC01).	
Postconditions	The changed fields are updated.	
Main success scenario	1-The Administrator provides the user identification number; 2-The system shows all the fields, including the modifiable ones; 3-The Administrator introduces new values into one or more fields; 4-The system validates the provided data and acknowledges the successful completion of the update operation.	

Table 56 provides some information about Use case UC04.

Table 56. Use case UC04

Abbreviated designation	UC04	
Extended designation	Delete user	
Primary actors	Administrator	
Stakeholders and interests	Administrator	Wants easy, accurate and fast option to delete users.
	Institution manager	Wants a repository that allows the deletion of users, because they are not anymore with active functions in the institution, or any other reason.
	Consumer/ Ontology engineer/ Opinion analyst/ Provider/Pedagogical adviser	Wants a system able to stop considering him as user when he is not anymore interested in its usage or able to use it.
Preconditions	The Administrator is identified and authenticated and the option to delete user was selected (UC01).	
Postconditions	The user status changes and he is not anymore able to use the system for operations that need previous authentication.	
Main success scenario	1-The Administrator provider the user identification number; 2-The system shows all the fields and asks for delete confirmation; 3-The Administrator confirms that the user is to be deleted; 4-The system changes the user status to inactive, and the user is not anymore able to log on. The system acknowledges the successful completion of the logical deletion operation.	

Table 57 describes Use case UC05.

Table 57. Use case UC05

Abbreviated designation	UC05	
Extended designation	Add user	
Primary actors	Administrator	
Stakeholders and interests	Administrator	Wants accurate, easy and fast entry of new users.
	Institution manager	Wants a repository that allows the addition of users in an accurate process.
	Consumer/ Ontology engineer/ Opinion analyst/ Provider/Pedagogical adviser	Wants to use the system and, thus, needs the necessary information to allow its proper usage.
Preconditions	The Administrator is identified and authenticated and the option to add user was selected (UC01).	
Postconditions	The system saves all the data related to the new user.	
Main success scenario	1-The system requisites mandatory and optional fields for user addition; 2-The Administrator fills in the fields; 3-The system validates the data and acknowledges the successful completion of the addition operation.	

Table 58 provides some information about Use case UC06.

Table 58. Use case UC06

Abbreviated designation	UC06	
Extended designation	Manage domain areas	
Primary actors	Administrator	
Stakeholders and interests	Administrator	Wants accurate, easy and fast management of domain areas.
	Institution manager	Wants a repository that allows the management of domain areas with interest to the institution with regard to the knowledge representation, but also an accurate process.
	Provider	Wants a system able to consider some other concepts to those provided when characterising resources or searching for resources of interest, in an easy and fast process that regards the relevant domain areas.
	Consumer	Wants a system able to considerer some other concepts to those provided when searching for resources of interest, in an easy and fast process that regards the relevant domain areas.
Preconditions	The Administrator is identified and authenticated.	
Postconditions	All the data pertinent to the performed operations are saved.	
Main success scenario	1-The Administrator chooses the option to manage domain areas; 2-The system shows a number of options available to the Administrator, which includes the options to manage domain areas (see UC07, UC08, UC09), but also to exit from the domain area management section; 3-The Administrator select the desired option and the system proceeds in accordance with his selection; 4-Steps 2 and 3 might be repeated until the Administrator chooses the exit option.	

Table 59 describes Use case UC07.

Table 59. Use case UC07

Abbreviated designation	UC07	
Extended designation	Add domain area	
Primary actors	Administrator	
Stakeholders and interests	Administrator	Wants accurate, easy and fast entry of new domain areas.
	Institution manager	Wants a repository that allows accurate addition of domain areas to assist the users in some repository functions.
	Provider	Wants a system able to consider some other concepts to those provided when characterising resources or searching for resources of interest, in an easy and fast process that regards all the relevant domain areas.
	Consumer	Wants a system able to considerer some other concepts to those provided when searching for resources of interest, in an easy and fast process that regards all the relevant domain areas.
Preconditions	The Administrator is identified and authenticated and the option to add domain area was selected (UC06).	
Postconditions	A new domain area was saved and it may start to have one or more ontologies associated.	
Main success scenario	1-The system requisites the domain area designation; 2-The Administrator provides the designation; 3-The system validates the provided data and acknowledges the successful completion of the addition operation.	

Table 60 provides some information about Use case UC08.

Table 60. Use case UC08

Abbreviated designation	UC08	
Extended designation	List domain areas	
Primary actors	Administrator and Ontology engineer	
Stakeholders and interests	Administrator/ Ontology engineer	Wants easy, accurate and fast access to the list of domain areas.
	Institution manager	Wants a repository that allows listing the domain areas, which could serve different purposes, such as finding uncovered areas.
	Provider	Wants a system able to consider some other concepts to those provided when characterising resources or searching for resources of interest, in an easy and fast process that regards the relevant domain areas.
	Consumer	Wants a system able to considerer some other concepts to those provided when searching for resources of interest, in an easy and fast process that regards the relevant domain areas.
Preconditions	The Administrator or Ontology engineer is identified and authenticated. The former should have selected the option to list domain areas from UC06.	
Postconditions	-	
Main success scenario	1-The system shows the sort options available to the Administrator or Ontology engineer; 2-The Administrator or Ontology engineer selects the desired option; 3-The system proceeds in accordance with the selected option and lists the domain areas.	

Table 61 describes Use case UC09.

Table 61. Use case UC09

Abbreviated designation	UC09	
Extended designation	Delete domain area	
Primary actors	Administrator	
Stakeholders and interests	Administrator	Wants accurate, easy and fast option to delete domain areas.
	Institution manager	Wants a repository that allows the deletion of domain areas, because they are not anymore directly or indirectly related to any course, or any other reason.
	Provider	Wants a system able to consider some other concepts to those provided when characterising resources or searching for resources of interest, in an easy and fast process that regards only the relevant domain areas.
	Consumer	Wants a system able to considerer some other concepts to those provided when searching for resources of interest, in an easy and fast process that regards only the relevant domain areas.
Preconditions	The Administrator is identified and authenticated and the option to delete domain area was selected (UC06).	
Postconditions	The domain area is not considered anymore and ontology engineer is not able to associate ontologies to the removed area anymore.	
Main success scenario	1-The system shows the domain areas; 2-The Administrator selects the domain area to be deleted; 3-The system asks for delete confirmation; 4-The Administrator confirms that the domain area is to be deleted; 5-The system changes the domain area status to inactive. The system acknowledges the successful completion of the logical deletion operation.	

Table 62 provides some information about Use case UC10.

Table 62. Use case UC10

Abbreviated designation	UC10	
Extended designation	Visualise own tags	
Primary actors	Consumer and Provider	
Stakeholders and interests	Consumer/Provider	Wants accurate, easy and fast visualisation of own tags previously assigned to resources.
	Institution manager	Wants a repository that allows the users to see their tags, ensuring a more direct access to the tagged resources in an accurate process.
Preconditions	The Consumer or Provider is identified and authenticated.	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1-The Consumer or Provider chooses the option to visualise own tags; 2-The system shows all the tags previously used by the Consumer or Provider; 3- The Consumer or Provider can choose one tag to then visualise the resources to which he assigned it (see UC11) or he can exit; 4-Steps 2 and 3 might be repeated until the Consumer or Provider chooses the exit option.	

Table 63 describes Use case UC11.

Table 63. Use case UC11

Abbreviated designation	UC11	
Extended designation	List tagged resources	
Primary actors	Consumer and Provider	
Stakeholders and interests	Consumer/Provider	Wants accurate, easy and fast visualisation of the resources to which he assigned a specific tag.
	Institution manager	Wants a repository that allows the users to see the resources to which they assigned a specific tag.
Preconditions	The Consumer or Provider is identified and authenticated, the option to visualise own tags was selected (UC10), and a specific tag was selected.	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1-The system shows the resources to which the Consumer or Provider assigned the tag; 2-The Consumer or Provider selects the resource to which he wants can see more details; 3-The system shows detailed information about the resource; 4-Steps 2 and 3 can be repeated 0 or more times, until the Consumer or Provider selects the exit option.	

Table 64 provides some information about Use case UC12.

Table 64. Use case UC12

Abbreviated designation	UC12	
Extended designation	Visualise tag cloud	
Primary actors	Consumer and Provider	
Stakeholders and interests	Consumer/Provider	Wants accurate, easy and fast visualisation of the tags he previously assigned to resources.
	Institution manager	Wants a repository that allows the users to see their tags, ensuring a more direct access to the tagged resources in an accurate process.
Preconditions	The Consumer or Provider is identified and authenticated.	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1-The Consumer or Provider chooses the option to visualise tag cloud or the process can be initiated from UC15 to a particular resource; 2-The system shows the tag cloud considering the most popular tags; 3-The Consumer or Provider selects a tag; 4- The system shows the resources to which the selected tag was applied (UC11); 5-Steps 3 and 4 might be chosen 0 or 1 time.	

Table 65 describes Use case UC13.

Table 65. Use case UC13

<table>
<tr><td>Abbreviated designation</td><td colspan="2">UC13</td></tr>
<tr><td>Extended designation</td><td colspan="2">Search</td></tr>
<tr><td>Primary actors</td><td colspan="2">Consumer and Provider</td></tr>
<tr><td rowspan="2">Stakeholders and interests</td><td>Consumer/Provider</td><td>Wants accurate, easy and fast search for resources.</td></tr>
<tr><td>Institution manager</td><td>Wants a repository that allows the users to submit queries and see the results when trying to find resources of interest.</td></tr>
<tr><td>Preconditions</td><td colspan="2">The Consumer or Provider is identified and authenticated.</td></tr>
<tr><td>Postconditions</td><td colspan="2">Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.</td></tr>
<tr><td>Main success scenario</td><td colspan="2">1-The Consumer or Provider chooses the option to search for resources;
2-The system shows the searchable fields to the Consumer or Provider, and the option to consider all inclusively or exclusively. The option to expand the keywords or not is also visible;
3-The Consumer or Provider fills the fields of interest;
4-If the option to expand keywords is selected then the system shows the related terms using the ontologies repository (UC21);
5-The user selects the keywords to be considered in addition to those initially provided;
6-The system proceeds in order to find the resources that match all the options (UC14).</td></tr>
</table>

Table 66 provides some information about Use case UC14.

Table 66. Use case UC14

<table>
<tr><td>Abbreviated designation</td><td colspan="2">UC14</td></tr>
<tr><td>Extended designation</td><td colspan="2">List resources</td></tr>
<tr><td>Primary actors</td><td colspan="2">Consumer and Provider</td></tr>
<tr><td rowspan="2">Stakeholders and interests</td><td>Consumer/Provider</td><td>Wants accurate, easy and fast access to a list of resources that might correspond to his information needs.</td></tr>
<tr><td>Institution manager</td><td>Wants a repository that allows the users to fulfil their information needs.</td></tr>
<tr><td>Preconditions</td><td colspan="2">The Consumer or Provider is identified and authenticated and the user previously filled the search fields and chose the options of interest (UC13).</td></tr>
<tr><td>Postconditions</td><td colspan="2">Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.</td></tr>
<tr><td>Main success scenario</td><td colspan="2">1- The system shows the list of resources that match the options previously specified (UC13);
2-The Consumer or Provider can select a resource to see detailed information about it (UC15) and further explore available functionalities or can choose to exit.</td></tr>
</table>

Table 67 describes Use case UC15

Table 67. Use case UC15

Abbreviated designation	UC15	
Extended designation	Explore resource (UC15)	
Primary actors	Consumer and Provider	
Stakeholders and interests	Consumer/Provider	Wants accurate, easy and fast access to detailed data about a resource that might correspond to his information needs, exploring it.
	Institution manager	Wants a repository that allows the users to explore resources that might correspond to their information needs.
Preconditions	The Consumer or Provider is identified and authenticated and the user previously selected a resource to explore (UC14).	
Postconditions	-	
Main success scenario	1- The system shows the metadata associated to the resource and also its most popular tags (UC12). A number of options are available to the Consumer or Provider (UC16, UC17, UC18, and UC20), and he can also choose to finalise the resource exploration; 2-The Consumer or Provider selects the desired option; 3-The system proceeds in accordance with the Consumer or Provider selection; 4-Steps 1 to 3 can be repeated many times.	

Table 68 provides some information about Use case UC16.

Table 68. Use case UC16

Abbreviated designation	UC16	
Extended designation	Tag resource	
Primary actors	Consumer and Provider	
Stakeholders and interests	Consumer/Provider	Wants accurate, easy and fast way to provide his metadata to resources, which can also be used as bookmarks.
	Institution manager	Wants a repository that allows wide characterisation of resources, with collaboration of different kinds of users.
Preconditions	The Consumer or Provider is identified and authenticated and the user previously selected the option to tag a specific resource (UC15).	
Postconditions	The system saves all the data and metadata associated to the tag(s) assignment.	
Main success scenario	1-The system shows the options available to the tag assignment or to exit; 2-The Consumer or Provider assigns a tag and chooses if it is public or private or he can choose to exit; 3-If a tag was provided, the system validates the data, ensuring that the tag is not in the list of banned ones. The system normalises the tag removing any spaces; 4-Steps 2 and 3 can be repeated many times until the Consumer or Provider chooses the exit option.	

Table 69 describes Use case UC17.

Table 69. Use case UC17

<table>
<tr><td>Abbreviated designation</td><td colspan="2">UC17</td></tr>
<tr><td>Extended designation</td><td colspan="2">Download resource</td></tr>
<tr><td>Primary actors</td><td colspan="2">Consumer and Provider</td></tr>
<tr><td rowspan="2">Stakeholders and interests</td><td>Consumer/Provider</td><td>Wants accurate, easy and fast access to resources and to download them.</td></tr>
<tr><td>Institution manager</td><td>Wants a repository that allows the users to download and use resources that might correspond to their information needs.</td></tr>
<tr><td>Preconditions</td><td colspan="2">The Consumer or Provider is identified and authenticated and the user previously selected the option to download a specific resource (UC15).</td></tr>
<tr><td>Postconditions</td><td colspan="2">The system saves information related to the resource download that might be used to generate statistics about resources usage.</td></tr>
<tr><td>Main success scenario</td><td colspan="2">1- The system shows the options available to the download of resources;
2-The Consumer or Provider provides the local directory to accommodate the resource;
3-The system proceeds to the download.</td></tr>
</table>

Table 70 provides some information about Use case UC18.

Table 70. Use case UC18

<table>
<tr><td>Abbreviated designation</td><td colspan="2">UC18</td></tr>
<tr><td>Extended designation</td><td colspan="2">List units of learning</td></tr>
<tr><td>Primary actors</td><td colspan="2">Consumer and Provider</td></tr>
<tr><td rowspan="2">Stakeholders and interests</td><td>Consumer/Provider</td><td>Wants accurate, easy and fast access to units of learning that can be explored for teaching or learning purposes.</td></tr>
<tr><td>Institution manager</td><td>Wants a repository that allows the users to explore resources facilitating its use in external platforms for learning or teaching purposes.</td></tr>
<tr><td>Preconditions</td><td colspan="2">The Consumer or Provider is identified and authenticated and the user previously selected the option to list the units of learning associated to a specific resource (UC15).</td></tr>
<tr><td>Postconditions</td><td colspan="2">-</td></tr>
<tr><td>Main success scenario</td><td colspan="2">1- The system lists all units of learning that encompass different ways to use the resource for learning or teaching purposes;
2-The Consumer or Provider might select one desired unit of learning and it can be then downloaded (UC19) or return to the main page;
3-Steps 1 and 2 can be repeated many times until the Consumer or Provider chooses to exit.</td></tr>
</table>

Table 71 provides some information about Use case UC19.

Table 71. Use case UC19

Abbreviated designation	UC19	
Extended designation	Download unit of learning	
Primary actors	Consumer and Provider	
Stakeholders and interests	Consumer/Provider	Wants accurate, easy and fast access to units of learning that can be explored for teaching or learning purposes.
	Institution manager	Wants a repository that allows the users to explore resources facilitating its use in external platforms for learning or teaching purposes.
Preconditions	The Consumer or Provider is identified and authenticated and the user previously selected the option to download a specific IMS LD file associated to a particular resource (UC18).	
Postconditions	The system saves information related to the unit of learning download that might be used to generate statistics reports.	
Main success scenario	1- The system shows the options available to the download of an unit of learning; 2-The Consumer or Provider provides the local directory to accommodate the file; 3-The system proceeds to the download.	

Table 72 describes Use case UC20.

Table 72. Use case UC20

Abbreviated designation	UC20	
Extended designation	Download package	
Primary actors	Consumer and Provider	
Stakeholders and interests	Consumer/Provider	Wants accurate, easy and fast access to IMS CP files that can be transposed to other platforms.
	Institution manager	Wants a repository that allows the users to explore resources in other platforms or facilitate the migration to other repository platforms.
Preconditions	The Consumer or Provider is identified and authenticated and the user previously selected the option to download the IMS CP file associated to a specific resource (UC15).	
Postconditions	The system saves information related to the package download that might be used to generate statistics reports.	
Main success scenario	1- The system shows the options available to the package download; 2-The Consumer or Provider provides the local directory to accommodate the package; 3-The system proceeds to the download.	

Table 73 describes Use case UC21.

Table 73. Use case UC21

Abbreviated designation	UC21	
Extended designation	Expand keyword	
Primary actors	Consumer and Provider	
Stakeholders and interests	Consumer/Provider	Wants accurate, easy and fast search for resources with assistance in the specification of keyword of interest.
	Institution manager	Wants a repository that allows the users to find resources of interest.
Preconditions	The Consumer or Provider is identified and authenticated and he previously specified keyword(s) and chose to expand keywords (UC13).	
Postconditions	The system saves data related to the usage of that functionality that can be used to generate statistics or to justify future changes.	
Main success scenario	1-The system shows the available options to expand keywords; 2-The Consumer or Provider specifies the desired options; 3-The system uses the ontology repository to find other related terms in accordance with the specified options, which are provided to the Consumer or Provider; 4-The consumer or provider chooses the desired keywords to be considered in UC13.	

Table 74 provides some information about Use case UC22.

Table 74. Use case UC22

Abbreviated designation	UC22	
Extended designation	List own template suggestions	
Primary actors	Consumer and Provider	
Stakeholders and interests	Consumer/Provider	Wants accurate, easy and fast access to all provided suggestions, inclusive those regarding pedagogical activities.
	Institution manager	Wants a repository that allows the users to actively collaborate in many activities in the repository and to manage their suggestions.
Preconditions	The Consumer or Provider is identified and authenticated.	
Postconditions	-	
Main success scenario	1- The system lists all templates that encompass a pedagogical idea of how to use resources that were suggested to be used by the Consumer or Provider. For each of them a summary is shown; 2-The Consumer or Provider might select one to see more detailed information (UC23) or can choose to exit from this functionality; 3-Steps 1 and 2 can be repeated many times until the Consumer or Provider chooses to exit.	

Table 75 describes Use case UC23.

Table 75. Use case UC23

<table>
<tr><td>Abbreviated designation</td><td colspan="2">UC23</td></tr>
<tr><td>Extended designation</td><td colspan="2">Visualise suggestion</td></tr>
<tr><td>Primary actors</td><td colspan="2">Consumer and Provider</td></tr>
<tr><td rowspan="2">Stakeholders and interests</td><td>Consumer/Provider</td><td>Wants accurate, easy and fast access to all provided suggestions, inclusive those regarding pedagogical activities.</td></tr>
<tr><td>Institution manager</td><td>Wants a repository that allows the users to actively collaborate in many activities in the repository and to manage their suggestions.</td></tr>
<tr><td>Preconditions</td><td colspan="2">The Consumer or Provider is identified and authenticated and the user previously selected the option to see his template suggestions (UC22) and selected one to see more details.</td></tr>
<tr><td>Postconditions</td><td colspan="2">The system saves information related to the package download that might be used to generate statistics reports.</td></tr>
<tr><td>Main success scenario</td><td colspan="2">1- The system lists all the activities and their associated data related to the specific template suggestion being analysed;
2-The Consumer or Provider chooses to exit from that functionality.</td></tr>
</table>

Table 76 provides some information about Use case UC24.

Table 76. Use case UC24

<table>
<tr><td>Abbreviated designation</td><td colspan="2">UC24</td></tr>
<tr><td>Extended designation</td><td colspan="2">Suggest a new template</td></tr>
<tr><td>Primary actors</td><td colspan="2">Consumer and Provider</td></tr>
<tr><td rowspan="2">Stakeholders and interests</td><td>Consumer/Provider</td><td>Wants accurate, easy and fast way to provide suggestions that might improve the overall functioning of the repository.</td></tr>
<tr><td>Institution manager</td><td>Wants a repository that allows the users to actively collaborate in many activities in the repository.</td></tr>
<tr><td>Preconditions</td><td colspan="2">The Consumer or Provider is identified and authenticated.</td></tr>
<tr><td>Postconditions</td><td colspan="2">A new template suggestion is saved, as well as its metadata.</td></tr>
<tr><td>Main success scenario</td><td colspan="2">1-The system initiates the process of suggesting a new template, by asking for its designation, that should encompass its main idea(s), and the number of activities;
2-The Consumer or Provider provides the necessary data;
3-The system validates all the data and for each activity asks for the pertinent data;
4-The Consumer or Provider provides the additional data;
5-The system validates and saves the data.</td></tr>
</table>

Table 77 describes Use case UC25.

Table 77. Use case UC25

<table>
<tr><td>Abbreviated designation</td><td colspan="2">UC25</td></tr>
<tr><td>Extended designation</td><td colspan="2">List in-use templates</td></tr>
<tr><td>Primary actors</td><td colspan="2">Consumer and Provider</td></tr>
<tr><td rowspan="2">Stakeholders and interests</td><td>Consumer/Provider</td><td>Wants accurate, easy and fast access to resources and related information available in the repository.</td></tr>
<tr><td>Institution manager</td><td>Wants a repository that allows the users to actively collaborate in many activities in the repository and to have access to resources and related information.</td></tr>
<tr><td>Preconditions</td><td colspan="2">The Consumer or Provider is identified and authenticated.</td></tr>
<tr><td>Postconditions</td><td colspan="2">-</td></tr>
<tr><td>Main success scenario</td><td colspan="2">1- The system lists all templates that are under usage in the repository;
2-The Consumer or Provider might select one to see more detailed information (UC26) or can choose to exit from this functionality;
3-Steps 1 and 2 can be repeated many times until the Consumer or Provider chooses to exit.</td></tr>
</table>

Table 78 provides some information about Use case UC26.

Table 78. Use case UC26

<table>
<tr><td>Abbreviated designation</td><td colspan="2">UC26</td></tr>
<tr><td>Extended designation</td><td colspan="2">Visualise template</td></tr>
<tr><td>Primary actors</td><td colspan="2">Consumer and Provider</td></tr>
<tr><td rowspan="2">Stakeholders and interests</td><td>Consumer/Provider</td><td>Wants accurate, easy and fast access to resources and related information available in the repository.</td></tr>
<tr><td>Institution manager</td><td>Wants a repository that allows the users to actively collaborate in many activities in the repository and to have access to resources and related information.</td></tr>
<tr><td>Preconditions</td><td colspan="2">The Consumer or Provider is identified and authenticated and the user previously selected the option to see the in-use templates (UC25) and selected one to see more details.</td></tr>
<tr><td>Postconditions</td><td colspan="2">The system saves information related to the package download that might be used to generate statistics reports.</td></tr>
<tr><td>Main success scenario</td><td colspan="2">1- The system lists all the activities and their associated data related to the specific template suggestion being analysed;
2-The Consumer or Provider chooses to exit from that functionality.</td></tr>
</table>

Table 79 describes Use case UC27.

Table 79. Use case UC27

Abbreviated designation	UC27	
Extended designation	Search for online ontologies	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants an accurate and easy option to find online ontologies to be considered for usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated.	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1- The system provides a form with many fields to be filled in; 2-The Ontology engineer selects the main domain area and provides a subarea. Then he has to specify the domain concepts (UC28) to proceed to the search for ontologies that cover these concepts; 3-The system proceeds to the search using a SWSE for that purpose, considering the configuration values; 4-The Ontology engineer can then analyse the results (UC29) under many aspects.	

Table 80 provides some information about Use case UC28.

Table 80. Use case UC28

Abbreviated designation	UC28	
Extended designation	Specify domain concepts	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants an accurate and easy option to find online ontologies to be considered for usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and a domain area and a subarea were previously specified (UC27).	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1- The system provides a form to be filled in with the concepts of interest; 2-The Ontology engineer specifies the concepts that characterises the domain subarea; 3-The system validates the data and for each concept some lexical variations are also considered in the composition of queries to be submitted to a SWSE also regarding the current configuration values. The queries are constructed and submitted with the goal to find firstly the ontologies that cover most of the concepts.	

Table 81 describes Use case UC29.

Table 81. Use case UC29

Abbreviated designation	UC29	
Extended designation	Analyse returned ontologies	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants an accurate and easy option to find online ontologies to be considered for usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and the search of ontologies was initiated (UC27).	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1- The system provides for each combination of the concepts a list of discovered ontologies. Some ontology metadata fields are displayed. However, the system disregards ontologies that cannot be processed by Pellet or that are no longer available. The maximum number of ontologies that are analysed is commanded by a configuration value. Also, for each concept the system shows some (probable) synonyms as provided by a dictionary tool. A number of options are available to facilitate a specific ontology analysis (UC30, UC31, UC32, UC33, UC38, UC39, and UC40); 2-The Ontology engineer selects one of the options to analyse ontologies or can choose to abandon this functionality.	

Table 82 provides some information about Use case UC30.

Table 82. Use case UC30

Abbreviated designation	UC30	
Extended designation	Visualise classes	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants an accurate and easy option to find online ontologies to be considered for usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and a specific ontology is under analysis (UC29).	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1-The system shows all the classes for a previously specified ontology; 2-The Ontology engineer can see the classes and return to UC29.	

Table 83 describes Use case UC31.

Table 83. Use case UC31

Abbreviated designation	UC31	
Extended designation	Visualise properties	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants an accurate and easy option to find online ontologies to be considered for usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and a specific ontology is under analysis (UC29).	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1-The system shows all the properties for a previously specified ontology; 2-The Ontology engineer can see the properties and return to UC29.	

Table 84 provides some information about Use case UC32.

Table 84. Use case UC32

Abbreviated designation	UC32	
Extended designation	Visualise instances	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants an accurate and easy option to find online ontologies to be considered for usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and a specific ontology is under analysis (UC29).	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1-The system shows all the instances for a previously specified ontology; 2-The Ontology engineer can see the instances and return to UC29.	

Table 85 describes Use case UC33.

Table 85. Use case UC33

Abbreviated designation	UC33	
Extended designation	Consider related terms	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants an accurate and easy option to find online ontologies to be considered for usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and a specific ontology is under analysis (UC29).	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1- The system provides two options to ontology analysis through UC34 or UC35; 2-The Ontology engineer selects the desired option; 3-The system proceeds in accordance with the Ontology engineer selection; 4-Steps 1-3 can be repeated 0 to 2 times.	

Table 86 provides some information about Use case UC34.

Table 86. Use case UC34

Abbreviated designation	UC34	
Extended designation	Consider similarity	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants an accurate and easy option to find online ontologies to be considered for usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and a specific ontology is under analysis (UC29).	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1- The system considers the threshold value for the syntactic similarity analysis and analyses the specified ontology to see if there are values above the configuration value for some concept regarding the ontology terms (only classes or classes, properties and instances, depending on the respective configuration value). The highest value for each concept is showed. A file is available with the detailed information about that process, which can be downloaded by the Ontology engineer; 2-The Ontology engineer exits from this functionality.	

Table 87 describes Use case UC35.

Table 87. Use case UC35

Abbreviated designation	UC35	
Extended designation	Consider synonyms	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants an accurate and easy option to find online ontologies to be considered for usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and a specific ontology is under analysis (UC29).	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1- The system considers the list of synonyms that were previously identified for each concept in order to determine which one(s) to use to verify if there are ontology terms syntactically similar to the relevant synonyms. There are two options (UC36 and UC37), but only the results provided by UC36 are authoritative and are used to automatically find syntactic similarities between the synonym and the ontologies terms. Depending on the synonyms reduction (configuration value), the similarity value is considered but decreased; 2-The Ontology engineer chooses the option to be considered, or can choose to exit; 3-The system proceeds in accordance with the Ontology engineer selection; 4-Steps 1-3 can be repeated 0 to 2 times until the Ontology engineer chooses to exit.	

Table 88 provides some information about Use case UC36.

Table 88. Use case UC36

Abbreviated designation	UC36	
Extended designation	Use online ontologies	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants an accurate and easy option to find online ontologies to be considered for usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and a specific ontology is under analysis (UC29) and the synonym consideration is being considered (UC35).	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1- The system tries to find ontologies with the synonym under consideration and the other concepts, regarding the some relation types and a certain inheritance level. The number of found relations are highlighted for each concept; 2-The Ontology engineer can see the obtained values and he can choose to proceed with the syntactic similarity values considering the synonyms with the most obtained relations and the ontology terms, or can choose to exit; 3-The system proceeds in accordance with the Ontology engineer choice.	

Table 89 describes Use case UC37.

Table 89. Use case UC37

Abbreviated designation	UC37	
Extended designation	Consider LOD databases	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants an accurate and easy option to find online ontologies to be considered for usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and a specific ontology is under analysis (UC29) and the synonym consideration is being considered (UC35).	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1- The system tries to find relation between the synonym under consideration and the other concepts, using some properties. The obtained results are highlighted; 2-The Ontology engineer can see the obtained values and he can see more detailed information about the process. Then he exits from that functionality.	

Table 90 provides some information about Use case UC38.

Table 90. Use case UC38

Abbreviated designation	UC38	
Extended designation	Consider other topics	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants an accurate and easy option to find online ontologies to be considered for usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated, some concepts were provided in UC28 and the returned ontologies are visible (UC29).	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1- The system has some of the results obtained from UC29 available. It could have also obtained results from UC35. The option to search for ontologies is again available; 2-The Ontology engineer can provide other concepts substituting the originally provided; 3-The system searches for online ontologies (see UC27), but for the concepts without substitutes specified by the Ontology engineer, the original concepts are considered. Then all processing from UC29 is available.	

Table 91 describes Use case UC39.

Table 91. Use case UC39

Abbreviated designation	UC39	
Extended designation	Estimate domain coverage	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants an accurate and easy option to find online ontologies to be considered for usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and a specific ontology is under analysis (UC29).	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1- The system provides a domain coverage value for the ontology under analysis (UC29), and possibly considering the values from other use cases if they were considered, and the related configuration values; 2-The Ontology engineer can choose to exit from that functionality.	

Table 92 provides some information about Use case UC40.

Table 92. Use case UC40

Abbreviated designation	UC40	
Extended designation	Submit ontology	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants an accurate and easy option to find online ontologies to be considered for usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated.	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements. A new ontology is made available in the temporary ontology repository.	
Main success scenario	1- The system provides an option to submit the ontology to the temporary ontology repository as it is, but only if the ontology is not already in the repository; 2-The Ontology engineer can choose to submit the ontology or can choose to exit from this functionality.	

Table 93 describes Use case UC41.

Table 93. Use case UC41

Abbreviated designation	UC41	
Extended designation	Visualise working properties	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants an accurate and easy option to find online ontologies to be considered for usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated.	
Postconditions	-	
Main success scenario	1- The system shows the current configuration values that affect how the discovery of online ontologies is carried out and also their analysis; 2-The Ontology engineer choose to exit from this functionality.	

Table 94 provides some information about Use case UC42.

Table 94. Use case UC42

Abbreviated designation	UC42	
Extended designation	Modify working properties	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants an accurate and easy option to find online ontologies to be considered for usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated.	
Postconditions	New configuration values are saved.	
Main success scenario	1- The system shows the current configuration values that affect how the discovery of online ontologies is carried out and also their analysis. They are modifiable; 2-The Ontology engineer provides the new values; 3-The system validates the new values and save them.	

Table 95 describes Use case UC43.

Table 95. Use case UC43

Abbreviated designation	UC43	
Extended designation	Manage in-use ontologies	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants easy, accurate and fast management of the ontologies under usage in the repository.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated.	
Postconditions	All the data pertinent to the performed operations are saved.	
Main success scenario	1-The Ontology engineer chooses the option to manage in-use ontologies; 2-The system shows a number of options available to the Ontology engineer, which includes the options to manage in-use ontologies (see UC44 and UC45), but also to exit from the in-use ontologies management section; 3-The Ontology engineer select the desired option; 4-The system proceeds in accordance with his selection; 5-Steps 2 to 4 might be repeated many times until the Ontology engineer chooses the exit option.	

Table 96 provides some information about Use case UC44.

Table 96. Use case UC44

Abbreviated designation	UC44	
Extended designation	List in-use ontologies	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants easy, accurate and fast access to the list of in-use ontologies.
	Institution manager	Wants a repository that allows listing the in-use ontologies, which could serve different purposes, such as finding insufficiently covered areas.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and selected the option to list in-use ontologies in UC43.	
Postconditions	-	
Main success scenario	1-The system shows the sort options available to the Ontology engineer; 2-The Ontology engineer selects the desired option; 3-The system proceeds in accordance with the selected option and lists the in-use ontologies.	

Table 97 describes Use case UC45.

Table 97. Use case UC45

Abbreviated designation	UC45	
Extended designation	Delete in-use ontology	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants accurate, easy and fast option to delete in-use ontologies.
	Institution manager	Wants a repository that allows the deletion of in-use ontologies, because they are not anymore directly or indirectly related to any course, or any other reason.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and the option to delete in-use ontology was selected (UC43).	
Postconditions	An ontology is deleted.	
Main success scenario	1-The system shows the in-use ontologies; 2-The Ontology engineer selects the ontology to be deleted; 3-The system asks for delete confirmation; 4-The Ontology engineer confirms that the ontology is to be deleted; 5-The system deletes the ontology and acknowledges the successful completion of the operation.	

Table 98 provides some information about Use case UC46.

Table 98. Use case UC46

Abbreviated designation	UC46	
Extended designation	Download in-use ontology	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants accurate, easy and fast options to download in-use ontologies.
	Institution manager	Wants a repository that allows the download of in-use ontologies to be further analysed in any external.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and the option to download in-use ontology was selected (UC43).	
Postconditions	An ontology is download and the related data is saved in order to support statistic analysis.	
Main success scenario	1-The system shows the in-use ontologies; 2-The Ontology engineer selects the ontology to be download; 3-The system asks for the local directory to accommodate the ontology; 4-The Ontology provides the local directory; 5-The system proceeds to download of the ontology and acknowledges the successful completion of the operation.	

Table 99 describes Use case UC47.

Table 99. Use case UC47

Abbreviated designation	UC47	
Extended designation	Manage temporary ontologies	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants easy, accurate and fast management of the temporary ontologies.
	Institution manager	Wants a repository that allows knowledge representation for domain areas of interest with accuracy.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated.	
Postconditions	All the data pertinent to the performed operations are saved.	
Main success scenario	1-The Ontology engineer chooses the option to manage; 2-The system shows a number of options available to the Ontology engineer, which includes sub-options to manage temporary ontologies (from UC48 to UC51), but also to exit from the temporary ontologies management section; 3-The Ontology engineer select the desired option; 4-The system proceeds in accordance with his selection; 5-Steps 2 to 4 might be repeated many times until the Ontology engineer chooses the exit option.	

Table 100 provides some information about Use case UC48.

Table 100. Use case UC48

Abbreviated designation	UC48	
Extended designation	List temporary ontologies	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants easy, accurate and fast access to the list of temporary ontologies.
	Institution manager	Wants a repository that allows listing the temporary ontologies, which could serve statistic purposes, among others.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated. The former should have selected the option to list temporary ontologies in UC47.	
Postconditions	-	
Main success scenario	1-The system shows the sort options available to the Ontology engineer; 2-The Ontology engineer selects the desired option; 3-The system proceeds in accordance with the selected option and lists the temporary ontologies.	

Table 101 describes Use case UC49.

Table 101. Use case UC49

Abbreviated designation	UC49	
Extended designation	Delete temporary ontology	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants accurate, easy and fast option to delete temporary ontologies.
	Institution manager	Wants a repository that allows the deletion of temporary ontologies, because they are not adequate, or any other reason.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and the option to delete temporary ontology was selected (UC47).	
Postconditions	An ontology is deleted.	
Main success scenario	1-The Ontology engineer selects the ontology to be deleted; 2-The system asks for delete confirmation; 3-The Ontology engineer confirms that the ontology is to be deleted; 4-The system deletes the ontology and acknowledges the successful completion of the operation.	

Table 102 provides some information about Use case UC50.

Table 102. Use case UC50

Abbreviated designation	UC50	
Extended designation	Put ontology in use	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants accurate, easy and fast option to put an ontology under usage.
	Institution manager	Wants a repository that allows the deletion of temporary ontologies, because they are not anymore directly or indirectly related to any course, or any other reason.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and the option to put ontology in use was selected (UC47).	
Postconditions	An ontology starts to be considered for the related functions in the repository.	
Main success scenario	1-The Ontology engineer selects the desired ontology; 2-The system asks for confirmation; 3-The Ontology engineer confirms that the ontology is to be added to the in-use ontologies repository; 4-The system adds the ontology to the in-use ontologies repository and acknowledges the successful completion of the operation.	

Table 103 describes Use case UC51.

Table 103. Use case UC51

Abbreviated designation	UC51	
Extended designation	Download temporary ontology	
Primary actors	Ontology engineer	
Stakeholders and interests	Ontology engineer	Wants accurate, easy and fast options to download temporary ontologies.
	Institution manager	Wants a repository that allows the download of ontologies to be further analysed in any external.
	Consumer	Wants a repository that provides some assistance in the specification of the information needs when searching for resources in the repository.
	Provider	Wants a repository that provides some assistance in the characterisation of the resources and in the specification of the information needs when searching for resources in the repository.
Preconditions	The Ontology engineer is identified and authenticated and the option to download temporary ontology was selected (UC47).	
Postconditions	An ontology is download and the related data is saved in order to support statistic analysis.	
Main success scenario	1-The system shows the temporary ontologies; 2-The Ontology engineer selects the ontology to be download; 3-The system asks for the local directory to accommodate the ontology; 4-The Ontology provides the local directory; 5-The system proceeds to download of the ontology and acknowledges the successful completion of the operation.	

Table 104 provides some information about Use case UC52.

Table 104. Use case UC52

Abbreviated designation	UC52	
Extended designation	Visualise opinion summaries	
Primary actors	Pedagogical adviser	
Stakeholders and interests	Pedagogical adviser	Wants accurate, easy and fast option to visualise opinion summaries.
	Institution manager	Wants a repository that allows the management of the opinions provided by their users.
	Consumer/Provider	Wants a repository that regards the opinions about the resources.
Preconditions	The Pedagogical adviser is identified and authenticated.	
Postconditions	-	
Main success scenario	1-The Pedagogical adviser selects the option to visualise opinion summaries; 2-The system shows the list of resources and the total number of positive, negative and neutral tags assigned to it; 3-The Pedagogical adviser chooses to exit.	

Table 105 describes Use case UC53.

Table 105. Use case UC53

Abbreviated designation	UC53	
Extended designation	Visualise resources' tags	
Primary actors	Pedagogical adviser	
Stakeholders and interests	Pedagogical adviser	Wants accurate, easy and fast option to visualise resources' tags.
	Institution manager	Wants a repository that allows the management of the opinions provided by their users.
	Consumer/Provider	Wants a repository that regards the opinions about the resources.
Preconditions	The Pedagogical adviser is identified and authenticated.	
Postconditions	-	
Main success scenario	1-The Pedagogical adviser selects the option to visualise resources' tags; 2-The system shows the list of resources and for each resource all the tags are visible as well as how they were regarded; 3-The Pedagogical adviser chooses to exit.	

Table 106 provides some information about Use case UC54.

Table 106. Use case UC54

Abbreviated designation	UC54	
Extended designation	Manage in-use templates	
Primary actors	Pedagogical adviser	
Stakeholders and interests	Pedagogical adviser	Wants accurate, easy and fast management of in-use templates.
	Institution manager	Wants a repository that allows the management of the activities that can be carried out with resources.
	Consumer/Provider	Wants a repository that regards different opinions about how to use resources.
Preconditions	The Pedagogical adviser is identified and authenticated.	
Postconditions	All the data pertinent to the performed operations are saved.	
Main success scenario	1-The Pedagogical adviser selects the option to manage in-use templates; 2-The system shows the sub-options available to manage in-use templates (from UC55 to UC58) and the option to exit; 3-The Pedagogical adviser selects the desired option; 4-The system proceeds in accordance with the selection of the Pedagogical adviser; 5-Steps 2 to 4 can be repeated many times until the Pedagogical adviser chooses to exit.	

Table 107 describes Use case UC55.

Table 107. Use case UC55

Abbreviated designation	UC55	
Extended designation	List in-use templates	
Primary actors	Pedagogical adviser	
Stakeholders and interests	Pedagogical adviser	Wants accurate, easy and fast management of in-use templates.
	Institution manager	Wants a repository that allows the management of the activities that can be carried out with resources.
	Consumer/Provider	Wants a repository that regards different opinions about how to use resources.
Preconditions	The Pedagogical adviser is identified and authenticated and the option to list in-use templates was selected (UC54).	
Postconditions	-	
Main success scenario	1-The system lists all the templates currently under usage with a short summary, which includes who submitted it and when, and the list of activities and related information. For each template the option to list related UoLs (UC56) is available. There is also an option to exit; 2-The Pedagogical adviser selects the desired option; 3-The system proceeds in accordance with the selection of the Pedagogical adviser; 4-Steps 1 to 4 can be repeated many times until the Pedagogical adviser chooses to exit.	

Table 108 provides some information about Use case UC56.

Table 108. Use case UC56

Abbreviated designation	UC56	
Extended designation	List related UoLs	
Primary actors	Pedagogical adviser	
Stakeholders and interests	Pedagogical adviser	Wants accurate, easy and fast management of in-use templates.
	Institution manager	Wants a repository that allows the management of the activities that can be carried out with resources.
	Consumer/Provider	Wants a repository that regards different opinions about how to use resources.
Preconditions	The Pedagogical adviser is identified and authenticated and the option to list related UoLs was previously selected for a specific template (UC55).	
Postconditions	-	
Main success scenario	1-The system lists all the UoLs generated from a specific template, with a short summary, which includes the number of downloads and a list of resources for which the generation of UoL were used; 2-The Pedagogical adviser chooses to exit.	

Table 109 describes Use case UC57.

Table 109. Use case UC57

Abbreviated designation	UC57	
Extended designation	Delete in-use template	
Primary actors	Pedagogical adviser	
Stakeholders and interests	Pedagogical adviser	Wants accurate, easy and fast management of in-use templates.
	Institution manager	Wants a repository that allows the management of the activities that can be carried out with resources.
	Consumer/Provider	Wants a repository that regards different opinions about how to use resources.
Preconditions	The Pedagogical adviser is identified and authenticated and the option to delete in-use template was selected (UC54).	
Postconditions	A template is removed.	
Main success scenario	1-The system lists all the templates currently under usage with a short summary; 2-The Pedagogical adviser selects the template to be deleted; 3-The system asks for confirmation; 4-The Pedagogical adviser confirms the template to be deleted; 5-The system proceeds with the delete operation.	

Table 110 provides some information about Use case UC58.

Table 110. Use case UC58

Abbreviated designation	UC58	
Extended designation	Add template	
Primary actors	Pedagogical adviser	
Stakeholders and interests	Pedagogical adviser	Wants accurate, easy and fast management of in-use templates.
	Institution manager	Wants a repository that allows the management of the activities that can be carried out with resources.
	Consumer/Provider	Wants a repository that regards different opinions about how to use resources.
Preconditions	The Pedagogical adviser is identified and authenticated and the option to add template was selected (UC54).	
Postconditions	A new template is saved, as well as its metadata.	
Main success scenario	1-The system asks for the template designation, that should encompass its main idea(s), and the number of activities; 2-The Pedagogical adviser provides the necessary data; 3-The system validates all the data and for each activity asks for the pertinent data; 4-The Pedagogical adviser provides the additional data; 5-The system validates and saves the data.	

Table 111 describes Use case UC59.

Table 111. Use case UC59

Abbreviated designation	UC59	
Extended designation	Manage temporary templates	
Primary actors	Pedagogical adviser	
Stakeholders and interests	Pedagogical adviser	Wants accurate, easy and fast management of temporary templates.
	Institution manager	Wants a repository that allows the management of the activities that can be carried out with resources.
	Consumer/Provider	Wants a repository that regards different opinions about how to use resources.
Preconditions	The Pedagogical adviser is identified and authenticated.	
Postconditions	All the data pertinent to the performed operations are saved.	
Main success scenario	1-The Pedagogical adviser selects the option to manage temporary templates; 2-The system shows the sub-options available to manage temporary templates (from UC60 to UC62) and the option to exit; 3-The Pedagogical adviser selects the desired option; 4-The system proceeds in accordance with the selection of the Pedagogical adviser; 5-Steps 2 to 4 can be repeated many times until the Pedagogical adviser chooses to exit.	

Table 112 provides some information about Use case UC60.

Table 112. Use case UC60

Abbreviated designation	UC60	
Extended designation	List unapproved templates	
Primary actors	Pedagogical adviser	
Stakeholders and interests	Pedagogical adviser	Wants accurate, easy and fast management of unapproved templates.
	Institution manager	Wants a repository that allows the management of the activities that can be carried out with resources.
	Consumer/Provider	Wants a repository that regards different opinions about how to use resources.
Preconditions	The Pedagogical adviser is identified and authenticated and the option to list unapproved templates was selected (UC59).	
Postconditions	-	
Main success scenario	1-The system lists all the unapproved templates with a short summary and the list of activities; 2-The Pedagogical adviser chooses to exit.	

Table 113 describes Use case UC61.

Table 113. Use case UC61

Abbreviated designation	UC61	
Extended designation	Evaluate template	
Primary actors	Pedagogical adviser	
Stakeholders and interests	Pedagogical adviser	Wants accurate, easy and fast management of in-use templates.
	Institution manager	Wants a repository that allows the management of the activities that can be carried out with resources.
	Consumer/Provider	Wants a repository that regards different opinions about how to use resources.
Preconditions	The Pedagogical adviser is identified and authenticated and the option to evaluate template was selected (UC59).	
Postconditions	-	
Main success scenario	1-The system lists all the unapproved templates with a short summary and the list of activities. 2-The Pedagogical adviser selects the desired template; 3-The system shows all the details about the template. The systems asks for the Pedagogical adviser evaluation; 4-The Pedagogical adviser selects the desired option (approve or reject); 5-The system proceeds in accordance with the selection. If the template is approved, it is put under usage in the repository; 6-Steps 1 to 4 can be repeated many times until the Pedagogical adviser chooses to exit.	

Table 114 provides some information about Use case UC62.

Table 114. Use case UC62

Abbreviated designation	UC62	
Extended designation	List rejected templates	
Primary actors	Pedagogical adviser	
Stakeholders and interests	Pedagogical adviser	Wants accurate, easy and fast management of rejected templates.
	Institution manager	Wants a repository that allows the management of the activities that can be carried out with resources.
	Consumer/Provider	Wants a repository that regards different opinions about how to use resources.
Preconditions	The Pedagogical adviser is identified and authenticated and the option to list rejected templates was selected (UC59).	
Postconditions	-	
Main success scenario	1-The system lists all the rejected templates with a short summary and the list of activities; 2-The Pedagogical adviser chooses to exit.	

Table 115 describes Use case UC63.

Table 115. Use case UC63

Abbreviated designation	UC63	
Extended designation	Submit a resource	
Primary actors	Provider	
Stakeholders and interests	Provider	Wants accurate, easy and fast management of resources.
	Institution manager	Wants a repository that allows the management of the resources with accuracy and efficiency.
	Consumer	Wants a repository well-populated with resources deeply characterised, which improves their findability.
Preconditions	The Provider is identified and authenticated.	
Postconditions	A new resource is uploaded.	
Main success scenario	1-The Provider selects the option to submit a resource; 2-The system starts by asking for the resource local in order to upload it, the associated community and collections; 3-The Provider supplies the data; 4-The system updloads the resource. Then the process of submitting its metadata initiated (UC64); 5-The system validates all the data. All the data and metadata are saved. The system acknowledges the success of the operation and shows the option to create a Unit of Learning for the uploaded resource (UC67); 6-The Provider chooses to provide a way to use the resource through UC67 and then exits from this functionality.	

Table 116 provides some information about Use case UC64.

Table 116. Use case UC64

Abbreviated designation	UC64	
Extended designation	Submit metadata	
Primary actors	Provider	
Stakeholders and interests	Provider	Wants accurate, easy and fast management of resources.
	Institution manager	Wants a repository that allows the management of the resources with accuracy and efficiency.
	Consumer	Wants a repository well-populated with resources deeply characterised, which improves their findability.
Preconditions	The Provider is identified and authenticated and a resource is being submitted (UC63).	
Postconditions	The metadata associated to the resource being uploaded are saved.	
Main success scenario	1-The system recognises the language of the resource and displays a form with the language field filled in, many others to be provided and also some that could be also automatically recognised (UC65), besides the language. For the keyword field there is an option to find related terms (UC66), using those automatically extracted or those directly provided; 2-The Provider chooses the desired options, and supply some fields or modify those automatically recognised; 3-The system validates all the data, ensuring that their format is correct but also that the mandatory fields were provided. All the data and metadata are saved.	

Table 117 describes Use case UC65.

Table 117. Use case UC65

Abbreviated designation	UC65	
Extended designation	Recognise metadata	
Primary actors	Provider	
Stakeholders and interests	Provider	Wants accurate, easy and fast management of resources.
	Institution manager	Wants a repository that allows the management of the resources with accuracy and efficiency.
	Consumer	Wants a repository well-populated with resources deeply characterised, which improves their findability.
Preconditions	The Provider is identified and authenticated and the option to recognise metadata was chosen (UC64).	
Postconditions	Some metadata fileds are temporarily saved.	
Main success scenario	1-The system recognises some metadata fields that are shown under UC64, but some others are not visible to the Provider, such as size and file type. These fields are temporarily saved.	

Table 118 provides some information about Use case UC66.

Table 118. Use case UC66

Abbreviated designation	UC66	
Extended designation	Expand keywords	
Primary actors	Provider	
Stakeholders and interests	Provider	Wants accurate, easy and fast management of resources.
	Institution manager	Wants a repository that allows the management of the resources with accuracy and efficiency.
	Consumer	Wants a repository well-populated with resources deeply characterised, which improves their findability.
Preconditions	The Provider is identified and authenticated and the option to expand keywords was chosen (UC64).	
Postconditions	Detailed information about the process, namely which terms were selected for which keyword, is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1-The system uses the values of the keyword fields to find some related terms using the ontologies associated with the community to which the resource being submitted is related to. The system tries to find related terms, if nothing is found, the ontologies associated to all communities are considered. The system shows the results, including the original keywords; 2-The Provider chooses the desired keywords to be considered under UC64.	

Table 119 describes Use case UC67.

Table 119. Use case UC67

Abbreviated designation	UC67	
Extended designation	Create a unit of learning	
Primary actors	Provider	
Stakeholders and interests	Provider	Wants accurate, easy and fast management of resources.
	Institution manager	Wants a repository that allows the management of the resources with accuracy and efficiency.
	Consumer	Wants a repository well-populated with resources deeply characterised, which improves their findability.
Preconditions	The Provider is identified and authenticated and the option to create a Unit of Learning was chosen (UC63).	
Postconditions	A new Unit of Learning is uploaded and the corresponding data and metadata are saved.	
Main success scenario	1-The system shows the available templates; 2-The Provider chooses the adequate template; 3-The system provides a form with the fields related to the template to be filled in; 4-The system validates all the data, and the UoL and the corresponding data and metadata are saved.	

Table 120 provides some information about Use case UC68.

Table 120. Use case UC68

Abbreviated designation	UC68	
Extended designation	List their resources	
Primary actors	Provider	
Stakeholders and interests	Provider	Wants accurate, easy and fast visualisation of how the resources were regarded by the other users.
	Institution manager	Wants a repository that allows the management of the resources with accuracy and efficiency, with a possibility of feedback provision to those users that submit resources.
	Consumer	Wants a repository well-populated with resources deeply characterised with his opinions considered.
Preconditions	The Provider is identified and authenticated.	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1-The Provider selects the option to list the resources that he submitted; 2-The system lists some data about the previous submissions and a summary with the number of positive, negative and neutral tags assigned to the each resource. There is an option that allows seeing all the public tags provided (UC69); 3-The Provider chooses to option to see the tagging details to a specific resource and can choose to exit; 4-The system proceeds in accordance with the provider choice; 5-Steps 2 to 4 can be repeated many times until the Provider chooses to exit.	

Table 121 describes Use case UC69.

Table 121. Use case UC69

Abbreviated designation	UC69	
Extended designation	List own resources' tags	
Primary actors	Provider	
Stakeholders and interests	Provider	Wants accurate, easy and fast visualisation of how the resources were regarded by the other users.
	Institution manager	Wants a repository that allows the management of the resources with accuracy and efficiency, with a possibility of feedback provision to the users that submit resources.
	Consumer	Wants a repository well-populated with resources deeply characterised with his opinions considered.
Preconditions	The Provider is identified and authenticated and the option to list own resources tags was previously selected (UC68).	
Postconditions	Detailed information about all the process is saved in order to support its in-depth analysis, statistical reports and future improvements.	
Main success scenario	1-The system lists the public tags assigned to the resource under consideration, and how they were regarded by the system (positive, negative or neutral); 2-The Provider chooses to option to exit.	

Table 122 provides some information about Use case UC70.

Table 122. Use case UC70

Abbreviated designation	UC70	
Extended designation	Manage global tags	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated.	
Postconditions	All the data pertinent to the performed operations are saved.	
Main success scenario	1-The Opinion analyst selects the option to manage global tags; 2- The system shows the sub-options available to manage global tags (from UC71 to UC74) and the option to exit; 3-The Opinion analyst selects the desired option; 4-The system proceeds in accordance with the selection of the Opinion analyst; 5-Steps 2 to 4 can be repeated many times until the Opinion analyst chooses to exit.	

Table 123 describes Use case UC71.

Table 123. Use case UC71

Abbreviated designation	UC71	
Extended designation	Add global tag	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated and he previously selected the option to add global tag (UC70).	
Postconditions	A new tag, the related data and metadata are saved.	
Main success scenario	1-The system displays a form asking for the tag and its global polarity; 2- The Opinion analyst fills in the form; 3-The system validates and saves the new tag and related data.	

Table 124 provides some information about Use case UC72.

Table 124. Use case UC72

Abbreviated designation	UC72	
Extended designation	Modify global polarity	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated and he previously selected the option to modify global tag (UC70).	
Postconditions	A new polarity and the related metadata are saved.	
Main success scenario	1-The Opinion analyst identifies the tag; 2- The system shows the tag's current polarity and solicits the new polarity; 3-The system saves the new polarity.	

Table 125 describes Use case UC73.

Table 125. Use case UC73

Abbreviated designation	UC73	
Extended designation	Delete global tag	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated and he previously selected the option to modify global tag (UC70).	
Postconditions	A tag is removed and the related metadata are saved in order to allow their in-depth analysis and improvement of the whole process.	
Main success scenario	1-The system solicits the tag to be deleted; 2-The Opinion analyst identifies the tag; 3-The system shows the tag's current polarity and solicits deletion confirmation; 4-The system verifies if the tag has a positive or negative polarity, and in that case, if it was not assigned to any resource. Then the tag is deleted and the system acknowledges the operation success.	

Table 126 provides some information about Use case UC74.

Table 126. Use case UC74

Abbreviated designation	UC74	
Extended designation	List global tags	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated and he previously selected the option to list global tags (UC70).	
Postconditions	-	
Main success scenario	1-The system shows the sort options available to the Opinion analyst; 2-The Opinion analyst selects the desired option; 3-The system lists the tags and their global polarity in accordance with the selected option.	

Table 127 describes Use case UC75.

Table 127. Use case UC75

Abbreviated designation	UC75	
Extended designation	Manage local tags	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated.	
Postconditions	All the data pertinent to the performed operations are saved.	
Main success scenario	1-The Opinion analyst selects the option to manage local tags; 2- The system shows the sub-options available to manage local tags (from UC76 to UC78) and the option to exit; 3-The Opinion analyst selects the desired option; 4-The system proceeds in accordance with the selection of the Opinion analyst; 5-Steps 2 to 4 can be repeated many times until the Opinion analyst chooses to exit.	

Table 128 provides some information about Use case UC76.

Table 128. Use case UC76

Abbreviated designation	UC76	
Extended designation	Modify local polarity	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated and he previously selected the option to modify local tag (UC70).	
Postconditions	A new polarity and the related metadata are saved.	
Main success scenario	1-The Opinion analyst identifies the resource; 2-The system shows the tags and their current polarity; 3-The Opinion analyst changes the polarity of one or more tags, and then chooses to exit; 4-The system saves the modifications.	

Table 129 describes Use case UC77.

Table 129. Use case UC77

Abbreviated designation	UC77	
Extended designation	Delete local tag	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated and he previously selected the option to modify local tag (UC70).	
Postconditions	A tag is removed and it is inserted in the list of improper ones. The related metadata are saved in order to allow their in-depth analysis and improvement of the whole process.	
Main success scenario	1-The Opinion analyst identifies the resource; 2-The system shows the tags and their current polarity; 3-The Opinion analyst selects the tag to be deleted; 4- The system deletion confirmation; 5-The Opinion analyst confirms the operation; 6-The system deletes the tag and adds it in the list of improper tags. Then it acknowledges the operation success.	

Table 130 provides some information about Use case UC78.

Table 130. Use case UC78

Abbreviated designation	UC78	
Extended designation	List local tags	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated and he previously selected the option to list local tags (UC70).	
Postconditions	-	
Main success scenario	1-The system shows the sort options and solicits the resource identification; 2-The Opinion analyst identifies the resource and selects the desired sort option; 3-The system shows the tags and their current polarity in accordance with the selected sort option; 4-The Opinion analyst chooses to exit.	

Table 131 describes Use case UC79.

Table 131. Use case UC79

Abbreviated designation	UC79	
Extended designation	Manage improper tags	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated.	
Postconditions	All the data pertinent to the performed operations are saved.	
Main success scenario	1-The Opinion analyst selects the option to manage improper tags; 2- The system shows the sub-options available to manage improper tags (from UC80 to UC81) and the option to exit; 3-The Opinion analyst selects the desired option; 4-The system proceeds in accordance with the selection of the Opinion analyst; 5-Steps 2 to 4 can be repeated many times until the Opinion analyst chooses to exit.	

Table 132 provides some information about Use case UC80.

Table 132. Use case UC80

Abbreviated designation	UC80	
Extended designation	Add improper tag	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated and he previously selected the option to add improper tag (UC70).	
Postconditions	A new improper tag, the related data and metadata are saved.	
Main success scenario	1-The system asks for the improper tag; 2- The Opinion analyst provides the improper tag; 3-The system validates and saves the new improper tag and related data.	

Table 133 describes Use case UC81.

Table 133. Use case UC81

Abbreviated designation	UC81	
Extended designation	Delete improper tag	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated and he previously selected the option to modify improper tag (UC70).	
Postconditions	An improper tag is removed and the related metadata are saved in order to allow their in-depth analysis and improvement of the whole process.	
Main success scenario	1-The system solicits the improper tag to be deleted; 2-The Opinion analyst identifies the tag; 3- The system solicits deletion confirmation; 4-The system acknowledges its deletion.	

Table 134 provides some information about Use case UC82.

Table 134. Use case UC82

Abbreviated designation	UC82	
Extended designation	List improper tags	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated and he previously selected the option to list improper tags (UC70).	
Postconditions	-	
Main success scenario	1-The system shows the sort options available to the Opinion analyst; 2-The Opinion analyst selects the desired option; 3-The system lists the improper tags in accordance with the selected option.	

Table 135 describes Use case UC83.

Table 135. Use case UC83

Abbreviated designation	UC83	
Extended designation	Manage modifiers	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags and related data.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated.	
Postconditions	All the data pertinent to the performed operations are saved.	
Main success scenario	1-The Opinion analyst selects the option to manage modifiers; 2- The system shows the sub-options available to manage modifiers (from UC84 to UC87) and the option to exit; 3-The Opinion analyst selects the desired option; 4-The system proceeds in accordance with the selection of the Opinion analyst; 5-Steps 2 to 4 can be repeated many times until the Opinion analyst chooses to exit.	

Table 136 provides some information about Use case UC84.

Table 136. Use case UC84

Abbreviated designation	UC84	
Extended designation	Add modifier	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags and related data.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated and he previously selected the option to add modifier (UC83).	
Postconditions	A new modifier, the related data and metadata operation are saved.	
Main success scenario	1-The system displays a form asking for the modifier and its type; 2- The Opinion analyst fills in the form; 3-The system validates and saves the new modifier and related data.	

Table 137 describes Use case UC85.

Table 137. Use case UC85

Abbreviated designation	UC85	
Extended designation	Modify modifier type	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated and he previously selected the option to modify modifier (UC83).	
Postconditions	A modifier characterisation is changed and the related data are saved.	
Main success scenario	1-The Opinion analyst identifies the modifier; 2- The system shows the modifier and its current type and solicits the new type; 3- The system verifies if the modifier was not used in a tag (positive or negative polarity) assigned to any resource. The system saves the new type and acknowledges the operation success.	

Table 138 provides some information about Use case UC86.

Table 138. Use case UC86

Abbreviated designation	UC86	
Extended designation	Delete modifier	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags and related data.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated and he previously selected the option to modify modifier (UC83).	
Postconditions	A modifier is removed and the related metadata are saved in order to allow their in-depth analysis and improvement of the whole process.	
Main success scenario	1-The system solicits the modifier to be deleted; 2-The Opinion analyst identifies the modifier; 3- The system shows the modifier and its type and solicits deletion confirmation; 4-The system verifies if the modifier is a negation indicator and in that case if it was not used in a tag (positive or negative polarity) assigned to any resource and acknowledges its deletion.	

Table 139 describes Use case UC87.

Table 139. Use case UC87

Abbreviated designation	UC87	
Extended designation	List modifiers	
Primary actors	Opinion analyst	
Stakeholders and interests	Opinion analyst	Wants accurate, easy and fast management of tags and related data.
	Institution manager	Wants a repository that allows the management and characterisation of the resources with accuracy and efficiency.
	Consumer/Provider	Wants a repository well-populated with resources deeply characterised with his opinions regarded.
Preconditions	The Opinion analyst is identified and authenticated and he previously selected the option to list modifiers (UC83).	
Postconditions	-	
Main success scenario	1-The system shows the sort options available to the Opinion analyst; 2-The Opinion analyst selects the desired option; 3-The system lists the modifiers and their type (intensifier indicator or negation indicator) in accordance with the selected option.	