Lengoaia eta Sistema Informatikoak Saila
Departamento de Lenguajes y Sistemas Informáticos

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Semi-Automatic Generation of Learning Domain Modules for Technology Supported Learning Systems

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy by

Miguel Larrañaga Olagaray

October 24, 2012

ISBN: 978-84-9860-781-9
D.L.-L.G.: BI-164-2013

Lengoaia eta Sistema Informatikoak Saila
Departamento de Lenguajes y Sistemas Informáticos

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Semi-Automatic Generation of Learning Domain Modules for Technology Supported Learning Systems

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy by Miguel Larrañaga Olagaray under the supervision of Dr. Jon Ander Elorriaga Arandia and Dr. Ana Arruarte Lasa

October 24, 2012

Acknowledgements

Now that I can glimpse the light at the end of the tunnel, the time to thank all the people who made this work feasible has arrived.

Joniri eta Anari lan hau ahalbidetzeagatik eta urteetan eskaini eta emandako laguntzagatik.

I want to thank all the members of the GaLan research work for all their suggestions, proposals and aids. In particular,

a Isabel, a ti se deben mis primeros pasos en el área y que los trabajos del grupo lleguen a buen puerto.

Bego Ferrerori, eten gabe emandako bultzadengaitik.

Ainhoari, buka dezadan estutzeagaitik. Apostuak ordaintzeko garaia dugu.

A Iñaki y Ángel, por vuestra colaboración en la última etapa.

A Maite Martín, Mikel Villamañe, Samara y David. Por animarme con vuestra alegría y vuestro humor.

Esker mila Ixa taldeko kideei, bereziki,

Arantzari, LNPan lehenengo urratsak egiten laguntzeagaitik eta emandako aholkuengaitik.

Maite Oronozi, gramatekin eta beste hainbat kontuekin laguntzeagaitik.

Enekori eta Aitorri, zuek gabe lan honen zati handia ezinezkoa izango zelako.

Kikeri, zuen tresnen erabilera eta zerbitzu batzuen garapena errazteagaitik.

Oierri, etorkizuneko lerro batzuk ezartzen laguntzeagaitik.

Rubeni, destorniak zain dituzu.

Dank u aan Erik, Katrien, Gonzalo, Joris, Bram, Sten, Nik en Martin voor de steun, hulp en advies die jullie mij gaven tijdens mijn verblijf aan de KULeuven.

Laguneei, muskideei, eskiatzen laguntzen didazuenei eta futbolekoei, noizbehinka honetaz ahazten laguntzeagatik.

A Beti, por su apoyo y optimismo.

A los compañeros del departamento, que habéis asumido mi docencia durante la estancia y congresos.

Y sobre todo, a mis padres. Por su paciencia y por aguantarme todo este tiempo.

Abstract

In a time when Technology Supported Learning Systems are being widely used, there is a lack of tools that allows their development in an automatic or semi-automatic way. Technology Supported Learning Systems require an appropriate *Domain Module*, i.e. the pedagogical representation of the domain to be mastered, in order to be effective. However, content authoring is a time and effort consuming task, therefore, efforts in automatising the *Domain Module* acquisition are necessary.

Traditionally, textbooks have been used as the main mechanism to maintain and transmit the knowledge of a certain subject or domain. Textbooks have been authored by domain experts who have organised the contents in a means that facilitate understanding and learning, considering pedagogical issues.

Given that textbooks are appropriate sources of information, they can be used to facilitate the development of the *Domain Module* allowing the identification of the topics to be mastered and the pedagogical relationships among them, as well as the extraction of Learning Objects, i.e. meaningful fragments of the textbook with educational purpose. Consequently, in this work *DOM-Sortze*, a framework for the semi-automatic construction of *Domain Modules* from electronic textbooks, has been developed. *DOM-Sortze* uses Natural Language Processing techniques, heuristic reasoning and ontologies to fulfill its work. *DOM-Sortze* has been designed and developed with the aim of automatising the development of the *Domain Module*, regardless of the subject, promoting the knowledge reuse and facilitating the collaboration of the users during the process. Although the approach is language independent, the Basque language has been chosen for the experimental work and

evaluation. Throughout this dissertation the design, development and evaluation of *DOM-Sortze* are described.

Laburpena

Informazio eta Komunikazio Teknologiek izandako iraultzak hezkuntzan ere eragin du, ikasketa eta ikaskuntza hobetzeko bideak eskainiz. Gaur egun, Teknologian Oinarritutako Hezkuntzarako Tresnak, esaterako Tutore Adimendunak eta Moodle[1] edo Blackboard[2] moduko hezkuntza kudeatzeko sistemak, ezinbestekoak bihurtu dira hainbat hezkuntza erakundetan (Parsad and Lewis, 2008). Are gehiago, Teknologian Oinarritutako Hezkuntzarako Tresnen erabilerak ikasleen motibazioan eta, horren eraginez, ikasketaren emaitzetan eragin dezake (Chen *et al.*, 2010).

Teknologian Oinarritutako Hezkuntzarako Tresnek *Domeinu-modulua* —hots, ikasi beharreko domeinuaren adierazpen pedagogikoa— behar dute. *Domeinu-modulua* da Teknologian Oinarritutako Hezkuntzarako edozein Tresnaren muina, hark adierazten baitu ikasleek ikasi beharreko ezagutza guztia (Anderson, 1988). Tutore Adimendunek, adibidez, osagai hori erabiltzen dute ikasleen ezagutza neurtzeko eta nola jokatu erabakitzeko. *Domeinu-modulua* ez bada egokia, ikasketa-prozesu eraginkorra burutzea ezinezkoa izango da (Anderson, 1988).

Domeinu-modulua sortzea ez da lan arina, ordea. Ikasi beharreko topikoak adierazteaz gain, identifikatu behar dira horien arteko erlazio pedagogikoak, ikasketa-saioak nola planifikatu zehazten dutenak, eta ikasteko erabiliko diren baliabideak. Hezkuntzara bideratutako liburu-egileek ere arazo horri aurre egin behar izaten diote liburuaren edukia egituratu eta antolatzeko orduan, alegia edukiak eta baliabide egokiak (deskribapenak, ariketak, etab.) identifikatu behar dituzte. Ezin ote genuke aurretik probetxugarriak suertatu diren liburuak berrerabili *Domeinu-modulua* automatikoki sortzeko?

[1] http://moodle.org
[2] http://www.blackboard.com

Azkeneko urte hauetan berrerabilpena bultzatzeko saiakerak burutu dira, hori ahalbidetzen duten estandarrak sortuz (LTSC, 2001) eta hezkuntzarako eduki berrerabilgarriak garatuz. Ikaste Objektu Biltegiak —ARIADNE (Duval *et al.*, 2001; Ternier *et al.*, 2009) edo Merlot (Cafolla, 2006)— eta GLOBE[3] bezalako biltegi sareak —ikastaro berriak sortzeko behar diren baliabideak eskaintzen dituztenak— gero eta ohikoagoak dira.

Lan honetan, *DOM-Sortze* aurkezten da, liburu elektronikoetatik *Domeinu-modulua* modu erdiautomatikoan sortzeko tresna.

Domeinu-moduluaren garapena

Hemen aurkeztutako proposamenean, Adimen Artifizialeko metodoak eta teknikak (adib. Hizkuntzaren Prozesamendua) erabiltzen dituen prozesu erdiautomatikoan datza *Domeinu-modulua*ren garapena. *Domeinu-moduluak* bi maila ezberdinetako ezagutza jasotzen du, Ikaste Domeinuaren Ontologia eta Ikaste Objektuen bilduma. *Domeinu-modulua* eraikitzeko hiru urrats egin behar dira (ikus 1. Irudia):

1. **Liburuaren aurreprozesua**: Hasteko, aukeratutako liburua informazio-erauzketarako egokitu behar da. Bai egokitutako liburua bai liburuaren aurkibidea erabiliko dira jarraian burutuko diren erauzketa-prozesuetan.

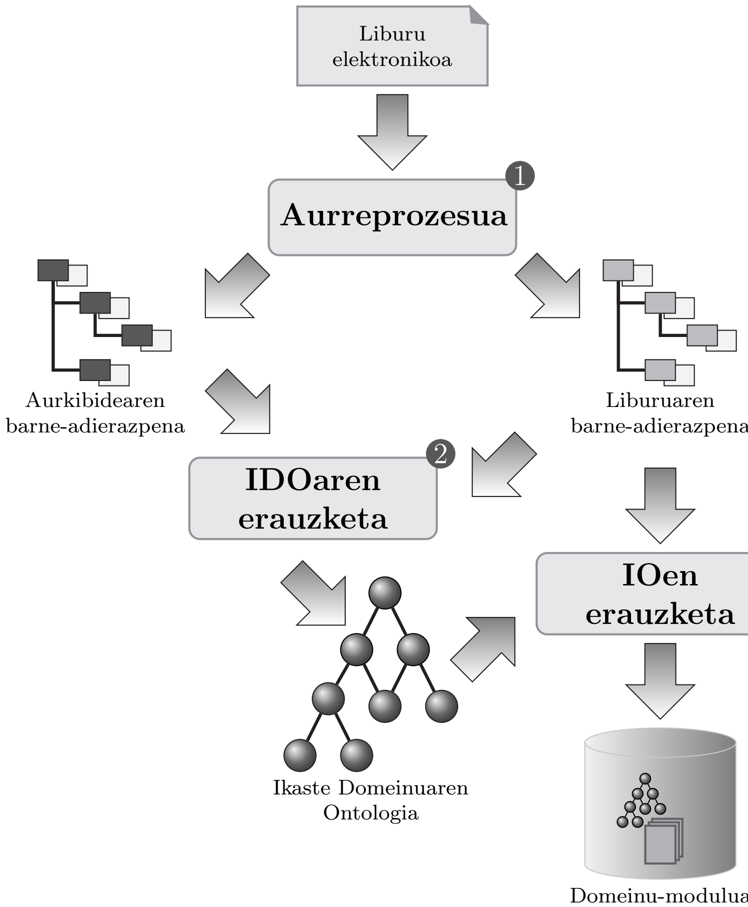
2. **Ikaste Domeinuaren Ontologiaren erauzketa (IDOaren erauzketa)**: Urrats honetan, ikasi beharreko topikoak zein beraien arteko erlazio pedagogikoak erauzten dira. Ikaste Domeinuaren Ontologiak Teknologian Oinarritutako Hezkuntzarako Tresnei ikasketa-saioen antolaketa bideratzea ahalbidetzen die. Horrez gain, ikasleek beraien kabuz ikasteko ere erabili dezakete.

3. **Ikaste Objektuen erauzketa (IOen erauzketa)**: Urrats honetan, ikasteko erabiliko diren edukiak (definizioak, adibideak, ariketak, ...) erauzten dira.

Prozesu erdiautomatiko horren emaitzak —bai Ikaste Domeinuaren Ontologia bai Ikast Objektuak— *Domeinu-modulu*aren egileek berrikustea ezinbestekoa iruditzen zaigu. Berrikuspena elkarlanaren bidez burutu ahal izateko Elkar-DOM (Larrañaga *et al.*, 2007) garatu da. Kontzeptu-mapetan oinarritutako tresna da Elkar-DOM, *Domeinu-modulu*a berrikustea ahalbidetzen duena. Tresna horren bidez emaitzak zuzentzeko edota egileen gustura egokitzeko aukera dago.

Jarraian, aipatutako urrats bakoitza xehetasun gehiagorekin azaltzen da.

[3] http://www.globe-info.org/

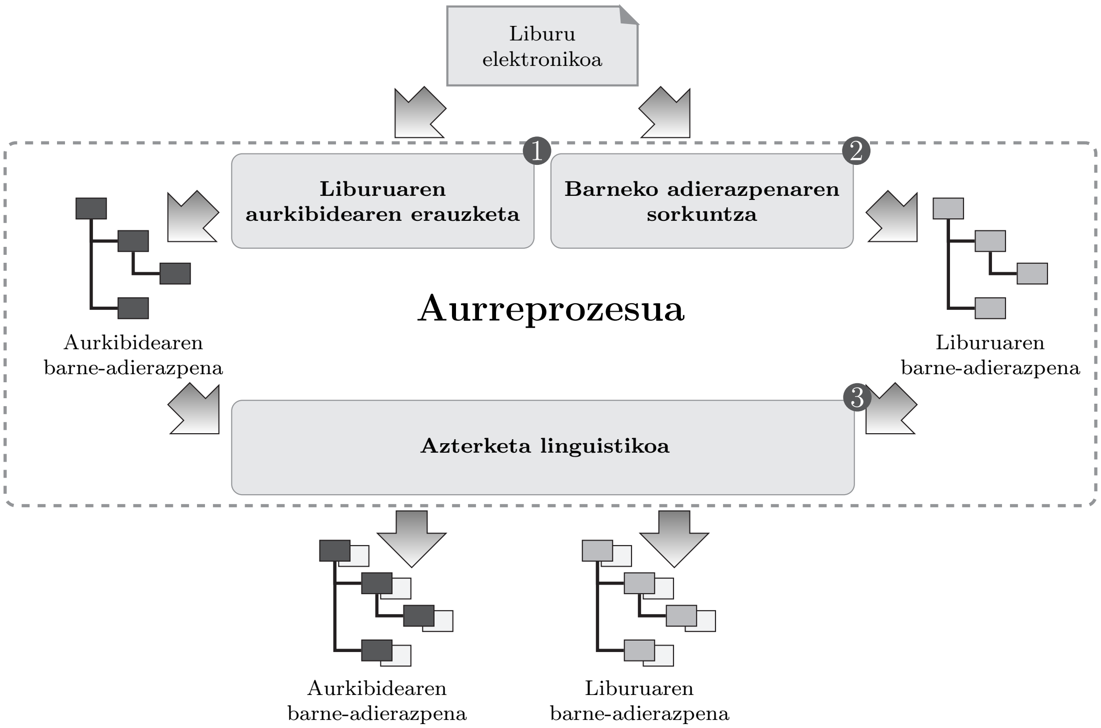


1. Irudia – Domeinu-moduluaren sorkuntza prozesua

Liburuen aurreprozesua

Liburuaren aurreprozesuak dokumentutik abiatu eta *Domeinu-modulu*aren erauzketarako egokia den adierazpen estandarra sortzen du (ikus 2. Irudia). Liburu elektronikoak formatu ezberdinetan (*pdf*, *doc*, etab.) egon daitezke, eta aurreprozesuak edozein dokumentu erauzketarako egokitzea ahalbidetzen du. Dokumentu mota horiek guztiak zuhaitz moduko egitura dute; liburuak ataletan banatzen dira, horiek azpiataletan, eta horrela hurrenez hurren. Urrats honetan, zuhaitz moduko egitura hori erauzten da eta informazio morfosintaktikoarekin aberasten da.

Liburuen aurkibidea —liburuaren eduki nagusiak eta egitura adierazten duena— liburuaren hasieran zein bukaeran topa daiteke. Gainera, zenbatzeko formatu desberdinak erabili ohi dira (zenbakiak, hizkiak, etab.). Arazo horiek gainditzeko, aurreprozesuan aurkibidearen barne-adierazpena ere sortzen da.



2. Irudia – Liburuaren aurreprozesua

Sortutako liburuaren eta aurkibidearen barne-adierazpenak azterketa linguistikoan lortutako informazio morfosintaktikoarekin aberasten dira. Informazio hori ezinbestekoa da erauzketa aurrera eramateko, batez ere Euskara bezalako hizkuntza eranskarietan.

Euskaraz, adibidez, lemei artikulua, numeroa eta deklinabideko kasua adierazten duten atzizkiak gehituz osatzen dira hitzak (Aduriz *et al.*, 1998). Hitzen esanahaia ulertzeko, ezinbestekoa da azterketa morfosintaktikoa, Euskararen kasuan EUSLEM (Aduriz *et al.*, 1996) lematizatzaile-etiketatzailea erabiliz burutzen dena.

Ikaste Domeinuaren Ontologiaren erauzketa

Ontologia-erauzketa —hots, domeinu-ontologien edukia hainbat iturburutik modu automatikoan edo erdiautomatikoan osatzea— modu ezberdinetan landu da dagoeneko (Buitelaar *et al.*, 2005b). Proiektu gehienek domeinu-ontologia bat sortzea edo zabaltzea dute helburu. Beste batzuek, ordea, Wordnet (Fellbaum, 1998) moduko ontologia lexikoak aberasteko asmoa dute. Ontologia-erauzketarako ikaste automatikoa eta Hizkuntzaren Prozesamendurako teknikak erabili ohi dira bai corpusetatik bai hiztegi elektroniko bai beste iturburu batzuetik ontologien edukia lortzeko.

Lan honetan, Ikaste Domeinuaren Ontologiak, domeinuko topiko nagusiez gain, beraien arteko erlazio pedagogikoak adierazten ditu. Erlazio pedagogikoak *egitura* (*isA* edo *partOf*) eta *ordena* (*prerequisite* edo *next*) adieraz dezakete. Adibidez, $X \xrightarrow{isA} Y$ erlazioak adierazten du X topikoa Yren mota zehatz bat dela. $X \xrightarrow{partOf} Y$ erlazioak, ordea, X topikoa Yren zati bat dela dio, hots, Y landutzat emateko ikasi behar diren topikoetako bat dela adierazten du. Halaber, $Y \xrightarrow{prerequisite} X$ erlazioak adierazten du X ikasten hasi aurretik ikasleak Y dagoeneko landuta izan behar duela eta $X \xrightarrow{next} Y$ erlazioak, berriz, X landu ondoren Y ikastea gomendatzen du.

Egitura sintaktikoen pean ezagutza semantikoa dagoelako premisan oinarritzen da ontologia-erauzketa. Adibidez, Text2Onto (Cimiano and Völker, 2005) tresnak Hearsten (1992) patroaik erabiltzen ditu erlazio taxonomikoak erauzteko, eta termino habiaratuak aztertzeko neurriak erabiltzen ditu termino hautagaiak bilatzeko (Frantzi *et al.*, 1998). OntoLT-k (Buitelaar *et al.*, 2004) —testuetatik ontologiak erauzteko Protégé[4]-rentzako pluginak— *genus et differentiam* egiturak baliatzen ditu erlazio taxonomikoak bilatzeko.

Ikaste Domeinuaren Ontologiaren erauzketa bi urratsetan burutzen da, Hizkuntzaren Prozesamendua eta arrazoibide heuristikoa erabiltzen dutenak: aurkibidearen azterketak hasierako ontologiaren bertsioa erauzten du, gero dokumentu osoaren azterketarekin aberastuko dena. Prozesu horretan zehar, ontologiaren barne-adierazpena ere erabiltzen da, edukiez gain informazioa erauzteko erabilitako heuristikoen informazioa adierazten duena. Prozesua bukatutakoan, *Domeinu-modulu*aren egileek ontologia berrikus dezakete Elkar-DOM (Larrañaga *et al.*, 2007) erabiliz.

Aurkibidearen azterketa

Hemen aurkeztutako ontologia eraikitzeko prozesuan, aurkibideak dira iturburu nagusiak. Aurkibideek eduki nagusiak laburbiltzen dituzte eta dokumentuaren egitura —hots, edukiak nola dauden antolatuta— adierazten dute. Egitura horien atzean irizpide pedagogikoak egon ohi dira; antolaketa erabakitzeko, adibidez, argi izan behar da zein eduki jakin behar den beste bat ulertzeko. Ezaugarri horiek dira aurkibideak baliabide aproposak bilakatzen dituztenak Ikaste Domeinuaren Ontologiaren erauzketa erdiautomatikorako. Erlazio pedagogikoak aurkibidean adierazita daudela kontuan hartuta, Hizkuntzaren Prozesamendurako tresnak eta heuristiko multzo bat erabiliko dira erlazio pedagogikoak erauzketarako.

[4] http://protege.stanford.edu/

Aurkibidearen azterketan zehar bi prozesu burutuko dira:

- **Oinarrizko azterketa**: Ataza honetan, aurkibidearen barne-adierazpenetik erauzten dira domeinuko topiko nagusiak eta beraien arteko erlazio pedagogikoak. Proposatutako prozeduraren arabera, aurkibideko elementu bakoitza topikotzat ematen da. Gainera, aurkibidearen egitura aintzat hartuz, erlazio pedagogikoak bilatzen dira. Aurkibide elementu baten umea bestearen zati edo kasu zehatz bat deskribatzeko erabiltzen da, beraz, *egitura* erlazio pedagogiko bat sortzen da aurkibide elementu bat eta bere ume bakoitzaren artean. Gainera, aurkibidearen antolaketak gomendatutako ordena adierazten du eta *ordena* erlazio pedagogikoak erauzteko balio du.

- **Azterketa heuristikoa**: Oinarrizko azterketan jasotako emaitzak birfintzen dira ataza honetan, eta erlazio berri batzuk —*prerequisite* erlazioa batez ere— erauzten dira heuristiko sorta bat baliatuz. Erauzitako erlazioak, motaz gain, bilaketan erabilitako heuristikoak eta horien ziurtasun mailak ere gordetzen dira. Azterketa heuristikoa bi fasetan egikaritzen da. Lehendabizi, *isA* eta *partOf* erlazioak erauzteko erabiltzen diren heuristikoak aplikatzen dira eta, jarraian, *next* eta *prerequisite* erlazioak topatzen dituztenak.

  Erabilitako heuristikoak aurkibide multzo batetik atera ziren eta, gerora, Euskal Herriko Unibertsitateko (UPV/EHU) 150 irakasgai ezberdinetako aurkibideekin probatu ziren emaitza onak lortuz (Larrañaga *et al.*, 2004).

Liburu osoaren azterketa

Aurkibidea aztertu ondoren, jasotako ontologia topiko zein erlazio berriekin osatzen da, horretarako liburu osoa aztertzen delarik.

Topiko berrien erauzketa

Urrats honetan, liburu osoa aztertzen da uneko Ikaste Domeinuaren Ontologia topiko berriekin aberasteko asmoz. Azken urte hauetan, Hizkuntzaren Prozesamendurako teknikak eta metodo estatistikoak nahasten dituzten prozedura hibridoak nagusitu dira. Lan askotan, lehendabizi patroiak erabiltzen dituzte termino hautagaiak topatzeko eta, neurri estatistikoak erabiliz, hautagai onenen sailkapen bat egiten dute (Justeson and Katz, 1995).

DOM-Sortze-n, topiko-erauzketa Erauzterm euskararako termino-erauzlearen bidez (Alegria *et al.*, 2004b) burutzen da. Erauzterm hitz bakarreko zein hitz anitzeko terminoak erauzteko gai da.

Topikoen arteko erlazio berrien erauzketa

Prozesu honek liburu osotik topikoen arteko erlazio berriak erauztea ahalbidetzen du patroiak erabiliz. Patroien bidez egitura sintaktikoetan erlazio pedagogikoak bilatzen dira. Erlazio horiek topatu ahal izateko, ontologiako topikoen agerraldiak etiketatzen dira liburuaren barne-adierazpenean, bertan baitago informazio morfosintaktikoa. Topatutako topikoen habiaraketa maila kontuan hartzen da *isA* erlazioa topatzeko. Adibidez, *Sirius izarra* topikoan *izar* dago, ontologiako beste topiko bat izan daitekeena. Ikus daitekeenez, *Sirius izarra* izar zehatz bat da, beraz, bi topiko horien artean *isA* erlazioa igar daiteke. Bukatzeko, topikoak etiketatuta dauzkaten esaldiak gramatika bat erabiliz aztertzen dira erlazio gehiago bilatzeko asmoz. Gramatika horrek Constraint Grammar formalismoa (Karlsson *et al.*, 1995) jarraitzen du, sintaxia aztertzeko modu zabalduenetako bat. 1. Taulan ikus daitezkeen moduko erregelak definituta daude gramatikan.

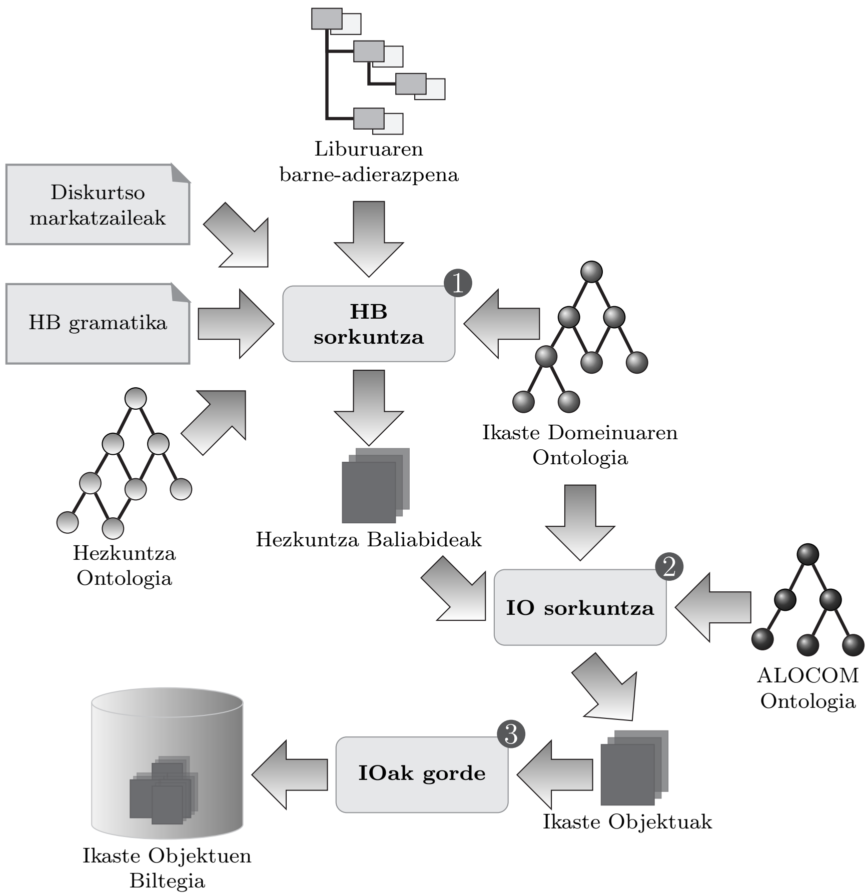
1. Taula – *isA* erlazioak topatzeko patroi batzuk

	Adibidea
Patroia	@Topic @Topic [*det*] **IZAN**
Adibidea	Lurra planeta *bat* **da**.
Patroia	@Topic **izeneko** @Topic
Adibidea	Esne bidea **izeneko** galaxiak 100 mila milioi izar dituela uste dute zientzialariek.

Ikaste Objektuen erauzketa

Liburuan dauden hezkuntzarako baliabideak, hots, ontologiako topiko bati edo gehiagori lotuta dauden zatiak eta helburu pedagogikoa daukatenak (adib. *definizioak*, *adibideak* edo *ariketak*) lortzean datza Ikaste Objektuen erauzketa. Horretarako, Hizkuntzaren Prozesamendurako teknikak eta ontologiak erabiltzen dira. Domeinuarekiko independentea izan nahi duen hurbilpena izateko asmoz sortu denez, domeinuari lotutako erabiltzen den informazio bakarra Ikaste Domeinuaren Ontologiak dauzkana da, aurreko fasean jasotakoa. Hemendik aurrera, Hezkuntza Baliabide terminoa erabiltzen denean, ikasteko asmoz erabiliko den liburu zati bati dagokio, adibidez *ariketa* edo *definizio* bat. Ikaste Objektua hezkuntzarako baliabide berrerabilgarria da, bilaketa eta erabilera errazten duten metadatuekin aberastua. Ikaste

Objektuen erauzketa *DOM-Sortze*-ren zati bat den *ErauzOnt* osagaia erabiliz burutzen da (Larrañaga *et al.*, 2011).



3. Irudia – Ikaste Objektuak liburutik sortzeko prozedura

Ikaste Objektuak sortzeko prozedura orokorra hiru atazatan burutzen da (ikus 3. Irudia). Lehendabizi, liburutik Hezkuntza Baliabideak erauzten dira (HB sorkuntza). Ondoren, baliabide horiek metadatuekin aberasten dira Ikaste Objektuak sortzeko (IO sorkuntza), eta bukaeran, Ikaste Objektuen Biltegian gordetzen dira (IOak gorde), berriro ere erabili ahal izateko.

Hezkuntza Baliabideen gramatikak (HB gramatikak) *definizioak*, *ariketak*, *teoriak*, etab. bilatzeko erregelak definitzen ditu eta testuko Hezkuntza Baliabideak topatzea ahalbidetzen du. Topatutako baliabideak bi modutan aberasten dira: alde batetik, jarraian dauden bi baliabide beste berri bat osatzeko batzen dira baldin eduki aldetik eta baliabide mota aldetik antzekoak badira (Larrañaga *et al.*, 2008c,a, 2011, 2012).

2. Taulan erakusten dira osatzeko hautagai diren bi baliabide jarraituak. Bestalde, diskurtso markatzaileak erabiltzen dira testuaren koherentzia mantendu ahal izateko.

2. Taula – Elkartu daitezkeen bi baliabide

Baliabidea	**Edukia**
HB_1	Planetak berezko argirik ez duten gorputzak dira, eta izar baten inguruan biraka mugitzen dira. Uste denez, Eguzki-Sistemako planetak Eguzkiarekin batera eratu ziren, eta pentsa daiteke antzeko planeta ugari izango direla beste izar batzuen inguruan.
HB_2	Lurra Planeta bat da.

Ondoren, Ikaste Objektuak sortzen dira Hezkuntza Baliabideetatik abiatuta. Horretarako, metadatuak modu automatikoan eransten zaizkie. Alde batetik, Samgi (Meire *et al.*, 2007) etiketatzailea erabiltzen da hasierako metadatuak lortzeko. Bestalde, gako-hitzak Ikaste Domeinuaren Ontologia erabiliz aberasten da topikoen erlazioak kontuan har daitezen. Samgi-k testu batean *Artizarra*, *Marte* eta *Lurra* aipatzen dituen testuan ez luke jakingo *planetei* buruz ari dela. Ontologia erabilita, aldiz, horrelako ezagutza kontuan hartzen da metadatuak hobetzeko.

Bukaeran, sortutako Ikaste Objektu guztiak Ikaste Objektuen Biltegian gordetzen dira. Biltegia ARIADNEren teknologian (Duval *et al.*, 2001; Ternier *et al.*, 2009) oinarritzen da.

Laburpena

Lan honetan *DOM-Sortze*, *Domeinu-modulua* liburu elektronikoetatik era erdiautomatikoan erauzteko tresna bat deskribatu da, domeinuarekiko independentea dena. Ontologietan, arrazonamendu heuristikoan eta Hizkuntzaren Prozesamendurako tekniketan oinarritzen da *DOM-Sortze*. Bai *DOM-Sortze* bai bere osagaiak hainbat dokumentu eta aurkibiderekin probatu dira emaitza onak jasoz (Larrañaga *et al.*, 2004, 2012, submitted). Tesian zehar emaitza hauek luze eta zabal aurkezten dira.

Contents

List of Figures

List of Tables

List of Listings

List of Algorithms

Acronyms

ADL	Advance Distributed Learning - US Department of Defense
AI	Artificial Intelligence
AICC	Aviation Industry CBT Committee
API	Application Programming Interface
CBT	Computer-Based Training
CEN	European Committee for Standardization
CEN/ISSS/LT	European Committee for Standardization/Information Society Standardisation System/Learning Technologies Workshop
CM	Concept Map
CP	IMS Content Packaging
DC	Dublin Core
DR	Didactic Resource
GEM	Gateway to Educational Materials
GESTALT	Getting Educational Systems Talking Across Leading Edge Technologies
ICT	Information and Communication Technologies

IMS	IMS Global Learning Consortium
ITS	Intelligent Tutoring System
LD	IMS Learning Design
LDO	Learning Domain Ontology
LMS	Learning Management System
LO	Learning Object
LOM	IEEE Learning Object Metadata
LOR	Learning Object Repository
LTSC	IEEE Learning Technology Standards Committee
NLP	Natural Language Processing
OAI	Open Archive Initiative
OAI-PMH	Open Archive Initiative Protocol for Metadata Harvesting
OL	Ontology Learning
PLQL	ProLearn Query Language
PROMETEUS	PROmoting Multimedia access to Education and Training in EUropean Society
QEL	Query Exchange Language
QTI	IMS Question and Test Interoperability
RTF	Relative Term Frequency
SCO	Sharable Content Object
SCORM	Sharable Content Object Reference Model
SPI	Simple Publishing Interface
SQI	Simple Query Interface
SS	IMS Simple Sequencing
TFIDF	Term Frequency Inverse Document Frequency
TSLS	Technology Supported Learning System

1 Introduction

Advances in the last few years have greatly increased the influence of new technologies in general, and Information and Communication Technologies (ICT) in particular, in technologically developed societies. On-line applications have become essential, they are continuously used for communication (e.g., instant messaging, mailing, phoning), consulting bank accounts, and so on.

This revolution has also affected education, providing means than enhance both teaching and learning. Years of research have facilitated the development of different kinds of Technology Supported Learning Systems (TSLSs) such as Learning Management Systems (LMSs), Intelligent Tutoring Systems (ITSs), Collaborative Learning Systems or Adaptive and Intelligent Web-based Educational Systems. LMSs such as Moodle[1] or WebCT\Blackboard[2] are currently being used at many academic institutions (Waits and Lewis, 2003; Parsad and Lewis, 2008). Furthermore, a positive relationship between the use of Web-based learning technology and student engagement and desirable learning outcomes has been observed (Chen *et al.*, 2010). ITSs have also proved to improve the achievements of students (Anderson *et al.*, 1995; Koedinger *et al.*, 1997; Corbett *et al.*, 1998; Mitrovic and Ohlsson, 1999; Arroyo *et al.*, 2001; Mitrovic *et al.*, 2004; Woolf *et al.*, 2006).

In order to facilitate learning, TSLSs require, regardless of the technology or the paradigm, an appropriate *Domain Module*, i.e., the pedagogical representation of the domain to be learnt. The *Domain Module* is considered the core of any TSLSs as it represents the knowledge about a subject matter to be communicated to the learner (Wenger, 1987; Anderson, 1988; Woolf, 2008; Nkambou *et al.*, 2010). The *Domain*

[1] http://moodle.org
[2] http://www.blackboard.com

Module enables either the students to learn by themselves, in the case of exploratory Learning Systems, or the TSLSs to plan the learning process in instructivist systems. For example, an ITS relies on the *Domain Module* to determine the content of the tutorial interaction, the selection of examples, questions and statements, and to evaluate the performance of the students (Stevens *et al.*, 1982; Wenger, 1987).

However, building TSLSs, especially their *Domain Modules*, is cost and labour-intense. In particular, Anderson (1988) estimated that in his experience developing ITSs 50% of the effort went on the development of the *Domain Module*. This proportion was even expected to increase, while the development of the other components has become more automated. This issue has slowed down the proliferation and broad use of ITSs at any academic institution, despite their proven usefulness.

Many authoring tools have been developed with the aim of getting the development of TSLSs more cost and effort-effective[3]. Nevertheless, there is a lack of average teacher oriented authoring tools; i.e., tools that teachers with low expertise in system development can use with the aim of developing TSLSs. Most authoring tools have been developed for computing literate users or knowledge engineers and, therefore, they become too complicated for average teachers, who may give up the development of their own TSLSs. Brusilovsky *et al.* (2003) claim that teachers should focus on *Domain Module* authoring while the development of the core of the TSLS should be carried out by expert developers. Technology Enhanced Learning Community should not expect the average teachers to contribute to the development of TSLSs any more than they should be expected to author a textbook in their field (Murray, 1999).

Nevertheless, building the *Domain Module* is a hard task that might become easier by reusing existing materials (Casey and McAlpine, 2003). *Domain Module* authoring entails not only selecting the domain topics to be learnt, but also defining the pedagogical relationships among the learning sessions, content sequencing, etc. On the one hand, the proliferation of Learning Objects (LOs), i.e. reusable resources with educational purposes, and Learning Object Repositories (LORs) might help to reduce the development cost of the learning material to be used (Downes, 2003). On the other hand, electronic textbooks might be used as the source to build the *Domain Module*, reproducing how average teachers behave while preparing their subjects: they choose a set of reference books that provide the main Didactic Resources (DRs) - *definitions*, *examples*, *exercises*, ... - for the subject, and rely on them for scheduling their lectures. Textbook authors have to structure the document in a means that helps learning, and provide appropriate resources for learning. Therefore, these kinds of documents can be used to build the *Domain Module*.

[3] Murray *et al.* (2003) describe the most popular authoring tools for ITSs.

This thesis addresses the semi-automatic generation of the *Domain Module* from electronic textbooks. Gathering the domain knowledge of a TSLS from already existing documents might considerably reduce its development cost.

1.1 Motivation and Goals

The main objective of this thesis work is to semi-automatise the generation of TSLSs. More specifically, it aims at reproducing the way teachers prepare their lessons and the learning material they will use throughout the course. Teachers tend to choose one or more textbooks that cover the contents of their subjects, determine the topics to be addressed, and identify the parts of the textbooks which may be helpful for the students. Textbooks provide resources with educational purposes, e.g., *definitions*, *theories* or *theorems*, or *exercises*, which might be useful for learning. Moreover, the authors of the textbooks, experts in the domain, organise their content in the way they consider most appropriate for learning. Therefore, the structure of the textbook, which is reflected in its outline or table of contents, recommends the order in which topics should be addressed and even when a topic should be learnt prior to attempting another one.

Given that electronic textbooks are appropriate sources of information, this work aims to analyse, design, develop, and validate a framework, *DOM-Sortze*, for the construction of *Domain Modules*. The framework should be intended for average teachers rather than for knowledge engineers, so it should not require advanced computer science/engineering mastery in order to be used. Facilitating the development of *Domain Modules* may also make teachers less reluctant to profit from TSLSs in their lectures.

The framework must fulfill the following requirements:

- **Semi-automatic construction of the *Domain Module***: Content development is time and effort consuming, so the workload of the users while building *Domain Modules* should be lightened. In order to achieve that goal, the framework should rely on semi-automatic processes that minimise the need for the intervention of the users. Rather than authoring the *Domain Module* from scratch, the users are expected to play a different role. They will supervise the automatically gathered knowledge and adapt it to their preferences and needs.

- **Knowledge reuse**: Content reusing can significantly reduce the cost of the development of the *Domain Modules*. It has been one of the major concerns of the area during the last years. In particular, standards and specifications that enable the development and distribution of reusable components have been defined. The developed framework should facilitate and promote content reuse, to which end it will rely on these specifications and standards.

- **Collaborative work**: Teachers usually cooperate while preparing and planning their lessons, so they should be able to supervise and develop the *Domain Modules* in a similar way. Therefore, the developed framework should facilitate the collaborative work of the users and enable them to work together, discuss the contents to be used, etc. The framework must provide an easy means to supervise the automatically elicited knowledge that facilitates the collaborative work of the *Domain Module* authors.

- **Domain-independence**: The developed framework should not be constrained to a particular domain, as it may limit its usefulness. Although focusing on a concrete domain may result in more accurate techniques, those are not be applicable to other areas, so a domain-independent approach is pursued and, thus, previous domain-specific information should be not used.

- **Multilingualism. A language independent framework**: Nowadays, education in different languages is a reality in many academic institutions. Therefore, it is also desirable that the framework is able to support diverse languages or can easily be enhanced so that it could deal with new languages.

1.2 Working Methodology

This thesis describes *DOM-Sortze*, a framework that enables the semi-automatic generation of the *Domain Module* for TSLS from electronic textbooks. The semi-automatic generation of the *Domain Module* entails the identification and elicitation of knowledge from the documents to which end Natural Language Processing (NLP) techniques are combined with ontologies and heuristic reasoning.

The *DOM-Sortze* framework consists of several applications that perform the acquisition of the different components of the *Domain Module*. Its modular design has facilitated carrying out an incremental development of the framework addressing, different components, which deal with particular aspects of the *Domain Module* construction, in each interaction.

For every developed component, the following procedure has been conducted: first, an analysis of the *state-of-the-art* has been performed to determine which are the adequate approach and means to deal with its purpose. Given that most of the tasks entail knowledge acquisition processes, a reduced set of documents has been manually analysed to identify the means for the automatic knowledge elicitation. After its development, the component has been evaluated using a Gold-standard approach. An instructional designers team has defined the reference expected results, which have been compared to those automatically obtained by *DOM-Sortze*. In this work, documents in the Basque language have been used as the source of information.

Widely used Software Engineering methodologies and approaches have been used to incrementally develop a modular, flexible, and multiplatform framework, *DOM-Sortze*.

1.3 Context

The work here presented has been developed in the GaLan[4] research group. This research group, located in the University of the Basque Country (UPV/EHU), has been carrying out its research activity in the area of Computer-aided Education since the early 90's. The main activity of the group has been focused on the development of architectures and tools for educational purposes. The GaLan group includes a multidisciplinary team whose members belong to three departments of the UPV/EHU ("Computer Languages and Systems", "Computer Science and Artificial Intelligence" and "Social Psychology and Methodology of Behavioural Sciences"). Their particular backgrounds integrate different aspects relevant to the development of educational tools.

Among other research lines, the GaLan research group has worked on the development of authoring tools for TSLSs, in particular the construction of ITSs (Arruarte, 1998; Arruarte *et al.*, 2003), and on Concept Mapping either as a knowledge representation mechanism or as an aiding tool for learning (Rueda, 2009; Rueda *et al.*, 2009; Elorriaga *et al.*, 2011). These two research lines are the pillars which sustain the work presented in this dissertation, which aims to go a step beyond authoring tools by automatising the construction of the *Domain Modules*, and relies on Concept Maps as a means to enable the collaborative work of the *Domain Module* authors in the supervision of the gathered knowledge.

1.4 Outline

This dissertation is divided in 7 chapters. Chapter 2 describes the state of the art on *Domain Module* representation and acquisition, focusing on the uses of ontologies

[4] http://galan.ehu.es

and standards in the Technology Enhanced Learning Community. This chapter also describes some of the most common ontology learning techniques.

Chapter 3 introduces the proposed approach for the semi-automatic generation of the *Domain Module*, enumerating the steps that are followed to perform the task combining NLP techniques, ontologies and heuristic reasoning.

Chapter 4 presents how the ontology that represents the *Domain Module*, including the domain topics and the pedagogical relationships among them, is gathered from electronic textbooks.

Chapter 5 addresses the identification and extraction of didactic resources from the electronic textbooks. As content reuse is one of the major goals, this chapter also describes how Learning Objects, i.e., reusable didactic resources, are built from the extracted didactic resources including aspects such as the annotation that enables the search and retrieval of these kinds of resources.

The architecture of *DOM-Sortze* and the applications that it entails are described in Chapter 6.

An evaluation of the proposed approach is depicted in Chapter 7, and, finally, some conclusions and future work are presented in Chapter 8.

More detailed information about the resources used for the *Domain Module* acquisition are presented in the appendices.

2

Domain Module Representation and Acquisition: State of the Art

The *Domain Module*, which is also referred to as *Expert Module*, is considered the core of any TSLSs as it represents the knowledge about a subject matter to be communicated to the learner (Wenger, 1987; Anderson, 1988; Woolf, 2008; Nkambou *et al.*, 2010). It is used either as the source for the knowledge in learning/teaching sessions or as the basis for evaluating the performance of the students (Wenger, 1987). The *Domain Module* was the first aspect of the expertise of the teacher to be explicitly represented in TSLSs.

An instructional system can hardly be effective without an appropriate description of the domain knowledge. Furthermore, an incomplete *Domain Module* may result in a TSLSs which is only able to provide part of the instruction required on the domain (Anderson, 1988). The representation of the domain knowledge is known to be crucial in order to achieve effective teaching (Clancey, 1982; Dede, 1986; Lesgold, 1988). The nature of the stored knowledge determines not only the content of a tutorial interaction, but also the goal structure that governs the tutor's selection of examples, questions and statements at different points in the dialogue, the types of misconceptions that students could have and the way that tutors diagnose and correct these misconceptions based on students' errors (Stevens *et al.*, 1982). The domain knowledge has to be organised and represented in such a way that it enables simple, clear and effective teaching or learning of the domain contents (Gutiérrez, 1994).

In the past, many approaches have been followed to define the *Domain Module* in different TSLSs. Nevertheless, in most cases the *Domain Module* was developed *ad-hoc* for a particular TSLS and it could seldom be reused in a different one. Recent research has focused on overcoming this problem using the following two approaches which enable the interoperability of different TSLSs: ontological engineering (Mizoguchi and Bourdeau, 2000) and the use of standards (Duval, 2004).

This chapter is structured as follows: Ontological engineering and works on the application of ontologies on TSLSs are presented in Section 2.1. Section 2.2 describes how ontologies can be semi-automatically generated from texts or other resources. Section 2.3 describes the standards and specifications that have been defined in the last few years. Frameworks and systems that profit from the described technologies are presented in Section 2.4. Finally, some conclusions are presented.

2.1 Ontologies and Ontological Engineering

The main goal of Ontological Engineering is to "provide a basis for building models of all things in which computer science is interested" (Mizoguchi and Ikeda, 1997). Ontologies are used as a means to achieve that purpose, but before summarising the benefits of ontological engineering in TSLS development, some definitions and reflections on ontologies are referred.

The ontology term has been adopted from philosophy, where it is defined as the *"theory of existence"*. There are many definitions for ontologies in the Computer Science area. Neches *et al.* (1991) proposed the following definition: *"an ontology defines the basic terms and relations comprising the vocabulary of the topic area as well as the rules for combining terms and relations to define extensions to the vocabulary"*. However, Gruber (1993) made the most popular definition of ontologies, which states that *"an ontology is an explicit explanation of a conceptualization"*. This definition was slightly enhanced by Borst (1997), who referred to ontologies as *"formal specifications of a shared conceptualization"*.

According to Studer *et al.* (1998), *"conceptualization refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used, and the constraints on their use are explicitly defined. Formal refers to the fact that the ontology should be machine-readable. Shared reflects the notion that an ontology captures consensual knowledge, that is, it is not private to some individual, but accepted by a group"*.

Ontologies aim at capturing and describing domain knowledge in a generic way and providing a commonly agreed understanding of a domain, which may be reused and shared across applications and groups (Chandrasekaran *et al.*, 1999). They arose as a means to get shareable and reusable knowledge bases (Gruber, 1991) and are the core of the *Semantic Web* (Berners-Lee and Fischetti, 1999; Berners-Lee *et al.*, 2001).

Uschold and Gruninger (1996) summarised some of the possible applications of ontologies, and classified the benefits of using them into the following groups:

- **Communication**. People from different but related fields might use different terms to describe the same underlying ideas. Ontologies can be used as the unifying conceptual framework that represents the common ideas and the terms used to correspond to them, thus enabling the translation between the different perspectives.

- **Inter-operability** among systems, i.e., using the ontology as an inter-lingua to unify different languages and software tools.

- **System engineering:** Ontologies also assist in the process of building and maintaining software systems, both knowledge-based and otherwise. In particular, they promote:

  - *Re-usability:* The ontology, when represented in a formal language, can be (or become so by automatic translation) a re-usable and/or shared component in a software system.
  - *Reliability:* A formal representation facilitates automatic consistency checking.
  - *Specification:* The ontology can assist the process of identifying requirements and defining the specification for a software system.

Guarino (1997) distinguished four kinds of ontologies according to their dependence level on a task or a particular point of view (see Figure 2.1):

- *Top-level ontologies*: They describe very general, and thus domain independent, concepts such as time, space, matter, etc. The other kinds of ontologies may be built upon top-level ontologies.

- *Domain ontologies*: They describe concepts related to a particular domain such as medicine.

- *Task ontologies*: They describe the vocabulary related to a general task or activity.

- *Application ontologies*: Application ontologies describe vocabulary related to both a particular domain and a particular task.

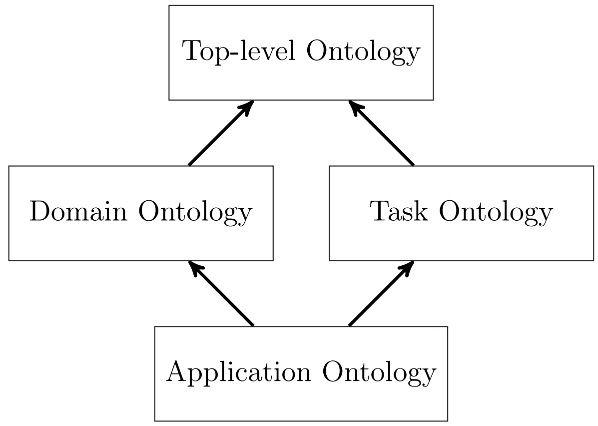


Figure 2.1 – Kinds of Ontologies (from Guarino, 1997)

Mizoguchi and Bourdeau (2000) advocate for the ontology-based development of TSLSs. They proposed the use of domain ontologies to model each individual domain and task ontologies to model pedagogy.

According to Sowa (2010), ontologies can be defined in three different ways:

- **Axiom-based definition**: Categories are distinguished by axioms and definitions.
- **Taxonomic-relation based definition**: Categories are defined based on *isA* relationships.
- **Prototype-based definition**: Categories are defined by typical instances. Semantic distances are used to compute the similarity between two instances.

Next, some applications of ontologies on TSLS are described.

2.1.1 Ontologies as a Means to Represent the Domain Module

Murray (1996) proposed the use of ontologies for the explicit representation of curricular and pedagogical knowledge for learner guidance in the learning process. The EON authoring tool (Murray, 1998, 2003) relies on the *topic ontology* as the basis for the domain module. The *topic ontology* defines the kinds of topics, the kinds of links that can be defined between topics, the topic levels, the topic properties, etc.

Martin and Mitrovic (2003) tested that relying on an existing domain model might ease the development of a new similar domain model for Constraint-based tutors (Ohlsson, 1993, 1994; Mitrovic *et al.*, 2007). It has been proved that a domain ontology can assist the manual composition of a constraint base for the domain that it describes (Suraweera *et al.*, 2004a) and that the ontology can even be used for the automatic acquisition of constraints (Suraweera *et al.*, 2004b; Martin *et al.*, 2008).

Cassel *et al.* (2008) have defined the ACM's Computer Ontology, which aims at representing Curriculum Representations and, thus, it might be the basis for the development of programs of study.

2.1.2 Ontologies as a Means to Achieve Interoperability Between Technology Supported Learning Systems

The integration of heterogeneous information sources is achievable using ontologies (Wache *et al.*, 2001). In order to attain semantic interoperability, the meaning of the information that is interchanged has to be understood by the systems involved. The use of ontologies for the representation of implicit and hidden knowledge is a possible approach to overcome the problem of semantic heterogeneity. Uschold and Gruninger (1996) mention interoperability as a key application of ontologies. Many approaches have relied on the use of ontologies to enable the integration of information, and, thus, achieve the interoperability between TSLSs.

The OntoAIMS (Denaux *et al.*, 2004, 2005a,c) adaptive information management system has been improved by integrating it with OWL-OLM (Denaux *et al.*, 2005b). Both OntoAIMS and OWL-OLM, which were developed as separate systems, represent the domain models using ontologies and model the user knowledge as ontology overlays. Using a shared ontology to model the domain module enables the exchange of the learners' model, which can be consistently interpreted by either of the two systems.

However, in most cases, the systems to integrate do not share the same ontology. Each system can use its own local ontology or representation. Integration in these cases is not so straightforward.

Dicheva and Aroyo (2004a) proposed an ontology-centered architecture for the integration of several TSLSs that supports the communication among the systems, allowing them to share and exchange information. The core of this architecture is the *Communication Ontology*, which defines the semantics of the communication. The communication is modeled in two layers: the *Interaction Protocol Ontology* describes the *message layer* specifying the terms related to message types, reasons, and preconditions; the *content layer* is specified in terms of the *Communication Content Ontology*. The latter has two parts: the domain-independent *Concept-based Informa-*

tion Model ontology, that includes the basic terms describing the information model (e.g.: *concept*, *relation type*, etc.), and the *Domain Ontology*, which represents the knowledge the systems have about the domain. As each TSLS uses its own ontology, it is responsible for mapping its internal ontology to the common *Domain Ontology* to enable knowledge sharing. A simplified version of this general architecture (Dicheva and Aroyo, 2004b; Aroyo and Dicheva, 2004) has been applied to integrate two systems, the previously mentioned OntoAIMS (Denaux *et al.*, 2004, 2005a,c) and TM4L (Dicheva *et al.*, 2004).

Mitrovic and Devedžić (2002) introduced M-OBLIGE, an architecture for the integration of several TSLSs. In this approach, each TSLS relies on its local ontology, which in turn refers to shared reference ontologies (see Figure 2.2). Sosnovsky *et al.* (2008b) integrated SQL-Tutor (Mitrovic, 1998; Mitrovic and Ohlsson, 1999; Mitrovic, 2003) and SQL-Guide (Sosnovsky *et al.*, 2008a; Brusilovsky *et al.*, 2008). These systems use different domain models and student modeling techniques. The SQL ontology has been used as a means to achieve the semantic integration of these systems, combining manual mappings and ontological inference.

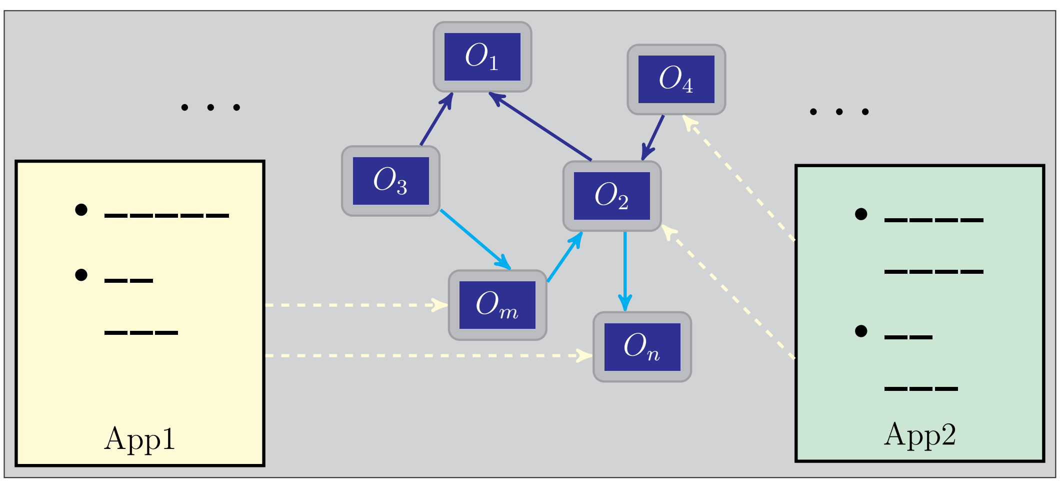


Figure 2.2 – Two Web-Based Intelligent Tutoring Systems Sharing Ontologies (from Mitrovic and Devedžić, 2002)

2.1.3 Other Uses of Ontologies in Technology Supported Learning Systems

Chen *et al.* (1998) developed an ontology-based intelligent authoring tool for Smart-Trainer, a computer based training system for substation operators in the electric power network.

Bourdeau *et al.* (2004) presented a theory-aware authoring environment. In this environment, instructional theories are described by ontologies so that every instructional designer can choose the theory that best fits in his/her system.

Ontologies can be used to build reusable and scrutable student models (Kay, 1999; Kay and Lum, 2004). A student model is reusable if it can be used by various applications. This kind of student model can be achieved, for example, by maintaining a database of models which is accessed by several applications. All these applications rely on an agreed ontology and representation so that the model can be understood by every system. Student model scrutability refers to the possibility of inspecting the model in order to control and even modify it. This scrutability is partially supported by an explanation subsystem. The explanations provided by this subsystem depend upon the student's level of knowledge and his/her preferences, which are described in the student model in terms of the ontology.

2.2 Ontology Learning - Semi-automatic Development of Ontologies

Ontology Learning (OL) refers to the application of a set of methods and techniques used to enable (semi-)automatic construction of ontologies from scratch. Different kinds of information sources, e.g., knowledge bases, corpora or web pages, can be used in OL.

OL entails several tasks: identifying terms that refer to domain topics, identifying synonyms among terms, relating terms to domain topics, identifying taxonomic relations between domain topics, gathering non-taxonomic relations and rule extraction (see Figure 2.3).

The following subsections provide a more detailed description of the most common approaches for term extraction, identification of synonyms, concept extraction, identification of relationships and rules.

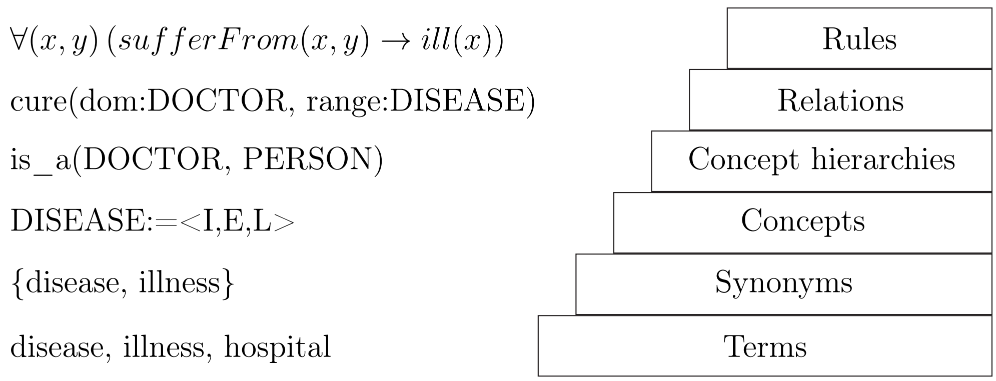


Figure 2.3 – Ontology Learning Layer Cake (from Buitelaar *et al.*, 2005a)

2.2.1 Term Extraction

Terms are linguistic realisations of domain-specific concepts (Sager *et al.*, 1980; Frantzi *et al.*, 1998) and are the basis for the other OL tasks. Term extraction combines NLP and information retrieval techniques. In most cases, term extraction entails running a *part-of-speech* tagger on the documents and, then, identifying a set of plausible terms based on some patterns (e.g., noun phrases, adjective-noun, proper names, . . .) and filter stop words, i.e., common words that do not carry any domain knowledge (e.g., *it*, *the*, . . .).

After that, candidate terms are weighted using some measure in order to determine their relevance or specificity in the domain and, thus, filter non appropriate terms. The term weighting methods vary from the simple Relative Term Frequency (RTF) method to methods such as Term Frequency - Inverse Document Frequency (TFIDF) (Salton and Buckley, 1988). While the first method measures the appearances of the candidate terms in the document, the second one also takes into account the frequency of the candidate terms in documents corresponding to other domains in order to measure the domain specificity. In addition, the C-value/NC-value method (Frantzi *et al.*, 1998, 2000) evaluates the nested terms to determine which terms are relevant for the domain.

2.2.2 Synonym Finding

The synonym level addresses the acquisition of semantic term variants. Some approaches also deal with the acquisition of *term translations*, i.e., equivalent terms in other languages. Much of the work in this area has focused on the integration of WordNet (Fellbaum, 1998) for the acquisition of English synonyms and EuroWord-

Net (Vossen, 1998) for bilingual or multilingual synonyms and term translations. Clustering techniques are also used to identify terms with similar meanings, based on Harris' distributional hypothesis (Harris, 1954), which states that words that occur in the same contexts tend to have similar meanings.

2.2.3 Concept Extraction

Although most of the research on concept extraction has been oriented to clustering related terms, which almost completely overlaps with term and synonyms extraction, Buitelaar *et al.* (2005a) states that concept instances and concept definitions should also be provided in this step. For example, the KnowItAll system (Etzioni *et al.*, 2004) aims to identify concept instances (e.g., movie actors) from the web. OntoLearn (Navigli and Velardi, 2004; Velardi *et al.*, 2005) searches for concept glosses, i.e., brief notations of the meaning of the term in a text, in the web.

2.2.4 Extraction of Taxonomic Relations

Taxonomic relationships (*isA* relationships) are essential for building domain ontologies, as they enable inheritance between concepts and automated reasoning (Corcho and Gómez-Pérez, 2000). The identification of such relationships is carried out in three different ways: (1) using syntactic patterns, (2) using clustering methods, or (3) employing subsumption based methods.

Next, the three ways for extracting taxonomic relations are described in more detail.

2.2.4.1 Pattern-Based Relation Extraction

This approach relies on the analysis of frequently used syntactic structures or patterns to identify underlying *isA* relationships. The most popular means is the use of Hearst's patterns (Hearst, 1992). They have been used in many ontology learning approaches to identify taxonomic relationships (*hyperonimes/hyponimes*) from syntactic structures like NP_0 *such as* NP_1, NP_2,... *(and/or)* NP_n, where NP_i refers to a noun-phrase corresponding to a domain topic. Text2Onto (Cimiano and Völker, 2005) uses these patterns to gather taxonomic relationships. OntoLT (Buitelaar *et al.*, 2004), a Protégé[1] plug-in for the extraction of ontologies from text, uses a pattern that identifies taxonomic relationships on nested terms, relying on the appearance of a term and some modifiers (*Genus et Differentiam*).

[1] http://protege.stanford.edu/

2.2.4.2 Clustering-Based Relation Extraction

The distributional hypothesis (Harris, 1954) states that terms that appear in similar contexts tend to be similar. The approaches that rely on the distributional hypothesis tend to build hierarchical clusters of terms, either in a bottom-up approach, where similar clusters are grouped in a more general cluster, or in a top-down approach, where clusters are refined into simpler clusters (Maedche, 2002). The ontologies gathered following this approach are, according to Sowa's classification, *prototype-based* ontologies (Sowa, 2010).

This approach has been followed by (Grefenstette, 1994; Cimiano *et al.*, 2005), and it has been used in OL systems such as ASIUM (Faure and Nédellec, 1998).

2.2.4.3 Subsumption-Based Relation Extraction

Term subsumption is a particular kind of term co-occurrence that is defined in the following way. Given two terms x and y, x is said to *subsume* y if the documents in which y occurs are a subset of the documents in which x appears – $P(x|y) = 1$ and $P(y|x) < 1$ – (Sanderson and Croft, 1999).

2.2.5 Extraction of Non-Taxonomic Relations

There is an agreement in the NLP community that relational information is, at sentence level, typically conveyed by verbs, while, in noun phrases this role is mainly played by prepositions. Prepositions only cover a limited set of domain-neutral relations such as parthood (*partOf*). Maedche and Staab (2000) used association rules to gather this kind of relationships. Text2Onto (Cimiano and Völker, 2005; Haase and Völker, 2008) identifies *verb-concept-concept* triples that are labelled, based on the verb, and ranked so that they can finally be reviewed by knowledge engineers so as to gather the relationships (Kavalec and Svátek, 2005).

2.2.6 Axiom and Rule Extraction

The extraction of rules and axioms is, by far, the least addressed task. It aims to gather knowledge that may enable reasoning and inferring new knowledge. Most works have been oriented to question answering (Lin and Pantel, 2001; Dagan *et al.*, 2005).

2.3 Achieving Interoperability through the Use of Standards and Specifications

The broad use of the Internet and its technological capabilities has brought about the spread of TSLSs. Nowadays, most universities, colleges and even schools make intensive use of LMSs or E-Learning platforms (Waits and Lewis, 2003; Parsad and Lewis, 2008). Initially, those TSLSs were developed *ad-hoc* to meet the requirements of a particular institution, and information exchange between those systems could hardly be achieved. However, reusing learning content could significantly reduce the cost of the development of new courses (Downes, 2001).

Nowadays, Learning Objects (LOs) are considered the core notion for learning content. The IEEE Learning Technology Standards Committee (LTSC) defines a Learning Object (LO) as "*any entity, digital or non-digital, which can be used, reused or referenced during technology supported learning*" (LTSC, 2001). However, as Wiley (2000) states, this definition may be too vague as almost everything matches it, i.e., the notes teachers use for their classes may be considered LOs since they can be referenced during the learning process, even though its reusability in an application is quite limited. Wiley (2000) instead recommends considering LOs as "*any digital resource that can be reused to support learning*".

LOs provide a means to facilitate knowledge reuse as they are "*reusable pieces of educational material intended to be strung together to form larger educational units such as activities, lessons or whole courses*" (Brooks *et al.*, 2003).

A crucial issue to facilitate the use and reuse of LOs is their availability and the possibility of retrieving the appropriate LO. Metadata, i.e., data that describes the LO, and Learning Object Repositories (LORs), that store and manage LOs and their metadata, enable the possibility of finding and using the appropriate LOs. LORs that only manage metadata and do not store LOs, are also referred to as LO Referatories. Nowadays, there are many available LORs, such as ARIADNE (Duval *et al.*, 2001; Ternier *et al.*, 2009), Merlot (Cafolla, 2006), Edna (Adcock *et al.*, 2000), or Edutella (Nejdl *et al.*, 2002). LORs may contain either domain-specific content or general content.

Many consortia and organisations have been involved in defining standards and specifications that may enable interoperability between TSLSs, allowing them to interact and to share information. The main contributors to this effort are: the LTSC, the IMS Global Learning Consortium (IMS), the Aviation Industry CBT Committee (AICC), the Advance Distributed Learning - US Department of Defense (ADL) initiative, the ARIADNE project, Getting Educational Systems Talking Across Leading

Edge Technologies (GESTALT), PROmoting Multimedia access to Education and Training in EUropean Society (PROMETEUS), the European Committee for Standardization (CEN), and the Gateway to Educational Materials (GEM) project. The standards and specifications can be classified in two main groups: standards and specifications for learning content, which focus on the description and representation of learning resources, and standards and specifications that enable the communication between several TSLSs and LORs. This section describes the most representative standards and specifications and, finally, presents ARIADNE and GLOBE which rely on the use of the described specifications.

2.3.1 Standards and Specifications for Learning Content

There are several aspects concerning learning content that are covered by different standards or specifications: the IEEE Learning Object Metadata (LOM) copes with the description of learning resources that enables the search and retrieval of such LOs. IMS Question and Test Interoperability (QTI) addresses the representation of tests and question based exercises. The structure of composite learning resources or LOs, considering both the sequencing and distribution, are considered by specifications such as IMS Simple Sequencing (SS) and IMS Content Packaging (CP). The next subsections present the standards and specifications referring to learning content, specifications for describing learning/teaching strategies and some reference models for TSLSs.

2.3.1.1 IEEE LOM - Metadata for Learning Resources

The capability of searching, evaluating and acquiring LOs is essential for the use (and the reuse) of LOs. Metadata, i.e., descriptive data about educational data and resources, allows cataloging the LOs so that they can be searched and retrieved. Therefore, the metadata has been one of the main focuses for the learning technology standardisation community during the last few years. The LTSC defined the IEEE Learning Object Metadata (LOM) standard (LTSC, 2001), which is based on the proposals of the IMS and the ARIADNE project. The LOM defines the syntax and semantics of metadata for LOs. These include properties to describe general information about the resource, terms of distribution, technical requirements, teaching or interaction style, etc. The LOM elements are classified into the following categories:

- **General**: information that describes the LO as a whole.

- **Lifecycle**: information about the current state and the historical evolution of the LO.
- **Meta-metadata**: information about the LOM instance.
- **Technical**: technical characteristics and requirements of the LO.
- **Educational**: educational and pedagogical characteristics of the LO.
- **Rights**: intellectual property rights and the conditions of use of the LO.
- **Relation**: relationship(s) between the described LO and other related LOs.
- **Annotation**: comments on the educational use of the LO, including when and by whom those comments were created.
- **Classification**: describes the LO according to a particular classification system.

The LOM also specifies how some of the metadata elements relate to the Dublin Core (DC) metadata standard (Dublin Core Metadata Initiative, 2004), which is aimed to describe any kind of networked resource.

2.3.1.2 IMS QTI: a Specification for Representing Content in Test and Question-Based Exercises

The IMS Question and Test Interoperability (QTI) specification (IMS Global Learning Consortium Inc., 2008) addresses the need to share assessment tools, such as test and question-based exercises. QTI describes a data model for the representation of questions, test data and the reports of their results. Although the data model is described abstractly, a binding to XML is provided and recommended.

2.3.1.3 IMS Simple Sequencing

The IMS Simple Sequencing (SS) specification (IMS Global Learning Consortium Inc., 2003b) defines a method for representing the intended behaviour of an authored learning experience so that any TSLS can sequence discrete learning activities, i.e., instructional event, appropriately. Content in SS is organised into a hierarchical structure. Each activity may include one or more child activities. Every activity has an associated set of sequencing behaviours, defined by the sequencing definition model, that describes how the parent activity or its children are used to create the desired learning experience. Using this specification, the instructional designer can declare the relative order in which the elements of the content will be presented to the learner, and the conditions under which a piece of the content will be selected, delivered or skipped. This specification includes rules that describe the flow and branching of learning activities according to the outcomes of the interactions of the learner with the content.

The specification is labelled as *simple* because it includes a limited number of widely used sequencing approaches, but does not (so far) address Artificial Intelligence (AI)-based sequencing, sequencing requiring data from closed external systems, collaborative learning or synchronisation between multiple parallel learning activities.

2.3.1.4 IMS Content Packaging

The IMS Content Packaging (CP) Information Model (IMS Global Learning Consortium Inc., 2007) describes a set of data structures that allow software developers and implementers to create instructional materials that can be exchanged between different systems - authoring tools, LMSs, etc. - aiming at importing, exporting, aggregating or disaggregating packages of learning content.

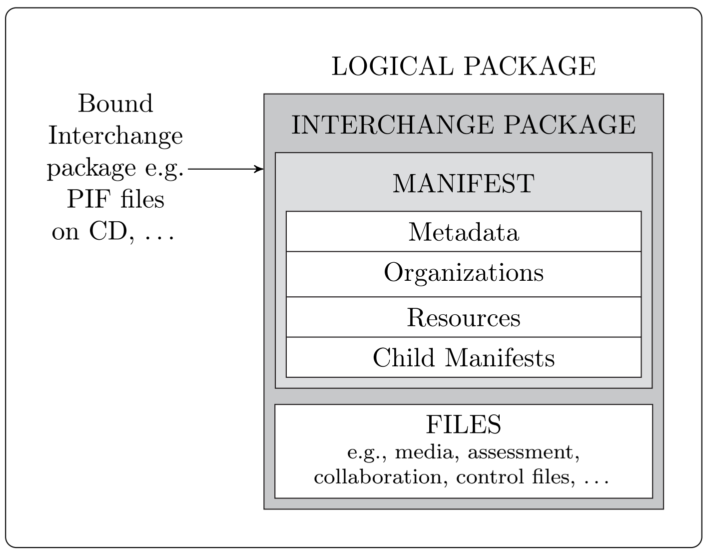


Figure 2.4 – IMS Content Packaging Conceptual Model (from IMS Global Learning Consortium Inc., 2007)

Figure 2.4 illustrates the structure of the CP information model. The *logical package* represents a unit of reusable content. It encompasses the whole set of components described by the *manifest* (including the remote referenced components). The *interchange package* is a logical directory that contains the set of components that are to be exchanged between systems, including the manifest and all the required files. The *manifest* is the component that describes the *logical package*. It may refer to remote elements and it may contain *child manifests*, each of them describing a *logical package* included in the content package. The *metadata* element contains descriptive information about the content packages, logical organisations, content or files. The *organizations*, which are specified using the SS (*cf.* Section 2.3.1.3), describe the sequential relationships among the units of the content. The *resources* element contains an inventory of the files used by the parent *manifest*. Besides, a set of files that embodies the content described by the logical package is included in the content package.

2.3.1.5 IMS Learning Design

The term *Learning Design* might be defined as "*the description of the teaching-learning process that takes place in a unit of learning (e.g., a course, a lesson or any other designed learning event)*" (Koper, 2006). The IMS Learning Design (LD) specification aims to provide a means to describe any design of teaching-learning process in a formal way (IMS Global Learning Consortium Inc., 2003a). A learning design is a description of a method enabling learners to achieve certain learning objectives by performing certain activities in a certain order in the context of a certain learning environment. LD allows specifying roles, learning activities, personalisation and adaptation issues. LD allows many people with different roles, and can be applied either for e-learning or blended learning. LD is not linked to a particular learning theory, so it can be used to described any learning process. In order to cope with the description of learning pocesses, LD refers to other specifications such as LOM (*cf.* Section 2.3.1.1), CP (*cf.* Section 2.3.1.4), QTI (*cf.* Section 2.3.1.2) or SS (*cf.* Section 2.3.1.3).

2.3.1.6 SCORM

The Sharable Content Object Reference Model (SCORM) (ADL, 2009a,b,c) is neither a standard nor a specification. Instead, it is a reference model for developing interoperable TSLSs that combines elements from the LTSC and IMS standards and specifications. The SCORM reference model comprises three major specifications:

- The SCORM Content Aggregation Model (ADL, 2009a) describes the components used in a learning session, how to package those components for exchange between systems, how to describe components to enable search and discovery, and how to define sequencing rules for these components. Figure 2.5 shows the UML representation of the SCORM Content Aggregation Model. The identified kind of learning components are:
  - **Asset**: Elementary block, such as text, images, sound, assessment object, etc., that can be used to build a learning resource.

- **Sharable Content Object (SCO)**: It represents the lowest level of granularity of a learning resource that can be tracked by an LMS using the SCORM Run-Time Environment data model. A SCO consists of one or more assets. The main difference between an asset and a SCO is that the latter communicates with a LMS using the IEEE ECMAScript Application Programming Interface for Content to Runtime-Services Communication Standard (LTSC, 2003).
- **Activity**: a Learning Activity can be described as a unit of instruction. Activities might be simple, i.e., entail a single learning resource (asset or SCO), or composed of several sub-activities.
- **Content Organization**: A Content Organization is a representation or map that defines the intended use of the content through structured units of instruction (activities). It describes how activities relate to one another. The sequencing of the activities is defined as part of the content organization.
- **Content Aggregation**: The Content Aggregation can be used to describe the composition of a set of functionally related content objects to be able to transfer them between systems.

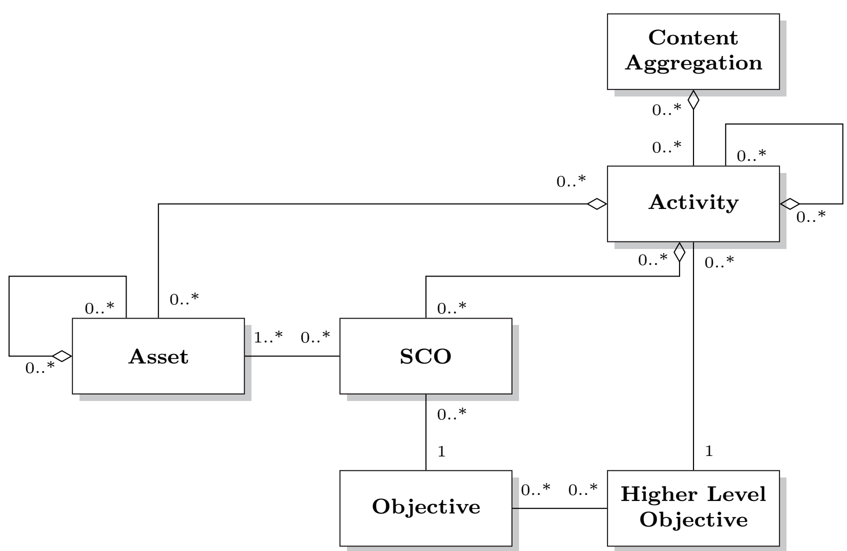


Figure 2.5 – UML Representation of the SCORM Content Aggregation Model (from Verbert, 2008)

- The SCORM Sequencing and Navigation (ADL, 2009c) book describes how SCORM-compliant content may be delivered to learners through a set of events initiated either by the learner or the system.
- The SCORM Run-Time Environment (ADL, 2009b) document describes the requirements for launching content objects, enabling the communication between SCOs and LMSs, and a standardised data model for tracking information relevant to the learner's experience with the content that can be exchanged between the SCOs and LMSs.

2.3.1.7 Common Cartridge

In a similar way to SCORM, IMS has its own proposal, Common Cartridge (IMS Global Learning Consortium Inc., 2011), a set of standards to assure the interoperability of TSLSs. It describes the format to exchange content, how to describe (annotate) the learning contents, etc. It also relies on the above described specifications: LOM, QTI, etc.

2.3.2 Standards and Specifications to Enable Communication between Technology Supported Learning Systems and Learning Object Repositorys

Currently, there are several LORs sharing LOs all around the world. However, those LORs are quite heterogeneous; different approaches to store the metadata have been used (e.g., relational databases, XML repositories). In order to allow searching LOs, those LORs provide query interfaces that allow searching for the desired LOs. In the last few years, LOR networks, such as GLOBE[2] and LORNet[3], are being promoted, which requires that the LORs are able to interact with others. Next, some of the initiatives that make this interoperation possible are described.

2.3.2.1 SQI: a Simple Query Interface

The Simple Query Interface (SQI) (Simon *et al.*, 2005) specification, promoted by the European Committee for Standardization/Information Society Standardisation System/Learning Technologies Workshop (CEN/ISSS/LT), is an Application Programming Interface (API) to enable query-posting and query answering to LORs. SQI does not rely on query languages or metadata standards. It supports different

[2] http://www.globe-info.org/
[3] http://www.lornet.ca/

query languages, such as ProLearn Query Language (PLQL) (Ternier *et al.*, 2008b), or the Query Exchange Language (QEL) (Nejdl *et al.*, 2002), and the vocabulary of LOM, DC or other specifications.

2.3.2.2 OAI-PMH

The Open Archive Initiative (OAI), that develops and promotes interoperability solutions with the aim of facilitating the efficient dissemination of content, defined the Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) (Lagoze and de Sompel, 2001). The OAI-PMH allows *service providers* to collect or harvest the metadata exposed by the *data providers*. Harvesting can be selective, i.e., set-based criteria or date-based criteria can be used to state which collection of metadata is to be collected or to allow incremental harvesting. The mandatory metadata format is DC, although other formats can also be supported.

2.3.3 ARIADNE and GLOBE: an Example of the Use of Standards and Specifications

ARIADNE (Duval *et al.*, 2001; Ternier *et al.*, 2009), which stands for *Association of Remote Instructional Authoring and Distribution Networks for Europe*, is a foundation that aims to promote the sharing and reusing of LOs. The core of the ARIADNE infrastructure is a distributed network of LOs, which relies on standards for distributed digital resource management in order to enable interoperability (Ternier and Duval, 2006). The ARIADNE repository supports the storage of LOs and LOM instances. LOs are described using the IEEE LOM standard (LTSC, 2001). The search interface of the repository is built on the SQI specification (Simon *et al.*, 2005). The publishing interface is based on the Simple Publishing Interface (SPI) specification (Ternier *et al.*, 2008a). The harvester collects metadata from external repositories in order to publish it in the ARIADNE repository. The harvester relies on the OAI-PMH (Lagoze and de Sompel, 2001). ARIADNE provides services such as the metadata validation service, which validates metadata against an application profile or Samgi (Meire *et al.*, 2007), an automatic metadata generator.

ARIADNE is part of the Global Learning Objects Brokering Exchange (GLOBE)[4] alliance of educational repositories together with Merlot (McMartin, 2004; Cafolla,

[4] http://www.globe-info.org/

2006), Lornet[5], KERIS[6], and Laclo[7] among others. GLOBE provides a distributed network of LORs built on the IEEE LOM, SQI and OAI-PMH standards. The federated Search Engine allows to query on the whole alliance.

The ARIADNE infrastructure has been adopted to develop networks for particular domains, such as Metadata of Architectural Contents (MACE) [8] and Metadata Ecology for Learning and Teaching (MELT)[9].

2.4 Domain Module Authoring Approaches

Automatic or semi-automatic approaches for developing TSLSs are required to lighten their development cost (Murray, 1999). This section presents some efforts aimed at covering diverse aspects of the TSLS development, from the construction of ITSs to the generation of reusable learning material.

2.4.1 KONGZI

KONGZI[10] (Lu *et al.*, 1995) is an authoring tool that was developed with the aim of automating the generation of ITSs. KONGZI was, probably, one of the first authoring tools that attempted to build ITSs automatically from documents. Later, KONGZI was enhanced to support the use of multimedia resources in the generated ITSs (Chen *et al.*, 1997). KONGZI gathers the information from a EBKDL+ program, i.e., a machine readable representation of the textbook using the formal Education Book Description Language (EBKDL+). An EBKDL+ program represents the book content as a sequence of chapters, either elementary chapters or complex chapters, that contain other chapters. Elementary chapters entail a sequence of sections, which have different learning goals (e.g., causes or effects of an event, or a classification of a concept).

Figure 2.6 shows the ITS construction process in KONGZI. The *Understanding System* extracts the knowledge from the EKBDL+ program and builds the *Domain Knowledge Base*, which contains the knowledge to be transferred to the student. The domain knowledge is represented by means of a semantic network. As it might not be appropriately organised for learning, the *Lesson Design* module processes the

[5] http://www.lornet.ca
[6] http://english.keris.or.kr
[7] http://www.lacklo.org
[8] http://www.mace-project.eu
[9] http://info.melt-project.eu
[10] Kǒng Zǐ is the name of the Chinese philosopher known as Confucius

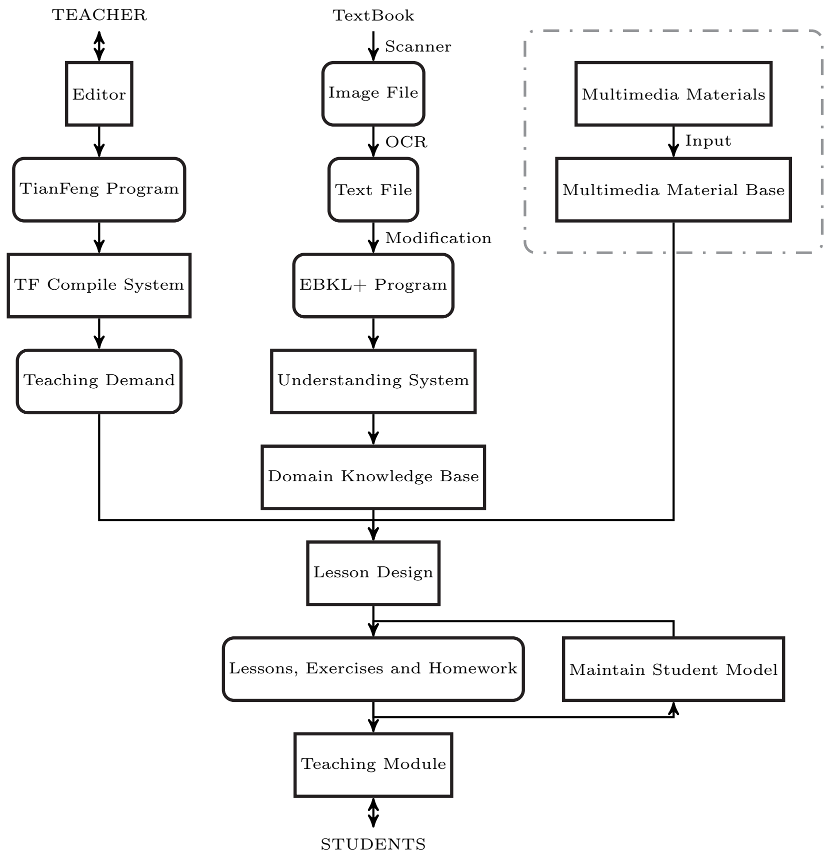


Figure 2.6 – Architecture of the Enhanced KONGZI (from Chen *et al.*, 1997)

Domain Knowledge base according to the *Teaching Demand*, i.e., the specification of the ITS, in order to organise the content in an appropriate lesson-oriented way. The content is enhanced with multimedia content obtained from the *Multimedia Material Base*. The *Teaching Demand*, which has been specified by the teacher, defines settings of the ITS including visual aspects and pedagogical strategies that are represented as rules.

KONGZI can automatically produce exercises and tests, whose solutions can be automatically assessed, for learner evaluation. It uses some heuristics to automatically produce the exercises. One of these heuristics consists of lining two or more concepts with a non-existing relationship and asking the students to point out the mistake. The *Student Model* is updated according to his or her performance, and it is used to plan the learning sessions.

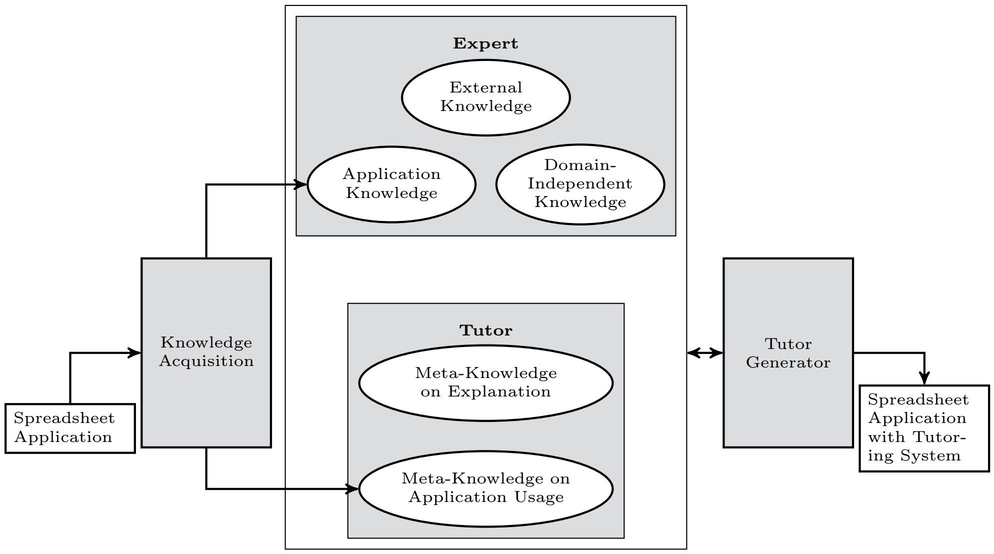


Figure 2.7 – System Architecture for the Automatic Generation of Tutors for Spreadsheet Applications (from Lentini *et al.*, 2000)

2.4.2 Generating Intelligent Tutoring Systems from Spreadsheets

Lentini *et al.* (1995, 2000) developed a system for automatic knowledge acquisition and tutor generation for spreadsheet applications. The system processes existing spreadsheets to extract the knowledge and improve the spreadsheet application with tutoring facilities. The architecture of the system, shown in Figure 2.7, relies on the classic architecture for ITSs. The *Expert Module* consists of the *Domain-Independent Knowledge*, the *Application Knowledge*, and the *External Knowledge*. The domain-independent knowledge is represented by general rules that define mathematical properties, i.e., definitions of operators/functions and simplification criteria for arithmetic expressions based on their properties (commutativity, distributivity and associativity). The application knowledge is extracted from the spreadsheet, and the external knowledge, which consists of information about the application domain, can be specified by the instructional designer during the acquisition process.

The generation of tutors consists of two stages: *Acquisition of the Knowledge* from the spreadsheet application, and the *Generation of the Tutoring Facilities*. Knowledge acquisition is performed in two steps. First, the application knowledge is gathered from the spreadsheet reconstructing the mathematical model coded into the spreadsheet scheme. The application knowledge is represented by a dependency graph, a directed acyclic graph. Next, the structure of the spreadsheet and the application knowledge are used to build the *Meta-knowledge on Application Usage*, a partition of the sheet into pieces that can be regarded as separate components of the overall scheme. This information is used by the *Tutor Generator Module* to enhance the processed spreadsheet with two kinds of tutoring support: a hypertext guide that describes the mathematical model coded in the spreadsheet, and an interactive tutor that supervises the end-user's activity.

2.4.3 IMAT

IMAT (de Hoog *et al.*, 1999) aimed at promoting the reuse of technical manuals, usually available in paper-based or electronic form, for Computer-Based Training (CBT). The training of maintenance is usually not part of the public curriculum, and therefore it is not an attractive market for educational publishers, which makes technical documents the only available source of information. However, technical documentation is designed for reference purposes but not for educational purposes, so it has to be revised to produce material for training purposes.

This project dealt with three domains: aircraft maintenance, car repair and programming and configuration of traffic control equipment. Technical manuals in these domains generally describe a system from a physical and functional point of view, and they often are structured in different parts: general description, part list, wiring diagrams, and operation manuals.

IMAT provides a set of tools to process the technical documents. The *Document Analysis Tool* breaks up the document, or its selected parts, into small parts or fragments, and indexes these fragments to facilitate their retrieval. The segmentation of the document relies on the original structure of the document (arrangement of chapters, sections, and paragraphs). To enable storage and retrieval, the document analysis is required to identify additional properties of the fragments such as the subject described in the fragment, the format of the fragment, and the way the information is represented (e.g., a list of parts or steps in a procedure).

Fragments are indexed using ontologies (Kabel *et al.*, 1999; Worring *et al.*, 2001; Kabel *et al.*, 2004b,a). Indexing the documents automatically relying on the use of ontologies assures the consistency of the annotation (Kabel *et al.*, 2004a). Technical manuals are highly structured, adhere to style rules, and have a limited vocabulary, which makes this task feasible (Worring *et al.*, 2001). The retrieved fragments can be copied&pasted into the authoring environment chosen by the author.

2.4.4 ALOCOM: a Disaggregation Framework

Although many efforts have been made to promote reuse, authors tend to reuse parts of LOs by copy&paste actions. Providing *on-the-fly* access to the components of the LOs, and supporting the partially automatic re-composition of LOs, can significantly enhance content authoring. The ALOCOM framework (Verbert *et al.*, 2008; Verbert, 2008) was developed to decompose composite LOs and make those components available for *on-the-fly* content reuse. LO content models define different kinds of LOs at different levels of granularity, which affects authoring deployment and repurposing (Duval and Hodgins, 2002), and rely on the assumption that independent and self-contained material can be created and that they can be used either alone or dynamically assembled to provide *just-in-time* learning. In order to deal with the existence of several different models and facilitate content model interoperability, the *Abstract Learning Object Content Model* (ALOCOM) was defined (Verbert and Duval, 2004; Verbert *et al.*, 2005). The ALOCOM model is represented by a global ontology that covers the content models, the local ontologies corresponding to each model, and the mappings between the ontologies.

The ALOCOM framework (Verbert *et al.*, 2008) transforms documents (e.g., Powerpoint presentations, Wikipedia Pages and SCORM Content Packages) into representations compliant to the ALOCOM model. In this transformation process, the framework decomposes LOs into content components that can be accessed and, therefore, reused in new LOs. To facilitate content reuse, the metadata for the decomposed content components is automatically generated by Samgi (Meire *et al.*, 2007). Content inclusion is controlled to avoid duplicates. Figure 2.8 illustrates the ALOCOM framework.

The client side applications enable content uploading to and component retrieval from the repository, based on the ARIADNE Knowledge Pool System (Duval *et al.*, 2001; Ternier *et al.*, 2009). The Disaggregation module is responsible for the decomposition of the LOs. The smallest granularity level for text components is the paragraph, as sentence fragments or words might not be appropriate for reuse (Rockley, 2002). Paragraphs may contain definitions, examples and references, which can be identified using patterns corresponding to broadly used syntactic structures. The decomposed content components are categorised according to their format (text, image, etc.), kind of content and structure. For every component, a preview thumbnail is also generated. The generated components are stored in the repository through the *Advanced Content Inserter*. The *Reuse Detector* determines if an analysed component is already stored in the repository, avoiding duplicates. The *Query service* enables the retrieval of components. Queries can be performed by searching for

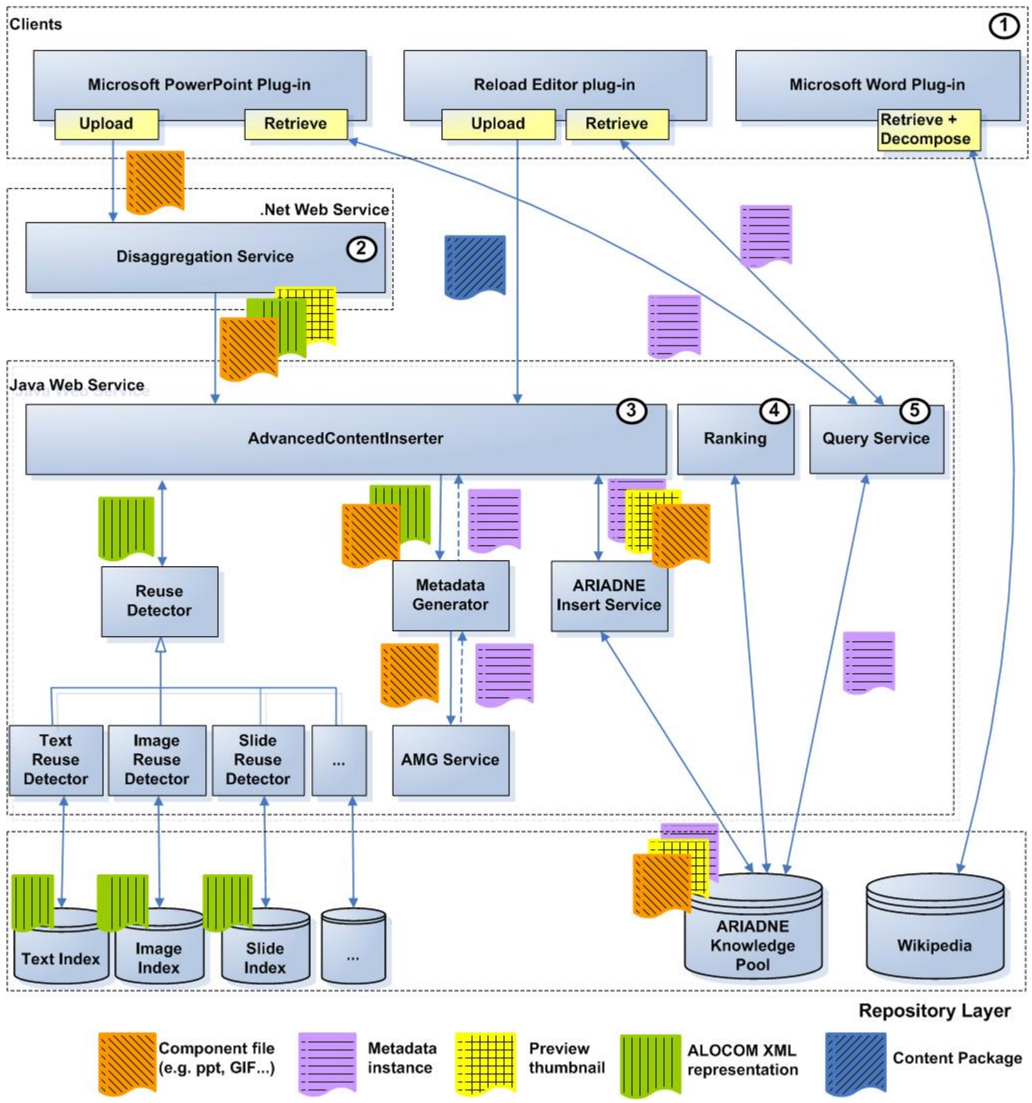


Figure 2.8 – The ALOCOM Framework (from Verbert, 2008)

descriptive keywords or component type, although more advanced queries are also supported. The *Ranking Module* assigns ordering values to components based on their reuse and enables ranking of components in a result list when a user searches for relevant objects, placing components with a high relevancy at the top of the list.

2.4.5 The Knowledge Puzzle Project - From Learning Objects to Learning Knowledge Objects

The Knowledge Puzzle Project is a framework which automatically composes instructional resources to fulfill a specific competence need *just-in-time* (Zouaq and Nkambou, 2009a). Wiley (2000) claims that LOs "*instructional design theory must be incorporated in any learning object implementation that aspires to facilitate learning*". To get such instructional theory-aware LOs, the knowledge representations used by ITSs have been combined with the LOs to obtain the so-called Learning Knowledge Objects (LKOs), i.e., active, independent and theory-aware LOs that can be considered tiny ITSs.

The core of the Knowledge Puzzle Project is the *Organizational Memory (OM)* a pool of knowledge in which LKOs can be retrieved through dynamic aggregation. The OM is composed of four layers:

- **Document Pool**: The set of documents that can be used to build the aggregated LKOs. Currently, only plain text documents, all of which belong to the same domain, can be automatically processed.

- **Ontology Layer**: This is the pillar of the OM, and contains a set of ontologies needed to index the resources and to dynamically compose the new LKOs. The ontologies are:

  - **Document Ontology**: Essential terms to describe a LO in terms of its structure (paragraph, sentences, figures, tables, etc.).
  - **Instructional Role Ontology**: The instructional roles (definition, introduction, etc.) corresponding to the LOs.
  - **Competence Ontology**: The kinds of skills on domain concepts. It is based on Bloom's Learning Objectives Taxonomy (Bloom *et al.*, 1956).
  - **Instructional Learning Theory Ontology**: Pedagogical theoretical knowledge that guides the dynamic composition of the LKOs. In this ontology, the instructional theories are defined as a sequence of instructional events by means of Semantic Web Rule Language (SWRL) (Horrocks *et al.*, 2004) rules. The curent instructional events are derived from theories of education such as Gagné's Theory of Instruction (Gagné *et al.*, 2004) or Merrill's Instructional Transaction Theory (Merrill, 1999). The Instructional Learning Theory Ontology contains a simple representation of the pedagogical strategies, but Zouaq and Nkambou (2009a) claim that

a real instructional theory ontology, such as OMNIBUS (Hayashi *et al.*, 2009), could be used instead to guide the composition process.

- **Domain Ontology**: This models the domain knowledge.

- **Resource Layer**: The Resource Layer stores the skills, the competences and the instructional theories, which are instances of the elements defined by the ontologies included in the Ontology Layer. The assets obtained through the semantic annotation, which describe the resources according to the content, structure and the pedagogical role, are also stored in the resource layer. The pedagogical role is manually annotated using the Knowledge Annotator.

- **Rule Layer**: This represents procedural knowledge that interacts with different ontologies at the Ontology Layer. The procedural knowledge is expressed by means of SWRL rules.

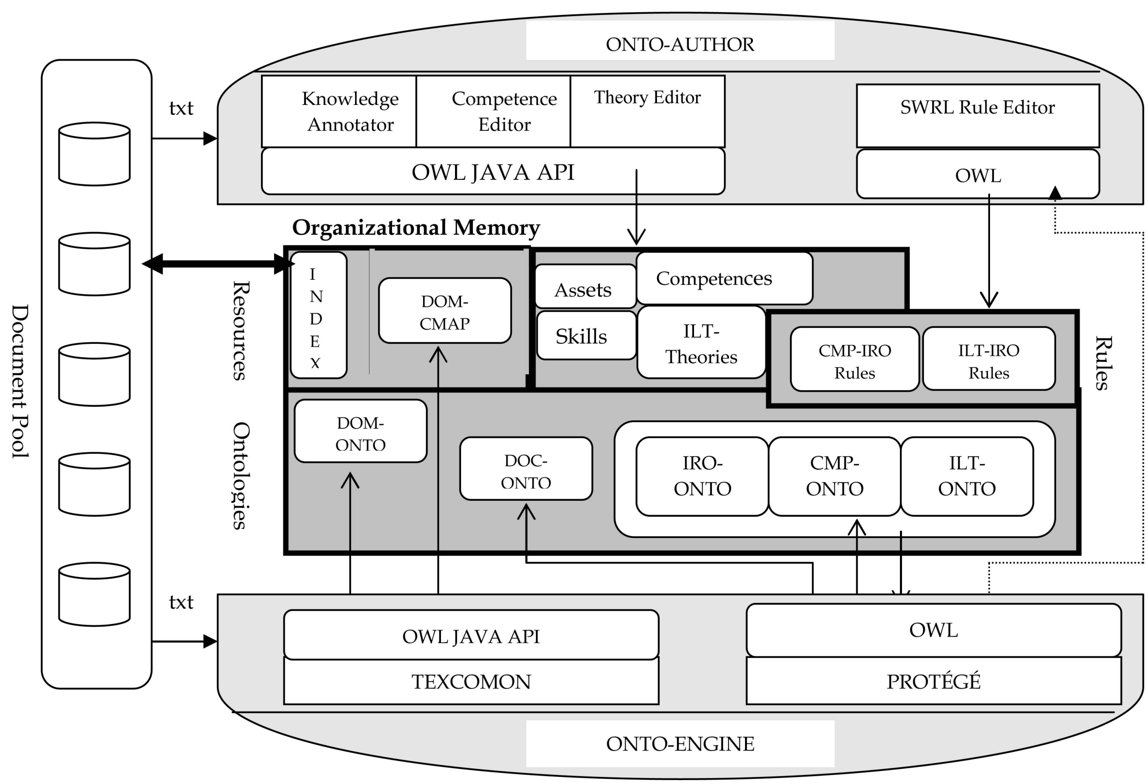


Figure 2.9 – Architecture of the Knowledge Puzzle (from Zouaq and Nkambou, 2009a)

The Knowledge Puzzle architecture, shown in Figure 2.9, includes two subsystems: the *Production Subsystem*, which enables the constitution of the components of the OM, and the *Exploitation Subsystem*, which dynamically composes LKOs from the OM.

The *Production Subsystem* is made up of two major tool suites, ONTO-ENGINE and ONTO-AUTHOR. The first, ONTO-ENGINE, is composed of TEXCOMON[11], which builds the domain ontology from the set of texts in the Document Pool (Zouaq and Nkambou, 2008) using the Protégé Ontology Environment. Both tools communicate through the Protégé OWL Java API. TEXCOMON applies text-mining techniques to build the domain ontology. First, the textual documents stored in the document pool are taken as input to build an index structure that is used to decompose each document into paragraphs and sentences and automate structure-related annotation. Next, the keyword extraction algorithm (Frank *et al.*, 1999) is applied to retrieve document keywords and identify key sentences. Then, these sentences are analysed using the Stanford Statistical Parser (Klein and Manning, 2003) to get a grammatical concept map, i.e., a typed dependency network that describes the grammatical structures of the analysed sentences (de Marneffe *et al.*, 2004). The grammatical structures are mined to identify instances of lexicon-syntactic patterns that are stored in the linguistic knowledge base. The lexicon-syntactic patterns are used to identify grammatical subgraphs corresponding to semantic links. Later, all the obtained semantic maps are merged to create a Domain Concept Map, which is finally formalised into the Domain Ontology.

ONTO-AUTHOR is an authoring environment that uses the ontologies to define and extract the resources layer of the OM. The *Instructional Role Annotator* is a semantic annotation tool used for the manual annotation of instructional roles in a LO, which produces the assets that will be used for LKO composition. The *Competence Editor* is used to define competences as skills that are linked to domain ontology concepts. The *SWRL Rule Editor* is used to define the memory rule layer. The *Theory Editor* enables the creation of a learning theory and links it to the set of related instructional events.

The *Exploitation Subsystem* uses the OM to dynamically generate the LKOs that best suit the learners needs. The learner profile, which is represented using IMS-ePortfolio (IMS Global Learning Consortium Inc., 2005), represents the mastered competencies in an overlay model. This profile can be used to identify competence needs. Once the needs are identified, a LKO to cover the knowledge gap is generated.

[11] The name TEXCOMON stands for TEXt-COncept Maps-ONtology, which states the stages of the Domain Ontology building process

The LKOs can be deployed to be run in a SCORM Runtime Environment, in an IMS-LD player or in an ITS.

2.4.6 Arikiturri - Automatic Generation of Exercises from Corpora

ArikIturri (Aldabe, 2011), a system for the automatic generation of exercises, uses NLP techniques to generate evaluation items from text corpora. ArikIturri is multilingual, it supports the generation of exercises in different languages, and has been tested in both Basque and English. ArikIturri supports the following kinds of exercises, which are illustrated with examples in both Basque and English:

- **Fill-in-the-blank**: This kind of exercise requires the student to complete an statement by supplying a brief response. These items entail a sentence with a blank to be completed. Table 2.1 shows an example of this kind of item, where the correct use of the verb is tested.

Table 2.1 – Example of a Fill-in-the-Blank Exercise

Basque	English
Sintomak honakoak . . . : aldarte txarra, estresa eta antsitatea.	*The symptoms . . . : bad-mood, stress and anxiety*

- **Word formation**: This kind of item consists of a sentence with a blank and a word whose form must be altered to fit into the gap. The Table 2.2 shows an item in which the students must find the correct tense and subject for the word *izan* (*to be*).

Table 2.2 – Example of a Word Formation Item

Basque	English
Sintomak hauek . . . (izan): aldarte txarra, estresa eta antsitatea.	*The symptoms . . . (to be): bad-mood, stress and anxiety.*

- **Multiple-choice questions**: These items entail a stem, i.e., a problem-statement that presents a item to be solved, and the set of possible answers (see Table 2.3). In order to build these kind of items, ArikIturri has to generate some *distractors* or incorrect answers (Aldabe and Maritxalar, 2010).

Table 2.3 – Example of a Multi-Choice Question Exercise

Basque	English
Sintomak honakoa hauek . . . : aldarte txarra, estresa eta antsietatea. *a) dira* *b) da* *c) daude*	*The symptoms . . . : bad-mood, stress and anxiety.* *a) are* *b) is* *c) are*[a]

[a] The Basque verbs *egon* (*daude*) and *izan* (*dira*) correspond to the English verb *to be*

- **Error correction questions**: These items consist of sentences with at least one error that students have to correct (see Table 2.4).

Table 2.4 – Example of an Error Correction Question

Basque	English
Sintomak hauek <u>da</u>: aldarte txarra, estresa eta antsitatea.	*The symptoms <u>is</u>: bad-mood, stress and anxiety.*

- **Short answer questions**: This kind of question requires the students to respond to a question by providing a brief text or response (see Table 2.5).

Table 2.5 – Example of a Short Answer Question

Basque	English
***Noiz** egiten dute karroza-desfile ikusgarria?*	***When** is an amazing float parade held?*

2.5 Conclusions

This chapter has presented two approaches aimed at achieving the interoperability among different TSLSs - ontology engineering and the adoption of standards and specifications - as well as some systems that attempt to lighten the construction of TSLSs in different ways. Ontologies are an appropriate formalism for describing the domain knowledge to be learnt, and they enable sharing knowledge among different systems. The standards and specifications described in Section 2.3 also enable the interoperability between TSLSs, as they provide a common vocabulary to describe diverse aspects of the learning process or to describe learning resources. Standards and specifications state how to describe LOs, represent assessment LOs, or structure composite LOs.

Ontologies and standards might also be used together. The specifications and standards are mostly modeled and represented upon XML-Schema, which lacks support to describe all the semantics of the elements of such conceptual models. This issue has been tackled by some researchers that propose the use of ontologies to overcome this problem. Ontologies define the basic terms to describe a domain and, therefore, they can be used for the annotation of the LOs assuring the consistency of their description (Kabel *et al.*, 1999; Worring *et al.*, 2001; Kabel *et al.*, 2004b,a). The *Learning Design Ontology* describes the elements of the IMS Learning Design (LD) (Lama *et al.*, 2005; Amorim *et al.*, 2006a,b) and a set of axioms that constraint the semantics of the LD conceptual model.

The construction of TSLSs and the authoring of learning content is time and effort consuming. Some efforts aimed at dealing with this fact have been presented in this chapter. Two approaches for generating ITSs from electronic documents have been described. While the first one, KONGZI (Lu *et al.*, 1995; Chen *et al.*, 1997), requires the document to be transcribed to a machine readable format using a formal descriptive language, the latter (Lentini *et al.*, 1995, 2000) is aimed at building math ITSs from spreadsheets. Some other approaches, IMAT (de Hoog *et al.*, 1999) and ALOCOM (Verbert *et al.*, 2008; Verbert, 2008), focus on the disaggregation and the reuse of electronic documents for building new learning material. The Knowledge Puzzle project (Zouaq and Nkambou, 2009a) was developed to improve LOs with instructional and domain knowledge, which is gathered from the LOs using NLP techniques. Arikiturri (Aldabe, 2011) is a tool for building exercises from text corpora based on NLP techniques. Both KONGZI and the system of Lentini *et al.* are restricted to a particular kind of TSLSs and do not allow interoperability with other systems. The other described systems are standards and specifications compliant, but do not support the authoring of the whole *Domain Module*. Therefore, a system that supports the semi-automatic generation of the *Domain Module* from electronic documents would be welcomed.

3 Semi Automatic Domain Module Generation

Traditionally, writen documents have been used as a means to transmit knowledge from one generation to the next. Textbooks, and books in general, collect years and years of knowledge about many different domains. They are structured in a means that facilitates learning and understanding. Being able to reuse that knowledge in TSLS would really be a great success. Although Murray (1999) pointed out the need for tools that facilitate the construction of the *Domain Module* in a semi-automatic way, not much effort has been yet oriented to fulfill this goal.

The final aim of the work here presented is to extract the domain knowledge for TSLS from existing documents. Advances in Artificial Intelligence methods and techniques from NLP and heuristic reasoning allow this task to be confronted. The approach here presented focuses on the semi-automatic generation of the *Domain Module* for standards and specifications aware TSLSs. This chapter describes the *Domain Module* authoring process from electronic textbooks, and the pre-process that prepares the documents for the knowledge acquisition tasks detailed in Chapter 4 and Chapter 5.

3.1 Process for the Semi Automatic Domain Module Generation

Building TSLSs, especially *Domain Modules*, is cost and labour-intensive, but gathering the domain knowledge of a TSLS from already existing documents in a semi-automatic way may considerably reduce the development cost. Artificial Intelligence methods and techniques such as NLP and heuristic reasoning can be applied to achieve the semi automatic generation of the *Domain Module*. In this way, *Domain*

Module authors - teachers or instructional designers - can select the documents to be used as source data, and later supervise the results to complete or adapt the automatically generated *Domain Module* to their requirements or teaching preferences.

In this work, the *Domain Module* encodes knowledge at two different levels: (1) the knowledge to be learnt, including the topics and the pedagogical relationships that enable planning and determining the learning sessions, which is described by the Learning Domain Ontology (LDO) and (2) the set of Learning Objects (LOs) that will be used for each domain topic. Using an ontology to describe the learning topics and the pedagogical relationships among the topics will facilitate reusing the described *Domain Module* in different TSLSs after the convenient (automatic) ontology mapping or translation (Uschold and Gruninger, 1996). The following steps are carried out to develop the *Domain Module* (see Figure 3.1):

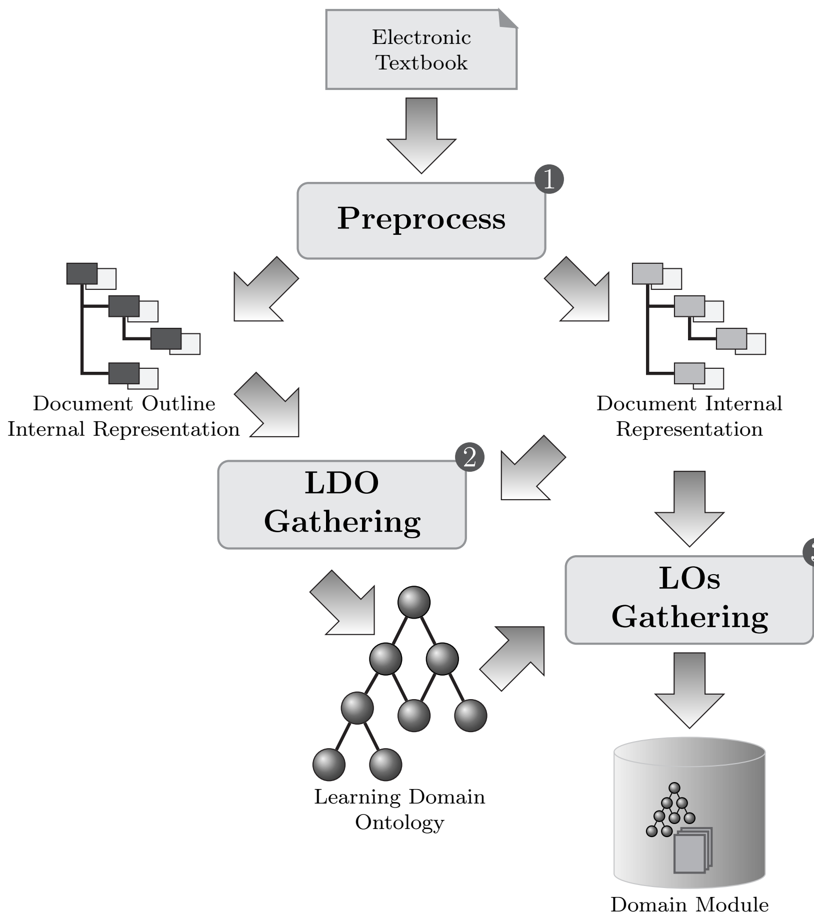


Figure 3.1 – Domain Module Building Process

1. **Document preprocessing**: First, the document must be prepared for the subsequent knowledge acquisition processes. The document preprocessing is described in Section 3.2, and the outcomes are then used to gather the two levels of knowledge encoded in the *Domain Module*. The outline of the document is suitable for the construction of the Learning Domain Ontology (LDO), while the content of the document is useful for both building the LDO and generating LOs.

2. **Gathering the LDO**: At this phase, the domain topics to be mastered, along with the pedagogical relationships among them, are identified and described in the LDO. The LDO ontology can be used in different ways for learning. On the one hand, instructivist TSLSs would use this information to plan the learning sessions. On the other hand, the students can rely on the LDO to guide them during the learning process.

3. **Gathering the LOs**: At this stage the LOs - definitions, examples, exercises, etc. - to be used during the learning process are identified and generated.

Both the acquisition of the LDO and the construction of LOs from the electronic textbook conduct automatic information extraction and, therefore, require human supervision to correct any error or to adapt the *Domain Module* to the preferences of the teachers. More details about the supervision process are presented in Chapter 6. The methodology followed to define the automatic information extraction processes - gathering the LDO and gathering the LOs - was manually analysing a set of documents and identifying patterns or heuristics that allow the knowledge acquisition. This analysis also allowed to empirically determine the confidences of the patterns and heuristics as well as some thresholds that control the processes.

3.2 Electronic Document Preprocessing

In this phase, the system prepares the electronic textbooks and gathers standardised representations of them, to later run the knowledge acquisition (see Figure 3.2). As electronic documents are available in many different formats (e.g., *pdf*, *rtf*, *doc*, or *odf*), a preprocess is carried out to prepare the document for the building process.

The content of electronic textbooks is organised using a hierarchical structure; documents contain chapters, which in turn are divided into sections, and so on. A tree-like format independent representation of the document is built in terms of the elements described in Figure 3.3 so that the rest of the task can be performed with

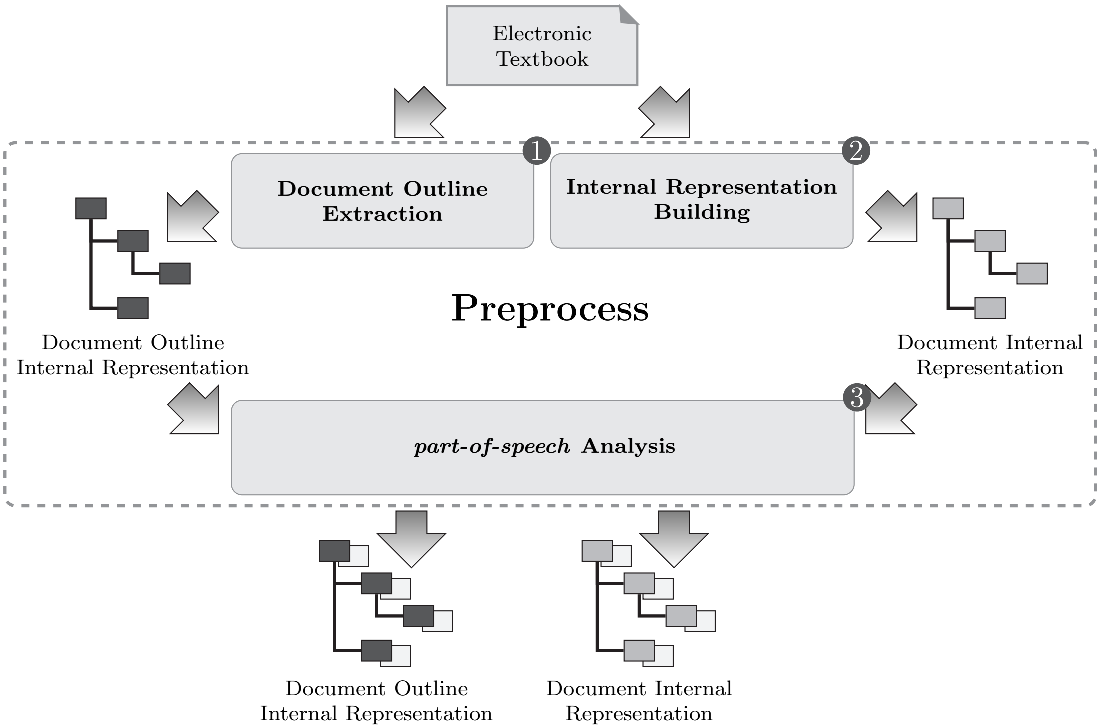


Figure 3.2 – Electronic Document Preprocessing

no dependance on the format the original document is stored in. In addition, the outline of the document, which might be located either at the beginning or the end, can also be numbered or indented in different ways showing its structure. Thus, a homogenised internal representation of the outline is also gathered in the preprocess.

The obtained internal representations for the outline and the whole document are then given a linguistic analysis to enhance them with the *part-of-speech* information that will be used in the next step to gather the expected elements - LDO and LOs - from the document. Linguistic analysis is essential, especially for agglutinative languages such as Basque where most words are formed by joining morphemes together to one lemma.

In the Basque language, for example, words are formed by adding the affixes to the dictionary entries. More specifically, the affixes corresponding to the determiner, number and declension case are taken in this order independently of each other (Aduriz *et al.*, 1998). As prepositional functions are realised by case suffixes inside word-forms, Basque presents a relative high power to generate inflected word-forms,

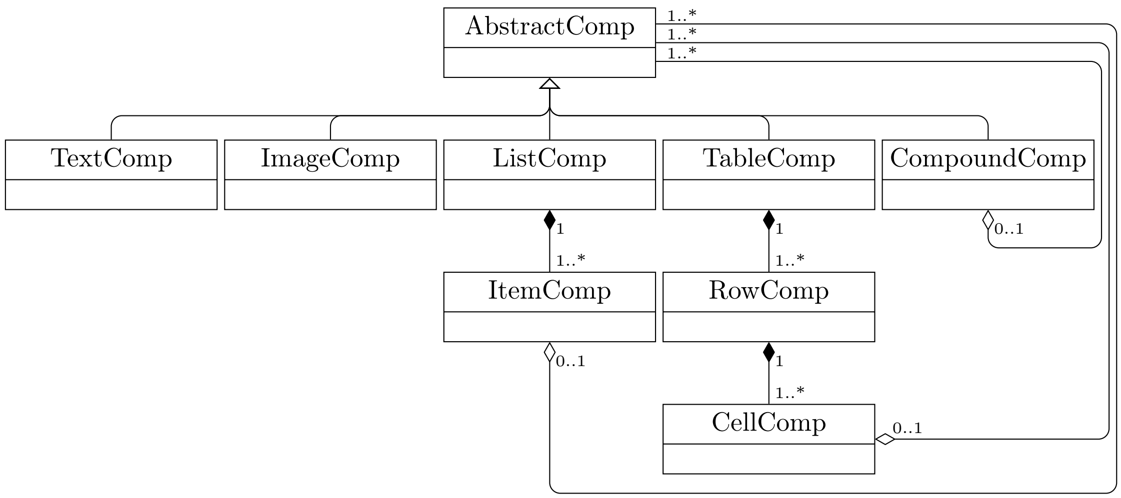


Figure 3.3 – Class Hierarchy for the Tree-Like Document Representation

which makes morphosyntactic analysis very important for being able to extract information from text fragments.

Listing 3.1 shows how the output of the linguistic analysis looks. For every word, the corresponding analysis is provided (indented lines). The first element of the analysis is the lemma or the dictionary entry for the processed word, followed by the *part-of-speech* tags. Some words may result in more than one analysis (ambiguity). This can occur because the word can be derived from different dictionary entries or because it might have different syntactic functions and the system was not able to determine which the correct one is.

Listing 3.1 – Structure of the Output of the Linguistic Analysis

```
 "<Word>"[Extra Information]
    "Lemma" TAG ...
"<Word>"[Extra Information]
    "Lemma" TAG ...
...
```

Listing 3.2 presents a fragment of the output of this phase for the sentence fragment *"Planetak berezko argirik ez duten gorputzak dira, eta ..."* (*"Planets are celestial bodies, which are not self-luminous, and ..."*). For each word of the sentence, the analysis containing the corresponding *part-of-speech* information is provided. This information determines among others the lemma, category, and syntactic function of the word. The fragment includes words with more than one analysis (ambiguity). For example, the word "*duten*" ("*have*") has two different analyses because it might have

Listing 3.2 – Excerpt of the *Part-of-Speech* Information for a Sentence

```
"<Planetak>"<HAS_MAI>"
   "planeta" IZE ARR BIZ- ABS NUMP MUGM ... @OBJ %SINT
   "planeta" IZE ARR BIZ- ABS NUMP MUGM ... @PRED %SINT
"<berezko>"
   "berezko" ADJ ARR IZAUR+ ZERO ...
"<argirik>"
   "argi" ADJ ARR IZAUR- PAR MG ...
   "argi" IZE ARR PAR MG ...
"<ez>"
   "ez" PRT EGI w898,L-A-PRT-53,lsfi1665 ...
"<duten>"
   "*edun" ADL ZHG A1 NOR_NORK NR_HURA NK_HAIEK-K ...
   "ukan" ADT PNT ZHG A1 NOR_NORK NR_HURA NK_HAIEK-K ...
"<gorputzak>"
   "gorputz" IZE ARR BIZ- ABS NUMP MUGM ...
"<dira>"
   "izan" ADL A1 NOR NR_HAIEK ...
   "izan" ADT PNT A1 NOR NR_HAIEK ...
"<,>"<PUNT_KOMA>"
   PUNT_KOMA
"<eta>"
   "eta" LOT JNT EMEN AORG w903,L-A-LOT-JNT-17 ...
...
```

two different syntactic functions. In all these analyses, the lemma is the transitive verb "*ukan*" ("*to have*"), which is conjugated in the third person plural (NK_HAIEK-K) and the object of the verb is the third person singular (NR_HURA).

The results of the preprocess step are the internal representations of the textbooks and their outlines enhanced with the *part-of-speech* information.

3.3 Gathering the Learning Domain Ontology

Once the preprocess has finished, its outcomes, i.e., the internal representations of the document and its outline, are processed to gather the LDO using heuristic reasoning and NLP techniques. This task entails both the identification of the relevant domain topics and defining the pedagogical relationships among the topics. While the textual content of the electronic textbooks is used to gather the domain topics, the structure of the document is valid as a source for identifying pedagogical relationships. The outline of the document contains a summarised representation of the structure of the document, so it is used as the initial source for gathering the LDO. This process is described in depth in Chapter 4.

3.4 Gathering the Learning Objects

The LDO is used to guide the identification of Didactic Resources (DRs) in the internal representation of the document using a pattern-based approach. The gathered DRs are intended to be useful for different *Domain Modules*, as long as they are standards and specifications compliant. The gathered LOs are annotated with metadata and stored in the LOR to facilitate their use. This phase is described in Chapter 5.

3.5 Summary

This chapter has depicted the semi-automatic process for the construction of the *Domain Module* from electronic textbooks focusing on the first phase: preprocessing. The described approach relies on ontologies and standards to assure that the *Domain Module* might be used in different TSLSs. The *Domain Module* is gathered from the electronic documents using ontologies, NLP techniques and heuristic reasoning.

The authoring of the *Domain Module* entails a preprocess, i.e., preparing the textbook for the knowledge acquisition tasks, the acquisition of the LDO that describes the domain to be learnt, including the domain topics and the pedagogical relationships, and the extraction of the LOs to be used in the learning sessions. This chapter has described the first phase and the next two chapters will focus on the others.

4 Gathering the Learning Domain Ontology

Ontology learning, i.e., gathering domain ontologies from different resources in an automatic or semiautomatic way has been addressed in many projects (Buitelaar *et al.*, 2005b). Most of these projects aim at building or extending a domain ontology or populating lexical ontologies such as Wordnet (Fellbaum, 1998) or EuroWordnet (Vossen, 1998). Ontology learning usually combines machine learning and NLP techniques to build domain ontologies or to enhance and populate some base ontologies. Different kinds of resources such as text corpora, document warehouses, machine readable dictionaries or lexical ontologies are broadly used as information sources for ontology learning.

In the approach here presented, the *Domain Module* is described by means of the Learning Domain Ontology (LDO), which contains the main domain topics and the pedagogical relationships among them. Pedagogical relationships can be *structural* - *isA* and *partOf* - or *sequential* - *prerequisite* and *next*. Figure 4.1 shows a fragment of a LDO to illustrate the semantics of these relationships. Topics t_1, t_2, and t_3 are *partOf* the topic t_0, i.e., they are lower granularity elements that are constituents of the more general topic t_0. Both t_4, and t_5 are related to t_2 by the *isA* relationship; in other words, they are particular instances or kinds of the t_1 topic. The relationship between t_2 and t_3 expresses that the latter is recommended to be learnt immediately after mastering the former. t_2 has to be mastered before attempting to learn t_1, as expressed by the *prerequisite* relationship between those topics.

Ontology learning relies on the assumption that there is semantic knowledge underlying syntactic structures. For example, Text2Onto (Cimiano and Völker, 2005) uses Hearst's patterns (Hearst, 1992) to gather taxonomic relationships, and nested terms-based methods (Frantzi *et al.*, 1998) to identify the set of candidate domain

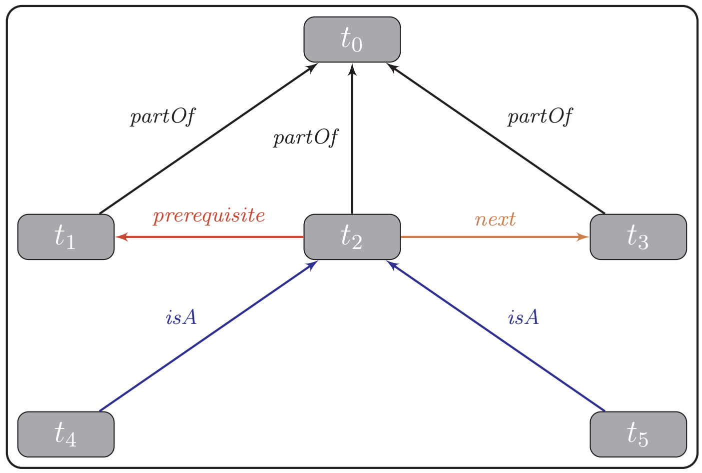


Figure 4.1 – Example of a Learning Domain Ontology Fragment

topics. OntoLT (Buitelaar *et al.*, 2004), a Protégé[1] plug-in for the extraction of ontologies from texts, identifies taxonomic relationships on nested terms, relying on the appearance of a term and some modifiers (*Genus et Differentiam*).

The LDO describes a certain domain for learning purposes; it is an *application ontology* according to Guarino's classification (Guarino, 1997). It describes the domain and also some pedagogical knowledge.

The LDO Acquisition entails two main NLP and heuristic reasoning-based steps: outline analysis, which results in an initial version of the LDO and the whole document analysis, which enhances the ontology with new topics and relationships (Figure 4.2).

4.1 Outline Analysis

Document outlines are useful sources of information for acquiring the *Domain Module* in a semi-automatic way as they are usually well-structured and contain the main topics of the domain. Besides, they are considerably summarised, so a lot of useful information can be extracted with a low cost process. The authors of textbooks have previously analysed the domain and decided how to organise the content according to pedagogical principles. The organisation of the textbook is reflected in the outline. Thus, most of the implicit pedagogical relations can be inferred from the outline by using NLP techniques and a collection of heuristics.

[1] http://protege.stanford.edu/

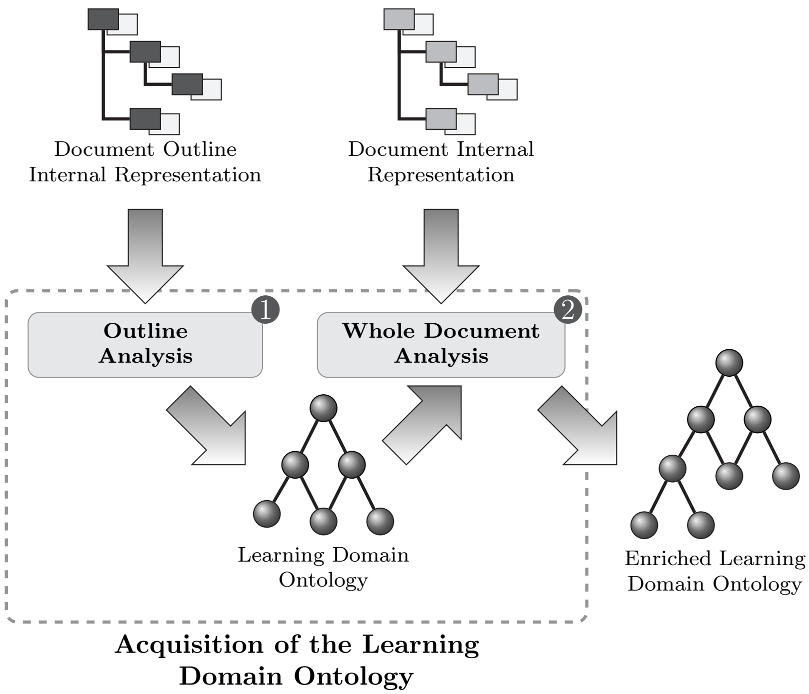


Figure 4.2 – Process for Gathering the Learning Domain Ontology

The outline analysis process consists of two phases:

Basic Analysis: In this task the main topics of the domain and the relationships among these topics are mined from the homogenised outline internal representation. In this approach, each index item is considered as a domain topic. Besides, the structure of the document outline is used as a means to gather pedagogical relationships. A subitem of a general topic is used to explain part of it or a particular case of it. Therefore, structural relationships are defined between every outline item and all its subitems. In addition, the order of the outline items reflects the recommended sequence for learning the domain topics. Thus, an initial set of sequential relationships is identified from the order of the outline items.

Heuristic Analysis: The results of the basic analysis are refined based on a set of heuristics that categorise the relationships identified in the previous step and mine new ones, mainly *prerequisite* relationships. The identified relationships are labeled with the inferred kind, the heuristic used, and the confidence in the inferred information. The heuristics entail the condition to be matched, and the post-condition, i.e., the relationships that are recognised (Figure 4.3). The heuristic analysis relies on the empirically gathered confidence of the heuristic, i.e. the % of times the heuristic fires correctly.

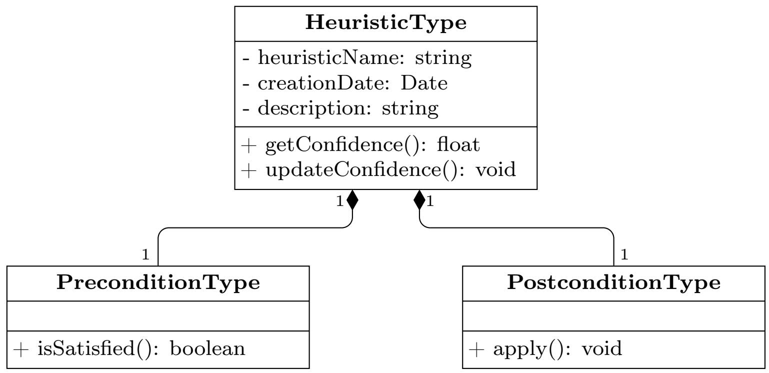


Figure 4.3 – Fragment of the UML Class Diagram of the Heuristics

A set of heuristics must be defined in order to perform this analysis. The set of heuristics was identified on a set of outlines of different subjects at the University of the Basque Country (UPV/EHU). The study allowed the identification of heuristics that recognise structural relationships (*isA* and *partOf*) and sequential relationships (*prerequisite* and *next*) (Larrañaga *et al.*, 2004). The following procedure was carried out in order to identify the heuristics:

1. A small set of 5 outlines related to Computer Science was analysed to detect some patterns that might help in the classification of relationships.
2. These heuristics were tested on a wider set of 150 outlines related to different domains. As a result of this experiment, the implicit pedagogical relationships were inferred from the outlines.
3. The relationships identified by the heuristics were contrasted with those which were real, i.e., manually identified relationships.
4. After analysing the results, paying special attention to the detected lack in the heuristics, some new heuristics were defined.
5. The performance of the improved heuristic set was then tested and improvements were found.

The performance of the heuristics for identifying pedagogical relationships is described in Section 7.2. The presented heuristics have been tested on outlines written in Basque. For the sake of readability, the examples that illustrate the heuristics will be presented in both Basque and English, although some information might be lost in translation.

4.1.1 Heuristics for Structural Relationships

The heuristics for structural relationships allow identifying the kind of relationship between an item and its subitems. The heuristic analysis works under the assumption that only one kind of structural relationship can exist between an outline item and all its subitems as this fact was observed in almost all the analysed outlines. The analysis of the outlines also showed that the most common structural relation is the *partOf* relationship. Therefore, by default, the structural relationships identified in the outline are labeled as *partOf*. In addition, some heuristics that detect the *isA* relationship or that reinforce the default hypothesis were found.

Some of the conditions of these heuristics have to be met by a particular subitem while others are checked on the general item or on the whole group of the subitems. Thus, the heuristics are classified into two groups according to the kind of conditions: individual heuristics and group heuristics. Next, the identification of structural relationships and the employed heuristics are described.

4.1.1.1 Individual Structural Heuristics

These heuristics check if an individual subitem meets a condition. The condition may also involve the general item. Next, the individual heuristics for identifying structural relationships are presented.

MultiWord Heuristic (MWH): MultiWord terms contain information to infer the *isA* relationship. *Genus et Differentiam* is one of the most common ways to define new topics, and this pattern has been used to gather taxonomic relationships among topics from dictionaries, thesauri or other sources of information (Vossen, 2001; Morin and Jaquemin, 1999; Byrd *et al.*, 1987). This pattern can be found in several ways: *noun + adjective*, *noun + noun phrase*, etc. If the noun that appears in these patterns (*agente* or *agent*) is the same as the general item (*agenteak* or *agents*), the *isA* relationship is more plausible (Table 4.1).

Table 4.1 – Fragment of an Outline in which the MWH Can Be Applied

Basque	English
3.4.2 **Agenteak**	3.4.2 **Agents**
3.4.2.1 **Agente** mugikorrak	3.4.2.1 Mobile **agents**
3.4.2.2 **Agente** estatikoak	3.4.2.2 Stationary **agents**

Entity Name Heuristic (ENH): Entity names are used to identify examples of a particular entity. When the subitems contain entity names, the relationship between the item and the subitems can be considered as the *isA* relationship. In Table 4.2, *Palm Os*, and *Windows CE* correspond to entity names, which are particular instances of *Laguntzaile Pertsonal Digitalentzako Sistema Eragileak* (*Operative Systems of Personal Digital Assistants*).

Table 4.2 – Fragment of an Outline in which the ENH Can Be Applied

Basque	**English**
3.7 Laguntzaile Pertsonal Digitalentzako Sistema Eragileak	3.7 Operative Systems for Personal Digital Assistants
3.7.1 **Palm OS**	3.7.1 **Palm OS**
3.7.2 **Windows CE**	3.7.2 **Windows CE**

Acronyms Heuristic (AH): Authors also use acronyms to refer to domain topics whose name is long and frequently used. When the subitems contain only acronyms, the structural relationship is likely to be the *isA* relation. In Table 4.3, the *XUL* and *jXUL* acronyms represent the names of some languages for designing graphical interfaces, thus there is an implicit *isA* relation between the item and its subitems.

Table 4.3 – Fragment of an Outline in which the AH Can Be Applied

Basque	**English**
4.2.2 Interfaze grafikoak sortzeko lengoaiak	4.2.2 Languages for building graphical interfaces
4.2.3.1 **XUL**	4.2.3.1 **XUL**
4.2.3.2 **jXUL**	4.2.3.2 **jXUL**
...	...

Head of the phrase + Multiword Heuristic (He-MWH): This heuristic checks if the subitem forms a multiword term including the head of the phrase of its parental outline item. Table 4.4 shows an example in which this heuristic can be applied, as *agente* (*agent*), the head of the phrase of *Agente laguntzaileak* (*Auxiliar agents*), is used to refer to the topic in a particular context. Therefore, the *isA* relationship can be inferred between those topics.

Table 4.4 – Fragment of an Outline in which the He+MWH Can Be Applied

Basque	English
11 **Agente** laguntzaileak	11 Auxiliar **Agents**
11.1 RMA **Agentea**	11.1 RMA **Agent**
11.2 DF **Agentea**	11.2 DF **Agent**

Acronyms + Multiword Heuristic (A+MWH): By merging the semantics of MWH and AH a new heuristic is created. Thus, this heuristic allows identifying *isA* relationships if the subitems form a multiword from the acronym of the upper level outline item (see Table 4.5).

Table 4.5 – Fragment of an Outline in which the A+MWH Can Be Applied

Basque	English
7 **Abonatu-Linea Digitala**	7 **Digital Subscriber Line**
7.1 **ALD** Simetrikoa	7.1 Symmetrical **DSL**
7.2 **ALD** Asimetrikoa	7.2 Asymmetrical **DSL**

Possessive Genitive Heuristic for Structural Relations (PGH1): Possessive genitives (*-en* suffix in Basque, *of* preposition in English) contain references to other contents. They are used to describe just parts of the content, so the hypothesis of an underlying *partOf* relationship between the general item and its subitems is reinforced by this heuristic (Table 4.6).

4.1.1.2 Group Structural Heuristics

Group structural heuristics check whether the general item or all its subitems match a condition. Two heuristics of this kind have been identified so far:

Keyword Heuristic (KH): This heuristic relies on a set of keywords - *adibideak* (*examples*), *elementuak* (*elements*), *motak* (*kinds*), . . . - to identify *isA* relationships

Table 4.6 – Fragment of an Outline in which the PGH1 Can Be Applied

Basque	English
4.5 **Inplementazioa**	4.5 **Implementation**
4.5.1 Aplikazioa**ren inplementazioa**	4.5.1 **Implementation of** the aplication
4.5.2 Agente**en inplementazioa**	4.5.2 **Implementation of** the agents

among an outline topic and its subtopics. If one of the keywords is the head of the phrase of the outline item, the heuristic triggers and thus the *isA* relationship is defined between the outline item and all its subitems.

Common Head + MultiWord Heuristic (CHe+MWH): This heuristic checks whether all the subitems of an outline item form a multiword term and share the head of the noun phrase, e.g., *clustering* in Table 4.7. This heuristic also identifies *isA* relationships among the outline item and its subitems.

Table 4.7 – Fragment of an Outline in which the CHe+MWH Can Be Applied

Basque	English
5.2 Zenbakizko taxonomia	5.2 Numeric classification
5.2.1 **Clustering** banatzailea	5.2.1 Exclusive **clustering**
5.2.2 **Clustering** hierarkikoa	5.2.2 Hierarchical **clustering**

4.1.1.3 Identification of Structural Relationships

The empirical analysis showed that different heuristics of this kind can be fired together in the same group of subitems. Table 4.8 shows a fragment in which the first subitem (*XUL*) is an acronym while the second (*Luxor*) is an entity name.

In order to classify the structural relationships, Algorithm 4.1 is applied to every outline item with subitems.

If a group heuristic triggers, then its postcondition is executed and the process finishes. Otherwise, for every subitem an individual heuristic that matches is looked

Table 4.8 – Fragment of an Outline with Different Patterns in the Subitems

Basque	English
4.2.2 Interfaze grafikoak sortzeko lengoa-iak	4.2.2 Languages for building graphical interfaces
4.2.3.1 **XUL**	4.2.3.1 **XUL**
4.2.3.2 **Luxor**	4.2.3.2 **Luxor**

Algorithm 4.1 Algorithm for Identifying Structural Relationships for an Outline Item

$gHeur \leftarrow findGroupHeuristic(outlineItem)$
if $gHeur \neq nil$ **then**
  $applyGroupHeuristic(outlineItem, gHeur)$
else
  $hList \leftarrow new\ List()$
  $subItems \leftarrow getSubItems(outline)$
  for all $subit\ in\ subItems$ **do**
    $iHeur \leftarrow findHeur(outlineItem, subit)$
    $add(hList, iHeur)$
  end for
  if $conf_{isA}(hList) > threshold$ **then**
    $addIsARel(outline, subItems, hRel)$
  else
    $addPartOfRel(outline, subItems, hRel)$
  end if
end if

for. In the case where several heuristics could be applied, the most confident one is returned; the *default* heuristic is returned when no other heuristic condition is met. Then, the list of applied heuristics is processed to get the confidence on an underlying *isA* relationship using (Equation (4.1)),

$$conf_{isA} = \frac{\sum\limits_{h \in H_i} f(h) \cdot c(h) - \sum\limits_{h \in H_p} f(h) \cdot c(h)}{n} \tag{4.1}$$

where h represents a heuristic, $f(h)$ is the number of times the heuristic h is triggered, $c(h)$ is the confidence on heuristic h, H_i the set of heuristics that identify *isA* relationships and H_p the set of heuristics that reinforce the hypothesis that the relationship is a *partOf* relationship, and n represents the number of subitems. If the $conf_{isA}$ value goes beyond a *threshold*, then the structural relationships are refined as *isA*, otherwise, the relationships are labeled as *partOf*. As mentioned above, every relationship is labeled with information about the heuristic that has been used to infer it.

Figure 4.4 shows a fragment of an outline and the inferred ontology fragment. Three *isA* relationships have been inferred between *Laguntzaile Pertsonal Digitalentzako Sistema Eragileak* and its subitems (*Palm OS*, *Windows CE*, and *Linux Familiar 0.5 Distribuzioa*). Both *Palm OS* and *Windows CE* have been recognised as entity names, so the *Entity Name* (*EH*) heuristic identifies the *isA* relationship among them and the upper level outline item (*Laguntzaile Pertsonal Digitalentzako Sistema Eragileak*). However, the *EH* heuristic does not meet the *Linux Familiar 0.5 Distribuzioa*, *Linux Familiar 0.5* is considered an entity name, but the whole text is not. Thus, this outline item does not trigger any heuristic except the *default* heuristic. Nevertheless, $conf_{isA}$ goes beyond the *threshold*. Therefore, this relationship is labeled as *isA*. However, the relationship specifies that, in this case, it has been inferred by the combination of the heuristics applied to the other subitems.

4.1.2 Heuristics for Sequential Relationships

Two kinds of sequential relationships can be found. The *next* relationship, which states that a learning topic is recommended to be learnt immediately after another one, appears between items at the same level, i.e., subitems of the same general item. The basic outline analysis only identifies relationships between items at the same nesting level, but the heuristic analysis allows identifying new sequential relationships between any pair of outline items. A *prerequisite* relation between two domain topics states that the former domain topic must be mastered before trying to learn the latter. *Prerequisite* relationships can be found between any outline items. The analysis of the outlines proved that the most common sequential relationship between items at the same level is the *next* relationship. Therefore, by default, any sequential relation identified in the outline is translated into *next*.

All the heuristics for sequential relationships check if a particular outline item and a previous one meet a condition to recognise a sequential relationship. When a heuristic for sequential relationship fires, its postcondition is applied, i.e., the

Basque	English
3.7 Laguntzaile Pertsonal Digitalentzako Sistema Eragileak	3.7 Operative Systems for Personal Digital Assistants
3.7.1 Palm OS	3.7.1 Palm OS
3.7.2 Windows CE	3.7.2 Windows CE
3.7.3 Linux Familiar 0.5 Distribuzioa	3.7.3 Linux Familiar 0.5 Distribution

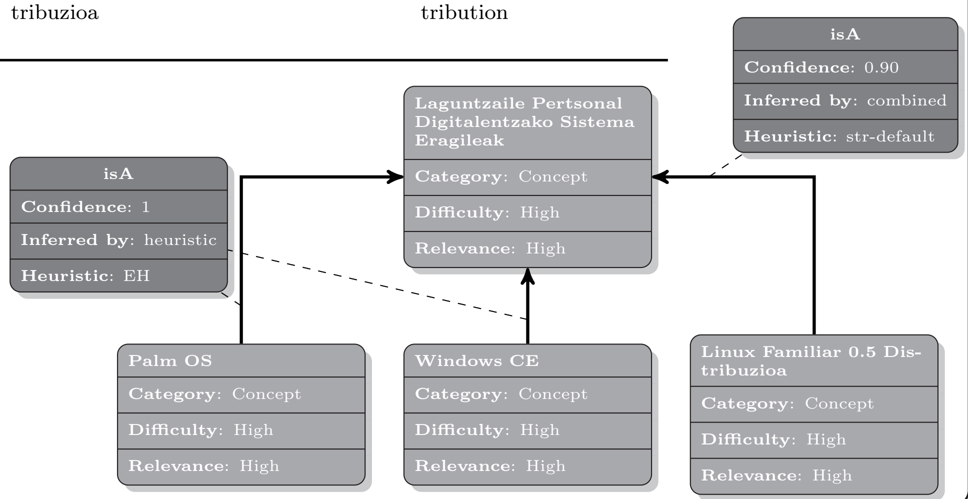


Figure 4.4 – Fragment of a Gathered Learning Domain Ontology

corresponding relationship is defined between the topics that triggered the heuristic. The following heuristics are used to identify *prerequisite* relationships:

Reference Heuristic (RH): This heuristic identifies a *prerequisite* relationship when an outline item refers to a previous topic (see Table 4.9).

Table 4.9 – Fragment of an Outline in which the RH Can Be Applied

Basque	English
4. **Sarrera/irteera**	4. **Input/output**
...	...
6. Eraginkortasun altuko **sarrera/irteera**	6. High performance **Input/output**

Head of the Phrase + Reference Heuristic (He+RH): References to the head of a previous outline item may allow recognising *prerequisite* relationships (see Table 4.10).

Table 4.10 – Fragment of an Outline in which the He+RH Can Be Applied

Basque	English
1.1 **Karga** elektrikoa	1.1 Electric **charge**
...	...
1.3 **Karga** kontserbazioa	1.3 **Charge** conservation

Acronym + Reference Heuristic (A+RH): When an outline item is formed by a reference to the acronym of a previous outline item, a *prerequisite* relationship is identified between those two outline items (see Table 4.11).

Table 4.11 – Fragment of an Outline in which the A+RH Can Be Applied

Basque	English
4.1 **Higidura Harmoniko Sinplea**	4.1 **Simple Harmonic Motion**
...	...
4.5 **HHS** eta higidura zirkular uniformea	4.5 **SHM** and the uniform circular movement

Possessive Genitives Heuristic for Sequential Relations (PGH2): Possessive genitives between outline items at the same nesting level can be used to identify *prerequisite* relationships (Table 4.12).

Head of the Phrase + Possessive Genitive Heuristic (He+PGH2): This heuristic is activated by outline items entailing a possessive genitive based on the head of the previous outline item at the same nesting level (see Table 4.13).

Acronym + Possessive Genitive Heuristic (A+PGH2): This heuristic is triggered by outline items formed by a possessive genitive based on the acronym of the previous outline item at the same nesting level (see Table 4.14).

Table 4.12 – Fragment of an Outline in which the PGH2 Can Be Applied

Basque	**English**
3.5.2. **Agenteak**	3.5.2 **Agents**
...	...
3.5.3. **Agente<u>en</u>** exekuzio eredua	3.5.3 Execution model **<u>of</u>** **agents**

Table 4.13 – Fragment of an Outline in which the He+PGH2 Can Be Applied

Basque	**English**
1.1 **Karga** elektrikoa	1.1 Electric **charge**
...	...
1.4 **Karga** zehatz bat**<u>en</u>** eremu elektrikoa	1.4 Electric field **<u>of</u>** a particular **charge**

4.2 Whole Document Analysis

At this stage, the initial LDO is enhanced with new topics and relationships gathered from the whole document. In order to achieve this goal, two processes are carried out: first, new topics are identified as described in Section 4.2.1, and later new pedagogical relationships among the topics are identified (*cf.* Section 4.2.2).

4.2.1 Identifying New Topics

This process aims at enhancing the LDO gathered in the previous phase with new domain topics. The whole document is analysed to get such new topics. In the last few years, the use of hybrid methods that combine NLP techniques and statistic methods has prevailed in term extraction. Many approaches use a set of patterns such as $((A|N)+|((A|N)*(NP)?(A|N)*)N$ to get the set of candidate terms, where A is an adjective, N is a noun, and P is a preposition, and then apply some termhood measures to rank the set of candidate terms and filter the most appropriate ones (Justeson and Katz, 1995).

Table 4.14 – Fragment of an Outline in which the A+PGH2 Can Be Applied

Basque	English
4.1 **Higidura Harmoniko Sinplea**	4.1 **Simple Harmonic Motion**
4.2 **HHSren** zinematika	4.2 Cinematics **of the SHM**

In the work here presented, term-extraction is carried out using Erauzterm (Alegria *et al.*, 2004b), a term extractor for Basque that looks for the most usual noun-phrase structures, to gather new domain topics. Erauzterm gathers either one-word or multiword terms which can be ranked using different measures to determine how related they are to the domain.

4.2.2 Identifying New Relationships among Topics

This process allows the identification of new pedagogical relationships from the electronic document using a pattern-based approach. These patterns recognise pedagogical relationships between domain topics based on the syntactic structures found in the sentences in which the topics appear. Therefore, the internal representation of the document is first annotated to label any domain topic appearance. Then, nested domain topics, i.e., domain topic constructed on other domain topics (*Genus et Differentiam*) are identified to propose *isA* relationships among them. For example, "*Sirius izarra*" (the "*Sirius star*") contains the topic "*izar*" ("*star*"). Thus, it can be inferred that "*Sirius izarra*" ("*Sirius izar*") is a "*izar*" ("*star*"). Finally, the document is given a grammar-driven analysis to identify a set of sentences which relate two or more domain topics. The grammar contains a set of rules describing syntactic structures corresponding to pedagogical relationships. The Constraint Grammar formalism, one of the most successful syntax analysing and disambiguation systems, has been used to develop and apply the grammar on the documents (Voutilainen and Tapanainen, 1993; Karlsson *et al.*, 1995; Tapanainen, 1996). Two kinds of rules are used in Constraint Grammar: *information removing rules* that are applied to delete wrong information, and *mapping rules*, which allow adding new information. In the current work, only *mapping rules* are used, as new information is to be added during the knowledge acquisition processes and removing information makes no sense at this step.

The grammar for identifying pedagogical relationships entails rules for recognising structural relationships - *isA* and *partOf* - and the *prerequisite* sequential relationship. The rules have been defined after an empirical analysis of a set of text fragments elicited from textbooks corresponding to primary school. 13 rules for the *isA* relationship, 6 for *partOf* and 1 for *prerequisite* are defined in the grammar. These rules include, among others, the equivalent for Hearst's patterns (Hearst, 1992) for the Basque language. Hearst patterns have been used in many ontology learning approaches to identify taxonomic relationships (*hypernyms/hyponyms*) from syntactic structures like NP_0 *such as* NP_1, NP_2,... *(and/or)* NP_n, where NP_i is a noun phrase.

Some of the rules for the identification of the pedagogical relationships are now described. A full list of the rules is provided in Appendix A. Table 4.15 shows an example in which an *isA* relationship can be identified when two ontology topics (@Topic) are related through the "*izan*" ("*to be*") verb.

Table 4.15 – Example of a Pattern that Allows Identifying *isA* Relationships

	Basque	English
Pattern	@Topic @Topic [*det*] **IZAN**	@Topic **TO BE** [*det*] @Topic
Example	Lurra planeta *bat* **da**.	The Earth **is** *a* planet.

When two domain topics are related by the *izeneko* (*referred to as*) expression, an *isA* relationship between those topics can be inferred, as shown in Table 4.16.

Table 4.16 – Example of a Pattern that Allows Identifying *isA* Relationships

	Basque	English
Pattern	@Topic **izeneko** @Topic	@Topic **referred to as** @Topic
Example	Esne bidea **izeneko** galaxiak 100 mila milioi izar dituela uste dute zientzalariek.	Sciencists believe that the galaxy **referred to as** the Milky Way has over 100 billion stars.

Table 4.17 shows a pattern which identifies *partOf* relatonships. This relationship can be identified in sentences in which a possessive genitive of a domain topic is found followed by *adreilu* or *osagaiak* (*components*) and later some other domain topics appear. Thus, *partOf* relationships can be defined among *unibertsoa* (*universe*) and *galaxiak* (*galaxies*).

Table 4.18 shows a rule that might identify a prerequisite relationship between two topics. To understand what "*satelitea*" ("*satellite*") is, the students must know about "*planeta*" ("*planet*").

Table 4.17 – Example of a Pattern that Allows Identifying *partOf* Relationships

	Basque	English
Pattern	@Topic (**PG**[a]) **adreilu/osagai Izan** @Topic	@Topic **To Be components of** @Topic
Example	Unibertso**aren** oinarrizko **adreiluak** galaxiak **dira**.	Galaxies **are** the main **components of** the Universe.

[a] The (PG) states that the @Topic is in possessive genitive case

Table 4.18 – Example of a Pattern that Allows Identifying *prerequisite* Relationships

	Basque	English
Pattern	@Topic (**PG**[a]) @Topic	@Topic **of** @Topic
Example	Planeta **baten** satelitea planetare inguruan biraka dabilen gorputz bat da..	The satellite **of** a planet is a celestial body that orbits around the planet .

[a] The (PG) states that the @Topic is in possessive genitive case

4.3 Supervision of the Learning Domain Ontology

The LDO automatically acquired following the process described above should be supervised by the *Domain Module* authors to correct any error or to adapt it to their preferences. Therefore, the *Domain Module* authors are provided with a graphical tool that facilitates this task. The approach selected for supervision is Concept Map-based. Concept Maps (CMs) (Novak, 1977, 1990) have proved to be useful to describe, organise and even to gather knowledge (Coffey *et al.*, 2004). CMs provide an intuitive and understandable description of the domain knowledge being modeled using graphic resources: nodes are used to represent the domain topics and arcs to express the relationships among them. For each domain topic or pedagogical relationship, the inferred information (e.g., the kind of relationship) and the heuristic that inferred the information can be shown. The *Domain Module* authors can review and modify the LDO either by removing topics or relationships, adding new ones or modifying the identified ones. See Chapter 6 for more details on the supervision of the LDO.

4.4 Summary

This chapter has described a NLP-based heuristic approach for building the LDO from electronic textbooks. The LDO acquisition entails the identification of topics and relationships between the topics. The proposed process is carried out in two main phases: the analysis of the textbook outline facilitates the construction of the initial LDO identifying both the set of main topics and relationships. The initial LDO is enhanced with new domain topics and pedagogical relationships gathered by the analysis of the whole document.

As the identification of topics and pedagogical relationships is carried out by automatic processes, the outcome must be supervised to correct errors or adapt the LDO to the *Domain Module* authors' preferences. A CM-based tool enables the *Domain Module* authors to carry out the supervision of the LDO.

5 Gathering Learning Objects from Electronic Textbooks

The last step for building the *Domain Module* entails defining and providing a set of Didactic Resources (DRs) that will be used during the learning sessions, either to present the domain topics to the learner or to assess the learner's knowledge level. Textbooks, traditionally used as a means to transfer knowledge, contain plenty of resources (e.g., *definitions*, *problem statements*, etc.) that could be extracted and used to complete the *Domain Module*.

Document authors tend to use very similar syntactic structures to present topic definitions, examples, images and other kinds of DRs. Some of those patterns have already been used to identify educational material, e.g., definitions, in electronic documents written in English (Liu *et al.*, 2003; Verbert, 2008).

The identified DRs, besides being useful for the pretended *Domain Module*, could be reused for other new *Domain Modules*. *Domain Module* generation can, therefore, profit from the already developed resources. For the sake of facilitating the reuse of the DRs, they must be annotated with descriptive metadata that allows their localization and retrieval from the Learning Object Repository (LOR) they are stored in. This chapter focuses on the generation of Learning Objects (LOs) from the electronic textbooks, presenting a general view of the process in Section 5.1. The generation of DRs is depicted in Section 5.2 and Section 5.3. The construction of LOs from the DRs is described in Section 5.4. Details about the storage of the LOs to promote their reuse are provided in Section 5.5.

5.1 General View of the Process

LOs are digital resources that can be reused to support learning (Wiley, 2000). The usefulness of a LO is determined by both the quality of its content and the facility of the LO to be found and retrieved from a large set or a LOR, which is highly influenced by the metadata that describes it (Cardinaels, 2007).

The generation of LOs from the electronic textbooks entails identifying and extracting the relevant DRs i.e., fragments of the document related to one or more topics with a particular educational purpose, their annotation with LOM and storage in the LOR. The gathered LOs are mainly text-based, although they may also contain some of the images used to illustrate the domain topics in the textbook. LOs are gathered from the electronic textbook by carrying out the process described in Figure 5.1. From now on, a DR will refer to a piece of the document aimed at being used for learning (e.g., definition, exercise, . . .) while a LO refers to a reusable DR which has been enriched with metadata.

The LO generation here described aims to be domain-independent. Therefore, the only domain-specific knowledge used is the LDO, which has been gathered from the electronic textbook in the previous phase. The process that identifies and extracts the DRs is performed following a pattern-based approach, i.e., identifying the syntactic structures that are often used, for example, to introduce domain topics or propose exercises. The searched text fragments are restricted to domain topics described in the LDO. The gathered DRs are aimed at being coherent and cohesioned. NLP techniques that combine a DR grammar and discourse markers are used, together with a didactic ontology (Meder, 2000; Leidig, 2001), i.e. an ontology that describes the different kinds of DRs than can be used in learning sessions, to achieve this goal. DR identification and extraction is described in Section 5.2.

Once the DRs have been identified and gathered (*cf.* Section 5.2 and 5.3), LOs are built from them. For each DR, the final file is generated and the resource is annotated. As mentioned above, the metadata that describes the LOs is essential to assure that the LOs can be found and retrieved from the LOR they are stored in. Many DRs can be gathered from the electronic textbook, and the *Domain Module* authors cannot be expected to spend much of their time in describing each gathered fragment. Therefore, automatic metadata generation will be used to build the LOs from the identified DRs. The LDO and the ALOCOM ontology (Verbert *et al.*, 2005) are used, as will be described in Section 5.4. Finally, the built LOs are stored in the LOR so that they can be reused either for the *Domain Module* being developed or any future TSLS (see Section 5.5).

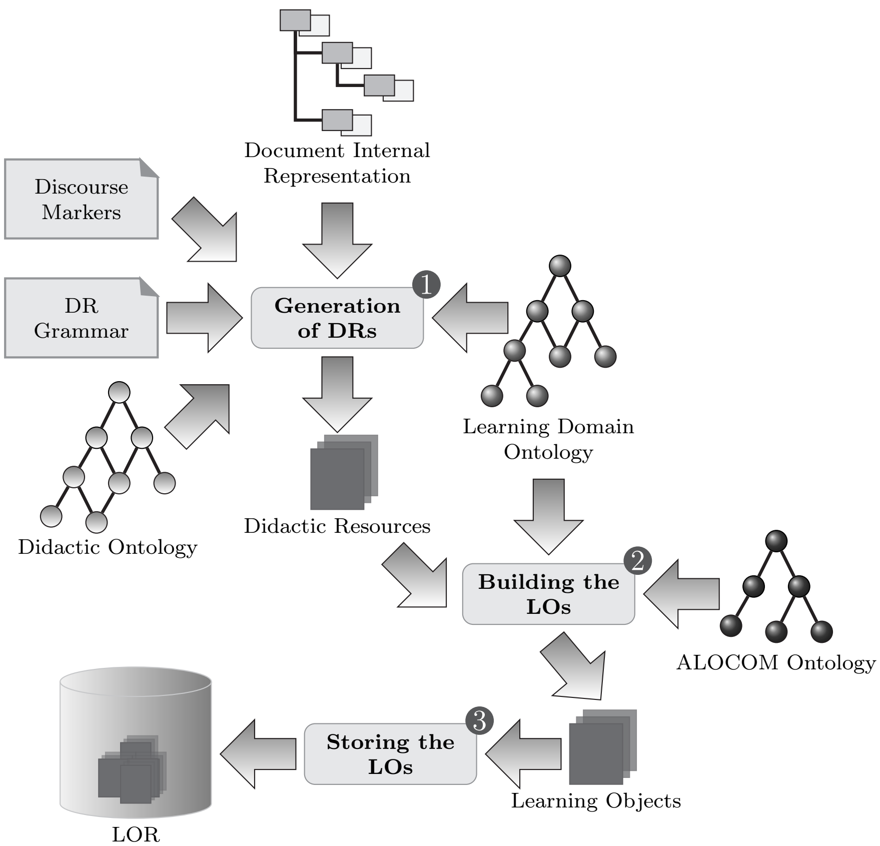


Figure 5.1 – Generation of Learning Objects from Textbooks.

5.2 Generation of Didactic Resources

The identification of the DRs is carried out by finding relevant text fragments for the LDO topics. Document authors use quite similar patterns (syntactic structures) when defining new topics, describing theorems or proposing exercises. These patterns may be used to gather some of the kinds of DRs described in the didactic ontology, namely, *definitions*, *examples*, *facts*, *theories*, *principle statements*, and *problem statements*, from the electronic documents.

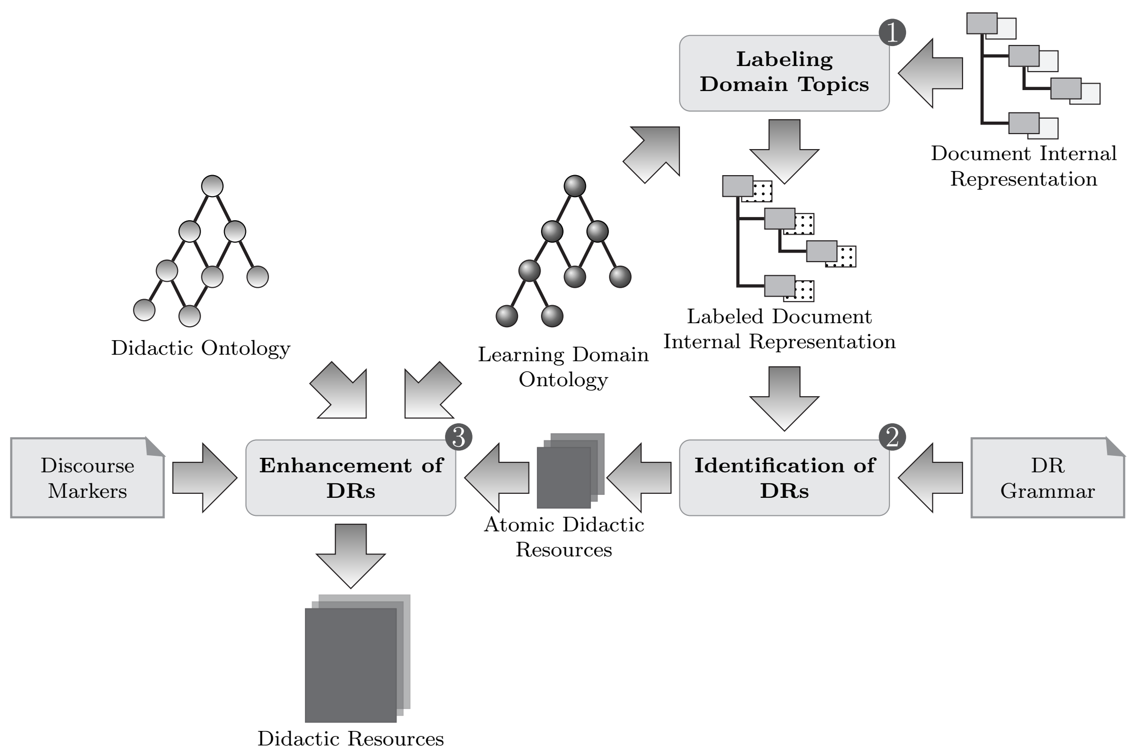


Figure 5.2 – Generation of Didactic Resources

The process described in Figure 5.2 is carried out to extract the DRs. The appearances of the LDO topics are labeled in the textbook internal representation built in the preprocess of the document. Next, the DR grammar is used to find text fragments that might contain appropriate resources. The DR grammar entails a set of rules that recognise the syntactic structures used to present the different kinds of DRs, e.g., topic definitions, examples, etc. Similar patterns are used for English in (Liu *et al.*, 2003; Verbert, 2008) to look for definitions. The grammar for gathering the DRs from the electronic document has also been developed using the Constraint Grammar formalism.

The DR grammar was tested on electronic textbooks to observe its performance. Some of the initially defined rules were removed from the final version of the DR grammar, as they had low precision. The precision of the grammar rules is used to determine the confidence in these rules.

The identified atomic DRs contain the sentence that triggered the rule for the corresponding DR and all the sentences that follow it, as long as they refer to the same topic(s). Every DR is labeled with the domain topics and with the rules of the DR grammar that identified it. This information is used later in the LO annotation process (see Section 5.4.2).

The gathered DRs are then processed and enhanced in order to get more appropriate DRs and to assure the coherence and cohesion of their content. As a result of this process, some of the DRs might be combined with consecutive DRs or text fragments. The composite DRs are built as an aggregation of DRs of lower granularity and keep the information about why they were composed (cohesion maintenance or DR similarity) and the similarity rates. Besides, the referred topics and the DR grammar rules used to identify the DR are also kept in every DR (Figure 5.3). This process is described in Section 5.3.

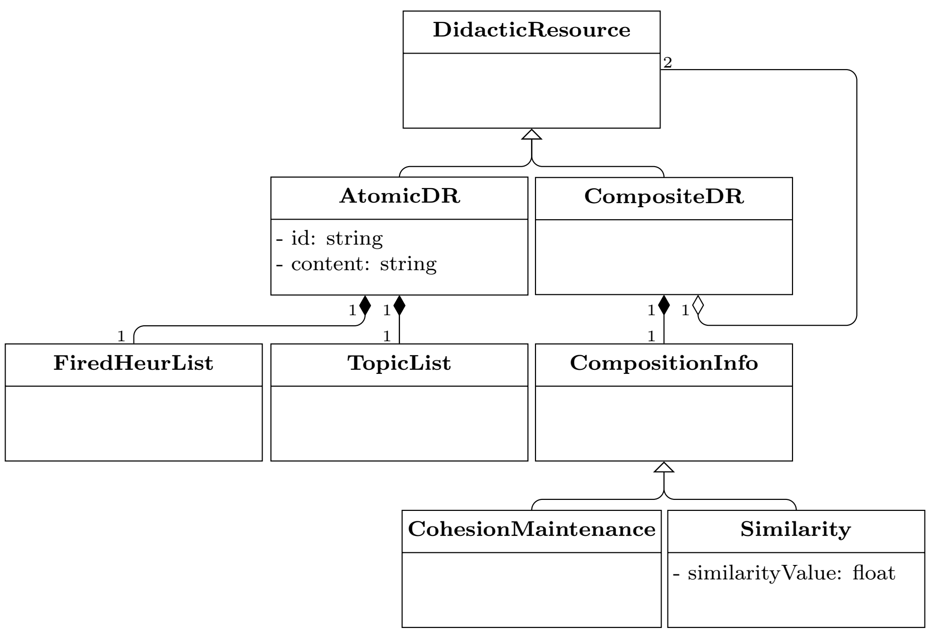


Figure 5.3 – Classes for the Internal Representation of Didactic Resources

5.2.1 Identification of the Didactic Resources

The DR grammar allows locating the sentences that use any of the identified syntactic structures referring to LDO topics. Text-based DRs are built from the sentences selected by the grammar. For each selected sentence, an atomic DR is built. The atomic DRs also contain the sentences that follow the selected one as long as they are not identified as other DR by the grammar, and they are content related, i.e., they are considered similar from the content perspective. Content similarity is measured considering the domain topics referred in the text. Textbook authors may also include some sentences that do not necessarily include the domain topics but that connect different sentences that do refer to domain topics. An empirically gathered number of consecutive sentences[1] of this kind are also allowed while building the atomic DRs, with the aim of being as complete and coherent as possible. Besides, every image found in the textbook is also considered a DR that requires no deeper processing.

Figure 5.4 shows a fragment of a document where some DRs can be detected[2]. Three DRs are identified and constructed, the first is an image, and the last two are consecutive definitions. The pattern used to identify them is underlined. The definition of the "*planetak*" ("*planets*") entails two sentences. The second one was added as it is related to similar domain topics, while the last sentence - "*Lurra planeta bat da.*" ("*The Earth is a planet.*") - contains the definition of another topic, so a different DR has been built from it. The next subsections describe some of the patterns used in the DR grammar to identify the text-based DRs. A full list is presented in Appendix B.

5.2.1.1 Patterns for Definitions

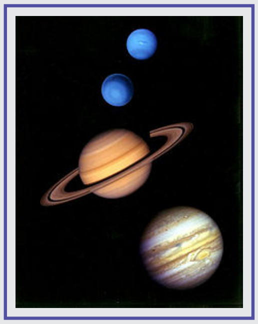
A definition is a passage that explains the meaning of a term (a word, phrase or other set of symbols). Textbook authors frequently use similar text expressions to provide topic definitions. Table 5.1 shows one of those expressions. Similar patterns have been also used to identify definitions in electronic documents written in English (Liu *et al.*, 2003; Verbert, 2008).

Some topics represent measures, and are frequently described by means of equivalent values as can be observed in the example presented in Table 5.2.

Expressions that contain a question such as the example shown in Table 5.3 also allow finding definitions. Although the text shown does not describe any topic, the following sentences, which might provide the answer to the question, are likely to define the referred topic.

[1] This number is by default 3, as this value attained the most accurate results
[2] The examples are presented in both Basque and English to facilitate understanding

...



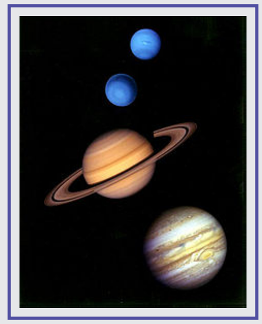
...

Planetak *berezko argirik ez duten gorputzak* dira, *eta izar baten inguruan biraka mugitzen dira.* Uste denez, Eguzki-Sistemako planetak Eguzkiarekin batera eratu ziren, eta pentsa daiteke antzeko planeta ugari izango direla beste izar batzuen inguruan.

Lurra *planeta bat* da.

...

...



...

Planets are n*on-self-luminous celestial bodies orbiting around a star.* Solar system planets are thought to be formed together with the Sun, and it is also conceived that many similar planets might be found orbiting other stars.

The **Earth** is a *planet.*

...

Figure 5.4 – Examples of Gathered Didactic Resources

5.2.1.2 Patterns for Examples

An *example* is something characteristic of its kind or that illustrates something being described. This kind of DRs usually contain a fragment of text that enumerate individual(s) of a particular domain topic. Table 5.4 shows how an example can be identified based on typical expressions such as "*adibidez*" ("*for instance*").

References to domain topics in the superlative case are also useful for finding examples in text fragments (see Table 5.5).

Table 5.1 – Fist Example of a Pattern that Allows Identifying Definitions

	Basque	English
Pattern	<u>@Topic</u> *definition* (DAT[a]) **deitu**	*Definition* **To Be called** <u>@Topic</u>
Example	<u>Unibertsoa</u> *astro guztien multzoari eta betetzen duten espazio***ari deitzen zaio**.	*The whole set of celestial bodies and the space they fill* **is called** <u>Universe</u>.

[a] The (DAT) states that the text of the description is in dative case, which is expressed with the **-ri** suffix in the example

Table 5.2 – Second Example of a Pattern that Allows Identifying Definitions

	Basque	English
Pattern	<u>@Topic</u> = *definition*	<u>@Topic</u> = *definition*
Example	1 <u>Unitate Astronomikoa (u.a.)</u> = *149.579.870 km*	1 <u>Astronomical Unit (AU)</u> = *149,579,870 km*

5.2.1.3 Patterns for Facts

A *fact* is a piece of information used as evidence or as part of a report or news article. Table 5.6 presents an example of a fact, an event associated to a particular date. It obviously is not the only way a fact can be found, but it is the only way that was found in the documents that was generalisable without domain knowledge.

5.2.1.4 Patterns for Theories

Theories are suppositions or a collection of ideas intended to explain something. Table 5.7 shows some examples of theories or theorems that could be identified by means of some keywords.

Table 5.3 – Third Example of a Pattern that Allows Identifying Definitions

	Basque	English
Pattern	**Zer Izan** <u>@Topic</u> [*det*]**?**	**What To Be** <u>@Topic</u> [*det*]**?**
Example	Baina, **zer dira** <u>behatoki astronomiko</u> horiek**?**	But, **what are** those <u>astronomical observatories</u>**?**

Table 5.4 – First Example of a Pattern that Allows Identifying Examples

	Basque	English
Pattern	**adibidez,** @Topic	**for instance,** @Topic
Example	Uretan, **adibidez,** hidrogeno eta oxigeno atomoak daude.	**For instance,** there are hydrogen and oxygen atoms in water.

Table 5.5 – Second Example of a Pattern that Allows Identifying Examples

	Basque	English
Pattern	@Topic text @Topic ADJ (SUP[a])	@Topic ADJ (SUP[a]) @Topic
Example	Betelgeuse ezagutzen diren izar **disdiratsuenetako eta handienetako** bat da.	Betelgeuse is **one of the largest and most luminous** stars known.

[a] The (SUP) states that the adjective is in the superlative case

5.2.1.5 Patterns for Principle Statements

This kind of DR tries to provide descriptions of how or why some phenomena, e.g., an eclipse, happens. Principle statements are found by identifying expressions that express causality (see Table 5.8).

5.2.1.6 Patterns for Problem Statements

Most of the patterns for problem statements rely on identifying the imperative case of the verb, which can be carried out using the *part-of-speech* information. Besides, keywords such as "*ariketa*" ("*exercise*") or the patterns shown in Tables 5.9 and 5.10 can be used.

Table 5.6 – Example of a Pattern that Allows Identifying Facts

	Basque	English
Pattern	**Date:** @Topic	**Date:** @Topic
Example	**1957:** Lurraren lehen satelite artifiziala (Sputnik I)	***1957*:** Sputnik I, was launched to orbit the Earth.

Table 5.7 – Example of a Pattern that Allows Identifying Theories

	Basque	English
Pattern	@Topic **teoria** *text*	@Topic **theory** *text*
Example	Eztanda handiaren **teoria** unibertsoaren sorrera azaltzen duen eredua da.	The Big Bang **theory** is the prevailing cosmological model of the early development of the universe.

Table 5.8 – Example of a Pattern that Allows Identifying Principle Statements

	Basque	English
Pattern	@Topic @Topic **oinarri izan**	@Topic **to be based on** @Topic
Example	Unibertsoko gainontzeko astroak bezala, Eguzkia, Lurra eta Ilargia mugitu egiten dira, eta era bat baino gehiagoko mugimenduak egiten dituzte, gainera. Lurreko fenomeno askok, esaterako eguna eta gaua, eklipseak, edo itsasaldiak, mugimendu horietan **dute beren oinarria**.	Like any celestial body, the Sun, the Earth and the Moon move in different ways. Many of the phenomena in Earth, e.g., the day, the night, eclipses and tides, **are based on** such movements.

Many problem statements start with expressions like "*erantzun galdera hauei*" ("*answer these questions*") as shown in Table 5.9.

Table 5.9 – First Example of a Pattern that Allows Identifying Problem Statements

	Basque	English
Pattern	**Erantzun galdera** [*det*]	**Answer** [*det*] **question**
Example	**Erantzun galdera** hau:	**Answer** this **question**:

In many cases, the students are requested to provide or summarise their knowledge about a topic. Those exercises usually start with the expression shown in Table 5.10.

Table 5.10 – Second Example of a Pattern that Allows Identifying Problem Statements

	Basque	English
Pattern	**Idatzi** <u>@Topic</u> (DAT[a]) **buruz jakin**	**Write know about** <u>@Topic</u>
Example	**Idatzi** <u>eclipseei</u> **buruz dakizun** guztia.	**Write** all **you know about** <u>eclipses</u>.

[a] The (DAT) states that the @Topic is in the dative case

5.3 Enhancement of the Didactic Resources

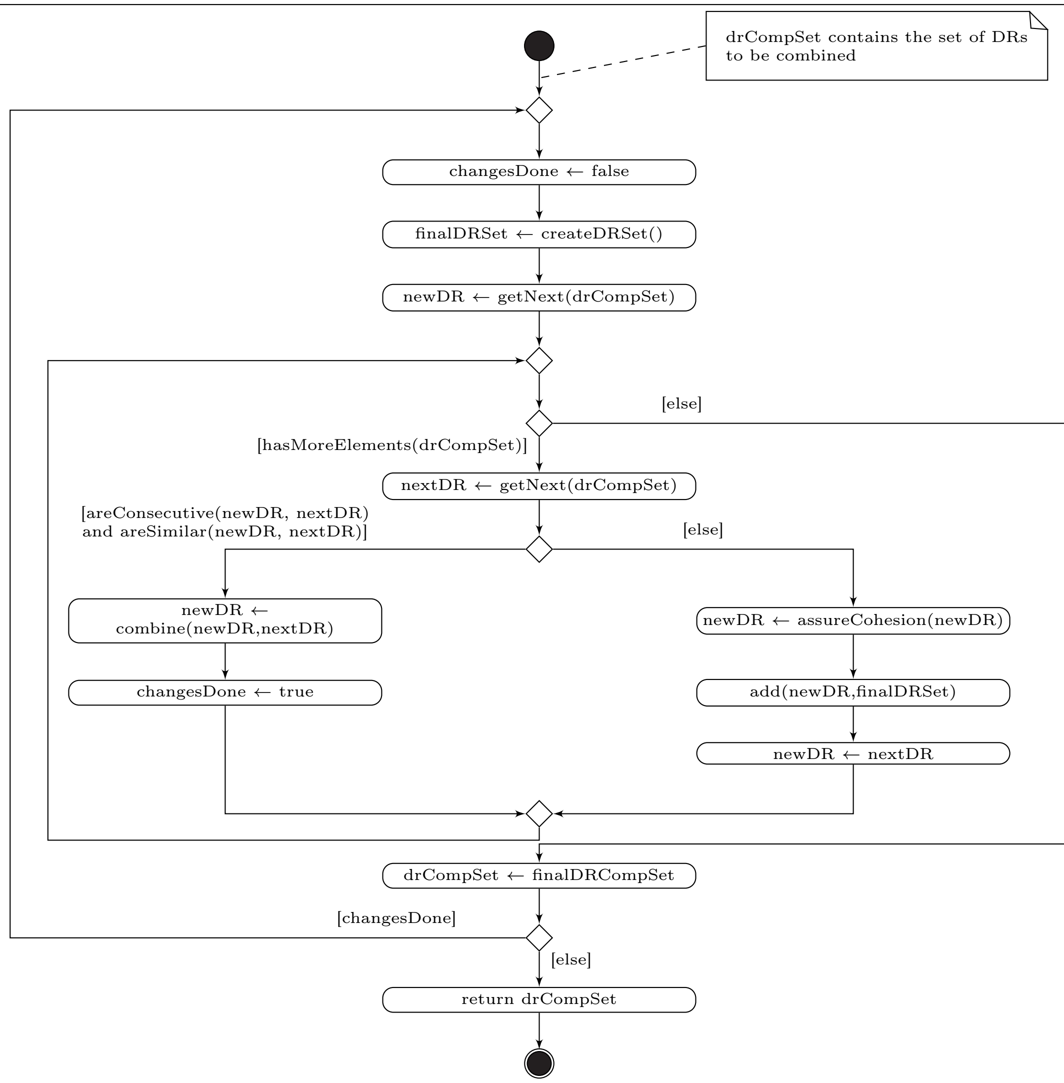
The DRs identified by the grammar are usually quite simple, they entail a set of sentences about a particular domain topic or a group of domain topics. Those DRs can be enhanced in two ways in order to meet the principles for determining the granularity of the DRs stated by Schoonenboom (2006). On the one hand, combining two consecutive DRs, such as those shown in Table 5.11, may result in more useful DRs than the atomic ones. On the other hand, and to keep the cohesion of the DRs, previous fragments are added to a DR that contains references to those fragments. If the referenced previous fragment is part of a DR, then both DRs are combined. Discourse markers, i.e., words or phrases that are used to link sentences, are employed to determine which DRs must be enhanced.

Table 5.11 – Examples of Consecutive Atomic Didactic Resources that Can Be Combined

	Basque	English
DR_1	Planetak berezko argirik ez duten gorputzak dira, eta izar baten inguruan biraka mugitzen dira. Uste denez, Eguzki-Sistemako planetak Eguzkiarekin batera eratu ziren, eta pentsa daiteke antzeko planeta ugari izango direla beste izar batzuen inguruan.	Planets are non-self-luminous celestial bodies orbiting around a star. Solar system planets are thought to be formed together with the Sun, and it is also conceived that many similar planets might be found orbiting other stars.
DR_2	Lurra Planeta bat da.	The Earth is a planet.

The enhancement of the DRs is crucial to obtain really reusable DRs, and is achieved following the algorithm presented in Figure 5.1 and based on similarity measuring methods. Every pair of consecutive DRs is tested to determine their resemblance. If they are considered similar, they are combined in a new DR that comprises them. Once the composition step has finished, the DRs undergo a cohesion assuring process. This process is repeated as long as changes are made on the identified set of DRs.

Algorithm 5.1 Algorithm for the Composition of Didactic Resources



5.3.1 Similarity Measuring Methods

Determining if two consecutive DRs are close enough is a key issue to obtain more accurate DRs. Two aspects are considered to determine if two DRs are suitable for combination. On the one hand, the content of the DRs is analysed to measure their relatedness. On the other hand, the kind of DRs is considered. For instance, examples may enrich a topic definition, and thus, their combination may result in a better DR. However, problem statements are seldom combined with other DRs, unless a whole unit is expected to be built. Thus DR similarity or relatedness measuring methods for each of these aspects have been defined. These methods return a value in the [0, 1] range; the higher the value, the closer the DRs. Two DRs are considered similar if the obtained content similarity and the DR type similarity are beyond the corresponding threshold values or the combined similarity score is beyond the threshold, depending on the user's preferences. More details on the experiments conducted to validate the similarity measures and which allow to empirically determine the threshold are presented in Chapter 7.

5.3.1.1 Content Similarity Measuring Methods

Content similarity measuring methods determine the resemblance of two DRs according to their content, i.e., the topics of the domain they reference. Four methods have been implemented and tested:

Same Topics Method: This is the first implemented and tested topic similarity measuring method. It determines that two DRs are similar if both refer to the same domain topics, i.e., the first one references all the domain topics that are mentioned in the second DR and vice versa. This method proved to be too restrictive, as not all the content related DRs always refer explicitly to all the associated domain topics.

Share Topic Method: Two DRs are considered similar if the first one makes reference at least to one of the domain topics mentioned in the second DR. This method was tested to check if it overcame the problem of the previous one, but it proved to be just the opposite, too loose. The DRs were too likely to be considered similar.

Cosine Method: The cosine similarity (Salton, 1992) calculates the similarity between two vectors by measuring the cosine of the angle between them. The result of the Cosine function is equal to 1 when the angle is 0, which states that both vectors point in the same direction, and it is less than 1 otherwise. This method is broadly used to compare text fragments; in this case, the vectors are built on the words that may appear in the text fragments.

This method can be used, in a similar way, to compare two text-based DRs based on the domain topics they refer to. The vectors corresponding to each DR contain how many times the corresponding domain topic has been referenced in the referring DR, and the similarity value is obtained using Equation (5.1), where $\vec{d}$ and $\vec{d'}$ are the vectors used to model each DR.

$$\cos\left(\vec{d}, \vec{d'}\right) = \frac{\vec{d} \times \vec{d'}}{|\vec{d}||\vec{d'}|} \tag{5.1}$$

Although this method obtains accurate similarity rates according to the explicitly referred domain topics, it does not consider the semantic relationships among the domain topics.

Ontology Based Method: The Cosine Method does not consider the semantic relationships among the domain topics, even though they may provide useful information. The ontology-based method uses the UKB tool (Agirre *et al.*, 2009a,b; Agirre and Soroa, 2009), an ontology based lexical similarity measuring application similar to Hughes and Ramage's Wordnet-based approach (Hughes and Ramage, 2007). For every analysed fragment, UKB returns the stationary distribution of the LDO topics considering both the semantic relatonships in the ontology and the topics referred in the analysed fragment. The similarity is obtained using Equation (5.1) on the stationary distributions of the compared fragments. This method proved to obtain the most accurate results compared to the instructional designers criteria.

5.3.1.2 Didactic Resource Type Similarity Measuring Methods

Two different means of determining the similarity of two DRs considering the type of resource (example, definition, etc.) have been used:

Same Resource Type Method: This is the first implemented and tested DR type resemblance measuring. Each detected DR has a list of patterns that have been used to gather it and identify its type. This method considers that two DRs are similar if the first pattern of the list of each DR assigns them the same category.

Didactic Ontology Method: It has been observed that textbook authors combine different types of DRs while building new ones (e.g., definitions and examples).

Thus, a new way to determine when two DRs can be combined according to their category has been developed. This method is similar to the Ontology-Based content similarity measure method but using the kinds of DRs instead of the domain topics. It uses a didactic ontology (Meder, 2000; Leidig, 2001), which represents the different kinds of DRs and relationships between those types, to compute the similarity between two DRs. Figure 5.5 shows a fragment of this ontology.

5.3.2 Assuring Cohesion in the Didactic Resource Enhancement

Discourse markers, i.e. words or expressions that connect part of a text with its context, are known to be related to the rhetorical relationships that govern the structure of the narratives (Knott and Dale, 1994; Taboada, 2006; Iruskieta *et al.*, 2010). Therefore, they can be used as a means to assure (or at least try to assure) the cohesion in the gathered DRs. Sentences starting with particular discourse markers are likely to be related to the previous sentence or sentences. Therefore, DRs starting with a particular discourse marker will be enhanced by adding previous sentences or even the previous DR that the previous sentence or sentences are part of to assure the cohesion of the text. If the previous sentences are included in another DR both are combined in a new one.

Kind	**Basque**	**English**
References	Hau, hura, Horiek, Horri, …	This, That, These, Those, …
Single	Gainera, Horretarako, bestalde, …	Besides, Therefore, However, …
Complex	*Alde batetik*, … *Bestetik* …	*On the one hand*, … *On the other hand*, …
	Hasteko, … *Bukatzeko* …	*First*, … *Finally*, …

Table 5.12 – Discoures Markers for Basque

Discourse markers are classified, independently of the related rhetorical relation, into three categories considering how the DRs that contain them have to be enhanced: single, complex and references. Single discourse markers - *Gainera* (*Besides*) or *Horretarako* (*Therefore*) - and references connect the sentences with the previous

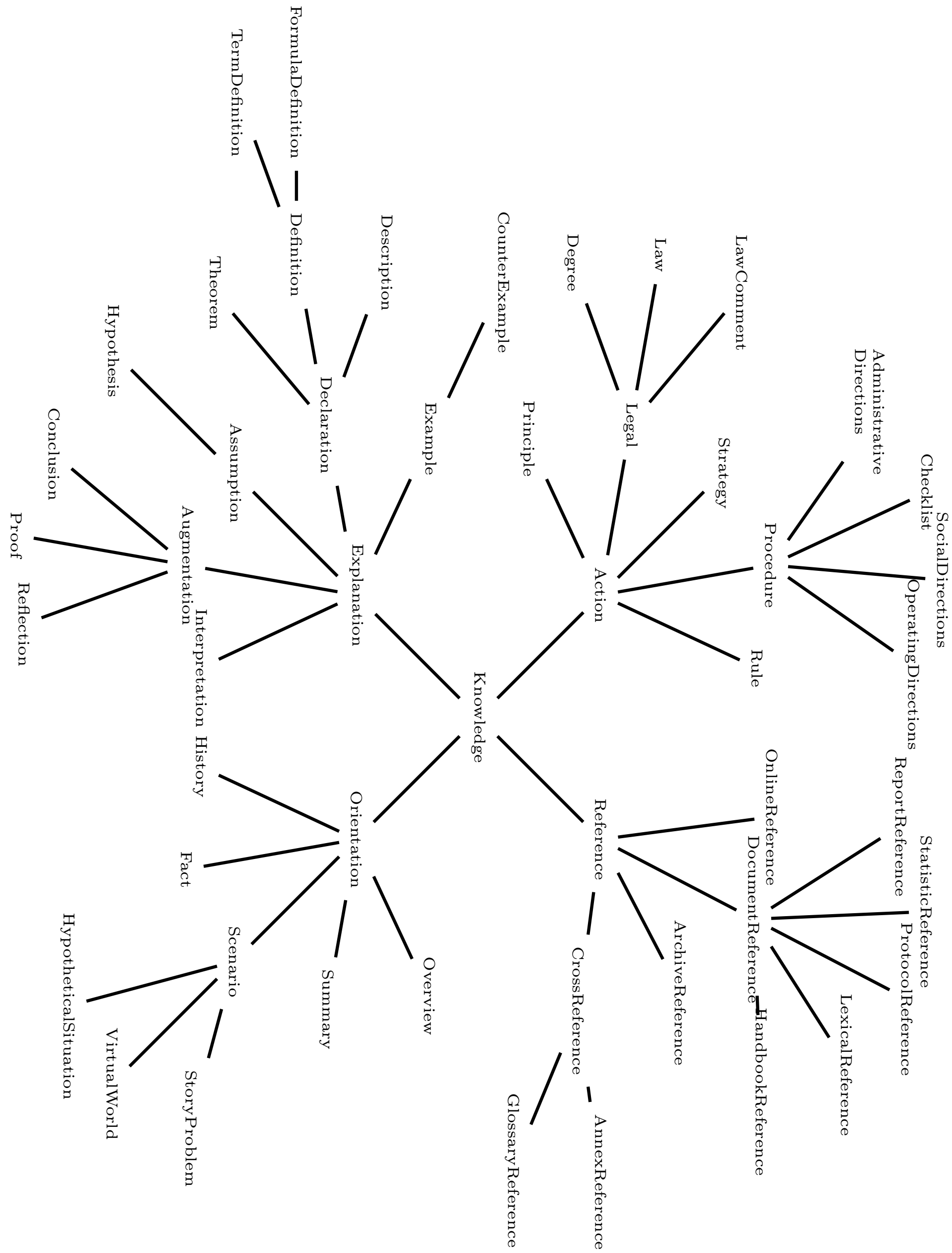


Figure 5.5 – Fragment of the Didactic Ontology (from Leidig, 2001)

sentences. Complex discourse markers require a starting expression - for example, *Hasteko . . . Bukatzeko . . .* (*First, . . . Finally, . . .*). The system deals differently with each kind of discourse markers. If the DR starts with the second part of a complex discourse marker (e.g., *Bukatzeko,. . .*), it will add all the necessary sentences until the initial part (*Hasteko,. . .*) is included. References and single discourse markers usually regard up to an empirically gathered maximum number of sentences; thus, at most the maximum number[3] sentences are added in this case. A sample of the discourse markers for Basque and their translation to English, in order to facilitate understanding, are shown in Table 5.12.

5.4 From Didactic Resources to Learning Objects

The gathered DRs might be not only useful for the *Domain Module* being developed from the processed electronic textbook, but also for other *Domain Modules*. Thus, to facilitate their reuse, LOs, i.e. reusable digital educational resources, are built from the gathered DRs. Building reusable DRs entails two aspects: using an appropriate format to store and represent the content, and also describing it (annotating it) with LOM to allow searching in and retrieving those LOs from the LOR.

The generation of DRs might gather resources of different granularity ranging from atomic DRs to composite DRs that comprise finer grained DRs. Although the composite DRs might be more appropriate for a certain context, the entailed DRs might be also used in other contexts, so LOs are also built from the components of the composite DRs.

5.4.1 Learning Object File Format

The presentation format of the LO may also affect its reusability. Presentation formats such as *html*, *pdf*, *doc*, and *odf* are suitable for final presentation, but are not appropriate for flexible content reuse, as the components cannot be easily accessed (Verbert *et al.*, 2008; Verbert, 2008). The ALOCOM framework (Verbert, 2008; Verbert *et al.*, 2009) was developed to overcome this problem and facilitate the decomposition of composite LOs and make those components available for *on-the-fly* content reuse. This framework relies on the ALOCOM ontology (Verbert *et al.*, 2005), which represents a content model for LOs and their components.

[3] The performed experiments showed that adding up to three previous sentences provided the best results. However, this value is configurable.

Listing 5.1 – Example of a Learning Object

```xml
<?xml version="1.0" encoding="UTF-8"?>
<ALOCOMComponent id="467c3115-e0a6-11dd-aa6f-1b45350a80e7" type="definition">
  <ALOCOMComponent type="definition">
    <ALOCOMComponent type="paragraph">
      <ALOCOMComponent type="text">
          Planetak berezko argirik ez duten
          gorputzak dira, eta izar baten
          inguruan mugitzen dira.
      </ALOCOMComponent>
    </ALOCOMComponent>
  </ALOCOMComponent>
  <ALOCOMComponent type="example">
    <ALOCOMComponent type="paragraph">
      <ALOCOMComponent type="text">
          Lurra planeta bat da.
      </ALOCOMComponent>
    </ALOCOMComponent>
  </ALOCOMComponent>
</ALOCOMComponent>
```

The generated DRs are stored in a ZIP file that contains the XML file for the LO, based on the ALOCOM formalism, as well as the referenced images or other resources. Listing 5.1 shows an example of a LO using the ALOCOM format. The ALOCOM ontology, which had to be enhanced to support *theorems* as they were not considered in the previous version, is used to categorise the LO (see Figure 5.6). Besides, for every LO a preview file in *rtf* format is generated so that the user may have an approximate idea of the content of the LO while looking for resources about a certain topic.

5.4.2 Learning Object Annotation

The likelihood to retrieve the desired LO from a large set or a LOR is a key issue to promote the use and reuse of LOs. This selection is highly influenced by the appropriateness of the metadata that describes the LO. While the manual creation of metadata can be considered for annotation of a single LO, it is not an option for larger deployments where a considerable number of LOs have to be managed (Duval and Hodgins, 2002; Cardinaels *et al.*, 2005; Duval and Hodgins, 2004). Furthermore, semi-automatic metadata generation can overcome metadata inconsistency problems by using ontologies (Kabel *et al.*, 1999, 2004a,b).

After an analysis of the LOM elements, and considering the kind of documents being processed, these elements were classified and it was concluded that only some

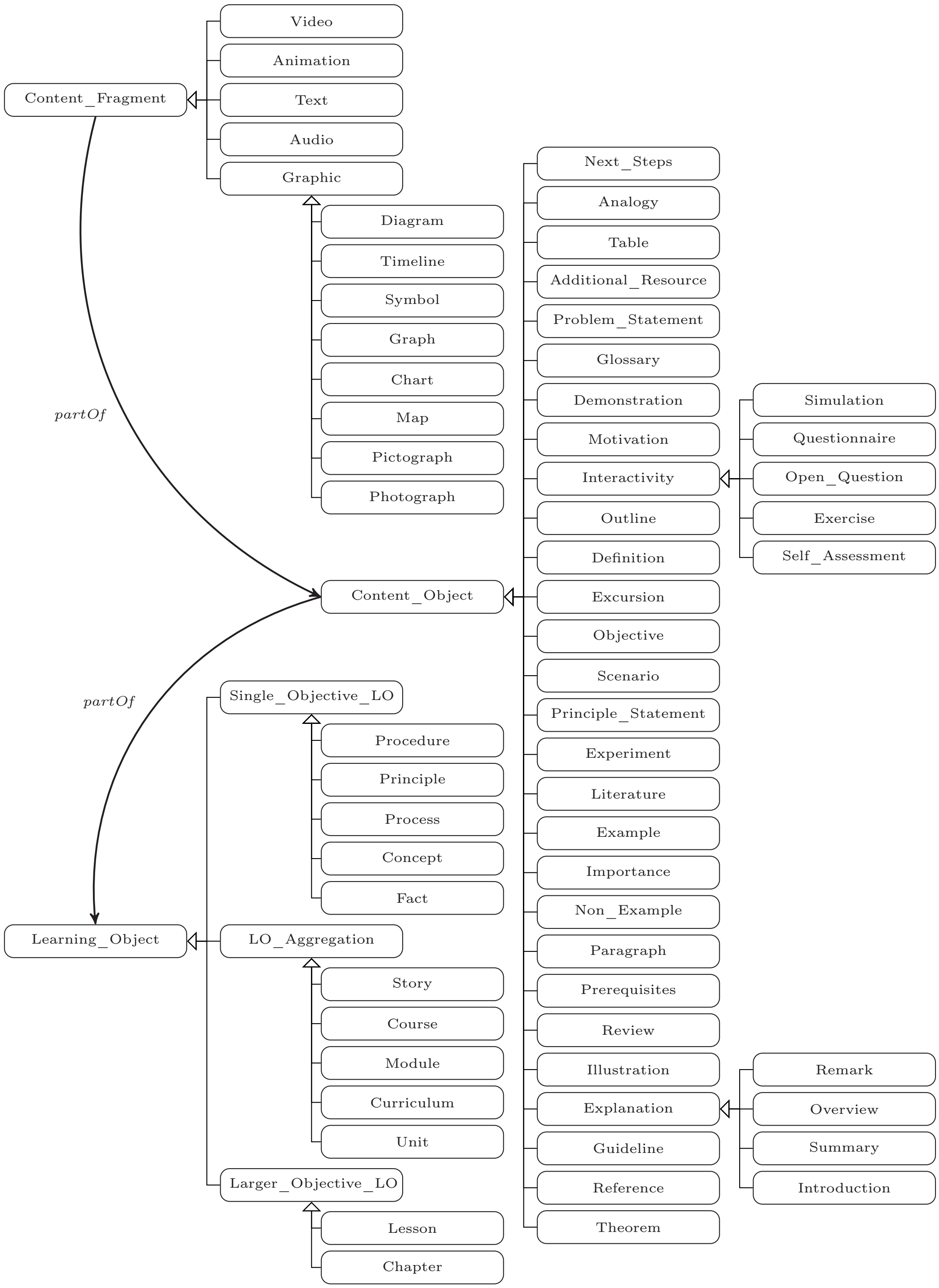


Figure 5.6 – Enhanced ALOCOM Ontology

of them differed from one of the gathered LOs to another, while most of them had similar values (Larrañaga *et al.*, 2008b). Thus, the metadata generation is carried out in the following way. The initial metadata elements are automatically generated from the electronic textbook, using Samgi (Meire *et al.*, 2007), an automatic metadata generator. Then, the metadata is enhanced with more information that has been extracted during the DR generation to improve some elements (*keywords* or *Learning Resource Type*). Most keyword annotation applications use statistical methods and rely on the frequency of the terms in the analysed text, but do not consider semantic relationships among the topics. For example, a keyword extractor may identify *Earth*, *Mars*, *Mercury*, and *Venus* in a fragment of text if they appear in it, but it would not consider that all of them are planets, and therefore it would not infer *planet* as a keyword, as it is not aware of the semantic relationships among these topics. Thus, the LDO and the identified domain topics in the LO are used to get a more accurate keyword list, as the semantics relationships are taken into account.

The *Learning Resource Type* is also specified in terms of the ALOCOM ontology (Verbert *et al.*, 2005), which represents a content model for the LOs and its components. The generated LOs can be images, definitions, examples, principle statements, problem statements, theories, and facts, which are kinds of LOs described in the ALOCOM ontology. To determine the *Learning Resource Type*, the rules of the DR grammar met by the content of the DR are used. As these rules may identify different kinds of DRs, the precision of the rules (% of times that the rule correctly identifies a DR) is used to determine which is the most plausible kind, which is therefore selected as the *Learning Resource Type* for the annotated LO.

5.5 Learning Object Storage

In order to make the LOs available, they are stored in LORs. Two repositories are employed, one for the LOs (both resources and metadata) and another one for keeping the preview files for the LOs. Both repositories are built on the ARIADNE Knowledge Pool System (Duval *et al.*, 2001; Ternier *et al.*, 2009).

Once the LOs and their preview files have been generated, they are pushed to the LOR to allow their retrieval and use in new TSLSs. The LO publishing service is based on the SPI specification (Ternier *et al.*, 2008a). The LOR can be queried to find the appropriate LOs using SQI (Simon *et al.*, 2005). When the LO is composed, its components are also annotated with LOM and stored in the LOR, as they might be useful in certain contexts.

5.6 Summary

This chapter has described the construction of LOs from electronic textbooks. The proposed approach relies on a DR grammar to recognise fragments of the document related to one or more domain topics with a particular educational purpose. In order to build coherent and cohesion LOs, the identified DRs, fragments of the document, are enhanced in two ways. On the one hand, two consecutive DRs are combined if they are related. On the other hand, discourse markers are considered to maintain the cohesion of the text, and, therefore, previous fragments are added to a DR that references previous DRs or sentences.

LOs are built from the gathered DRs. The content of the LOs is represented using the formalism proposed by the ALOCOM framework (Verbert *et al.*, 2009), that enables content reuse. LOs are annotated using LOM and stored in a LOR so that they can be reused to build *Domain Modules*.

6

Architecture and Implementation Issues of DOM-Sortze

The previous chapters have described a semi-automatic approach for building the *Domain Module* from electronic textbooks using ontologies, NLP techniques and heuristic reasoning. The proposed approach aims at facilitating the development of *Domain Modules* for TSLSs by identifying and extracting topics, pedagogical relationships and learning material.

A system for building *Domain Modules* that implements the described approach, *DOM-Sortze*, has been developed. This system supports both the knowledge acquisition process and the supervision process. This chapter describes *DOM-Sortze*.

6.1 Architecture of DOM-Sortze

DOM-Sortze is a suite of applications and web-services that cope with different tasks of the *Domain Module* generating process. Figure 6.1 shows the architecture of *DOM-Sortze*. The rounded boxes represent web services, while the applications or modules are represented by rectangular boxes. This web-service oriented approach makes *DOM-Sortze* flexible and platform-independent. Although it uses some platform-specific applications (mainly NLP tools), these are used only by the web-services. Therefore, the client side applications are platform independent.

DOM-Sortze entails four main applications – the *Preprocessor*, the *LDO Builder*, *ErauzOnt*, and *Elkar-DOM* – that carry out the tasks for building the *Domain Module*. The first three carry out the textbook processing tasks described in previous chapters and the latter facilitates the intervention of the *Domain Module* authors,

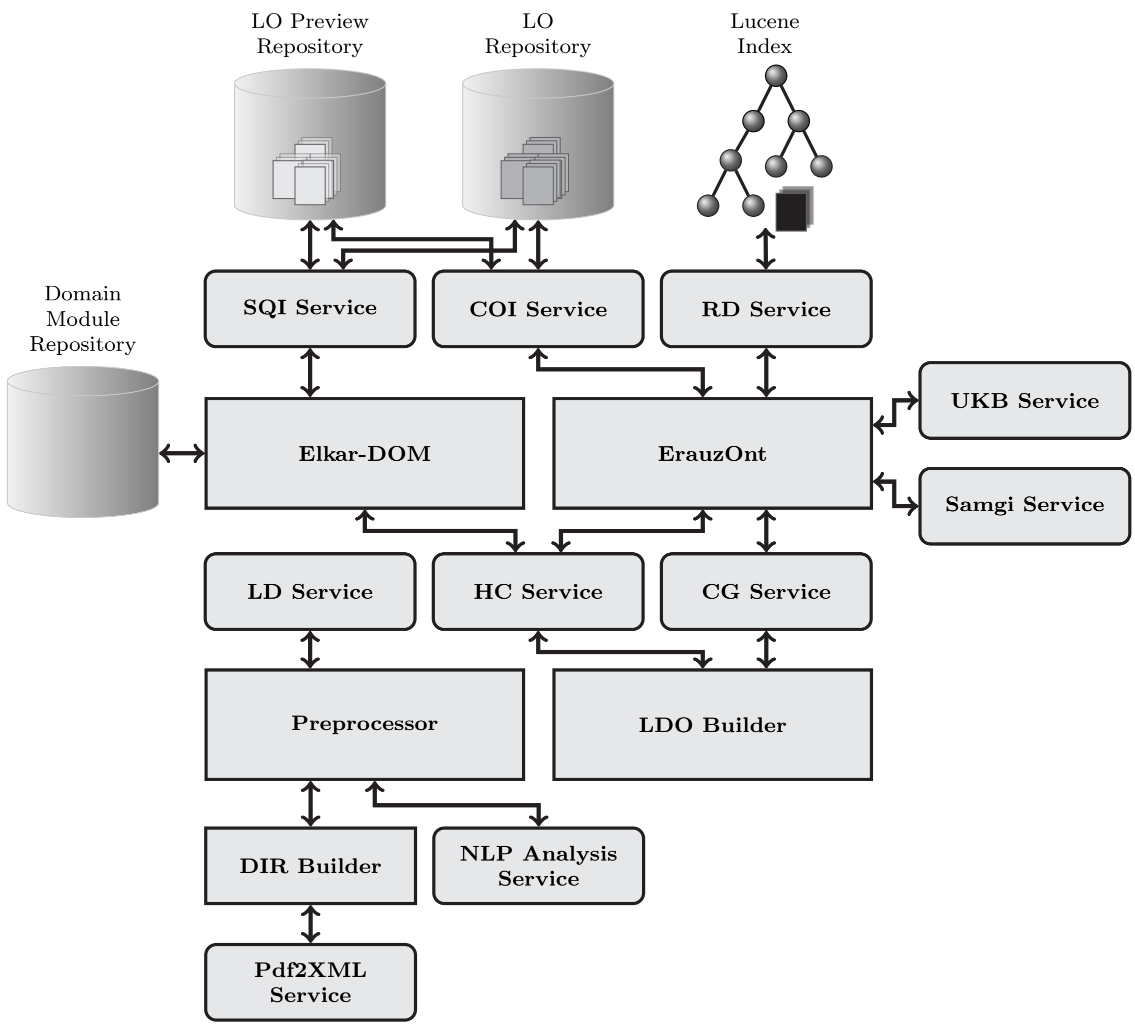


Figure 6.1 – General Architecture of DOM-Sortze

either instructional designers or teachers, to supervise the results. These applications take advantage of some web-services to perform their job.

Two repositories allow storing and providing LOs for the construction of *Domain Modules*; the *LO Repository* stores the LOs (resources and metadata) and the *LO Preview Repository* keeps the preview files for the LOs. The preview files might be helpful for the users to determine whether or not the LO fits their requests. The ALOCOM framework (Verbert *et al.*, 2008) also uses this feature as a means to facilitate the search and retrieval of LOs. The *SQI Service* is used to query for LOs from the LORs, and the *COI Service* allows adding new LOs to the LOR.

The *Replicate Detection Service* (*RD Service*) determines whether a document or a fragment of the document has been processed before. This service uses MD5 hash codes (Rivest, 1992) to compare new resources with the resources that have previously been analysed. A lucene[1] index keeps the information about the processed resources, i.e., the identifier and the hash code of the LO. A copy of the processed resources is tracked for safety, so that the lucene index can be restored if needed (e.g., when the index is corrupted or removed). The *RD Service* prevents documents or fragments from being processed more than once.

The contents of *pdf* files are extracted by the *Pdf2XML service*, which provides a ZIP file that contains an XML file and the images used in the document. This service allows the *DIR Builder* to gather the internal representations of the document and its outline. The *NLP Analysis service* returns the *part-of-speech* information for a text. The *CG Service* is used to carry out the grammar-based analysis and the *HC Service* returns the confidence of the heuristics used during the analysis of the textbook. The *UKB Service* allows to obtain the similarity measures for the ontology-based methods described in Chapter 5. The automatic annotation of the generated LOs is facilitated by the *Samgi Service*.

The *Domain Module Repository* maintains all the *Domain Modules* built with *DOM-Sortze*.

Following, the applications that entail *DOM-Sortze* - the *PreProcessor*, the *LDO Builder*, *ErauzOnt* and *Elkar-DOM*- are described.

6.2 Preprocessor

Textbooks can be found in diverse forms. Documents are available in different electronic formats (e.g., *pdf*, *doc*, *html*, etc.), although usually all the documents are structured in a hierarchical structure (chapters, section, ...). Besides, some documents have their outline at the beginning of the document while others place it at the end. In addition, authors or publishing companies use different numbering or structuring styles. Therefore, the textbooks have to be prepared before proceeding with the knowledge acquisition tasks.

The *Preprocessor* is responsible for carrying out the preprocess of the document. It relies on the *DIR Builder* module to gather the internal representation of both the textbook to be analysed and its outline. The *DIR Builder* module allows the *Preprocessor* to perform its work independently of the format of the document. It currently supports *pdf* documents to which end it delegates on the *Pdf2XML Service*

[1] http://lucene.apache.org/

to build the internal representations of the document and its outline. Nevertheless, it can be extended to support new formats. The *Pdf2XML Service* provides a ZIP file containing the XML representation of the *pdf* document and the image resources found in it. The *LD Service* identifies the language a document is written in. The *NLP Analysis Service* is used to get the *part-of-speech* information of the text fragments. Currently the Basque language is supported, and this service uses EUSLEM (Aduriz *et al.*, 1996) to perform the linguistic analysis.

6.3 LDO Builder

The LDO is gathered by the *LDO Builder* from the internal representations of the electronic textbook and the outline as described in Chapter 4. The LDO is elicited from the internal representations of the electronic textbook and its outline.

The topics of the LDO are gathered from the outline of the textbook and from the whole document. The identification of the topics from the whole document is conducted by using Erauzterm (Alegria *et al.*, 2004a,b; Gurrutxaga *et al.*, 2005).

The identification of the pedagogical relationships is also achieved following a pattern-recognition approach. Some pedagogical relationships are identified from the outline by the heuristics and the inference engine described in Section 4.1, while others are recognised by the analysis of the whole textbook. The elicitation of pedagogical relationships is managed using the *Constraint Grammar Service* (*CG Service*), which relies on the Constraint Grammar formalism (Voutilainen and Tapanainen, 1993; Karlsson *et al.*, 1995; Tapanainen, 1996). The reliability of the employed heuristics varies from one to another. Thus, the *HC Service* is used to get the confidence of the patterns.

The LDO uses an XML-based formalism to describe the gathered LDO. Listing 6.1 shows a fragment of an XML file that describes a LDO fragment of a LDO in which some topics and a relationship are described. As can be observed, the information about the heuristic used and the confidence on that heuristic are also included to facilitate the supervision process depicted later. The formalism for describing the LDO also supports the description of the kind of topic, the relevance of the topic, and the difficulty level, although these features are not currently elicited from the textbooks.

Listing 6.1 – Fragment of the Learning Domain Ontology

```xml
<?xml version="1.0" encoding="UTF-8"?>
<TopicSet>
  ...
  <Topic>
     <ItemId>T2</ItemId>
     <ItemContent>Gailu logiko programagarriak (PLDak)</ItemContent>
     <DRS/>
  </Topic>
  <Topic>
     <ItemId>T21</ItemId>
     <ItemContent>PAL</ItemContent>
     <DRS/>
  </Topic>
  ...
</TopicSet>
<RelationSet>
  <Relation>
     <RelationID>IS-A36</RelationID>
     <Target>T2</Target>
     <Source>T21</Source>
     <Category>
        <InferredCategory>Is-A</InferredCategory>
        <InferredBy>
            <UsedHeuristic>
                <HeuristicName>AH</HeuristicName>
            </UsedHeuristic>
        </InferredBy>
        <Confidence>0.9</Confidence>
        </Category>
     </Relation>
  ...
</RelationSet>
```

6.4 ErauzOnt

ErauzOnt (Larrañaga *et al.*, 2011) is responsible for gathering the LOs from the electronic textbook following the process presented in Chapter 5. Figure 6.2 shows the architecture of *ErauzOnt*. The *Learning Object Extractor and Generator* is the core of the *ErauzOnt* framework, and it is responsible for generating new LOs from the internal representation of the electronic textbook using the server-side web-services and the client-side modules. It uses the *CG Service* to identify the fragments of the text that may contain DRs and the *HC Service* to get the confidence of the employed heuristics. The *UKB Service* provides the resemblance for the ontology-based similarity measuring methods used to determine whether or not two DRs should be combined.

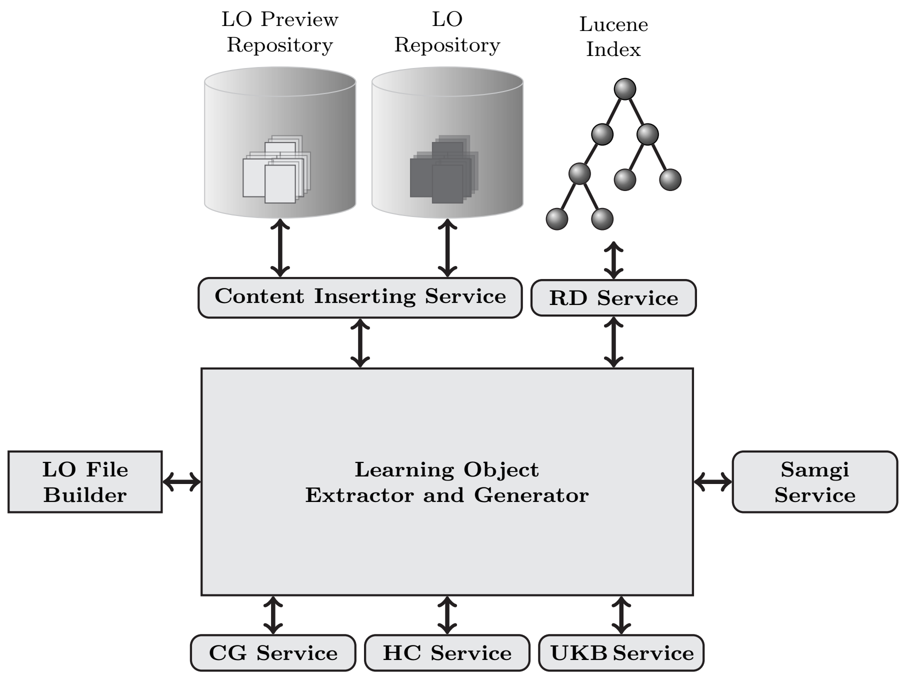


Figure 6.2 – Architecture of ErauzOnt

The *Replicate Detection Service* (*RD Service*) determines whether a document or a fragment of the document has been processed before.

The *LO File Builder* module generates the final archive for the generated LOs, as well as the preview files. Presentation formats such as *pdf*, *doc* and *odf* are suitable for final presentation, but are not appropriate for flexible content reuse, as the components cannot be easily accessed (manual *copy&paste* has been used so far). The ALOCOM framework (Verbert *et al.*, 2008) was developed to decompose composite LOs and make those components available for on-the-fly content reuse. This framework relies on the ALOCOM ontology (Verbert *et al.*, 2005), which represents a content model for LOs and their components. The generated LOs are stored in a ZIP file that contains the XML-based representation of the LO based on the ALOCOM formalism, as well as the referenced images. The XML-based ALOCOM document format allows content reusing and even versioning (Brooks *et al.*, 2005).

The *Samgi service* is responsible for generating the metadata for each LO automatically, and the *Content Inserting Service* supports the publishing of both LOs and preview files in the repositories. Both repositories are built on the ARIADNE Knowledge Pool System (Duval *et al.*, 2001; Ternier *et al.*, 2009) for the storage and retrieval of resources and their metadata.

6.5 Elkar-DOM: a Concept Map-Based Tool for Supervising the Domain Module Authoring

Elkar-DOM is a graphical tool that fulfills two main goals. On the one hand, it enables the supervision and modification of the LDO. On the other hand, it allows the instructional designers or teachers to select the most appropriate LOs for each domain topic. *Elkar-DOM* is a concept-map based tool, built on Elkar-CM (Arellano *et al.*, 2006; Elorriaga *et al.*, 2011), for supervising the *Domain Module* authoring in a collaborative way. Concept maps (Novak, 1977, 1990), which are comprised of nodes and links, are graphical means for representing and organising knowledge. The nodes of the concept map represent topics, and the links the relationships among them.

Elkar-DOM has been developed with the aim of enhancing collaboration in the domain knowledge building process. It allows synchronous collaboration based on token-passing. Several users can be working at the same time seeing the current state of the domain ontology but only one of them at a time can perform operations on it. When a user wants to modify the *Domain Module*, he or she must request the token. Once obtained, the user has a limited time to work on the ontology. The duration of the turns for edition is configurable.

Elkar-DOM addresses two kinds of users: the system administrator and the *Domain Module* authors, i.e., instructional designers or teachers. Depending on their responsibility in the authoring process, the authors may have two roles: supervisor or contributor. Each *Domain Module* has at least one supervisor and a set of contributors that collaborate on its development. The supervisors administer the collaborators of the *Domain Module* authoring, who can edit or modify it and, even, the duration of the editing turns.

6.5.1 Supervision of the Learning Domain Ontology

In the approach presented here, the *Domain Module* entails the LDO, which describes the domain topics to be mastered, and the pedagogical relationships between them. The LDO is semi-automatically gathered, i.e., domain topics and relationships are elicited from the analysed textbook, and it must be reviewed by the *Domain Module* authors to correct any error and adapt it to their preferences.

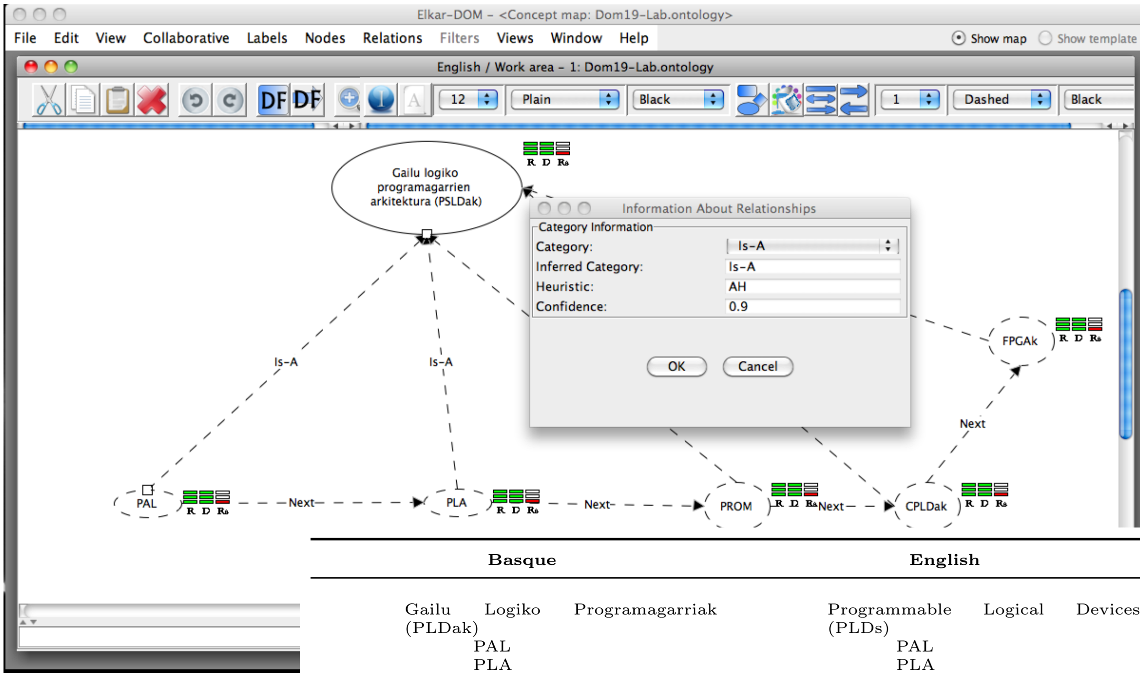
A graphical tool is essential to facilitate the inspection and refinement of the great amount of data that the gathered ontology may contain. Concept maps have been successfully used to allow knowledge elicitation and exchange (Coffey *et al.*, 2004). Therefore, they also can be a powerful tool to facilitate the work of users in *Domain Module* authoring. Moreover, Suthers (2005) observed that concept maps also facilitate the interaction in collaborative tasks, such as collaborative learning. Therefore, concept maps might be an appropriate means for *Domain Module* authors to cooperate on the supervision of the *Domain Module* authoring in the same way they collaborate to prepare the material and the schedule for their courses.

Elkar-DOM enables the collaborative supervision of the LDO gathered by the *LDO Builder*, which contains the domain topics, the pedagogical relationships and the information of the heuristics used for the acquisition.

Figure 6.3 shows a screen-capture of *Elkar-DOM* during the supervision of the LDO gathered from the illustrated outline fragment. A piece of such an LDO was presented before in Listing 6.1. As can be observed, the *Domain Module* author who holds the token is analysing a relationship that has been elicited by the *Acronyms Heuristic* (*AH*).

Elkar-DOM represents the domain topics by means of nodes and the pedagogical relationships with links between the corresponding nodes. It uses graphical resources such as color, forms and line style to provide information about each element of the LDO. For example, the node form distinguishes the kind of topic (e.g., oval nodes represent *concepts*) and dashed lines indicate that the topic or the relationship has not yet been reviewed. The category of the pedagogical relationship is specified in the label of the link. The relationships and topics identified by heuristics with low confidence are highlighted in red so that the users know which contents or relations are more likely to be changed.

Graphic flags are used to visualise, in a summarised way, information about the topic relevance, difficulty or even the quantity of LOs associated to the particular topic (when reviewing the results of the second phase of the *Domain Module* construction) without overwhelming the users with too much information. Although the *difficulty* and the *relevance* of the topics are not yet elicited from the textbooks, the



Basque	English
Gailu Logiko Programagarriak (PLDak)	Programmable Logical Devices (PLDs)
PAL	PAL
PLA	PLA
PROM	PROM
CPLD	CPLD
FPGA	FPGA

Figure 6.3 – Snapshot of Elkar-DOM

Domain Module authors can specify these values through *Elkar-DOM*, which will be able to provide the information about the inference of such information whenever the *LDO Builder* is able to deal with it.

The inferred information about a topic or a relationship is prompted when the user that has the token double-clicks on it. The presented window shows the inferred category, the heuristic that has inferred it, and the confidence on the decision (Figure 6.3). If the user has double-clicked a domain topic, the window will also contain information about the relevance and difficulty. The user cannot modify the inferred information but he or she may specify a different value for a particular characteristic by selecting it in the corresponding combobox. The concept map immediately reflects any modification either with different node shapes, relationship labels, thickness or flags. Once a relationship or a topic has been inspected, it is drawn with a continuous black line.

The acquisition of the LDO relies on a set of heuristics and their confidences for eliciting the pedagogical relationships. The initial confidence of each heuristic was empirically determined from the percentage of correctly identified relationships from a sample of document outlines or text fragments. The more precise the confidence is, the more accurate the identification of the pedagogical relationships might be. The supervision of the LDO allows dynamically adjusting the confidence of the used heuristics. Once the users have finished supervising the LDO, *Elkar-DOM* increases the correct hits of every heuristic for each correct triggering and the wrong hits for each failed triggering in the gathered LDO through the *HC service*.

In addition, *Elkar-DOM* provides two mechanisms to facilitate the work of the *Domain Module* authors and try to avoid information overload. On the one hand, the user who is editing the *Domain Module* can contract and expand parts of the LDO represented by the concept map. On the other hand, a filter mechanism that allows the users to work with parts of the LDO is available: they can supervise only the structural relationships or the sequential relationships by selecting the corresponding view. The users may also use or define their own filters to work with the domain contents they are interested in (e.g., more relevant topics . . .).

When the users finish the supervision of the LDO, it is exported to Web Ontology Language (OWL) (Bechhofer *et al.*, 2004), so that it can be used in a broader set of TSLSs. Listing 6.2 shows a fragment of a LDO in OWL.

Listing 6.2 – Fragment of the Learning Domain Ontology in OWL

```
<?xml version="1.0"?>
...
      <LearningDomain:Concept rdf:about="#T90">
        <rdf:type rdf:resource="&owl;Thing"/>
        <rdfs:label xml:lang="eu"
             >Teleskopio optiko</rdfs:label>
        <LearningDomain:hasDifficultyLevel rdf:resource="&
            LearningDomain;LOW"/>
        <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
            MEDIUM"/>
        <LearningDomain:isA rdf:resource="#T20"/>
        <LearningDomain:isA rdf:resource="#T53"/>
        <LearningDomain:hasDR rdf:resource="#DR1"/>
    </LearningDomain:Concept>
...
</rdf:RDF>
```

6.5.2 Selecting the Learning Objects

To obtain a complete *Domain Module*, the LOs to be used during the learning sessions must be provided for every domain topic. *Elkar-DOM* facilitates this task to the *Domain Module* authors, as it allows the search and retrieval of the LOs from the LOR through the *SQI Service*, built on SQI (Simon *et al.*, 2005), and SPI (Ternier *et al.*, 2008a).

Adding LOs to a particular LDO topic is quite simple. The *Domain Module* author who holds the token has to double-click on the topic node. Then, the properties of the topic are shown (see Figure 6.4). In the bottom side of the window, a table summarises the resources that have been linked with the topic, including their identifiers, the kind of LO and the path of the cached file. The *Domain Module* authors can add new LOs or, even, remove some of the current ones. In order to add new LOs, the user has to select the kind of LO he or she is interested in and click the *Find new DRs* button.

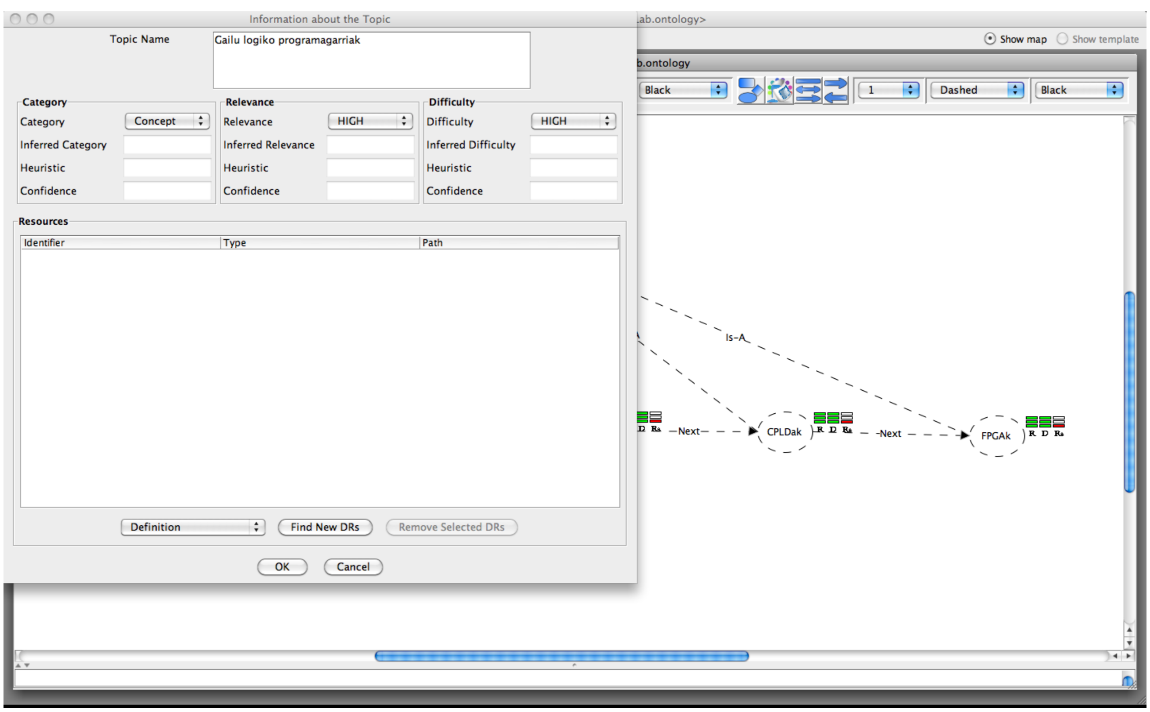


Figure 6.4 – Search of a Learning Object using Elkar-DOM

The window shown in Figure 6.5 provides the preview and the descriptive metadata of the LOs found, allowing to determine which LOs might be used to master the topic and, therefore, included in the *Domain Module*.

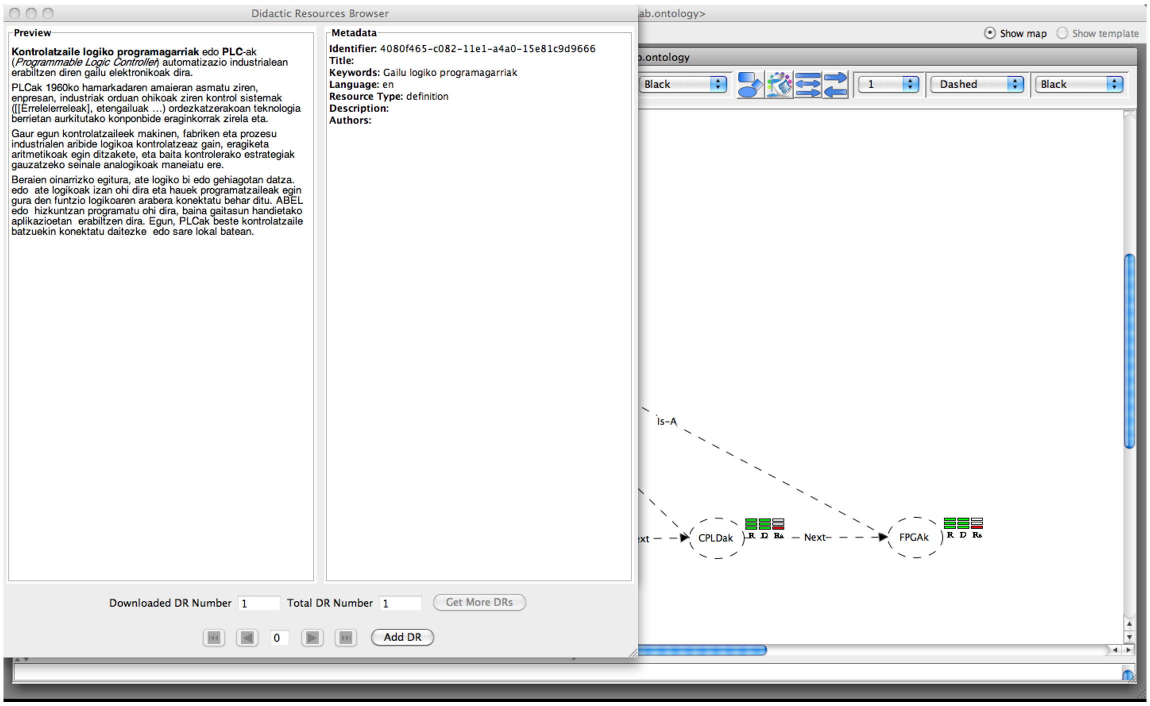


Figure 6.5 – Previewing a Learning Object with Elkar-DOM

Once the authors have selected the LOs for every domain topic, the *Domain Module* is ready to be used in TSLSs. Then, *Elkar-DOM* builds the *Domain Module* package, i.e., a zip file containing the LDO and the LOs used. Figure 6.6 shows the structure of the *Domain Module* packages.

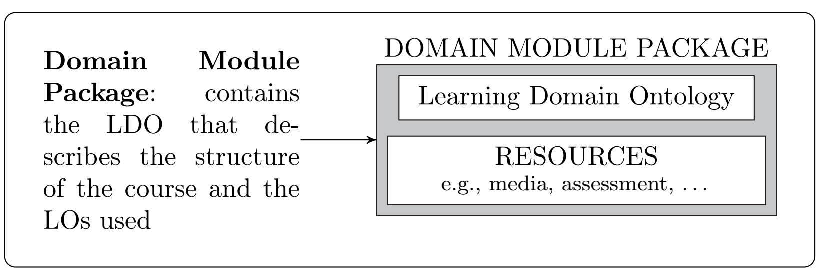


Figure 6.6 – Structure of the Domain Module Package

6.5.3 Architecture of Elkar-DOM

Elkar-DOM relies on a client-server architecture that entails the *Elkar-DOM Server* and two kinds of clients: the *Domain Module Authoring Tools*, which is used for building *Domain Modules* collaboratively and the *Server Management Client* (Figure 6.7).

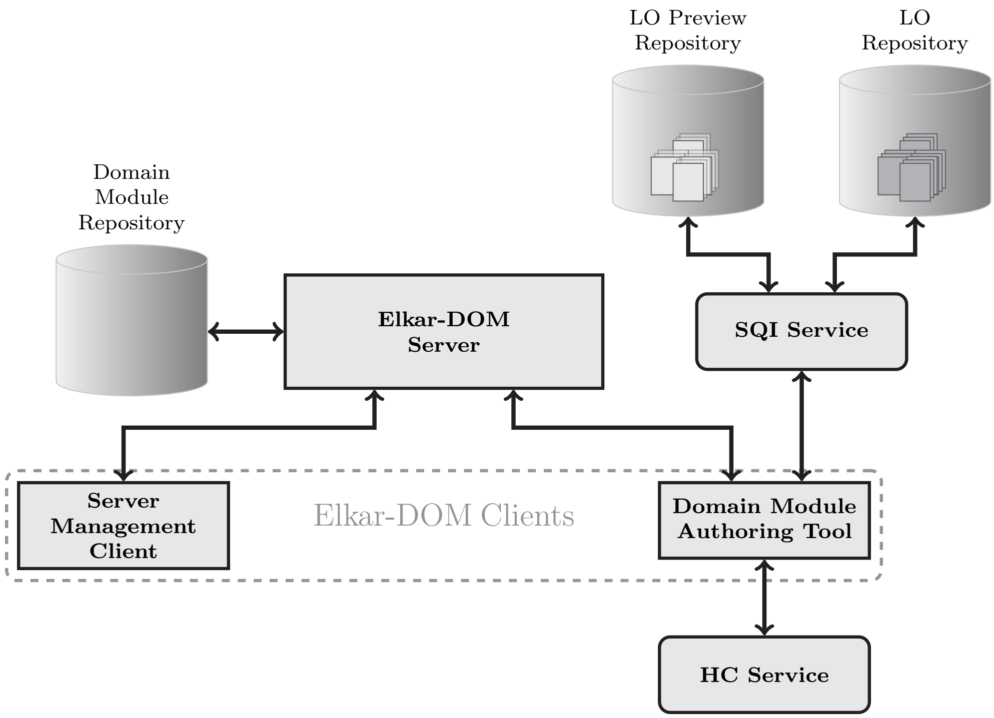


Figure 6.7 – Architecture of Elkar-DOM

The *Elkar-DOM Server* is the module responsible for knowledge sharing between the community of users. It manages the *Domain Module Repository*, which stores the *Domain Modules*, and supports the communication among the users facilitating their collaboration. The *Elkar-DOM Server* communicates any modification performed on the LDO to the *Domain Module Authoring Tools* so that they always show the updated information. Besides, it sends an updated LDO to any *Domain Module Authoring Tool* that joins an domain authoring session.

The *Server Management client* allows to control the server remotely. The administrator can use this application to manage the users of the system, manage the group of authors of a *Domain Module*, configure the server (e.g., specifying the time limit for the editing turns), and stop the server. Only one instance at a time of this application can be running, although it is not required to be running for *Elkar-DOM* to work.

The *Domain Module Authoring Tool* allows the collaborative edition of the *Domain Module*, facilitating both the supervision of the LDO and selecting the most appropriate LOs. Besides the functionality described above, the *Domain Module Authoring Tool* provides a chat tool to enable the synchronous communication among the *Domain Module* authors, and supports the asynchronous communication by means of notes, which can be added to either nodes or links.

The development of the *Domain Module* is an incremental process in which the ontology is continuously refined and modified as a result of the argumentations that the participants state. Therefore, it is very important to be able to represent, in an intuitive and inspectable way, these argumentations. Elkar-DOM is able to record the stages during the domain ontology development and provide a dynamic reproduction that shows the evolution of the *Domain Module*. This feature provides the user with a means to replay the different snapshots of the LDO using video-like buttons. Besides, the set of chat messages corresponding to each operation can be reviewed.

The *Domain Module Authoring Tool* can search and retrieve LOs for the domain topics from the repositories through the *SQI Service*.

6.6 Extendability of DOM-Sortze

Currently *DOM-Sortze* supports the construction of the *Domain Module* from electronic textbooks in *pdf* format written in Basque. However, it is not tightly coupled to a particular language or electronic document format and, therefore, it is flexible enough to provide support for new languages and document formats without major effort.

The only document-format dependent step is the construction of the internal document and outline representations, which is performed during the preprocess. Document formats such as *doc*, *docx* and *odf* use XML internally to store the information. XML files have an inherent tree-structure, so taking this into account, and the availability of libraries to work with these kinds of documents, tree builders for other document formats can quite easily be plugged into the *DIR Builder*, which would build the internal representation of the document from those kinds of documents.

To support a new language, the discourse markers for that language and new versions of the grammars and the heuristics for analysing the outline must be developed. Both the *LDO Builder* and *ErauzOnt* used the appropriate resources based on the language.

Currently, the *NLP Analysis Service* only supports the Basque language, which is carried out using the EUSLEM lemmatiser/tagger for Basque (Aduriz *et al.*, 1996). However, this web-service can be enhanced to use new NLP parsers for the new languages with minor effort. Once again, the language of the document will determine which tool would be used. For example, using Freeling (Atserias *et al.*, 2002) could be used to support English, Spanish and Catalan. The Stanford Parser (Klein and Manning, 2002, 2003) has been used for English, German, Arabic and even Chinese (Levy and Manning, 2003), so it can also significantly contribute to support new languages.

6.7 Summary

In this chapter, the architecture of *DOM-Sortze* and its components have been described. *DOM-Sortze* has been designed and developed with the aim of facilitating the semi-automatic development of *Domain Modules* following a domain-independent approach. Moreover, it is not coupled to a particular language and can be extended to support additional document formats and languages. *DOM-Sortze* has a service-oriented architecture which entails four applications. The *preprocessor* is responsible for preparing the analysed textbook for the *Domain Module* acquisition. The *LDO Builder* enables the acquisition of the LDO from textbooks and *ErauzOnt* builds LOs from the document. Finally, *Elkar-DOM* facilitates the supervision and completion of the *Domain Module* to *Domain Module* authors.

7 Evaluation of DOM-Sortze

Previous chapters described the proposed approach for building the *Domain Module* from electronic textbooks. *DOM-Sortze*, a suite of applications and web-services was developed to support the proposed *Domain Module* construction procedure.

This chapter presents the most relevant experiments conducted for evaluating the proposed approach. First, the evaluation of the performance of the *LDO Builder* on outlines is presented. Next, the evaluation of *ErauzOnt*, the subsystem that extracts LOs from the textbooks is described. Finally, the assessment of *DOM-Sortze*, which carries out the whole process for building the *Domain Module*, is reported.

7.1 Introduction

The development of the *Domain Module* using *DOM-Sortze* entails several tasks: preparing the textbook for the process, gathering the LDO which is carried out by analysing both the outline and the whole textbook, and the generation of LOs from the textbook. *DOM-Sortze* was developed incrementally, coping with the acquisition of the LDO from the outline of the textbooks first, later the extraction of LOs from the documents and, finally, the whole *Domain Module* construction process.

At each stage of the development of *DOM-Sortze*, an evaluation was carried out to verify the correct performance and to measure how much it can help the authors when developing new *Domain Modules*. Every conducted evaluation has been performed using the *Gold Standard* approach using the LDOs and sets of DRs defined by instructional designers as reference.

7.2 Evaluation of the LDO Builder on Outlines

The *LDO Builder* is the subsystem responsible for eliciting the LDO from electronic textbooks. It gathers the contents of the LDO from both the outline and the whole textbook. Nevertheless, at the time this evaluation was conducted, the LDO only supported the analysis of the outlines of documents and, therefore, only that feature was tested.

The acquisition of the LDO from the document outlines works under the assumption that every outline item represents a unique domain topic and relies on the use of heuristics for the identification of pedagogical relationships. In order to verify the appropriateness of the proposal, the performance of the heuristic-based acquisition of the pedagogical relationships of the LDO from the outlines was tested. This evaluation was conducted on 150 outlines of different subjects offered in several grades at the University of the Basque Country (UPV/EHU), measuring the automatically identified pedagogical relationships. Every analysed index was preprocessed to get a homogenised internal representation of the outline (*cf.* Section 3.2) using a graphical tool that also allows the correction of errors before the analysis.

The automatically gathered pedagogical relationships were evaluated by comparing them to the pedagogical relationships defined in the LDOs collaboratively developed by three instructional designers. In order to simplify the evaluation, the instructional designers were requested to use only the domain topics referred in the outlines for building their agreed LDOs. The pedagogical relationships were, therefore, restricted to relationships among those topics.

Table 7.1 – Summary of the Learning Domain Ontology Relationships in the Analysed Outlines

	Structural		**Sequential**		
	partOf	***isA***	***next***	**prereq.**	**Total**
Real	3,668	325	2,827	468	**7,288**
Found	3,704	289	2,834	445	**7,272**
Correct	3,637	258	2,823	435	**7,153**

Table 7.1 summarises the information about the acquisition of the pedagogical relationships from the textbook outlines. The LDOs built from the analysed outlines contained 7,288 pedagogical relationships (3,668 *partOf*, 325 *isA*, 2,827 *next*, and

468 *prerequisite*). The *LDO Builder* correctly identified 7,153 of the 7,272 found pedagogical relationships (3,637 of 3,704 *partOf*, 258 of 289 *isA*, 2,823 of 2,834 *next*, and 435 of 445 *prerequisite*).

Table 7.2 presents the statistics on the generation of the LDOs from the analysed outlines, including the *recall*, the *precision* and the *f-measure*, i.e., the harmonic mean of the other two measures. The *LDO Builder* achieved an overall *recall* of 98.15%, and 98.36% *precision*, hence, the *f-measure* was 98.26%. As can be observed, the results are quite similar for *partOf* and *next* relationships, while they are a little lower for *prerequisite*, and especially *isA* relationships, which require deeper domain knowledge in order to be identified. However, the lowest values are 79.38% for *recall* and 89.27% for *precision*.

Table 7.2 – Statistics of the Automatic Acquisition of the Learning Domain Ontology Relationships from the Outlines

	Structural		**Sequential**		
	partOf	***isA***	***next***	**prereq.**	**Total**
Recall (%)	99.15	79.38	99.86	92.95	**98.15**
Precision (%)	98.19	89.27	99.61	97.75	**98.36**
F-measure (%)	98.67	84.04	99.74	95.29	**98.26**

To analyse the results in more detail, the outlines were classified into three main knowledge areas to determine if the domain affects the outcomes of the LDO acquisition. The identified areas are:

- **Engineering**: Outlines corresponding to subjects of computer or industrial engineering.

- **Economics**: Outlines corresponding to economics and business management, including legal aspects.

- **Social**: Outlines corresponding to social sciences such as philosophy, psychopedagogy, etc.

As can be observed in Figure 7.1, the overall *recall*, *precision*, and *f-measure* rates are quite similar in the three areas, but significant variation was found in *isA* relationships. Although the heuristics for identifying the *isA* relationship got high *precision* rates, the *recall* for engineering outlines was much lower, as instances or

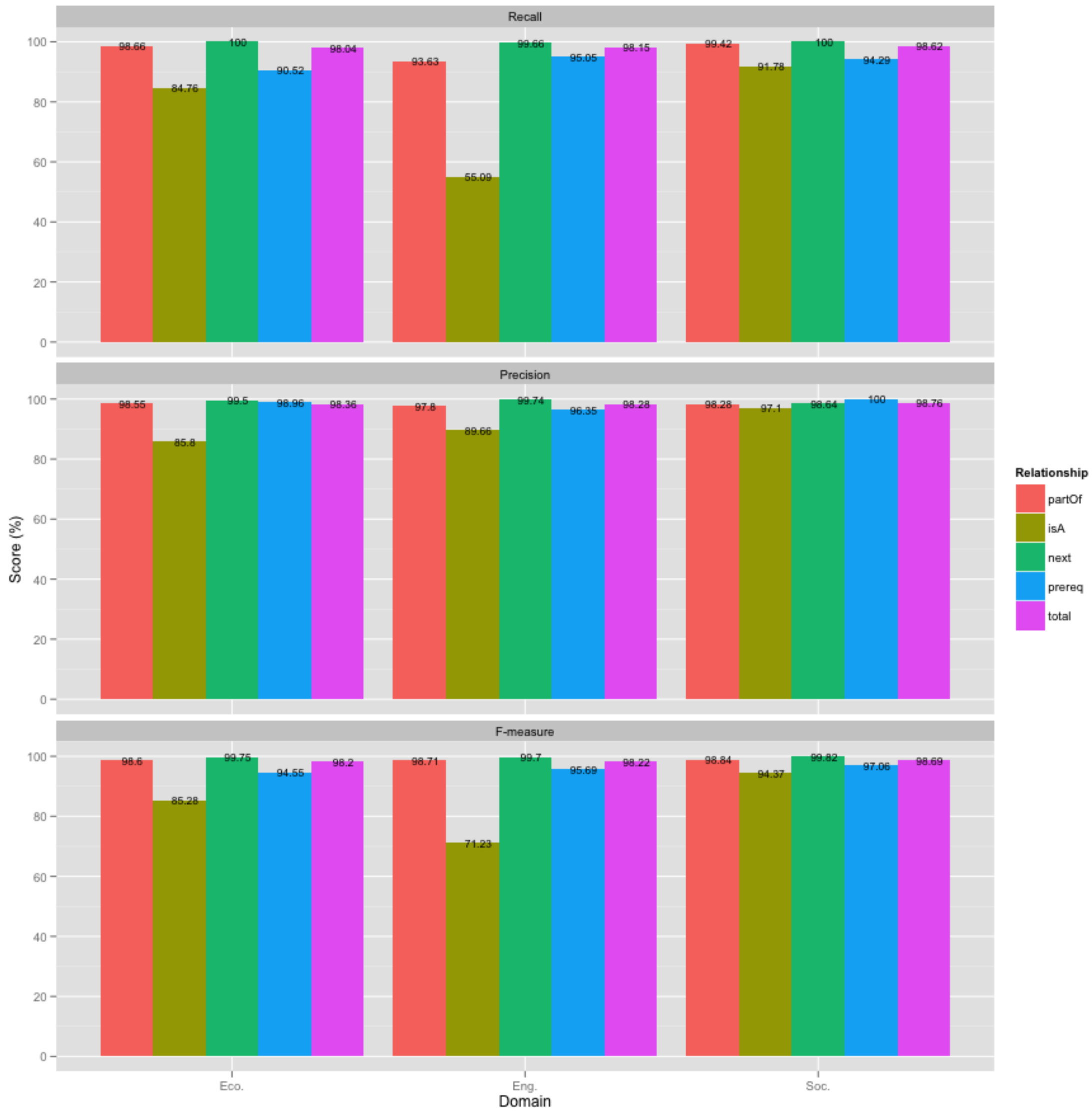


Figure 7.1 – Statistics of the Acquisition of the Learning Domain Ontology per Domain

particular cases of a topic were represented by means of common nouns instead of using proper nouns, acronyms or multi-word terms.

The LDO ontologies used as reference for the evaluation were limited to the topic in the outlines and, hence, there were relationships that were not defined in the

LDOs, which facilitates the high *recall* scores obtained. Nevertheless, the remarkable *precision* achieved proved that the outline analysis is an accurate and appropriate means to elicit pedagogical information from electronic textbooks.

7.3 Evaluation of ErauzOnt

ErauzOnt, the framework for gathering the LOs from electronic textbooks, was tested with the aim of validating its performance. *ErauzOnt* uses a DR grammar, to identify meaningful fragments of the document, and similarity measuring methods that determine the resemblance of DRs, to decide if two DRs should be combined or not (Chapter 5). A preliminary test was conducted to determine which similarity measures produced more accurate results in order to set up *ErauzOnt*.

Four textbooks provided by the Euskal Herriko Ikastola (EHI)[1], one of the main publishers for Basque-medium education at all educational levels in compulsory education in the Basque Country, were used for the evaluation of *ErauzOnt*. The textbooks are used in the *Nature Sciences* subject in the first course of secondary education. They cover different aspects of the subject, from geology to biology. The first book has 36 pages and 7,900 words, the second 20 pages and 5,100 words, the third 29 pages and 7,800 words, and the fourth 21 pages and 5,500 words. The evaluation was aimed at the assessment of text-based LO extraction, even though *ErauzOnt* also supports the identification and extraction of LOs containing images. Therefore, adapted versions of the textbooks in which the images were removed were analysed instead of the original ones.

The experiment for evaluating *ErauzOnt* was carried out in the following way: four instructional designers manually analysed the electronic textbooks, and collaboratively defined the LDOs that describe the learning domain for each document. The LDOs included the topics to be mastered and the pedagogical relationships among the topics. Once the instructional designers had reached a consensus on the LDOs, they were requested to identify and classify the fragments of the documents related to the domain topics. The instructional designers collaboratively identified *definitions*, *principle statements*, *examples*, *problem statements*, and *combined resources*, i.e., composite resources that entail more than one kind of DR. The identified set of DRs was then used as the reference for the evaluation of the performance of *ErauzOnt*, which relied on the LDOs defined by the instructional designers for gathering the LOs from the textbooks.

[1] http://www.ikastola.net

In the conducted experiment two aspects of the generation of the LOs were assessed. On the one hand, the adequacy of the DR grammar used for the identification of the relevant text fragments was measured. On the other hand, the appropriateness of the generated LOs was tested against the set of DRs defined by the instructional designers.

Next, the preliminary test conducted to set up *ErauzOnt* and the results of the evaluation of *ErauzOnt* are presented.

7.3.1 Preliminary experiments. Setting up the system

Before using *ErauzOnt* for the extraction of LOs from electronic textbooks, it has to be set up. *ErauzOnt* requires a DR grammar and the similarity measures that are used for the combination of consecutive related LOs into new composite LOs. Although *ErauzOnt* is configurable, a preliminary experiment was conducted to determine which are the most accurate similarity measuring methods and, therefore, which methods will be used by default. This preliminary test also allowed to determine the default values of the thresholds used to determine the similarity of the DRs, etc.

The experiment was conducted over a sample of documents and the results were reported in (Larrañaga *et al.*, 2008a,c). Table 7.3 and Table 7.4 show the results of the evaluation of the similarity measuring methods described in Section 5.3.1. As can be observed, all the combinations between the similarity measuring methods were proved. The column *SRTM* details the results obtained with the *Same Resource Type Method* while the *DOM* column contains the performance of the *Didactic Ontology Method*. The rows of *Same Topics Method*, *Share Topic Method*, *Cosine Method* and *Ontology Based Method* detail the results for those topic similarity measuring methods. Table 7.3 includes the aspects considered positive for the methods, i.e., the percentage of DRs for which the related topics were correctly identified (*Topics %*) and the percentage of DRs that were considered appropriate by the instructional designers, who considered both the content and if *ErauzOnt* correctly determined the type of DR, (*Valid DR %*). Table 7.4 shows the negative aspects, i.e., the percentage of DRs that should be enhanced by combining them with preceding fragments (*Enhanced %*) and the percentage of DRs that were joined and should be split (*Split %)*.

The results of *Same Resource Type Method* (*SRTM*) combined with any of the content similarity measuring methods are quite satisfying if the *Valid DR %* and the *Topic %* scores are considered. The obtained scores for *Topic %* range between 82.63%, which is achieved by the *Same Topics Method*, and 91.49%, which was ob-

Table 7.3 – Performance of the Similarity Measuring Methods. Positive Aspects.

		SRTM[a]	**DOM**[b]
Same Topics Method	**Topic %**	82.63	88.64
	Valid DR %	82.63	83.57
Shared Topic Method	**Topic %**	91.49	92.50
	Valid DR %	91.40	94.17
Cosine Method	**Topic %**	88.89	90.00
	Valid DR %	88.02	90.83
Ontology Based Method	**Topic %**	90.29	92.22
	Valid DR %	91.03	92.86

[a] SRMT: *Same Resource Type Method*
[b] DOM: *Didactic Ontology Method*

tained by the *Share Topic Method*. The *Valid DR %* varies between 82.63% (*Same Topics Method*) and 91.40% (*Share Topic Method*). The *Split %* was quite low, ranging from 0.00% for the *Same Topics Method* to 5.41% for the *Ontology Based Method*. However, it was observed that many of the obtained DRs should be enhanced (12.56% - 32.39%) when comparing the results with the instructional designers output.

The *Didactic Ontology Method* (*DOM*) obtained better results compared with the *SRTM*. All the content similarity measuring methods improved the results, either by getting a better *Valid DR %* and *Topic %* rates or by reducing the *Enhanced %* and *Split %* rates. The *Cosine Method* and the *Share Topic Method* obtained better classification rates (*Valid DR %* and *Topic %*). However, the DRs composition and organisation was not the most accurate, i.e., they got higher negative scores (*Split %* and *Enhanced %*). The *SRTM* proved to be too restrictive; it considers two DRs similar if they both are the same kind. However, *DOM* uses a Didactic Ontology to decide if two DRs are similar and allows more DRs to be composed.

The most precise method was the *Ontology Based Method* for measuring content similarity combined with the *Didactic Ontology Method*. Even though it did not achieve the *Valid DR %* score of the *Shared Topics Method* (92.86% vs. 94.17%), the obtained DRs better suit the preferences of the human instructors as it can be deduced from the low *Enhanced %* (7.59% vs. 9.87%) and *Split %* rates (4.12% vs. 6.90%). The *Ontology Based Method* not only considers topic references but also the semantic relationships among the topics. In the same way, the *Didactic Ontology*

Table 7.4 – Performance of the Similarity Measuring Methods. Negative Aspects.

		SRTM[a]	**DOM**[b]
Same Topics Method	**Enhanced %**	32.39	30.39
	Split %	0.00	0.00
Shared Topic Method	**Enhanced %**	14.03	9.87
	Split %	3.92	6.90
Cosine Method	**Enhanced %**	19.35	17.43
	Split %	4.48	7.48
Ontology Based Method	**Enhanced %**	12.56	7.59
	Split %	5.41	4.12

[a] SRMT: *Same Resource Type Method*
[b] DOM: *Didactic Ontology Method*

Method considers the closeness among the different kinds of DRs. The combination of these two similarity measuring methods provides DRs with high cohesion and fit output of the human instructors in a better way. Therefore this combination is going to be used in the following experiments.

7.3.2 Evaluation of the DR Grammar

The atomic LOs, the finer grained LOs built from the identified atomic DRs, were checked as a means to assess the DR grammar and, thus, the pattern-based approach for gathering the LOs from electronic textbooks.

Table 7.5 – Accuracy of the DR Grammar

	Def.	**Prin.**	**Examp.**	**Theo.**	**Prob.**	**Total**
Rule Firing	103	22	35	1	342	**504**
Correct	43	10	10	0	308	**371**
Accuracy (%)	41.75	45.45	28.57	0.00	90.06	**73.61**

Table 7.5 summarises the information about the performance of the DR grammar. In this experiment, the rules of the DR grammar fired 504 times; the rules for *definitions* triggered 103 times, the *principle statement* rules 22 times, the *example* rules 35 times, the *theory* rules once, and finally the *problem statement* rules 342

Table 7.6 – Statistics on the Identified Atomic Learning Objects

	Def.	**Prin.**	**Examp.**	**Theo.**	**Prob.**	**Fact**	**Total**
Found	106	22	37	1	341	1	**508**
Correct	46	10	13	0	305	0	**374**
%	43.40	45.45	35.14	0.00	98.44	0.00	**73.62**

times. Although the DR grammar can also identify *facts*, none of the rules that recognises such a kind of DRs triggered in this experiment.

371 of the triggered rules performed properly, i.e., they identified the correct kind of DR, so 73.61% of the gathered text fragments contained valid DR fragments and were correctly categorised. *Problem statement* rules obtained the highest success rate, as 90.06% (308 of 341) of the rule activations found valid *problem statement* fragments. *Problem statements* mostly entail imperative sentences, which in the Basque language are expressed by an auxiliary verb, or sentences that begin with verb phrases such as *Erantzun ezazu galdera ...* (*Answer this/these question/s*). These patterns make *problem statements* easy to find. However, *examples* and other kinds of DRs rely on patterns that recognise syntactic structures that might be found in sentences without educational value, so they are more susceptible to fail and, therefore, have lower success rates. *Definition* rules correctly fired 41.75% (43 of 103) of the times, *principle statements* 45.45% (10 of 22), and *examples* 28.57% (10 of 35) of the times. *Theory* rules fired one time, but the instructional designers did not identify *theorem* or *theory* descriptions, so these rules had a success rate of 0%.

The atomic LOs were inspected to obtain the percentage of correctly identified fragments. In fact, more than one rule can trigger in one sentence, and the kind of DR is determined by the most confident ones. The accuracy of the atomic LOs, 73.62%, was quite similar as can be observed in Table 7.6.

7.3.3 Evaluation of the Gathered Learning Objects

Finally, the gathered LOs were evaluated. This evaluation was carried out comparing the manually identified DRs with those that were automatically elicited. Many of the manually identified DRs also were composite fragments that contained finer grain resources. An aspect to be considered when evaluating the gathered LOs is that while a LO might be the most appropriate in a particular context, one of its components or a more complex LO (a composite LO that comprises it) might fit better and, therefore be more reusable, in other situations.

Table 7.7 – Distribution of the Didactic Resources Identified by the Instructional Designers

	Def.	**Prin.**	**Examp.**	**Prob.**	**Comb.**	**Total**
Quantity	67	12	8	105	37	**229**

The instructional designers found 229 LOs in the analysed electronic textbooks, 67 *definitions*, 12 *principle statements*, 8 *examples*, 105 *problem-statements*, and 37 *combined* LOs (see Table 7.7). The *combined* LOs aggregate LOs of different types, while the rest only contain resources of one kind, either atomic or composite. The instructional designers did not recognise any *theory* description or *fact*, which *ErauzOnt* is also capable of recognising.

Table 7.8 – Recall of the Identified Learning Objects

	Def.	**Prin.**	**Examp.**	**Prob.**	**Comb.**	**Total**
Real	67	12	8	105	37	**229**
Found	40	6	7	86	22	**161**
Recall (%)	59.70	50.00	87.50	81.90	59.46	**70.31**

Table 7.8 shows the information about the *recall* of the LO acquisition, i.e., the percentage of the DRs identified by the instructional designers and also recognised by *ErauzOnt*. *ErauzOnt* correctly identified 40 *definitions* (59.70%), 6 *principle statements* (50%), 7 *examples* (87.50%), 86 *problem-statements* (81.90%), and 22 *combined* LOs (59.46%). The overall *recall* achieved in this experiment was 70.31%.

Moreover, the *precision*, i.e., the percentage of the LOs gathered which were considered valid, was measured. The information about the *precision* can be observed in Table 7.9. 94.15% of the *definitions*, 96.30% of the *principle statements*, 100% of the *examples*, 85.55% of the *problem-statements*, and 97.84% *combined* LOs were valid resources, so *ErauzOnt* achieved a global *precision* of 91.88% .Although one rule for recognising *theories* fired in the experiments, another rule with higher confidence that identifies another kind of LO triggered in the same fragment. Therefore, no *theory* LOs were gathered from the analysed textbooks.

Table 7.10 summarises the statistics of the conducted experiment showing the *recall*, the *precision*, and the *f-measure*. The achieved *f-measure* was 72.14% for *def-*

Table 7.9 – Precision of the Identified Learning Objects

	Def.	**Prin.**	**Examp.**	**Prob.**	**Comb.**	**Total**
Found	158	27	34	664	324	**1207**
Correct	144	26	34	588	317	**1109**
Precision (%)	91.14	96.30	100.00	88.55	97.84	**91.88**

initions, 65.82% for *principle statements*, 93.33% for *examples*, 85.10% for *problem-statements*, and 73.97% for *combined* LOs, which result in a global 79.66% score.

Table 7.10 – F-Measure of the Identified Learning Objects

	Def.	**Prin.**	**Examp.**	**Prob.**	**Comb.**	**Total**
Recall (%)	59.70	50.00	87.50	81.90	59.46	**70.31**
Precision (%)	91.14	96.30	100.00	88.55	97.84	**91.88**
F-measure (%)	72.14	65.82	93.33	85.10	73.97	**79.66**

7.4 Evaluation of DOM-Sortze

DOM-Sortze was tested, with the intention of validating it, with a textbook provided by the Euskal Herriko Ikastola (EHI)[2]. The textbook used, *Izarrei Begira* (*Looking at the Stars*), is one of a five book series used in the *Nature Sciences* subject in the first course of secondary education. The main goal of this experiment was to evaluate how *DOM-Sortze* helps the authors to build the *Domain Module* by measuring how much knowledge was automatically elicited from the textbook, both to define the LDO or to generate the LOs. For the experiment, only text-based LOs were considered. Hence, an adapted version of the electronic textbook in which the images were removed was used. The analysed textbook has 30 pages and 8,495 words. The outline of the document has two levels: 3 main items, each having 5 subitems, except for the second item, which has 4 subitems.

In order to evaluate the process of the generation of the *Domain Module* using *DOM-Sortze*, a reference LDO and DR set were needed to compare the obtained results. Therefore, three instructional designers collaborated to manually develop the LDO and identify fragments of the documents to be used as DRs for the identified

[2] http://www.ikastola.net

domain topics. The instructional designers reached a consensus about the relevant domain topics and the pedagogical relationships among them in order to define the LDO, as well to determine which DRs should be used.

The process for generating the LDO was evaluated based on the amount of automatically gathered knowledge, i.e., domain topics and pedagogical relationships and the correctness of the proposed topics and relationships. The contents of the LDO were gathered first from the outline of the textbook and, then, from the whole document. In this evaluation, the gathered LDO elements - topics and relationships - were examined. The details of this evaluation are presented in Section 7.4.1.

The evaluation of the LO generation considered both the adequacy of the identified LOs and the *recall* of the LO acquisition process, to which end the automatically gathered LOs were compared to the manually identified DRs. This last evaluation is presented in Section 7.4.2.

7.4.1 Evaluation of the Gathered Learning Domain Ontology

The evaluation of the LDO acquisition allowed to measure the appropriateness of the proposed approach, testing the performance of (1) the heuristics for eliciting pedagogical relationships from the document outline, (2) the identification of topics, and, (3) the grammar that allows recognising pedagogical relationships from the whole text.

Table 7.11 – Content Distribution of the Learning Domain Ontology Developed by Instructional Designers

		Level 1	**Level 2**	**Total**
Topics		38	45	**83**
Relations		108	27	**135**
	partOf	36	7	**43**
	isA	51	15	**66**
	prerequisite	13	0	**13**
	next	8	5	**13**

The evaluation of the construction of the LDO was carried out by comparing the automatically identified domain topics and pedagogical relationships to the reference LDO, i.e., the LDO collaboratively defined by the instructional designers. The reference LDO entailed 83 domain topics and 135 pedagogical relationships (43 *partOf*, 66 *isA*, 13 *prerequisite* and 13 *next*). The instructional designers also classified the LDO elements in two levels according to their relevance. Level 1 entailed the most relevant domain topics as well as all the pedagogical relationships in which at least one of the topics is Level 1. Table 7.11 shows the LDO content distribution: 38 of the 83 topics and 108 of the 135 relationships were considered Level 1 elements. The rest were classified as Level 2 elements. The classification of the topics and relationships in two levels was aimed at observing if *DOM-Sortze* was more accurate on the relevant topics and relationships or not.

Table 7.12 – Recall of the Learning Domain Ontology Topics

	Level 1 (%)	Level 2 (%)	Total (%)
Outline	28.95	13.33	**20.48**
Whole Document	63.16	71.11	**67.47**
Total	92.11	84.44	**87.95**

As can be observed in Table 7.12, 87.95% of the LDO topics were automatically gathered from the electronic document, 20.48% by the outline analysis and 67.47% from the whole document analysis. It is remarkable that the *recall* of the most relevant topics, Level 1 topics, rises up to 92.11%. In addition, it can be observed that most of the topics are detected when analysing the whole document. This issue is influenced by the simplicity of the outline, which entailed 3 items, which have 5 subitems each, with the exception of the second item, which has 4 subitems. The structure of the outline, as will be explained later, also limited the identification of the pedagogical relationships.

Table 7.13 – Precision of the Learning Domain Ontology Topics

	Level 1 (%)	Level 2 (%)	Total (%)
Outline	100.00	100.00	**100.00**
Whole Document	100.00	14.18	**16.36**
Total	100.00	14.60	**17.48**

The *precision* of the LDO identification process is summarised in Table 7.13. The LDO acquisition achieved a 17.48% *precision*, being the score 100.00% for Level 1 topics. In order to distinguish Level 1 topics from Level 2 topics, the classification stated by the instructional designers was used. Any proposal that was not considered by the instructional designers as a Level 1 topic, was labeled as a Level 2 topic for these statistics. The *recall* of the LDO topics (Table 7.12) was given priority over the *precision* in order to facilitate the acquisition of the pedagogical relationships, so many topic proposals were identified and, therefore, the *precision* score was low.

Table 7.14 presents the *f-measure* of the elicitation process of the LDO topics. Although a 29.19% score was achieved, a high *f-measure* score for Level 1 topics, 95.65%, was obtained. Both the *recall* and the *precision* were remarkably high for the most relevant domain topics, which suggests that the *LDO Builder* is able to accurately identify, at least, the most relevant topics,

Table 7.14 – F-Measure of the Learning Domain Ontology Topics

	Level 1 (%)	Level 2 (%)	Total (%)
Outline	44.90	23.53	**34.00**
Whole Document	77.42	23.65	**26.34**
Total	95.85	24.89	**29.16**

The identification of the pedagogical relationships was measured both by kind of relationship and by considering all the relationships together. Table 7.15 shows the information about the recall of the identification of the *partOf* relationships. 55.81% of the *partOf* relationships defined in the LDO were automatically identified, 37.21% from the document outline, and 18.60% from the whole document. The percentage of identified Level 1 *partOf* relationships was 58.33%, 36.11% from the document outline, while 22.22% were found in the whole document.

Table 7.15 – Recall of the *partOf* Relationship

	Level 1 (%)	Level 2 (%)	Total (%)
Outline	36.11	42.86	**37.21**
Whole Document	22.22	0.00	**18.60**
Total	58.33	42.86	**55.81**

The accuracy of the heuristics that allow the identification of the *partOf* relationships is summarised in Table 7.16. All the triggered heuristics correctly identified the *partOf* relationships, either Level 1 relationships or Level 2 relationships. As can be observed, there was no proposal of Level 2 *partOf* relationships on the analysis of the whole document.

Table 7.16 – Precision of the *partOf* Relationship

	Level 1 (%)	Level 2 (%)	Total (%)
Outline	100.00	100.00	**100.00**
Whole Document	100.00	N/A	**100.00**
Total	100.00	100.00	**100.00**

The acquisition of the *partOf* relationships is summarised in Table 7.17, which shows the *f-measure*, i.e., the harmonic mean of the scores presented above. The achieved *f-measure* score was, 71.67%, being higher for the score of Level 1 relationships compared with the score of Level 2.

Table 7.17 – F-Measure of the *partOf* Relationship

	Level 1 (%)	Level 2 (%)	Total (%)
Outline	53.06	60.00	**54.24**
Whole Document	36.26	N/A	**31.37**
Total	73.68	60.00	**71.64**

Table 7.18 – Recall of the *isA* Relationship

	Level 1 (%)	Level 2 (%)	Total (%)
Outline	0.00	0.00	**0.00**
Whole Document	27.45	40.00	**30.30**
Total	27.45	40.00	**30.30**

As can be observed in Table 7.18, 30.30% of the 66 *isA* relationships were identified, all of them from the whole document. No *isA* relationship was identified from the document outline, mainly due to the simplicity of the outline. The heuristics achieved a 50.00% *precision* for Level 1 *isA* relationships and 77.78% for Level 2 relationships, which results in a 56.10% overall *precision* (see Table 7.19).

The *f-measure* for the *isA* relationships was much lower than the score for *partOf*. On the one hand, the accuracy of the heuristics for the *isA* is lower because domain knowledge is usually required to realise that one topic is an example or an instance of another one. On the other hand, the size of the analysed textbook and the simplicity of the outline might also have affected the *recall* of this kind of relationships. Had the outline been richer, the *recall* for the *isA* would have been higher considering the accuracy of the heuristics for identifying the pedagogical relationships from the outline.

Table 7.19 – Precision of the *isA* Relationship

	Level 1 (%)	Level 2 (%)	Total (%)
Outline	N/A	N/A	**N/A**
Whole Document	50.00	77.78	**56.10**
Total	50.00	77.78	**56.10**

Table 7.20 – F-measure of the *isA* Relationship

	Level 1 (%)	Level 2 (%)	Total (%)
Outline	N/A	N/A	**N/A**
Whole Document	34.44	52.83	**39.35**
Total	34.44	82.83	**39.35**

Although most of the relationships could be identified either from the document outline or the whole document, the *next* relationship, which expresses that a topic is recommended to be learnt just after mastering another particular topic, is observable in the document outline but more difficult to identify in the whole document. In fact, no rule for identifying *next* relationships was defined in the relationship grammar, so that relationship can be only identified from the document outline. Nevertheless, the identification of this kind of relationships was satisfactory. 61.54% of the defined *next* relationships were identified, 75% of the Level 1 *next* relationships (see Table 7.21).

66.67% of the proposed *next* relationships were correct. The *precision* for Level 1 *next* relationships was 60.00% while the *precision* for Level 2 was 100.00% (see Table 7.22).

Table 7.21 – Recall of the *next* Relationship

	Level 1 (%)	Level 2 (%)	Total (%)
Outline	75.00	40.00	**61.54**
Whole Document	0.00	0.00	**0.00**
Total	75.00	40.00	**61.54**

Table 7.22 – Precision of the *next* Relationship

	Level 1 (%)	Level 2 (%)	Total (%)
Outline	60.00	100.00	**66.67**
Whole Document	N/A	N/A	**N/A**
Total	60.00	100.00	**66.67**

The acquisition of the *next* relationships is summarised in Table 7.23, which presents the *f-measure* for this kind of relationship. A 64.00% score was obtained, being higher for the Level 1 relationships than for Level 2, despite the fact that the *precision* was lower for the former. The heuristics for identifying sequential relationships wrongly classified some *prerequisite* relationships as *next*. The *prerequisite* relationship also requires domain knowledge that might not be explicitly expressed in the analysed text fragments and, therefore, the heuristics applied on the outline analysis may fail to classify them.

Table 7.23 – F-measure of the *next* Relationship

	Level 1 (%)	Level 2 (%)	Total (%)
Outline	66.67	57.14	**64.00**
Whole Document	N/A	N/A	**N/A**
Total	66.67	57.14	**64.00**

23.08% of the *prerequisite* relationships were identified, all of them from the document outline (see Table 7.24). All the *prerequisite* relationships identified by the instructional designers were Level 1.

Table 7.24 – Recall of the *prerequisite* Relationship

	Level 1 (%)	Level 2 (%)	Total (%)
Outline	23.08	N/A	**23.08**
Whole Document	0.00	N/A	**0.00**
Total	23.08	N/A	**23.08**

All the *prerequisite* relationships identified by the heuristics were correct, so a 100.00% *precision* was achieved for this kind of relationships (Table 7.25).

Table 7.25 – Precision of the *prerequisite* Relationship

	Level 1 (%)	Level 2 (%)	Total (%)
Outline	100.00	N/A	**100.00**
Whole Document	N/A	N/A	**N/A**
Total	100.00	N/A	**100.00**

The *prerequisite* relationship, as can be observed in Table 7.26, is the most difficult to identify. On the one hand, domain knowledge is required to determine that a topic should be learnt before addressing another one. The order of the topics on the outline might help, but in many cases it shows in which sequence the authors want the topics to be tackled. On the other hand, the grammar defines rules that allow the identification of this kind of relationships, but in many cases not all the *prerequisite* are referred in the sentence. For example, if the topics *Ilargi-eklipse* (*Moon eclipse*) and *Ilargi* (*Moon*) were in the outline, the Reference Heuristic (RH) might have identified a *prerequisite* relationship between those topics. Nevertheless, *prerequisite* relations between other relevant topics for describing the lunar eclipse phenomena - *Lurra* (*Earth*), *ilargiaren translazio mugimendu* (*orbital movement of the Moon*) - could hardly be identified.

In summary, 40.74% of the relations were automatically gathered, 20.00% from the document outline, and 20.74% from the whole document with a 72.50% overall *precision*, which increased to 80.10% for the relationships gathered from the outline (Table 7.27). The heuristics for the analysis of the outline achieved higher performance than the heuristics used in the analysis of the whole document, where whole sentences were analysed and the patterns observed proved to be less accurate.

Table 7.26 – F-measure of the *prerequisite* Relationship

	Level 1 (%)	Level 2 (%)	Total (%)
Outline	37.50	N/A	**37.50**
Whole Document	N/A	N/A	**N/A**
Total	37.50	N/A	**37.50**

Table 7.27 – Statistics on the Acquisition of Pedagogical Relationships

	Outline (%)	Whole Document (%)	Total (%)
Recall	20.00	20.74	**40.74**
Precision	80.10	63.27	**72.50**
F-measure	32.53	31.24	**52.17**

As can be observed in Table 7.27, the statistics on the acquisition of pedagogical relationships were not too impressive, even if the *precision* of the employed heuristics was satisfactory. The characteristics of the analysed book might have considerably affected the LDO acquisition. The simplicity of the outline limits the acquisition of the pedagogical relationships, although the heuristics proved to be accurate. The length of the text also limits the appearances of text fragments that may allow the identification of pedagogical relationships. However, it can also be concluded that the LDO builder is helpful even in short documents such as that employed for this evaluation and the proposed relationships are quite accurate.

7.4.2 Evaluation of the Gathered Learning Objects

The evaluation of the LO acquisition was conducted following the same procedure described in Section 7.3. Both the performance of the DR grammar and the gathered LOs were tested. The performance of the DR grammar was checked as a means to observe whether or not the proposed approach behaves similarly in different documents as was expected. Next the evaluation of the DR grammar performance and the evaluation of the gathered LOs are presented.

Table 7.28 – Accuracy of the DR Grammar

	Def.	**Prin.**	**Examp.**	**Theo.**	**Prob.**	**Fact**	**Total**
Rule Firing	86	16	21	8	94	5	**230**
Correct	50	10	7	5	86	3	**161**
Accuracy (%)	58.14	62.50	33.33	62.50	91.49	60.00	**70.00**

Table 7.29 – Statistics on the Identified Atomic Learning Objects

	Def.	**Prin.**	**Examp.**	**Theo.**	**Prob.**	**Fact**	**Total**
Found	67	11	15	6	68	5	**172**
Correct	38	9	6	3	63	3	**122**
%	56.72	81.82	40.00	50.00	92.65	60.00	**70.93**

7.4.2.1 Evaluation of the DR Grammar

The atomic LOs, the finer grained LOs built from the identified atomic DRs, were checked as a means to assess the pattern-based approach for gathering LOs from the electronic textbooks, testing the performance of the DR grammar rules.

Table 7.28 summarises the information about the performance of the DR grammar. In this experiment, the rules (heuristics) of the DR grammar fired 230 times; the rules for *definitions* triggered 86 times, the *principle statement* rules 16 times, the *example* rules 21 times, the *theory* rules 8 times, the *problem-statement* rules 94, and finally the *fact* rules 5 times. 161 of the triggered rules performed properly, i.e., they identified the correct kind of DR, so 70% of the gathered text fragments contained valid DR fragments, and were correctly categorised. The success rate varies from 91.49% (86 of 94) achieved by the *problem-statement* rules to 33.33% (7 of 21) achieved by the *example* identification rules.

Table 7.29 summarises the performance of the identification of atomic LOs. 70.93% (122 of 172) of the gathered atomic LOs are valid and correctly categorised LOs. Considering each kind of LO, the rate varies from 92.65% (63 of 68) in *problem-statements* to the lower rate 40% (6 of 15) achieved by the *examples*. 56.72% (38 of 67) of the identified *definitions* were correct, while the success rate was 81.82% (9 of 11) for *principle-statements*, 50% (3 of 6) for *theories*, and 60% (3 of 5) for *facts*.

As can be observed, the accuracy of the atomic LOs is slightly higher than the accuracy of the DR grammar, because in some text fragments rules that identify different kinds of DRs triggered at the same time. While building the DR, the

Table 7.30 – Distribution of the Didactic Resources Identified by the Instructional Designers

	Def.	**Theo.**	**Prob.**	**Fact.**	**Comb.**	**Total**
Quantity	27	6	21	6	6	**66**

Table 7.31 – Recall of the Identified Learning Objects

	Def.	**Theo.**	**Prob.**	**Fact.**	**Comb.**	**Total**
Real	27	6	21	6	6	**66**
Found	20	2	20	3	6	**51**
%	74.07	33.33	95.24	50.00	100.00	**77.27**

probability of being a DR kind is calculated based on the confidence of the fired rules, and the kind with the highest score is selected. This way, some incorrect classifications are prevented and, thus, the accuracy of the identification of the atomic LOs is a little bit higher.

7.4.2.2 Evaluation of the Gathered Learning Objects

Finally, the gathered LOs were evaluated. This evaluation was carried out, on the one hand, by comparing the manually identified DRs with the automatically gathered ones to measure the *recall*, and, on the other hand, analysing the gathered LOs to measure which ones were acceptable. Many of the manually identified DRs also were composite fragments that contained finer grain resources.

The instructional designers identified 66 DRs: 27 *definitions*, 6 *theories*, 21 *problem-statement*s, 6 *facts*, and 6 *combined* LOs (Table 7.30). The combined LOs aggregate different kinds of LOs, while the rest only contain resources of one kind, either atomic or composite.

Table 7.31 shows the information about the evaluation of the LO acquisition process. DOM-Sortze found 77.27% of the LOs that correspond to the DRs identified by the instructional designers, the recall being especially high for the *problem statements* and the *combined LOs*. The high *recall* might be expected, as the DR grammar has proved to be very effective for this kind of DRs and *problem statements* appear in limited and homogeneous forms. The high *recall* for *combined LOs* manifests that the combination of related DRs performed accurately if compared to the way instructional designers combined them.

Table 7.32 – Precision of the Identified Learning Objects

	Def.	**Theo.**	**Prob.**	**Fact.**	**Comb.**	**Total**
Real	66	3	90	7	87	**271**
Found	52	3	84	5	75	**229**
%	78.79	100	93.33	71.43	86.21	**84.50**

Table 7.33 – F-Measure the Identified Learning Objects

	Def.	**Theo.**	**Prob.**	**Fact.**	**Comb.**	**Total**
Recall %	74.07	33.00	95.24	50.00	100.00	**77.27**
Precision %	78.79	100.00	93.33	71.43	86.21	**84.50**
F-measure %	76.36	50.00	94.28	58.82	92.59	**80.73**

Table 7.32 presents the *precision* of the LO acquisition process. The process achieved a 84.50% *precision*. Both *principle statements* and *examples* were gathered by *ErauzOnt* but were not identified by the instructional designers and selected for the reference DR set. Moreover, these LOs were actually components of some composite LOs, so these kinds of LOs are not presented in the table but were considered to obtain the overall *precision*. As can be observed, the *precision* is satisfactory for any kind of identified LO.

Table 7.33 summarises the statistics of the LO acquisition process. The process obtained a 80.73% *f-measure* score, achieving remarkable scores for *Problem statements* and *combined LOs*. Although the analysed document covered a different matter, the results were similar to those obtained in the evaluation of *ErauzOnt* (Table 7.10).

7.5 Conclusions

This chapter has presented the conducted experiments for evaluating *DOM-Sortze*. *DOM-Sortze* was developed and tested in a incremental way. First, the LDO Builder was tested to assess the acquisition of the LDO from the outlines of the documents. *ErauzOnt*, the framework for extracting LOs from textbooks was tested after. Finally, *DOM-Sortze*, the whole framework for gathering the *Domain Module* from textbooks, was evaluated. These evaluations were carried out following a Gold Standard approach, i.e., the LDOs and the sets of DRs developed by instructional designers were considered the best solution and the results of the tested systems were compared to the reference LDOs or DR sets they collaborativelly defined.

The acquisition of the LDO from the outlines was tested on 150 outlines of different subjects offered by the University of the Basque Country (UPV/EHU). To simplify the evaluation of the LDOs, their content was limited to the topics referred in the outlines and the pedagogical relationships among those topics. The acquisition of the LDO from the outlines achieved 98.15% *recall*, 98.36% *precision* (98.26% *f-measure*). Considering that the content of the LDOs was restricted to the topics referred in the outlines, the *recall* might not be very relevant, but the *precision* was remarkable. Nevertheless, this test proved that the outlines can be appropriate sources of information and, therefore, detailed outlines can have positive influence on the LDO acquisition. The analysed outlines were classified into three main areas, and the results were quite similar. Therefore, the proposed approach can be considered to be domain-independent.

ErauzOnt was tested over 4 secondary education textbooks on *geology* and *biology* to measure its performance. The evaluation of the gathered LOs was carried out comparing the manually identified DRs with those that were automatically gathered. Many of the manually identified DRs also were combined fragments that contain finer grain resources of different kinds. An aspect to be considered to evaluate the gathered LOs is that while a LO might be more appropriate in a particular context, one of its components or a composite LO that comprises it might fit better and, therefore, be more reusable in other situations. The LO acquisition process achieved a 70.31% *recall*, a 91.88% *precision* and a 79.66% *f-measure*. LO acquisition achieved satisfying results. The *definitions*, *principle statements* and the *composite LOs* are more difficult to identify than *problem statements*, which appeared in limited forms and are, therefore, easy to identify and classify. This experiment proved that the proposed approach can be helpful as it has an acceptable *recall* and a high *precision*, i.e., most of the gathered LOs are useful for *Domain Module* authoring.

ErauzOnt was developed with the aim of promoting the content reuse to build new LOs and, thus, reduce the workload of content authoring. In the conducted experiments, *ErauzOnt* was able to gather LOs reusing from 55% to 65% of the textbooks used and 70% of the textbook analysed when evaluating *DOM-Sortze*. The more topics the LDO defines, the higher the percentage of the document that may be reused.

The complete process carried out by *DOM-Sortze* has been tested using a electronic textbook and comparing the automatically generated elements against the *Domain Module* manually developed by instructional designers. The aim was to evaluate how *DOM-Sortze* contributes to *Domain Module* authoring. A textbook on *Nature Sciences* in Basque for the first course of mandatory secondary education was used. As the experiment aimed to measure the knowledge acquisition from text, a version without images of the document was used as the source of data.

In this case, both the set of LOs and the LDO were gathered. The results of the LO extraction were similar, on average, to the results obtained when evaluating *ErauzOnt* on its own. Therefore, it could be deduced that the LO extraction is not limited to a particular kind of Domain. Although the analysed textbook differed on the subject and the kinds of DRs used, the evaluation obtained similar results to the evaluation of *ErauzOnt* if the overall accuracy of the DR grammar and the statistics of the LO acquisition are considered.

The acquisition of the LDO was not so successful, but this was mainly due to the characteristics of the analysed document. This textbook was developed for primary education students, it is quite short and presents a very limited outline, which affected the LDO acquisition. The *precision* of the heuristics for the identification of the pedagogical relationships suggests that better results would have been obtained from textbooks with more detailed outlines. Longer documents might probably increase the *recall* of relationships from the whole document.

Other attempts to semi-automatically gather domain ontologies from diverse sources (e.g., machine readable dictionaries, corpora, etc.) have obtained quite diverse results in their evaluation. TEXCOMON (Zouaq and Nkambou, 2009b) gathers a domain ontology from a set of text-based LOs with the aim of enhancing them with more knowledge. The authors report several experiments where the achieved recalls vary from 86.65% to 90.84% for classes, 75.28% to 84.33% for taxonomic relationships and 80.08% to 93.12% for conceptual relationships. TEXCOMON was also tested against TEXT-TO-ONTO (Maedche and Staab, 2001), which obtained 73.06% success rates for classes, 47.53% for hierarchical relationships and 0.31% to 53.03% for conceptual relationships. OntoLearn (Velardi *et al.*, 2005) has been used to develop ontologies for tourism and economy. It uses on-line corpora, glossaries, and documents as sources for the ontology learning process, and has reported a 79.3% recall of the relationships for the tourism ontology and 45.5% for economy. The terminology recall was 55%.

DOM-Sortze is not aimed at building an exhaustive domain ontology, but providing aids to build an ontology for didactic purposes. While most Ontology Learning approaches combine many resources or are restricted to certain particular domains, *DOM-Sortze* is domain-independent, and relies exclusively on the electronic textbook provided.

All in all, the *Domain Module* developing process presented in this dissertation follows a semi-automatic approach. Therefore, the automatically gathered LDO can be supervised and enhanced using Elkar-DOM to obtain the desired contents and structure.

8 Conclusions and Future Work

This chapter summarises the work carried out towards the semi-automatic construction of *Domain Modules* for TSLSs using electronic textbooks as the source of information. The main contributions and results are pointed out throughout the chapter. In addition, some future lines of the work described are also presented.

8.1 Summary

In a time when Technology Supported Learning Systems (TSLSs) are being widely used, there is a lack of tools that allows their development in an automatic or semi-automatic way. TSLSs require an appropriate representation of the knowledge and skills to be mastered in order to be effective, but content authoring is a time and effort consuming task. Although authoring tools have been developed in the past, most of them were intended for knowledge engineers and, therefore, they are not suitable for average teachers without advanced computer engineering skills. Moreover, most of the authoring tools require the *Domain Module* to be built from scratch and do not provide facilities to reuse already developed contents.

Traditionally, textbooks have been used as the main mechanism to maintain and transmit the knowledge of a certain subject or domain. Textbooks have been authored by domain experts who have organised the contents in a means that facilitate understanding and learning, considering the recommended order to address the topics or which topics should be mastered before attempting others. Specifically, the outlines of the textbooks implicitly contain that kind of pedagogical information. In addition, textbooks provide several resources with educational purposes that can be used as part of lecture sessions.

Content reuse has been one of the major concerns during the last few years, and standards and specifications have been defined to support the reusability of the learning material. The use of Learning Objects (LOs), i.e., reusable resources with educational purposes, may contribute to lighten the development cost of the *Domain Modules*.

Throughout this dissertation, the analysis, design and evaluation of *DOM-Sortze*, a framework for the semi-automatic development of *Domain Modules* from electronic textbooks, have been described. *DOM-Sortze* employs NLP techniques, heuristic reasoning, and ontologies for the knowledge acquisition processes.

In *DOM-Sortze*, the *Domain Module* is described by the Learning Domain Ontology (LDO), which contains the main domain topics and the pedagogical relationships among them, and the LOs. The acquisition of the *Domain Module* with *DOM-Sortze* is carried out in three steps:

1. **Preprocess**: First, the document is prepared for the subsequent knowledge elicitation tasks. The results of this step are the internal representation of the document and the internal representation of the outline of the document. Both contain the *part-of-speech* information that will be used to gather the components of the *Domain Module*.

2. **Acquisition of the LDO**: The second step implies the identification of the domain topics to be mastered and the pedagogical relationships among the topics, which will allow instructional TSLSs to plan the learning sessions or might be used by the learners to determine their learning schedule. The LDO is gathered from the internal representations of the textbook and its outline.

3. **Acquisition of the LOs**: Finally, the LOs that might be used for mastering the domain topics are extracted from the internal representation of the textbook. This step is guided by the LDO built in the previous step.

DOM-Sortze entails a suite of applications that allow these steps to be carried out. Besides, as the knowledge acquisition processes are conducted automatically, the results must be supervised by the authors of the *Domain Module* to assure their correctness. *DOM-Sortze* provides a tool for the collaborative supervision of the results, *Elkar-DOM*. *Elkar-DOM* allows the users to adapt the automatically elicited LDO to their preferences and needs or to look for and choose the most appropriate LOs for each domain topic.

8.2 Results and Contributions

The main results of this thesis work can be summarised in a proposal of a method for the semi-automatic generation of *Domain Modules* from electronic textbooks based on NLP techniques, heuristic reasoning and ontologies, and the development of the framework that carries out such a method, *DOM-Sortze*.

Next, the contributions of this work will be presented considering the degree of achievement of the requirements for *DOM-Sortze* stated in Section 1.1.

8.2.1 Semi-automatic Construction of the Domain Module

Given that electronic textbooks might be appropriate sources of information, the main goal of this work was to semi-automatise the construction of the *Domain Modules*, allowing the teachers to play the role of supervisors instead of requiring them to author all the content. *DOM-Sortze* was created to fulfill this objective.

In the approach here presented, the development of the *Domain Module* entails gathering the LDO, which defines the domain topics to be learnt and the pedagogical relationships between the topics, and extracting LOs, i.e. reusable Didactic Resources (DRs), from the textbook. The LDO is first automatically gathered from the analysed document and, then, supervised by the *Domain Module* authors to adapt it to their teaching preferences and requirements. Next, the LOs are extracted from the electronic textbook and stored in the LOR to allow their use. Finally, the LDO is completed by selecting the LOs to be used for each domain topic.

DOM-Sortze and all its components have been tested to measure the percentage of the *Domain Module* that can be automatically elicited. *DOM-Sortze* was developed and evaluated incrementally. First, the *LDO Builder*, the subsystem for gathering the LDO, was tested on a set of outlines. Later, *ErauzOnt*, the subsystem for gathering the LOs, was evaluated. Finally, the whole process of generating the *Domain Module* carried out by *DOM-Sortze* was evaluated. All the evaluations, carried out over documents written in the Basque language, followed the Gold-standard approach in which a team of instructional designers defined the reference output that was compared to the output of the automatic knowledge acquisition process.

The experiment conducted to evaluate the acquisition of the LDO from the document outlines, to which end 150 outlines of different subjects taught at the University of the Basque Country (UPV/EHU) were used, proved that the approach applied is an appropriate means to gather the LDO (Larrañaga *et al.*, 2004). The remarkable *precision* score achieved by the *LDO Builder*, 98.36%, must be noted. Given that for this experiment the instructional designers were requested to limit the Gold Standard LDOs to the domain topics referred in the outlines, the high *recall* score might be ignored. It can be concluded that the heuristics used on the analysis of the outlines of the electronic textbooks are an appropriate means to elicit pedagogical relationships from the outlines of the textbooks. Moreover, due to the educational purpose of the LDO, it defines sequential relationships- *next* and *prerequisite* - that constrain the order in which some topics might be addressed. Ontology Learning usually deals with domain ontologies and rarely deals with relationships such as *prerequisite* or *next*, which are extremely difficult to infer from sources of information other than outlines of textbooks.

The extraction of LOs from the electronic textbooks (Larrañaga *et al.*, 2008a,c), which is carried out by *ErauzOnt*, was also evaluated (Larrañaga *et al.*, 2012). The experiment conducted on 4 textbooks on *geology* and *biology* for secondary education concluded that *ErauzOnt* is an appropriate tool to elicit reusable DRs from the documents and, therefore, might be helpful to lighten the content authoring workload.

Once the main components of *DOM-Sortze* were individually evaluated, the whole system was tested on a new textbook on *astronomy*, also for secondary education, to assess the process of constructing the *Domain Module*. The results obtained in the experiment conducted to evaluate *DOM-Sortze* also proved that it is a valuable tool for building *Domain Modules*. In the extraction of the LOs, the results were similar to those obtained in the experiment conducted to evaluate *ErauzOnt*. *DOM-Sortze* also helped in the acquisition of the LDO, and the *precision* achieved suggests that the results might be even better using more complex textbooks. Therefore, *DOM-Sortze* has proved to facilitate the semi-automatic construction of the *Domain Modules* from electronic textbooks (Larrañaga *et al.*, submitted).

8.2.2 Knowledge Reuse

Promoting the reuse of knowledge, one of the main concerns of the Technology Enhanced Learning Community, was another of the requirements for this work. Bearing this in mind, methods to facilitate the reuse of information have been employed. The use of standards and ontologies contributes to accomplish this objective.

The work presented here facilitates the reuse of knowledge in two different levels. On the one hand, the learning material extracted from the analysed textbooks can be used to build new learning resources. On the other hand, the *Domain Modules* built can be used in the development of new ones.

DOM-Sortze, and *ErauzOnt* (Larrañaga *et al.*, 2011) in particular, rely on the LOM standard (LTSC, 2001) for the annotation of the gathered LOs, which facilitates their search and retrieval from the LORs, built on the ARIADNE knowledge pool system (Duval *et al.*, 2001; Ternier *et al.*, 2009).

All the *Domain Modules* built using *DOM-Sortze* are stored in the *Domain Module Repository*. *Domain Module* authors may reuse any of the stored *Domain Modules* to build a new one which is based on them by using Elkar-DOM (Larrañaga *et al.*, 2005, 2007), either by modifying the domain topics or by adding other LOs.

As stated by Uschold and Gruninger (1996), ontologies facilitate the re-usability as they can be translated, either manually or automatically, to another ontology or representation formalism suitable for a new system. Accordingly, the LDO can be automatically mapped to another formalism for describing the domain to be learnt. Hence, the *Domain Module* might be useful for different TSLSs.

8.2.3 Collaborative Work

Teachers usually collaborate while preparing the scheduling materials for the subjects they lecture in. *Elkar-DOM* (Larrañaga *et al.*, 2005, 2007), the concept map-based tool for supervising the construction of the *Domain Module*, provides *Domain Module* authors with a means to cooperate, as they can collaboratively determine the topics to be mastered and the resources to be used as well.

Domain Module authors can communicate with each other using a chat tool or even interact synchronously on the LDO being developed. Suthers (2005) pointed out that, usually, users prefer to interact on the CM instead of using other communication means.

Besides, the *Domain Module* authors can observe the evolution of the LDO, viewing what their colleagues modified, and even analyse the discussion carried out in the chat.

8.2.4 Domain-Independence

In the past, some efforts have been aimed at automatising the construction of a specific domain. For example, Lentini *et al.* (1995, 2000) developed an authoring tool for the automatic construction of ITSs from spreadsheets. Obviously, this approach is restricted to the domain of Mathematics.

However, the approach presented in this work is aimed at being domain-independent.

The experiments conducted to evaluate the performance of *DOM-Sortze* and its components combined documents of diverse domains. In particular, 150 outlines of diverse subjects and educational levels were used for the evaluation of the *LDO*

Builder on outlines (Larrañaga *et al.*, 2004). The outlines were classified on 3 main areas and the results obtained, in general, were similar (Section 7.2).

The textbooks used for the evaluation of *ErauzOnt* covered different fields of the *Nature Sciences* - *geology*, *biology* and *astronomy*, and the results obtained were alike. Besides, a first attempt using *ErauzOnt* on a textbook on *Object Oriented Programming* has been completed and, again, almost identical results were achieved (Conde *et al.*, 2012). Thus, it can be concluded that *DOM-Sortze* can process different domains.

8.2.5 Multilingualism

DOM-Sortze is intended to be able to deal with different languages. *DOM-Sortze* does not strongly depend on a concrete language, even though it has been initially applied on textbooks written in the Basque language.

DOM-Sortze combines NLP techniques with heuristic reasoning and ontologies to construct the *Domain Modules*. In particular, it uses a set of heuristics and patterns based on syntactic information that allow the identification of meaningful pieces of knowledge from which the LDO and the LOs are built. Besides, the discourse markers are used to obtain cohesioned resources. The heuristics, the patterns, and the discourse markers depend on the language, but in *DOM-Sortze* they are defined in external files instead of being hardcoded.

Moreover, it has been observed that similar or equivalent patterns exist for other languages such as English (Liu *et al.*, 2003; Verbert, 2008). Thus, *DOM-Sortze* can easily be enhanced to deal with a new language. It needs the heuristics for the outlines, the patterns for identifying the relationships and the DRs for that language, and to enhance the *NLP Analysis Service* with a NLP analyser for the new language.

In fact, *ErauzOnt* has already been enhanced to work on textbooks on English. As mentioned above, it has been tested on a textbook about *Object Oriented Programming* obtaining similar results to those obtained for Basque (Conde *et al.*, 2012). As expected, the only component of *DOM-Sortze* that had to be improved was the *NLP Analysis Service*, which was extended to use FreeLing (Atserias *et al.*, 2002) for the linguistic analysis of English. Freeling supports more languages besides English. In addition, the English version of the DR Grammar and the discourse markers for the English were defined. *ErauzOnt* is capable of using the appropriate grammars and discourse markers according to the language the document is written in. Ergo, *DOM-Sortze* can easily be adapted to work on textbooks provided in different languages.

8.2.6 Other Contributions

DOM-Sortze can self-improve its performance. As described in the previous chapters, it relies on a set of heuristics or syntactic patters that allow the identification of different elements of the *Domain Module*. Both the heuristics and the patterns have certain confidence values which are exploited by *DOM-Sortze* to resolve conflicts between the patterns and heuristics when more than one is applicable.

Elkar-DOM considers the changes proposed by the users and updates the confidence of the heuristics and the patterns to become more accurate according to their performance. Hence, *DOM-Sortze* is not a static tool but it can improve its knowledge and performance.

DOM-Sortze has some other interesting features. It has been built on a flexible architecture that can be easily extended to support either new languages or document formats which are different from the originally supported *pdf*.

8.3 Future Research Lines

This section depicts the future research lines opened by this thesis work, some of which are currently being addressed. Future research lines range from improvements in the functionality of the applications that entail *DOM-Sortze* to the integration of *DOM-Sortze* with other tools.

8.3.1 Generation of Multilingual Domain Modules

Education in multilingual contexts is nowadays a reality. Many teachers lecture the same or similar subjects in more than one language. Those teachers would be grateful for a framework that allows them to semi-automatically build the multilingual *Domain Module* for the subjects they teach, as they could use similar resources in the languages they lecture in.

A multilingual *Domain Module* entails a LDO whose topics are labeled in diverse languages and LOs in those languages. Thus, the same *Domain Module* could be used to teach in more than one language.

Nowadays, ontology topics can be labeled in different languages. In fact, the OWL-based formalism for describing the LDO allows the topics to be defined in diverse languages. However, the XML-representation that is used during the acquisition of the LDO should be enhanced to be able to deal with multilingualism.

The multilingual acquisition of the *Domain Module* can be dealt with in different ways. On the one hand, the framework could analyse parallel documents, i.e., documents that have been translated from one language to another, to build the *Domain Module*. On the other hand, an incremental procedure could be followed. The *Domain Module* could be initially gathered and built from a textbook written in a concrete language and, the machine translation, and even multilingual resources such as EuroWordnet (Vossen, 1998), could be used to translate the LDO and obtain basic translations of the LOs. These translations would be used to look for and retrieve similar resources that might be used for the multilingual *Domain Module*.

8.3.2 Machine Learning for Identifying New Patterns

Machine Learning has been widely used for Ontology Learning and other knowledge acquisition purposes. *DOM-Sortze* could be enhanced combining machine learning with the current pattern-based approach. In this way, *DOM-Sortze* would learn new ways to identify either components of the LDO or new LOs.

Machine learning techniques would allow *DOM-Sortze* to learn from the supervised *Domain Modules*, especially from the LDOs, to periodically review and enrich the set of patterns that allow the identification and elicitation of *Domain Module* components.

8.3.3 Library of Patterns for Particular Domains

Although one of the main goals for the work presented here was the domain-independence, the use of domain-specific knowledge could improve the results, so a hybrid approach should be worked on.

Enhancing *DOM-Sortze* with a library of patterns for particular domains, e.g., mathematics and history, could improve the process for generating the *Domain Module* in a semi-automatic way. *DOM-Sortze* could determine if the document being analysed describes one of the domains included in the library and, subsequently, apply the corresponding patterns. If *DOM-Sortze* is not capable of classifying the new document in one of those domains, it will apply the generic patterns, i.e., the patterns described along this work.

8.3.4 Versioning and Maintainance

Domains are continuously evolving and new textbooks and documents are constantly being published. Therefore, mechanisms to keep the *Domain Modules* updated should be provided. *DOM-Sortze* should allow versioning the *Domain Modules* and building up-to-date versions of *Domain Modules* by enriching their LDOs and the LOs used.

The following incremental procedure could be followed to get a newer version of a *Domain Module*. First, the most recent material would be analysed to get a new LDO that describes the domain. Then, the previous LDO and the new one would be integrated. Finally, the new material would be analysed by *ErauzOnt*, using the new LDO, to extract new LOs for the *Domain Module*. This procedure can be repeated whenever a new textbook is available and the teachers want to adapt their lessons to the new contents.

8.3.5 Integrating DOM-Sortze with Other Applications

Another observed future line envisages the integration of *DOM-Sortze* with other systems.

Arikiturri (Aldabe, 2011) is a system for the automatic generation of assessable exercises from text corpora. Integrating Arikiturri with *DOM-Sortze* would result in a system that is not only capable of building the *Domain Module* by extracting the contents explicitly represented in the textbook but also generating new exercises from corpora.

As mentioned above, the *Domain Modules* built using *DOM-Sortze* could be translated to other *Domain Module* representation formalisms to be used in other systems. In order to bring *DOM-Sortze* to a wider community of users, translators for different TSLSs, e.g., the Moodle LMS, should be developed.

8.3.6 Promoting Collaboration

Elkar-DOM was developed to enable the collaborative supervision of the *Domain Module* being developed. However, the interactions are currently limited to the direct manipulation of the Concept Map that describes the *Domain Module* and to the chat communication.

Elkar-DOM could be enhanced with some decision making support. For example, the information about the opinion of the authors about a particular LO might be useful to select or discard it. This could be achieved by implementing a Model of Trust similar to that proposed for Virtual Learning Communities by Chamba Eras *et al.* (2011).

Bibliography

Adcock, G., Ip, A., and Mason, J. (2000). «Modelling Information to Support Value-Adding: EdNA Online». *WebNet Journal: Internet Technologies, Applications & Issues*, **vol. 2(3)**, pp. 38–45. URL `http://www.editlib.org/p/8002`.

ADL (2009a). «SCORM 2004 Content Agregation Model version 1.1». `http://www.adlnet.gov/Technologies/scorm/SCORMSDocuments/2004%204th%20Edition/Documentation.aspx`.

ADL (2009b). «SCORM 2004 Run-Time Environment version 1.1». `http://www.adlnet.gov/Technologies/scorm/SCORMSDocuments/2004%204th%20Edition/Documentation.aspx`.

ADL (2009c). «SCORM 2004 Sequencing and Navigation version 1.1». `http://www.adlnet.gov/Technologies/scorm/SCORMSDocuments/2004%204th%20Edition/Documentation.aspx`.

Aduriz, I., Agirre, E., Aldezabal, I., Alegria, I., Ansa, O., Arregi, X., Arriola, J.M., Artola, X., Díaz de Ilarraza, A., Ezeiza, N., Gojenola, K., Maritxalar, A., Maritxalar, M., Oronoz, M., Sarasola, K., Soroa, A., Urizar, R., and Urkia, M. (1998). «A Framework for the Automatic Processing of Basque». In *Proceedings of the 1st International Conference on Language Resources and Evaluation, LREC 1998*. Granada.

Aduriz, I., Aldezabal, I., Alegria, I., Artola, X., Ezeiza, N., and Urizar, R. (1996). «EUSLEM: A Lemmatiser / Tagger for Basque». In *Proceedings of the 7th EURALEX International Congress on Lexicography, EURALEX 1996*, vol. 1, pp. 17–26. Göteborg, Sweden.

Agirre, E., Alfonseca, E., Hall, K., Kravalova, J., Pasca, M., and Soroa, A. (2009a). «A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches». In *Proceedings of NAACL-HLT 09*. Association for Computational Linguistics, Boulder, USA.

Agirre, E., Lopez de Lacalle, O., and Soroa, A. (2009b). «Knowledge-Based WSD and Specific Domains: Performing Better than Generic Supervised WSD». In C. Boutilier (Ed.), *Proceedings of the 21st International Joint Conference on Artificial Intelligence, IJCAI 2009*, pp. 1501–1506. Pasadena, California, USA.

Agirre, E. and Soroa, A. (2009). «Personalizing PageRank for Word Sense Disambiguation». In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2009*, pp. 33–41. The Association for Computer Linguistics, Athens, Greece.

Aldabe, I. (2011). *Automatic Exercise Generation Based on Corpora and Natural Lenguaje Processing Techniques*. Ph.D. thesis, University of the Basque Country (UPV/EHU).

Aldabe, I. and Maritxalar, M. (2010). «Automatic Distractor Generation for Domain Specific Texts». In H. Loftsson, E. Rögnvaldsson, and S. Helgadóttir (Eds.), *ICETAL-2010*, 6233, pp. 27–38. Springer, Reykjavik, Iceland.

Alegria, I., Gurrutxaga, A., Lizaso, P., Saralegi, X., Ugartetxea, S., and Urizar, R. (2004a). «Linguistic and Statitstical Approaches to Basque Term Extraction». In *GLAT-2004: The Production of Specialized Texts*.

Alegria, I., Gurrutxaga, A., Lizaso, P., Saralegi, X., Ugartetxea, S., and Urizar, R. (2004b). «An Xml-Based Term Extraction Tool for Basque». In *Proceedings of the 4rd International Conference on Language Resources and Evaluation, LREC 2004*. ELRA, Lisboa, Portugal.

Amorim, R.R., Lama, M., Sánchez Vila, E., Riera Alvarez, A., and Vila Sobrino, X.A. (2006a). «A Learning Design Ontology based on the IMS Specification». *Journal of Educational Technology & Society*, **vol. 9(1)**, pp. 38–57.

Amorim, R.R., Sánchez, E., and Lama, M. (2006b). «The Problem of LD Execution and the Need for an IMS LD Level B Ontology». In W. Nejdl and K. Tochtermann (Eds.), *Innovative Approaches for Learning and Knowledge Sharing, Proceedings of the 1st European Conference on Technology Enhanced Learning, EC-TEL 2006*, *Lecture Notes in Computer Science*, vol. 4227, pp. 537–542. Springer, Crete, Greece.

Anderson, J.R. (1988). «The Expert Module». In M.C. Polson and J.J. Richardson (Eds.), *Foundations of Intelligent Tutoring Systems*, pp. 21–54. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.

Anderson, J.R., Corbett, A.T., Koedinger, K.R., and Pelletier, R. (1995). «Cognitive Tutors: Lessons Learned». *The Journal of the Learning Sciences*, **vol. 4(2)**, pp. 167–207.

Arellano, C., Rueda, U., Niebla, I., Larrañaga, M., Arruarte, A., and Elorriaga, J.A. (2006). «Elkar-CM: a Multilingual Collaborative Concept Map Editor». In *Proceedings of the 6th IEEE International Conference on Advanced Learning Technologies, ICALT 2006*, pp. 482–483. IEEE Computer Society, Kerkrade, The Netherlands.

Aroyo, L. and Dicheva, D. (2004). «Authoring support in concept-based web information systems for educational applications». *International Journal of Continuing Engineering Education and Lifelong Learning*, **vol. 14(3)**, pp. 297–312.

Arroyo, I., Schapira, A., and Woolf, B.P. (2001). «Authoring and Sharing Word Problems with AWE». In J. Moore, C. Redfield, and W.L. Johnson (Eds.), *Proceedings of the 10th International Conference on Artificial Intelligence in Education AIED-2001*, pp. 527–529. IOS Press, San Antonio, USA.

Arruarte, A. (1998). *Fundamentos y diseño de IRIS: Un entorno para la generación de Sistemas de Enseñanza Inteligentes*. Ph.D. thesis, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU).

Arruarte, A., Ferrero, B., Fernández de Castro, I., Urretavizcaya, M., Álvarez, A., and Greer, J.E. (2003). «The IRIS Authoring Tool». In Murray *et al.* (2003), pp. 233–268.

Atserias, J., Casas, B., Comelles, E., González, M., Padró, L., and Padró, M. (2002). «FreeLing 1.3: Syntactic and semantic services in an open-source NLP library». In *Proceedings of the 3rd International Conference on Language Resources and Evaluation, LREC 2002*. ELRA, Las Palmas, Canary Islands, Spain. URL `http://www.lsi.upc.edu/~nlp/freeling`.

Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D.L., Patel-Schneider, P.F., and Stein, L.A. (2004). «OWL Web Ontology Language Reference». http://www.w3.org/TR/owl-ref/.

Berners-Lee, T. and Fischetti, M. (1999). *Weaving the Web: The Past, Present and Futeure of the World Wide Web by its Inventor*. Harper Business.

Berners-Lee, T., Hendler, J.A., and Lassila, O. (2001). «The Semantic Web». *Scientific American*.

Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., and Krathwohl, D.R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals: Handbook I, Cognitive Domain*. Longman.

Borst, W.N. (1997). *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. Ph.D. thesis, University of Twente.

Bourdeau, J., Mizoguchi, R., Psyché, V., and Nkambou, R. (2004). «Selecting Theories in an Ontology-Based ITS Authoring Environment». In J.C. Lester, R.M. Vicari, and F. Paraguaçu (Eds.), *Proceedings of the 7th International Conference on Intelligent Tutoring Systems, ITS 2004*, *Lecture Notes in Computer Science*, vol. 3220, pp. 150–161. Springer, Maceiò, Alagoas, Brazil.

Brooks, C., Cooke, J., and Vassileva, J. (2003). «Versioning of Learning Objects». In V. Devedžić, J.M. Spector, D.G. Sampson, and Kinshuk (Eds.), *Proceedings of the 3rd IEEE International Conference on Advanced Learning Technologies, ICALT 2003*, pp. 296–297. IEEE Computer Society, Athens, Greece.

Brooks, C., McCalla, G.I., and Winter, M. (2005). «Flexible Learning Object Metadata». In L. Aroyo and D. Dicheva (Eds.), *SW-EL'05: International Workshop on Applications of the Semantic Web for E-Learning Technologies at the 12th International Conference on Artificial Intelligence in Education, AIED 2005*, pp. 1–8. SW-EL, Amsterdam, The Netherlands. URL `http://www.win.tue.nl/SW-EL/2005/swel05-aied05/proceedings/2-Brooks-final-full.pdf`.

Brusilovsky, P., Sosnovsky, S.A., Lee, D.H., Yudelson, M., Zadorozhny, V., and Zhou, X. (2008). «An open integrated exploratorium for database courses». In J. Amillo, C. Laxer, E.M. Ruiz, and A. Young (Eds.), *Proceedings of the 13th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education, ITiCSE 2008*, pp. 22–26. ACM, Madrid, Spain.

Brusilovsky, P., Sosnovsky, S.A., Yudelson, M., and Chavan, G. (2003). «Interactive Authoring Support for Adaptive Educational Systems». In U. Hoppe, F. Verdejo, and J. Kay (Eds.), *Shaping the Future of Learning through Intelligent Technologies, Proceedings of the 11th International Conference on Artificial Intelligence in Education, AIED 2003*, *Frontiers in Artificial Intelligence and Applications*, vol. 97, pp. 97–103. IOS Press, Sidney, Australia.

Buitelaar, P., Cimiano, P., and Magnini, B. (2005a). «Ontology Learning from Text: An Overview». In Buitelaar *et al.* (2005b), pp. 3–12.

Buitelaar, P., Cimiano, P., and Magnini, B. (Eds.) (2005b). *Ontology Learning from Text: Methods, Applications and Evaluation*, *Frontiers in Artificial Intelligence and Applications*, vol. 123. IOS Press.

Buitelaar, P., Olejnik, D., and Sintek, M. (2004). «A Protégé Plug-In for Ontology Extraction from Text Based on Linguistic Analysis». In C. Bussler, J. Davies, D. Fensel, and R. Studer (Eds.), *The Semantic Web: Research and Applications, First European Semantic Web Symposium, ESWS 2004*, *Lecture Notes in Computer Science*, vol. 3053, pp. 31–44. Springer, Heraklion, Crete, Greece.

Byrd, R.J., Calzolari, N., Chodorow, M.S., Klavans, J.L., Neff, M.S., and Rizk, O.A. (1987). «Tools and Methods for Computational Lexicology». *Computational Linguistics*, **vol. 13(3-4)**, pp. 219–240.

Cafolla, R. (2006). «Project Merlot: Bringing Peer Review to Web-based Educational Resources». *Journal of Technology and Teacher Education*, **vol. 14(2)**, pp. 313–323. URL `http://www.editlib.org/p/4623`.

Cardinaels, K. (2007). *A Dynamic Learning Object Life Cycle and its Implications for Automatic Metadata Generation.* Ph.D. thesis, Faculteit Ingenieurswetenschappen, Katholieke Universiteit Leuven.

Cardinaels, K., Meire, M., and Duval, E. (2005). «Automating Metadata Generation: the Simple Indexing Interface». In A. Ellis and T. Hagino (Eds.), *Proceedings* *of the 14th International Conference on World Wide Web, WWW 2005*. ACM, Chiba, Japan.

Casey, J. and McAlpine, M. (2003). «Writing and Using Reusable Educational Material - A Beginner's guide». http://assessment.cetis.ac.uk/groups/20010809144711/FR20020507112554.

Cassel, L.N., Davies, G., LeBlanc, R., Snyder, L., and Topi, H. (2008). «Using a Computing Ontology as a Foundation for Curriculum Development». In D. Dicheva, A. Harrer, and R. Mizoguchi (Eds.), *SW-EL'08: Ontologies and Semantic Web for Intelligent Educational Systems at 9th International Conference on Intelligent Tutoring Systems, ITS 2008*. Montreal, Canada.

Chamba Eras, L.A., Arruarte, A., and Elorriaga, J.A. (2011). «Modelo de Confianza para Comunidades Virtuales de Aprendizaje». In J. Sánchez Ilabaca (Ed.), *Nuevas Ideas en Informática Educativa, Memorias del XVI Congreso Internacional de Informática Educativa, TISE*, vol. 7, pp. 80–87.

Chandrasekaran, B., Josephson, J.R., and Benjamins, R.V. (1999). «What Are Ontologies, and Why Do We Need Them?» *IEEE Intelligent Systems*, **vol. 14(1)**, pp. 20–26.

Chen, P.S.D., Lambert, A.D., and Guidry, K.R. (2010). «Engaging online learners: The impact od Web-based learning technology on college student engagement». *Computers & Education*, **vol. 54(4)**, pp. 1222–1232.

Chen, W., Hayashi, Y., Jin, L., Ikeda, M., and Mizoguchi, R. (1998). «An Ontology-based Intelligent Authoring Tool». In *Proceedings of the 6th International Conference on Computers in Education, ICCE 1998*, pp. 41–49.

Chen, W., Lu, R., Zhang, W., and Du, H. (1997). «A Tool for Automatic Generation of Multimedia ICAI Systems». In B. du Boulay and R. Mizoguchi (Eds.), *Knowledge and Media in Learning Systems. Proceedings of the 8th International Conference on Artificial Intelligence in Education, AIED-1997*, *Frontiers in Artificial Intelligence and Applications*, vol. 39, pp. 571–573. IOS Press, Kobe, Japan.

Cimiano, P., Hotho, A., and Staab, S. (2005). «Learning Concept Hierarchies from Text Corpora using Formal Concept Analysis». *Journal of Artificial Intelligence Research*, **vol. 24**, pp. 305–339.

Cimiano, P. and Völker, J. (2005). «Text2Onto - A Framework for Ontology Learning and Data-driven Change Discovery». In A. Montoyo, R. Muñoz, and E. Métais (Eds.), *10th International Conference on Applications of Natural Language to Information Systems, NLDB 2005*, *Lecture Notes in Computer Science*, vol. 3513, pp. 227–238. Springer, Alicante, Spain.

Clancey, W.J. (1982). «Tutoring rules for guiding a case method dialogue». In D. Sleeman and J. Brown (Eds.), *Intelligent Tutoring Systems*, pp. 201–225. Academic Press.

Coffey, J.W., Eskridge, T.C., and Sanchez, D.P. (2004). «A Case Study in Knowledge Elicitation for Institutional Memory Preservation Using Concept Maps». In A. Cañas, J.D. Novak, and F. González (Eds.), *Concept Maps: Theory, Methodology, Technology. Proceedings of the 1st International Conference on Concept Mapping, CMC 2004*, vol. I, pp. 151–157. Pamplona, Spain.

Conde, A., Larrañaga, M., Calvo, I., Elorriaga, J.A., and Arruarte, A. (2012). «Automating the Authoring of Learning Material in Computer Engineering Education». In *Proceedings of 2012 Frontier in Education Conference (FIE 2012)*, pp. 1376–1381.

Corbett, A.T., Trask, H.J., Scarpinatto, K.C., and Hadley, W.S. (1998). «A Formative Evaluation of the PACT Algebra II Tutor: Support for Simple Hierarchical Reasoning». In B.P. Goettl, H.M. Halff, C.L. Redfield, and V.J. Shute (Eds.), *Proceedings of the 4th International Conference on Intelligent Tutoring Systems, ITS 1998*, *Lecture Notes in Computer Science*, vol. 1452, pp. 374–383. Springer, San Antonio, USA.

Corcho, Ó. and Gómez-Pérez, A. (2000). «A Roadmap to Ontology Specification Languages». In R. Dieng and O. Corby (Eds.), *Proceedings of the 12th International Conference on Knowledge Engineering and Management, EKAW 2000*, *Lecture Notes in Computer Science*, vol. 1937, pp. 80–96. Springer, Juan-les-Pins, France.

Dagan, I., Glickman, O., and Magnini, B. (2005). «The PASCAL Recognising Textual Entailment Challenge». In J.Q. Candela, I. Dagan, B. Magnini, and F. d'Alché Buc (Eds.), *MLCW*, *Lecture Notes in Computer Science*, vol. 3944, pp. 177–190. Springer, Southampton, UK.

Dede, C. (1986). «A Review and Synthesis of Recent Research in Intelligent Computer-Assisted Instruction». *International Journal of Man-Machine Studies*, **vol. 24(4)**, pp. 329–353.

Denaux, R., Aroyo, L., and Dimitrova, V. (2005a). «An Approach for Ontology-Based Elicitation of User Models to Enable Personalization on the Semantic Web». In A. Ellis and T. Hagino (Eds.), *Proceedings of the 14th International Conference on World Wide Web, WWW 2005 - Special interest tracks and posters*, pp. 1170–1171. ACM Press, Chiba, Japan.

Denaux, R., Aroyo, L., and Dimitrova, V. (2005b). «OWL-OLM: Interactive Ontology-based Elicitation of User Models». In *Proceedings of Workshop on Personalization for the Semantic Web (PerSWeb'05) at 10th International Conference on User Modeling, UM 2005*, pp. 34–46. Edinburgh, Scotland.

Denaux, R., Dimitrova, V., and Aroyo, L. (2004). «Interactive Ontology-Based User Modelling for Personalized Learning Content Management». In L. Aroyo, D. Dicheva, P. Dolog, and M. Wolpers (Eds.), *SW-EL'04: Semantic Web for E-Learning. Applications of Semantic Web Technologies for Educational Adaptive Hypermedia Workshop at International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, AH 2004*. Eindhoven, The Netherlands.

Denaux, R., Dimitrova, V., and Aroyo, L. (2005c). «Integrating Open User Modeling and Learning Content Management for the Semantic Web». In L. Ardissono, P. Brna, and A. Mitrovic (Eds.), *Proceedings of the 10th International Conference on User Modeling, UM 2005*, *Lecture Notes in Computer Science*, vol. 3538, pp. 9–18. Springer, Edinburgh, Scotland.

Dicheva, D. and Aroyo, L. (2004a). «General Architecture Supporting Component-based EIS Interoperability». In L. Aroyo, D. Dicheva, A. Mitrovic, and V. Devedžić (Eds.), *SW-EL'04: Applications of Semantic Web Technologies for Web-based ITS at 7th International Conference on Intelligent Tutoring Systems, ITS 2004*, pp. 21–28. Maceió, Alagoas, Brazil.

Dicheva, D. and Aroyo, L. (2004b). «A Pragmatic Approach to Support Concept-based Educational Information Systems Communication». In L. Aroyo, D. Dicheva, R. Mizoguchi, and Y. Itoh (Eds.), *SW-EL'04: Semantic Web for E-Learning. Applications of Semantic Web Technologies for E-learning Workshop at International Semantic Conference, ISWC 2004*, pp. 23–30. Hiroshima, Japan.

Dicheva, D., Dichev, C., Sun, Y., and Nao, S. (2004). «Authoring Topic Maps-based Digital Course Libraries». In L. Aroyo, D. Dicheva, P. Dolog, and M. Wolpers (Eds.), *SW-EL'04: Semantic Web for E-Learning. Applications of Semantic Web Technologies for Educational Adaptive Hypermedia Workshop at International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, AH 2004*. Eindhoven, The Netherlands. URL `http://www.win.tue.nl/SW-EL/2004/AH-SWEL-Camera-ready/SWEL04-AH-PDF/%2311-Dicheva-Dichev-SWEL-AH.pdf`.

Downes, S. (2001). «Learning Objects: Resources for Distance Education Worldwide». *International Review of Research in Open and Distance Learning*, **vol. 2(1)**.

Downes, S. (2003). «Design and Reusability of Learning Objects in an Academic Context: A new Economy of Education?» *UDSLA Journal*, **vol. 17(1)**.

Dublin Core Metadata Initiative (2004). «Dublin Core Metadata Element Set, Version 1.1: Reference Description».

Duval, E. (2004). «Learning Technology Standardization: Making Sense of it All». *Computer Science and Information Systems*, **vol. 1**.

Duval, E., Forte, E., Cardinaels, K., Verhoeven, B., Durm, R.V., Hendrikx, K., Forte, M.W., Ebel, N., Macowicz, M., Warkentyne, K., and Haenni, F. (2001). «The ARIADNE Knowledge Pool System». *Communications of the ACM*, **vol. 44(5)**, pp. 72–78.

Duval, E. and Hodgins, H.W. (2002). «A LOM Research Agenda». In *Proceedings of the 11th International Conference on World Wide Web, WWW 2002*, pp. 1–9. ACM Press, Hawaii, USA.

Duval, E. and Hodgins, H.W. (2004). «Making Metadata Go Away: "Hiding Everything but the Benefits"». In *Proceedings of the International Conference on Dublin Core and Metadata Applications*, pp. 29–35. Shanghai, China.

Elorriaga, J.A., Arruarte, A., Calvo, I., Larrañaga, M., Rueda, U., and Herran, E. (2011). «Collaborative Concept Mapping Activities in a Classroom Scenario». *Behaviour & Information Technology*, pp. 1–13. URL `http://www.tandfonline.com/doi/abs/10.1080/0144929X.2011.632649`.

Etzioni, O., Cafarella, M.J., Downey, D., Kok, S., Popescu, A.M., Shaked, T., Soderland, S., Weld, D.S., and Yates, A. (2004). «Web-scale information extraction in knowitall: (preliminary results)». In S.I. Feldman, M. Uretsky, M. Najork, and C.E. Wills (Eds.), *WWW*, pp. 100–110. ACM, New York, NY, USA.

Faure, D. and Nédellec, C. (1998). «A Corpus-based Conceptual Clustering Method for Verb Frames and Ontology Acquisition». In *Proceedings of the LREC Workshop on Adapting Lexical and Corpus Resources to Sublanguages and Applications*.

Fellbaum, C. (Ed.) (1998). *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Frank, E., Paynter, G.W., Witten, I.H., Gutwin, C., and Nevill-Manning, C.G. (1999). «Domain-Specific Keyphrase Extraction». In T. Dean (Ed.), *Proceedings of the 16th International Joint Conference on Artificial Intelligence, IJCAI 1999*, pp. 668–673. Morgan Kaufmann, Stockholm, Sweden.

Frantzi, K.T., Ananiadou, S., and Mima, H. (2000). «Natural Language Processing for Digital Libraries: Automatic Recognition of Multi-Word Terms:. the C-value/NC-value method». *International Journal on Digital Libraries*, **vol. 3(2)**, pp. 115–130.

Frantzi, K.T., Ananiadou, S., and Tsujii, J. (1998). «The C-value/NC-value Method of Automatic Recognition for Multi-Word Terms». In C. Nikolaou and C. Stephanidis (Eds.), *Proceedings of the 2nd European Conference on Research and Advanced Technology for Digital Libraries, ECDL '98*, *Lecture Notes in Computer Science*, vol. 1513, pp. 585–604. Springer, Heraklion, Crete, Greece.

Gagné, R.M., Wagner, W.W., Golas, K., and Keller, J.M. (2004). *Principles of Instructional Design*. Wadsworth Publishing, 5th edn.

Grefenstette, G. (1994). *Explorations in Automatic Thesaurus Discovery*. Kluwer Academic Publishers.

Gruber, T.R. (1991). «The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases». In J.F. Allen, R. Fikes, and E. Sandewall (Eds.), *Proceedings of the 2nd International Conference on Principles of*

Knowledge Representation and Reasoning, KR'91, pp. 601–602. Morgan Kauffmann Publishers, Inc., Cambridge, MA, USA.

Gruber, T.R. (1993). «A translation approach to portable ontologies». *Knowledge Acquisition*, **vol. 5(2)**, pp. 199–220.

Guarino, N. (1997). «Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration». In M.T. Pazienza (Ed.), *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology, International Summer School, SCIE-97*, *Lecture Notes in Computer Science*, vol. 1299, pp. 139–170. Springer, Frascati, Italy.

Gurrutxaga, A., Saralegi, X., Ugartetxea, S., and Alegria, I. (2005). «Erauzterm: euskarazko terminoak erauzteko tresna erdiautomatikoa». In *IX. Jardunaldiak: Euskera zientifiko-teknikoa*. Mendebalde Kultur Alkartea.

Gutiérrez, J. (1994). *INTZA: un Sistema Tutor Inteligente para Entrenamiento en Entornos Industriales*. Ph.D. thesis, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU).

Haase, P. and Völker, J. (2008). «Ontology Learning and Reasoning - Dealing with Uncertainty and Inconsistency». In P.C.G. da Costa, C. d'Amato, N. Fanizzi, K.B. Laskey, K.J. Laskey, T. Lukasiewicz, M. Nickles, and M. Pool (Eds.), *Uncertainty Reasoning for the Semantic Web I*, *Lecture Notes in Computer Science*, vol. 5327, pp. 366–384. Springer.

Harris, Z. (1954). «Distributional Structure». *Word*, **vol. 10(32)**, pp. 146–162.

Hayashi, Y., Bourdeau, J., and Mizoguchi, R. (2009). «Using Ontological Engineering to Organize Learning/Instructional Theories and Build a Theory-Aware Authoring System». *International Journal on Artificial Intelligence in Education*, **vol. 19(2)**, pp. 211–252.

Hearst, M.A. (1992). «Automatic Acquisition of Hyponyms from Large Text Corpora». In *Proceedings of the 14th Conference on Computational Linguistics, COLING-92*, pp. 539–545. Association for Computational Linguistics, Nantes, France.

de Hoog, R., Barnard, Y., and Wielinga, B.J. (1999). «IMAT: Re-using Multimedia Electronic Technical Documentation for Training». In J.Y. Roger,

B. Stanford-Smith, and P.T. Kidd (Eds.), *Business and Work in the Information Society: New Technologies and Applications*, pp. 415–421. IOS Press.

Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B., and Dean, M. (2004). «SWRL: a Semantic Web Rule Language Combining OWL and RuleML». http://www.w3.org/Submission/SWRL/.

Hughes, T. and Ramage, D. (2007). «Lexical Semantic Relatedness with Random Graph Walks». In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP-CONLL 2007*, pp. 581–589. Association for Computational Linguistics, Prague, Czech Republic.

IMS Global Learning Consortium Inc. (2003a). «IMS Learning Design Information Model». `http://www.imsglobal.org/learningdesign/index.html`.

IMS Global Learning Consortium Inc. (2003b). «IMS Simple Sequencing Behavior and Information Model v1.0 Final Specification». `http://www.imsglobal.org/simplesequencing/`.

IMS Global Learning Consortium Inc. (2005). «IMS ePortfolio Specification». `http://www.imsglobal.org/ep/`.

IMS Global Learning Consortium Inc. (2007). «IMS Content Packaging Information Model Version 1.2 Public Draft v2.0». `http://www.imsglobal.org/content/packaging/cpv1p2pd2/imscp_infov1p2pd2.html`.

IMS Global Learning Consortium Inc. (2008). «IMS Question and Test Interoperability». `http://www.imsglobal.org/question/index.html`.

IMS Global Learning Consortium Inc. (2011). «IMS GLC Common Cartridge Specification». `http://www.imsglobal.org/cc/index.html`.

Iruskieta, M., Díaz de Ilarraza, A., and Lersundi, M. (2010). «Correlaciones en Euskera entre las Relaciones Retóricas y los Marcadores de Discurso». In R. Caballero and M.J. Pinar (Eds.), *Ways and Modes of Human Communication*. Universidas de Castilla-La Mancha, Cuenca.

Justeson, J.S. and Katz, S.M. (1995). «Technical Terminology: Some Linguistic Properties and an Algorithm for Identification of Terms in Text». *Natural Language Engineering*, **vol. 1(1)**, pp. 9–27.

Kabel, S.C., de Hoog, R., Wielinga, B.J., and Anjewierden, A. (2004a). «Indexing Learning Objects: Vocabularies and Empirical Investigation of Consistency». *Journal of Educational Multimedia and Hypermedia*, **vol. 13(4)**, pp. 405–425.

Kabel, S.C., de Hoog, R., Wielinga, B.J., and Anjewierden, A. (2004b). «The Added Value of Task and Ontology Based Mark-Up for Information Retrieval». *Journal of the American Society for Information Science and Technology*, **vol. 55(4)**, pp. 348–362.

Kabel, S.C., Wielinga, B.J., and de Hoog, R. (1999). «Ontologies for Indexing Technical Manuals for Instruction». In *Workshop on Ontologies for Intelligent Educational Systems at 9th International Conference on Artificial Intelligence in Education, AI-ED' 99*, pp. 44–53. Le Mans, France.

Karlsson, F., Voutilainen, A., and Heikkila, J. (Eds.) (1995). *Constraint Grammar: Language-independent System for Parsing Unrestricted Text*. No. 4 in Natural Language Processing. Mouton de Gruyter.

Kavalec, M. and Svátek, V. (2005). «A Study on Automated Relation Labelling in Ontology Learning». In Buitelaar *et al.* (2005b), pp. 44–58.

Kay, J. (1999). «Ontologies for Reusable and Scrutable Student Models». In *Workshop on Ontologies for Intelligent Educational Systems at 9th International Conference on Artificial Intelligence in Education, AI-ED' 99*. Le Mans, France.

Kay, J. and Lum, A. (2004). «Ontologies for Scrutable Student Modelling in Adaptive E-Learning». In L. Aroyo, D. Dicheva, P. Dolog, and M. Wolpers (Eds.), *SW-EL'04: Semantic Web for E-Learning. Applications of Semantic Web Technologies for Educational Adaptive Hypermedia Workshop at International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, AH 2004*. Eindhoven, The Netherlands.

Klein, D. and Manning, C.D. (2002). «Fast Exact Inference with a Factored Model for Natural Language Parsing». In S. Becker, S. Thrun, and K. Obermayer (Eds.), *NIPS*, pp. 3–10. MIT Press, Vancouver, British Columbia, Canada.

Klein, D. and Manning, C.D. (2003). «Accurate Unlexicalized Parsing». In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, ACL 2003*, pp. 423–430. ACL.

Knott, A. and Dale, R. (1994). «Using Linguistic Phenomena to Motivate a Set of Coherence Relations». *Discourse Processes*, **vol. 18(1)**, pp. 35–62.

Koedinger, K.R., Anderson, J.R., Hadley, W.H., and Mark, M.A. (1997). «Intelligent Tutoring Goes To School in the Big City». *International Journal on Artificial Intelligence in Education*, **vol. 8**, pp. 30–43.

Koper, R. (2006). «Current Research in Learning Design». *Educational Technology & Society*, **vol. 9(1)**, pp. 13–22.

Lagoze, C. and de Sompel, H.V. (2001). «The open archives initiative: building a low-barrier interoperability framework». In *Proceddings of the 1st ACM/IEEE Joint Conference on Digital Libraries, JCDL 2001*, pp. 54–62. ACM, Roanoke, Virginia, USA.

Lama, M., Sánchez, E., Amorim, R.R., and Vila, X.A. (2005). «Semantic Description of the IMS Learning Design Specification». In L. Aroyo and D. Dicheva (Eds.), *SW-EL'05: International Workshop on Applications of the Semantic Web for E-Learning Technologies at the 12th International Conference on Artificial Intelligence in Education, AIED 2005*, pp. 37–46. SW-EL, Amsterdam, The Netherlands.

Larrañaga, M., Calvo, I., Elorriaga, J.A., Arruarte, A., Verbert, K., and Duval, E. (2011). «ErauzOnt: A Framework for Gathering Learning Objects from Electronic Documents». In *Proceedings of the 11th IEEE International Conference on Advanced Learning Technologies, ICALT 2011*, pp. 656–658. IEEE Computer Society, Athens, Georgia, USA.

Larrañaga, M., Conde, A., Calvo, I., Arruarte, A., and Elorriaga, J.A. (2012). «Evaluating the Automatic Extraction of Learning Objects from Electronic Textbooks Using ErauzOnt». In S.A. Cerri, W.J. Clancey, G. Papadourakis, and K. Panourgia (Eds.), *Proceedings of the 11th International Conference on Intelligent Tutoring Systems, ITS 2012*, *Lecture Notes in Computer Science*, vol. 7315, pp. 655–656. Springer, Crete, Greece.

Larrañaga, M., Conde, A., Calvo, I., Elorriaga, J.A., and Arruarte, A. (submitted). «Automatic Generation of the Domain Module from Electronic Textbooks. Method & Validation». *IEEE Transactions on Knowledge and Data Engineering.*

Larrañaga, M., Elorriaga, J.A., and Arruarte, A. (2008a). «A Heuristic NLP Based Approach for Getting Didactic Resources from Electronic Documents». In P. Dillenbourg and M. Specht (Eds.), *Times of Convergence. Technologies Across Learning Contexts, Proceedings of the 3rd European Conference on Technology Enhanced Learning, EC-TEL 2008*, *Lecture Notes in Computer Science*, vol. 5192, pp. 197–202. Springer, Maastricht, The Netherlands.

Larrañaga, M., Elorriaga, J.A., and Arruarte, A. (2008b). «Building Learning Objects from Electronic Documents». In *Proceedings of the 16th International Conference on Computers in Education, ICCE 2008*, pp. 141–146. Asia-Pacific Society for Computers in Education, Taipei, Taiwan.

Larrañaga, M., Elorriaga, J.A., and Arruarte, A. (2008c). «Semi Automatic Generation of Didactic Resources from Existing Documents». In B.P. Woolf, E. Aïmeur, R. Nkambou, and S.P. Lajoie (Eds.), *Proceedings of the 9th International Conference on Intelligent Tutoring Systems, ITS 2008*, *Lecture Notes in Computer Science*, vol. 5091, pp. 728–730. Springer, Montreal, Canada.

Larrañaga, M., Niebla, I., Rueda, U., Elorriaga, J.A., and Arruarte, A. (2007). «Towards Collaborative Domain Module Authoring». In J.M. Spector, D.G. Sampson, T. Okamoto, Kinshuk, S.A. Cerri, M. Ueno, and A. Kashihara (Eds.), *Proceedings of the 7th IEEE International Conference on Advanced Learning Technologies, ICALT 2007*, pp. 814–818. IEEE Computer Society, Niigata, Japan.

Larrañaga, M., Rueda, U., Elorriaga, J.A., and Arruarte, A. (2004). «Acquisition of the Domain Structure from Document Indexes Using Heuristic Reasoning». In J.C. Lester, R.M. Vicari, and F. Paraguaçu (Eds.), *Proceedings of the 7th International Conference on Intelligent Tutoring Systems, ITS 2004*, *Lecture Notes in Computer Science*, vol. 3220, pp. 175–186. Springer, Maceiò, Alagoas, Brazil.

Larrañaga, M., Rueda, U., Kerejeta, M., Elorriaga, J.A., and Arruarte, A. (2005). «CM-DOM: a Concept Map Based Tool for Supervising Domain Acquisition». In C.K. Looi, G.I. McCalla, B. Bredeweg, and J. Breuker (Eds.), *Artificial Intelligence in Education - Supporting Learning through Intelligent and Socially Informed Technology, Proceedings of the 12th International Conference on Artificial Intelligence in Education, AIED 2005*, *Frontiers in*

Artificial Intelligence and Applications, vol. 125, pp. 857–859. IOS Press, Amsterdam, The Netherlands.

Leidig, T. (2001). «L3–Towards an Open Learning Environment». *ACM Journal of Educational Resources in Computing*, **vol. 1(1)**, pp. 5–11.

Lentini, M., Nardi, D., and Simonetta, A. (1995). «Automatic Generation of Tutors for Spreadsheet Applications». In J.E. Greer (Ed.), *Proceedings of the 7th International Conference on Artificial Intelligence in Education, AIED 1995*, pp. 59–66. AACE, Washington, DC.

Lentini, M., Nardi, D., and Simonetta, A. (2000). «Self-instructive spreadsheets: an environment for automatic knowledge acquisition and tutor generation». *International Journal on Human-Computer Studies*, **vol. 52(5)**, pp. 775–803.

Lesgold, A. (1988). «Toward a Theory of Curriculum for Use in Designing Intelligent Instructional Systems». In H. Mandl and A. Lesgold (Eds.), *Learning Issues for Intelligent Tutoring Systems*, pp. 114–137. Springer.

Levy, R. and Manning, C.D. (2003). «Is it Harder to Parse Chinese, or the Chinese Treebank?» In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, ACL 2003*, pp. 439–446. ACL.

Lin, D. and Pantel, P. (2001). «DIRT - Discovery of Inference Rules from Text». In *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'01*, pp. 323–328.

Liu, B., Chin, C.W., and Ng, H.T. (2003). «Mining Topic-specific Concepts and Definitions on the Web». In I. King and T. Maray (Eds.), *Proceedings of the 12th International Conference on World Wide Web, WWW 2003*, pp. 251–260. ACM Press, Budapest, Hungary. URL `http://dx.doi.org/10.1145/775152.775188`.

LTSC (2001). «1484.12.1 IEEE LTSC Draft Standard for Learning Object Metadata». URL `http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf`.

LTSC (2003). «IEEE P1484.11.2 Standard for Learning Technology - ECMAScript Application Programming Interface for Content to Runtime Services Communication». `http://ltsc.ieee.org/wg11/files/IEEE_P1484-11-2_ballot-d2.pdf`.

Lu, R., Cao, C., Chen, Y., and Han, Z. (1995). «On Automatic Generation of Intelligent Tutoring Systems». In J.E. Greer (Ed.), *Proceedings of the 7th International Conference on Artificial Intelligence in Education, AIED 1995*, pp. 67–74. AACE, Washington, DC.

Maedche, A. (2002). *Ontology Learning for the Semantic Web*. Kluwer Academic Publishers.

Maedche, A. and Staab, S. (2000). «Discovering Conceptual Relations from Text». In W. Horn (Ed.), *Proceedings of the 14th European Conference on Artificial Intelligence, ECAI 2000*, pp. 321–325. IOS Press, Berlin, Germany.

Maedche, A. and Staab, S. (2001). «Ontology Learning for the Semantic Web». *IEEE Intelligent Systems*, **vol. 16(2)**, pp. 72–79.

de Marneffe, M.C., MacCartney, B., and Manning, C.D. (2004). «Generating Typed Dependency Parsers fromm Phrase Structure Parsers». In *Proceedings of the 4rd International Conference on Language Resources and Evaluation, LREC 2004*, pp. 449–454. ELRA, Lisboa, Portugal.

Martin, B. and Mitrovic, A. (2003). «ITS Domain Modelling: Art or Science?» In U. Hoppe, F. Verdejo, and J. Kay (Eds.), *Shaping the Future of Learning through Intelligent Technologies, Proceedings of the 11th International Conference on Artificial Intelligence in Education, AIED 2003*, *Frontiers in Artificial Intelligence and Applications*, vol. 97, pp. 183–190. IOS Press, Sidney, Australia.

Martin, B., Mitrovic, A., and Suraweera, P. (2008). «ITS Domain Modelling with Ontology». *Journal of Universal Computer Science*, **vol. 14(17)**, pp. 2758–2776.

McMartin, F.P. (2004). «MERLOT: A Model for User Involvement in Digital Library Design and Implementation». *Journal of Digital Information*, **vol. 5(3)**.

Meder, N. (2000). «Didaktische ontologien». In H.P. Ohly, G. Rahmstorf, and A. Sige (Eds.), *Globalisierung und Wissensorganisation: Neue Aspekte für Wissen, Wissenschaft und Informationssysteme*, vol. 6, pp. 401–416.

Meire, M., Ochoa, X., and Duval, E. (2007). «SAmgI: Automatic Metadata Generation v2.0». In C. Montgomerie and J. Seale (Eds.), *Proceedings of the World*

Conference on Educational Multimedia, Hypermedia and Telecommunications 2007, ED-MEDIA 2007, pp. 1195–1204. AACE, Vancouver, Canada. URL `http://www.editlib.org/p/25528`.

Merrill, M.D. (1999). «Instructional Transaction Theory (ITT): Instructional Design Based on Kowledge Objects». In C.M. Reigeluth (Ed.), *Instructional Design Theories and Models: A New Paradigm of Instructional Theory*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, USA.

Mitrovic, A. (1998). «A Knowledge-Based Teaching System for SQL». In T. Ottmann and I. Tomek (Eds.), *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications 1998, ED-MEDIA 1998*, pp. 1027–1032. AACE.

Mitrovic, A. (2003). «An Intelligent SQL Tutor on the Web». *International Journal on Artificial Intelligence in Education*, **vol. 13**, pp. 171–195.

Mitrovic, A. and Devedžić, V. (2002). «A Model of Multitutor Ontology-Based Leaning Environments». In Kinshuk, R. Lewis, K. Akahori, R. Kemp, T. Okamoto, L. Henderson, and C. Lee (Eds.), *Proceedings of the 10th International Conference on Computers in Education, ICCE 2002*. IEEE Computer Society, Auckland, New Zeland.

Mitrovic, A., Martin, B., and Suraweera, P. (2007). «Intelligent Tutors for All: The Constraint-Based Approach». *IEEE Intelligent Systems*, **vol. 22(4)**, pp. 38–45.

Mitrovic, A. and Ohlsson, S. (1999). «Evaluation of a Constraint-Based Tutor for a Database Language». *International Journal on Artificial Intelligence in Education*, **vol. 10(1)**, pp. 238–256.

Mitrovic, A., Suraweera, P., Martin, B., and Weerasinghe, A. (2004). «DB-suite: Experiences with Three Intelligent, Web-based Database Tutors». *Journal of Interactive Learning Research*, **vol. 15(4)**, pp. 409–432.

Mizoguchi, R. and Bourdeau, J. (2000). «Using Ontological Engineering to Overcome AI-ED Problems». *International Journal on Artificial Intelligence in Education*, **vol. 11(2)**, pp. 107–121.

Mizoguchi, R. and Ikeda, M. (1997). «Towards Ontology Engineering». In *Proceedings of the Joint Pacific Asian Conference on Expert Systems/Singapore*

International Conference on Intelligent Systems, PACES/SPICIS'97, pp. 259–266. Singapore.

Morin, E. and Jaquemin, C. (1999). «Projecting Corpus-Based Semantic Links on a Thesaurus». In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, ACL 1999*, pp. 389–396.

Murray, T. (1996). «Special Purpose Ontologies and the Representation of Pedagogical Knowledge». In *Proceedings of the 2nd International Conference on Learning Sciences, ICLS 1996*, pp. 235–242. AACE, International Society of the Learning Sciences, Evanston, Illinois, USA.

Murray, T. (1998). «Authoring Knowledge-Based Tutors: Tools for Content, Instructional Strategy, Student Model and Interface Design». *The Journal of the Learning Sciences*, **vol. 7(1)**, pp. 5–64.

Murray, T. (1999). «Authoring Intelligent Tutoring Systems: An analysis of the state of the art». *International Journal on Artificial Intelligence in Education*, **vol. 10**, pp. 98–129.

Murray, T. (2003). «EON: Knowledge Based Tutor Authoring Tools for Content, Instructional, Strategy, student Model and Interface Design». In Murray *et al.* (2003).

Murray, T., Blessing, S., and Ainsworth, S. (Eds.) (2003). *Authoring Tools for Advanced Technology Learning Environments*. Kluwer Academic Publishers.

Navigli, R. and Velardi, P. (2004). «Learning Domain Ontologies from Document Warehouses and Dedicated Web Sites». *Computational Linguistics*, **vol. 30(2)**, pp. 151–179.

Neches, R., Fikes, R., Finin, T., Gruber, T.R., Senator, T., and Swartout, W. (1991). «Enabling Technology for Knowledge Sharing». *AI Magazine*, **vol. 12(3)**, pp. 36–56.

Nejdl, W., Wolf, B., Qu, C., Decker, S., Nilsson, M., Palmer, M., and Risch, T. (2002). «Edutella: A P2P Networking Infrastructure Based on RDF». In *Proceedings of the 11th International Conference on World Wide Web, WWW 2002*. ACM Press, Hawaii, USA.

Nkambou, R., Bourdeau, J., and Mizoguchi, R. (Eds.) (2010). *Advances in Intelligent Tutoring Systems*, *Studies in Computational Intelligence*, vol. 308. Springer.

Novak, J.D. (1977). *A Theory of Education*. Cornell University Press, cornell university, ithaca, ny edn.

Novak, J.D. (1990). «Concept Mapping: a Useful Tool for Science Education». *Journal of Research in Science Teaching*, **vol. 27(10)**, pp. 937–949.

Ohlsson, S. (1993). «Constraint-Based Student Modeling». *International Journal on Artificial Intelligence in Education*, **vol. 3(4)**, pp. 429–447.

Ohlsson, S. (1994). «Constraint-Based Student Modeling». In J.E. Greer and G.I. McCalla (Eds.), *Student Modelling: The Key to Individualized Knowledge-Based Instruction*, pp. 167–189. Springer.

Parsad, B. and Lewis, L. (2008). «Distance Education at Degree-Granting Postsecondary Institutions: 2006–07». Tech. rep., National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC.

Rivest, R.L. (1992). «The MD5 Message-Digest Algorithm (RFC 1321)». `http://www.ietf.org/rfc/rfc1321.txt`.

Rockley, A. (2002). *Managing Enterprise Content: A Unified Content Strategy*. New Riders Press.

Rueda, U. (2009). *Desarrollo de Software Reutilizable de Mapas Conceptuales y Estudio de su Aplicación en Contextos de Aprendizaje y en los Modelos Abiertos de Estudiante*. Ph.D. thesis, University of the Basque Country (UPV/EHU).

Rueda, U., Arruarte, A., Elorriaga, J.A., and Herrán, E. (2009). «Learning the Attachment Theory with the CM-ED Concept Map Editor». *Computers & Education*, **vol. 52(2)**, pp. 460–469.

Sager, J.C., Dungsworth, D., and McDonald, P.F. (1980). *English Special Languages: Principles and Practice in Science and Technology*. Brandstetter Verlag.

Salton, G. (1992). «The State of Retrieval System Evaluation». *Information Processing & Management*, **vol. 28(4)**, pp. 441–450.

Salton, G. and Buckley, C. (1988). «Term-Weighting Approaches in Automatic Text Retrieval». *Information Processing & Management*, **vol. 24(5)**, pp. 513–523.

Sanderson, M. and Croft, B. (1999). «Deriving Concept Hierarchies from Text». In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'99*, pp. 206–213. ACM Press, Berkeley, California, USA.

Schoonenboom, J. (2006). «A Model for Determining the Size of Learning Objects». In *Proceedings of the 6th IEEE International Conference on Advanced Learning Technologies, ICALT 2006*, pp. 46– 50. IEEE Computer Society, Kerkrade, The Netherlands.

Simon, B., Massart, D., Assche, F.V., Ternier, S., Duval, E., Brantner, S., Olmedilla, D., and Miklós, Z. (2005). «A Simple Query Interface for Interoperable Learning Repositories». In *Proceedings of the 1st Workshop on Interoperability of Web-based Educational Systems at the 14th International World Wide Web Conference, WWW 2005*, pp. 11–18. CEUR, Chiba, Japan.

Sosnovsky, S.A., Brusilovsky, P., Lee, D.H., Zadorozhny, V., and Zhou, X. (2008a). «Re-assessing the Value of Adaptive Navigation Support in E-Learning Context». In W. Nejdl, J. Kay, P. Pu, and E. Herder (Eds.), *AH*, *Lecture Notes in Computer Science*, vol. 5149, pp. 193–203. Springer, Hannover, Germany.

Sosnovsky, S.A., Mitrovic, A., Lee, D., Brusilovsky, P., and Yudelson, M. (2008b). «Ontology-based Integration of Adaptive Educational Systems». In *Proceedings of the 16th International Conference on Computers in Education, ICCE 2008*, pp. 11–18. Asia-Pacific Society for Computers in Education, Taipei, Taiwan.

Sowa, J. (2010). «Ontology». `http://www.jfsowa.com/ontology`. URL `http://www.jfsowa.com/ontology`.

Stevens, A., Collins, A., and Goldin, S.E. (1982). «Misconceptions in Student's Understanding». In D. Sleeman and J. Brown (Eds.), *Intelligent Tutoring Systems*, pp. 13–24. Academic Press.

Studer, R., Benjamins, V.R., and Fensel, D. (1998). «Knowledge Engineering, Principles and Methods». *Data and Knowledge Engineering*, **vol. 25(1-2)**, pp. 161–197.

Suraweera, P., Mitrovic, A., and Martin, B. (2004a). «The Role of Domain Ontology in Knowledge Acquisition for ITSs». In J.C. Lester, R.M. Vicari, and F. Paraguaçu (Eds.), *Proceedings of the 7th International Conference on Intelligent Tutoring Systems, ITS 2004*, *LNCS*, vol. 3220, pp. 207–216. Springer, Maceió, Alagoas, Brazil.

Suraweera, P., Mitrovic, A., and Martin, B. (2004b). «The Use of Ontologies in ITS Domain Knowledge Authoring». In L. Aroyo, D. Dicheva, A. Mitrovic, and V. Devedžić (Eds.), *SW-EL'04: Applications of Semantic Web Technologies for Web-based ITS at 7th International Conference on Intelligent Tutoring Systems, ITS 2004*, pp. 41–49. Maceió, Alagoas, Brazil.

Suthers, D.D. (2005). «Collaborative Knowledge Construction through Shared Representations». In *Proceedings of the 38th Hawaii International Conference on System Sciences, HICSS 2005*, vol. 1, p. 5.1. IEEE Computer Society.

Taboada, M. (2006). «Discourse Markers as Signals (or Not) of Rethorical Relations». *Journal of Pragmatics*, **vol. 38(4)**, pp. 567–592.

Tapanainen, P. (1996). *The Constraint Grammar parser CG-2*. Publications of the University of Helsinki.

Ternier, S. and Duval, E. (2006). «Interoperability of Repositories: The Simple Query Interface in ARIADNE». *International Journal on E-Learning*, **vol. 5(1)**, pp. 161–166.

Ternier, S., Massart, D., Assche, F.V., Smith, N., Simon, B., and Duval, E. (2008a). «A Simple Publishing Interface for Learning Object Repositories». In *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications 2008, ED-MEDIA 2008*, pp. 1840–1845. AACE, Vienna, Austria.

Ternier, S., Massart, D., Campi, A., Guinea, S., Ceri, S., and Duval, E. (2008b). «Interoperability for Searching Learning Object Repositories: The ProLearn Query Language». *D-Lib Magazine*, **vol. 14(1)**.

Ternier, S., Verbert, K., Parra, G., Vandeputte, B., Klerkx, J., Duval, E., Ordoñez, V., and Ochoa, X. (2009). «The Ariadne Infrastructure for Managing and Storing Metadata». *IEEE Internet Computing*, **vol. 13(4)**, pp. 18–25.

Uschold, M. and Gruninger, M. (1996). «Ontologies: Principles, Methods and Applications». *Knowledge Engineering Review*, **vol. 11(2)**, pp. 93–136.

Velardi, P., Navigli, R., Cucchiarello, A., and Neri, F. (2005). «Evaluation of OntoLearn, a Methodology for Automatic Learning of Domain Ontologies». In Buitelaar *et al.* (2005b), pp. 92–106.

Verbert, K. (2008). *An Architecture and Framework for Flexible Reuse of Learning Object Components*. Ph.D. thesis, Faculteit Ingenieurswetenschappen, Katholieke Universiteit Leuven, Leuven, Belgium.

Verbert, K. and Duval, E. (2004). «Towards a Global Component Architecture for Learning Objects: A Comparative Analysis of Learning Object Content Models». In L. Cantoni and C. McLoughlin (Eds.), *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004, ED-MEDIA 2004*, pp. 202–208. AACE, Chesapeake, VA.

Verbert, K., Gašević, D., Jovanović, J., and Duval, E. (2005). «Ontology-Based Learning Content Repurposing». In A. Ellis and T. Hagino (Eds.), *Proceedings of the 14th International Conference on World Wide Web, WWW 2005 - Special interest tracks and posters*, pp. 1140–1141. ACM Press, Chiba, Japan.

Verbert, K., Ochoa, X., and Duval, E. (2008). «The ALOCOM Framework: Towards Scalable Content Reuse». *Journal of Digital Information*, **vol. 9(1)**.

Verbert, K., Wiley, D.A., and Duval, E. (2009). «A Methodology and Framework for the Semi-automatic Assembly of Learning Objects». In U. Cress, V. Dimitrova, and M. Specht (Eds.), *Learning in the Synergy of Multiple Disciplines, Proceedings of the 4th European Conference on Technology Enhanced Learning, EC-TEL 2009*, *Lecture Notes in Computer Science*, vol. 5794, pp. 757–762. Springer, Nice, France.

Vossen, P. (Ed.) (1998). *EuroWordNet: a multilingual database with lexical semantic networks*. Kluwer Academic Publishers, Norwell, MA, USA.

Vossen, P. (2001). «Extending, Trimming and Fusing WordNet for Technical Documents». In *NAACL-2001 Workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customications*.

Voutilainen, A. and Tapanainen, P. (1993). «Ambiguity Resolution in a Reductionistic Parser». In *Proceedings of the 6th Conference on European Chapter of the Association for Computational Linguistics, EACL 1993*, pp. 394–403. Association for Computational Linguistics, Utrecht, The Netherlands.

Wache, H., Vögele, T., Visser, U., Stuckenschmidt, H., Shuster, G., Neumann, H., and Hübner, S. (2001). «Ontology-Based Integration of Information - A Survey of Existing Approaches». In A. Gómez-Pérez, M. Gruninger, H. Stuckenschmidt, and M. Uschold (Eds.), *Workshop on Ontologies and Information Sharing at International Joint Conference on Artificial Intelligence, IJCAI 2001*, pp. 108–117. Seattle, USA.

Waits, T. and Lewis, L. (2003). «Distance Education at Degree-Granting Postsecondary Institutions: 2000–2001». Tech. rep., National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC.

Wenger, E. (1987). «Artificial Intelligence and Tutoring Systems». In *Artificial Intelligence and Tutoring Systems*. Morgan Kauffmann Publishers, Inc.

Wiley, D.A. (2000). «Connecting Learning Objects to Instructional Design Theory: A Definition, a Metaphor, and a Taxonomy». In D.A. Wiley (Ed.), *The Instructional Use of Learning Objects*, pp. 3–23. Association for Instructional Technology. URL `http://www.reusability.org/read/`.

Woolf, B.P. (2008). *Building Intelligent Interactive Tutors: Student-centered Strategies for Revolutioniizing E-learning*. Morgan Kauffmann Publishers, Inc.

Woolf, B.P., Arroyo, I., Beal, C.R., and Murray, T. (2006). «Gender and Cognitive Differences in Help Effectiveness During Problem Solving». *International Journal of Technology, Instruction, Cognition and Learning*, **vol. 3**, pp. 89–95.

Worring, M., Wielinga, B.J., Anjewierden, A., Verster, F., Todoran, L., Kabel, S.C., and de Hoog, R. (2001). «Automatic Indexing Of Text And Graphics In Technical Manuals». In *Proceedings of the 2001 IEEE International Conference on Multimedia and Expo, ICME 2001*, pp. 949–952. IEEE Computer Society, Tokyo, Japan.

Zouaq, A. and Nkambou, R. (2008). «Building Domain Ontologies from Text for Educational Purposes». *IEEE Transactions on Learning Technologies*, **vol. 1(1)**, pp. 49–62.

Zouaq, A. and Nkambou, R. (2009a). «Enhancing Learning Objects with an Ontology-Based Memory». *IEEE Transactions on Knowledge and Data Engineering*, **vol. 21(6)**, pp. 881–893.

Zouaq, A. and Nkambou, R. (2009b). «Evaluating the Generation of Domain Ontologies in the Knowledge Puzzle Project». *IEEE Transactions on Knowledge and Data Engineering*, **vol. 21(11)**, pp. 1559–1572.

Appendices

A

Complete List of Patterns for Gathering Relationships

This appendix presents the heuristics used by the *LDO Builder* to gather the pedagogical relationships from the outlines of the textbooks and the patterns defined in the grammar that is used to elicit the pedagogical relationships from the whole document.

A.1 Heuristic for Eliciting Pedagogical Relationships from the Outline

The outlines of the textbooks are appropriate sources of information for the elicitation of pedagogical relationships. This section presents the set of heuristics used for the analysis of the outlines, which identify the *isA* relationships (Table A.1 and Table A.2), the *partOf* relationships (Table A.3) and the PRE relationships (Table A.4). These tables provide examples of fragments of outlines in which the heuristics can be applied, the confidence (Conf.) of the heuristics and the XML code that implements the preconditions of the heuristics.

Table A.1 – Individual Heuristics for the Identification of the *isA* Relationship

Id.	**Conf.**	**Example in Basque**	**Example in English**	**Code**
MWH	**0.93**	3.4.2 **Agenteak** 3.4.2.1 **Agente** mugikorrak 3.4.2.2 **Agente** estatikoak	3.4.2 **Agents** 3.4.2.1 Mobile **agents** 3.4.2.2 Stationary **agents**	<**PreConditions**> <**CurrentItem**>current</**CurrentItem**> <**ExtraItem**>SuperItem</**ExtraItem**> <**Condition**> <**MultiWord**> <**TargetItem**> <**Item**>current</**Item**> </**TargetItem**> <**ReferenceItem**> <**Item**>SuperItem</**Item**> </**ReferenceItem**> </**MultiWord**> </**Condition**> </**PreConditions**>
ENH	**1.00**	3.4.2 **Agenteak** 3.4.2.1 **Agente** mugikorrak 3.4.2.2 **Agente** estatikoak	3.4.2 **Agents** 3.4.2.1 Mobile **agents** 3.4.2.2 Stationary **agents**	<**PreConditions**> <**CurrentItem**>current</**CurrentItem**> <**ExtraItem**>SuperItem</**ExtraItem**> <**Condition**> <**EntityName**> <**TargetItem**> <**Item**>current</**Item**> </**TargetItem**> </**EntityName**> </**Condition**> </**PreConditions**>
AH	**0.9**	4.2.2 Interfaze grafikoak sortzeko lengoaiak 4.2.3.1 **XUL** 4.2.3.2 **jXUL** ...	4.2.2 Languages for building graphical interfaces 4.2.3.1 **XUL** 4.2.3.2 **jXUL** ...	<**PreConditions**> <**CurrentItem**>current</**CurrentItem**> <**ExtraItem**>SuperItem</**ExtraItem**> <**Condition**> <**Acronyms**> <**TargetItem**> <**Item**>current</**Item**> </**TargetItem**> </**Acronyms**> </**Condition**> </**PreConditions**>

Table A.1 – Individual Heuristics for the Identification of the *isA* Relationship (Continued)

Id.	Conf.	Example in Basque	Example in English	Code
He+MWH	0.82	11 **Agente** laguntzaileak 11.1 RMA **Agentea** 11.2 DF **Agentea**	11 Auxiliar **Agents** 11.1 RMA **Agent** 11.2 DF **Agent**	<**PreConditions**> <**CurrentItem**>current</**CurrentItem**> <**ExtraItem**>SuperItem</**ExtraItem**> <**Condition**> <**MultiWord**> <**TargetItem**> <**Item**>current</**Item**> </**TargetItem**> <**ReferenceItem**> <**Head**>SuperItem</**Head**> </**ReferenceItem**> </**MultiWord**> </**Condition**> </**PreConditions**>
A+MWH	0.87	7 **Abonatu-Linea Digitala** 7.1 **ALD** Simetrikoa 7.2 **ALD** Asimetrikoa	7 **Digital Subscriber Line** 7.1 Symmetrical **DSL** 7.2 Asymmetrical **DSL**	<**PreConditions**> <**CurrentItem**>current</**CurrentItem**> <**ExtraItem**>SuperItem</**ExtraItem**> <**Condition**> <**MultiWord**> <**TargetItem**> <**Item**>current</**Item**> </**TargetItem**> <**ReferenceItem**> <**GetAcr**>SuperItem</**GetAcr**> </**ReferenceItem**> </**MultiWord**> </**Condition**> </**PreConditions**>

Table A.2 – Group Heuristics for the Identification of the *isA* Relationship

Id.	Conf.	Example in Basque	Example in English	Code
KH	0.77	4.4.2 Izarren **adibideak** 4.4.2.1 Sirius izarra 4.4.2.2 Ipar Izarra	4.4.2 **Examples of** stars 4.4.2.1 Sirius star 4.4.2.2 Nothern star	<**PreConditions**> <**CurrentItem**>current</**CurrentItem**> <**ExtraItem**>SubItems</**ExtraItem**> <**Condition**> <**Keyword**> <**TargetItem**> <**Item**>current</**Item**> </**TargetItem**> </**Keyword**> </**Condition**> </**PreConditions**>
CHe+MWH	0.89	5.2 Zenbakizko taxonomia 5.2.1 **Clustering** banatzailea 5.2.2 **Clustering** hierarkikoa	5.2 Numeric classification 5.2.1 Exclusive **clustering** 5.2.2 Hierarchical **clustering**	<**PreConditions**> <**CurrentItem**>current</**CurrentItem**> <**ExtraItem**>SubItems</**ExtraItem**> <**Condition**> <**ForEach**> <**Item**><**Item**>SubItem</**Item**></**Item** > <**Set**>SubItems</**Set**> <**Condition**> <**MultiWord**> <**TargetItem**> <**Item**>SubItem</**Item**> </**TargetItem**> <**ReferenceItem**> <**GetCommonHead**>Subitems </**GetCommonHead**> </**ReferenceItem**> </**MultiWord**> </**Condition**> </**ForEach**> </**Condition**> </**PreConditions**>

Table A.3 – Individual Heuristic for the Identification of the *partOf* Relationship

Id.	Conf.	Example in Basque	Example in English	Code
PGH1	1.00	4.5 **Inplementazioa** 4.5.1 Aplikazioa**ren inplementazioa** 4.5.2 Agente**en inplementazioa**	4.5 **Implementation** 4.5.1 **Implementation of** the application 4.5.2 **Implementation of** the agents	<**PreConditions**> <**CurrentItem**>current</**CurrentItem**> <**ExtraItem**>SuperItem</**ExtraItem**> <**Condition**> <**PossessiveGenitive**> <**TargetItem**> <**Item**>current</**Item**> </**TargetItem**> <**ReferenceItem**> <**Item**>SuperItem</**Item**> </**ReferenceItem**> </**PossessiveGenitive**> </**Condition**> </**PreConditions**>

Table A.4 – Individual Heuristics for the Identification of the *prerequisite* Relationship

Id.	Conf.	Example in Basque	Example in English	Code
RH	0.93	4. **Sarrera/irteera** ... 6. Eraginkortasun altuko **sarrera/irteera**	4. **Input/output** ... 6. High performance **Input/output**	<**PreConditions**> <**CurrentItem**>current</**CurrentItem**> <**ExtraItem**>Preceding</**ExtraItem**> <**Condition**> <**Reference**> <**TargetItem**> <**Item**>current</**Item**> </**TargetItem**> <**ReferenceItem**> <**Item**>Preceding</**Item**> </**ReferenceItem**> </**Reference**> </**Condition**> </**PreConditions**>
He+RH	0.81	1.1 **Karga** elektrikoa ... 1.3 **Karga** kontserbazioa	1.1 Electric **charge** ... 1.3 **Charge** conservation	<**PreConditions**> <**CurrentItem**>current</**CurrentItem**> <**ExtraItem**>Preceding</**ExtraItem**> <**Condition**> <**Reference**> <**TargetItem**> <**Item**>current</**Item**> </**TargetItem**> <**ReferenceItem**> <**Head**>Preceding</**Head**> </**ReferenceItem**> </**Reference**> </**Condition**> </**PreConditions**>
A+RH	0.88	4.1 **Higidura Harmoniko Sinplea** ... 4.5 **HHS** eta higidura zirkular uniformea	4.1 **Simple Harmonic Motion** ... 4.5 **SHM** and the uniform circular movement	<**PreConditions**> <**CurrentItem**>current</**CurrentItem**> <**ExtraItem**>Preceding</**ExtraItem**> <**Condition**> <**Reference**> <**TargetItem**> <**Item**>current</**Item**> </**TargetItem**> <**ReferenceItem**> <**GetAcr**>Preceding</**GetAcr**> </**ReferenceItem**> </**Reference**> </**Condition**> </**PreConditions**>

Table A.4 – Individual Heuristics for the Identification of the *prerequisite* Relationship (Continued)

Id.	Conf.	Example in Basque	Example in English	Code
PGH2	0.96	3.5.2. **Agenteak** ... 3.5.3. **Agenteen** exekuzio eredua	3.5.2 **Agents** ... 3.5.3 Execution model **of agents**	<**PreConditions**> <**CurrentItem**>current</**CurrentItem**> <**ExtraItem**>Preceding</**ExtraItem**> <**Condition**> <**PossessiveGenitive**> <**TargetItem**> <**Item**>current</**Item**> </**TargetItem**> <**ReferenceItem**> <**Item**>Preceding</**Item**> </**ReferenceItem**> </**PossessiveGenitive**> </**Condition**> </**PreConditions**>
He+PGH2	0.83	1.1 **Karga** elektrikoa ... 1.4 **Karga** zehatz baten eremu elektrikoa	1.1 Electric **charge** ... 1.4 Electric field **of** a particular **charge**	<**PreConditions**> <**CurrentItem**>current</**CurrentItem**> <**ExtraItem**>Preceding</**ExtraItem**> <**Condition**> <**Reference**> <**TargetItem**> <**Item**>current</**Item**> </**TargetItem**> <**ReferenceItem**> <**Head**>Preceding</**Head**> </**ReferenceItem**> </**Reference**> </**Condition**> </**PreConditions**>
A+PGH2	0.89	4.1 **Higidura Harmoniko Sinplea** 4.2 **HHSren** zinematika	4.1 **Simple Harmonic Motion** 4.2 Cinematics **of the SHM**	<**PreConditions**> <**CurrentItem**>current</**CurrentItem**> <**ExtraItem**>Preceding</**ExtraItem**> <**Condition**> <**Reference**> <**TargetItem**> <**Item**>current</**Item**> </**TargetItem**> <**ReferenceItem**> <**GetAcr**>Preceding</**GetAcr**> </**ReferenceItem**> </**Reference**> </**Condition**> </**PreConditions**>

A.2 Patterns Defined in the Grammar for the Acquisition of Pedagogical Relationships

Next, the patterns defined by the grammar for identifying the pedagogical relationships from the whole textbook. These patterns allow the identification of the *isA* (Table A.5), the *partOf* (Table A.6), and the *prerequisite* (Table A.7) relationships.

These tables present, besides the syntactic structures that allow the identification of the relationships, some examples and the confidence of the patterns.

Table A.5 – Patterns for the Identification of the *isA* Relationship

Id.	Conf.	Example in Basque	Pattern in Basque	Example in English	Pattern in English
ISA1	**0.67**	Esne-bidea **izeneko** galaxiak 100 mila milioi izar baino gehiago dituela uste dute zientzialariek.	@Topic **izeneko|deituriko** @Topic	Sciencist believe that the galaxy **referred to as** Milky Way has over 100 billion stars.	@Topic **called|refered to as** @Topic
ISA2	**0.72**	Lurra planeta *bat* **da**.	@Topic @Topic [*det*] **IZAN**	The Earth **is** *a* planet.	@Topic **TO BE** [*det*] @Topic
ISA3	**0.87**	Zeruan hainbat konstelazio ikus ditzakegu, **adibidez** hartz nagusia...	@Topic **, adibidez** @Topic	Many constellations**, such as** Ursa Major ... can be observed in the sky.	@Topic **, such as** @Topic
ISA4	**0.57**	Eiffel dorrea munduko eraikin ezagu**nenetako bat da**	@Topic @Topic [adj(SUP)[1]]**-etako** [*def*] **IZAN**	The Eiffel tower **is one of the most recognizable** buildings in the world	@Topic **TO BE one|some of[ADJ(SUP)[1]]** @Topic

[1] The (SUP) states that the adjective is in superlative case

Table A.6 – Patterns for the Identification of the *partOf* Relationship

Id.	Conf.	Example in Basque	Pattern in Basque	Example in English	Pattern in English
POF1	**0.69**	Galixia izarrez **osatuta** galaxiak **daude**.	@Topic @Topic **osatuta EGON**	Galaxy **consist of** stars.	@Topic **consist of** @Topic
POF2	**0.71**	Galaxiak **dira** unibertsoaren **osagai** nagusiak.	@Topic **IZAN** @Topic(GEN)[1] **osagai**	Galaxies **are** the main **components of** the universe.	@Topic **TO BE component(s) of** [*det*] @Topic

[1] The (GEN) states that the @Topic is in possessive genitive case

Table A.7 – Pattern for the Identification of the *prerequisite* Relationship

Id.	Conf.	Example in Basque	Pattern in Basque	Example in English	Pattern in English
PRE1	**-**	Planeten**ren** mugimenduak . . .	@Topic(GEN) @Topic	The movements **of** *the* **planets**. . .	@Topic **of** [*det*] @Topic

B

Complete List of Patterns for Identifying Didactic Resources

This appendix presents the patterns defined by the DR grammar. These patterns are used to identify *definitions* (Table B.1), *examples* (Table B.2), *facts* (Table B.3), *principle statements* (Table B.4), *theories* or *theorems* (Table B.5) and *problem statements* (Table B.6).

Table B.1 – Patterns for Identifying Definitions

Id.	Conf.	Example in Basque	Pattern in Basque	Example in English	Pattern in English
DEF1	**0.67**	Unibertsoa *astro guztien multzo***ari** *eta betetzen duten espazio***ari** **esaten zaio**.	@Topic *text*(DAT)[1] **ESAN|DEITU**	universe **refers to** *the whole set of celestial bodies and the space they fill.*	@Topic **REFER to** *text*
DEF2	**0.70**	Galaxiak *milioika izarrek osatutako multzoak* **dira**.	@Topic *text* **IZAN**	Galaxies **are** *sets of millions of stars.*	@Topic **TO BE** *text*
DEF3	**1.00**	Asteroideak *planeta txiki***tzak** **jo daitezke**	@Topic *text* **-tzat jo**	Asteroids **can be considered as** *small planets.*	@Topic **VERB considered as** *text*
DEF4	**0.50**	Cometak**,** *isats argidun astroak* Antzinotik ezagutzen dira.	@Topic **,** *text* **,**	Comets**,** *light tailed celestial bodies***,** have been known for ages.	@Topic **,** *text* **,**
DEF5	**1.00**	Nebula **deritze** *izarren artean hedatzen diren gas eta hautsezko masa lainotsuei.*	@Topic **IRITZI** *text*	Nebula **is defined as** *the interstellar cloud of dust, hydrogen gas and plasma.*	@Topic **TO BE defined as** *text*
DEF6	**0.34**	Galaxiak, bestalde, elkarturik ager daitezke Kumulu **izeneko** *multzotan.*	@Topic **IZENEKO** *text*	*Besides, Galaxies can be found as aggregated sets of celestial bodies* **called** cumulus.	*text* **called** @Topic
DEF7	**1.00**	Ia galaxien erdiak kiribilak dira, **hau da,** *nukleo batez eta nukleo horretatik irteten diren besoez osatutako galaxia*	@Topic **, hau da,** *text*	Nearly half of the galaxies are spiral**, i.e.** *consist of a rotating disk of stars and interstellar medium, along with a central bulge of generally older stars.*	@Topic **, i.e.** *text*
DEF8	**0.50**	1 unitate astronimikoa = *149.579.870 km*	@Topic = *text*	1 astronomical unit = *149,579,870 km*	@Topic = *text*
DEF9	**0.32**	*Hala ere, kontuan izan behar da planeta bakoitzak orbita osoa egiteko behar duen denbora***,** translazio-periodoa**,** Eguzkiarekiko duen distantziaren araberakoa dela.	*text* **,** @Topic**,**	*However, the time each planet takes to make its complete orbit***,** orbital period**,** depends on the distance to the Sun.	*text* **,** @Topic**,**
DEF10	**1.00**	**zer dira** behatoki astronomiko horiek **?** ...	**zer IZAN** @Topic **?**	**what are** those astronomical observatories**?**	**what TO BE** @Topic **?**
DEF11	**0.74**	*Lurra 24 zatitan* **edo** ordu sektoretan banatzen da	*text* **edo** @Topic	*The Earth is divided in in 24 parts* **or** time zones	*text* **or** @Topic
DEF12	**0.77**	Lurraren ardatza 23,5° inklinatuta dago orbita ekliptikoarekiko **(***Eguzkiaren inguruan Lurrak egiten duen orbitaren planoarekiko***)**	@Topic**(***text***)**	The Earth axis is 23.5° inclined in reference to the Ecliptical orbit **(***the plane of the orbit the Earth makes around the Sun***)**	@Topic**(***text***)**
DEF13	**0.75**	*Hilean beste bitan, ilbehera eta ilgora garaian, erakarpenaren eragina txikiagoa da, eta mareak apalagoak***:** marea hilak	*text* **:** @Topic	*When the moon is at first quarter, the forces induced by the Sun partially cancel those of the Moon. At these points in the lunar cycle, the tide's range is minimum***:** neap tide	*text* **:** @Topic
DEF14	**0.67**	Materia**ren** **definizioa** hau **da**: *espazioaren lekua hartzen duen masa.*	@Topic(GEN)[2] **definizioa izan** *text*	This the **definition of** the matter*: anything which both occupies space and has mass*	**definition of** @Topic *text*

[1] The (DAT) states that the text is in dative case

[2] The (GEN) states that the @Topic is in possessive genitive case

Table B.2 – Patterns for Identifying Examples

Id.	Conf.	Example in Basque	Pattern in Basque	Example in English	Pattern in English
ADIB1	**0.33**	Zeruan hainbat konstelazio ikus ditzakegu, **esaterako|adibidez,** Hartz Nagusia.	**,esaterako|adibidez,** @Topic	Many constellations can be observed in the sky**, for instance,** Ursus Major.	**,e.g.|for example|for instance,** @Topic
ADIB2	**0.25**	Eiffel dorrea munduko eraikin ezagun**enetako bat da**	@Topic @Topic [adj(SUP)[1]]-**etako** [*def*] **IZAN**	The Eiffel tower **is one of the most recognizable** buildings in the world	@Topic **TO BE one|some of[ADJ(SUP)[1]]** @Topic

[1] The (SUP) states that the adjective is in superlative case

Table B.3 – Patterns for Identifying Facts

Id.	Conf.	Example in Basque	Pattern in Basque	Example in English	Pattern in English
FACT1	**0.60**	**1957:** Lurraren lehen satelite artifiziala (Sputnik I).	**Date:** @Topic	**Date:** Sputnik I was launched to orbit around the Earth	**YEAR:** @Topic

Table B.4 – Patterns for Identifying Principle Statements

Id.	**Conf.**	**Example in Basque**	**Pattern in Basque**	**Example in English**	**Pattern in English**
PS1	**0.66**	Ilargiaren aldiak *Ilargiak Lurraren inguruan egiten duen mugimendu***aren** **ondorioz gertatzen dira**	@Topic *text*(GEN)[1] **ondorioz|eraginez GERTATU**	Lunar Phases **are caused by** *the orbital movement of the Moon around the Earth.*	@Topic **TO BE due to|caused by** *text*
PS2	**0.56**	Unibertsoko gainontzeko astroak bezala, Eguzkia, Lurra eta Ilargia mugitu egiten dira, eta era bat baino gehiagoko mugimenduak egiten dituzte, gainera. Lurreko fenomeno askok, esaterako eguna eta gaua, eklipseak, edo itsasaldiak, mugimendu horietan **dute beren oinarria**.	@Topic @Topic **oinarri IZAN**	Like any celestial body, the Sun, the Earth and the Moon move in different ways. Many of the phenomena in Earth, e.g., the day, the night, eclipses and tides, **are based on** such movements.	@Topic **TO BE based on** @Topic

[1] The (GEN) states that the text is in possessive genitive case

Table B.5 – Patterns for Identifying Theorems or Theories

Id.	Conf.	Example in Basque	Pattern in Basque	Example in English	Pattern in English
TEO1	**0.63**	Eztanda handiaren **teoria** *unibertsoaren sorrera azaltzen duen eredua da.*	@Topic **teoria|teorema** *text*	The Big Bang **theory** is the prevailing cosmological model of the early development of the universe.	@Topic **theory|theorem** *text*

Table B.6 – Patterns for Identifying Problem Statements

Id.	Conf.	Example in Basque	Pattern in Basque	Example in English	Pattern in English
PROBS1	**0.95**	**Idatzi** eklipseei **buruz dakizun** guztia.	**Idatzi** @Topic (DAT)[1] **buruz jakin**	**Write** all **you know about** eclipses.	**Write know about** @Topic
PROBS2	**0.97**	**Erantzun galdera** hau:	**Erantzun galdera** [*det*]	**Answer** *this* **question**	**Answer** [*det*] **question**
PROBS3	**0.83**	**Ariketa** ...	**Ariketa**	**Exercise** ...	**Exercise**

[1] The (DAT) states that the @Topic is in dative case

C Ontologies

DOM-Sortze uses some ontologies, the Didactic Ontology (Meder, 2000; Leidig, 2001) and the ALOCOM ontology (Verbert *et al.*, 2005; Verbert, 2008), during the *Domain Module* construction. The *Domain Module* is described by the Learning Domain Ontology (LDO). This appendix presents the didactic ontology (Section C.1), the ALOCOM ontology (Section C.2), and how the LDO is represented, to which end some examples are also provided (Section C.3).

C.1 Didactic Ontology

The Didactic Ontology, which represents the different kinds of DRs and relationships between those types, is based on the didactic knowledge type taxonomy (Meder, 2000; Leidig, 2001). Listing C.1 shows the implementation of the didactic ontology in OWL.

Listing C.1 – Didactic Ontology in OWL

```
<?xml version="1.0"?>

<!DOCTYPE rdf:RDF [
    <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
    <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
    <!ENTITY balTaxonomia "http://lsi.vc.ehu.es/balTaxonomia#" >
    <!ENTITY owl2xml "http://www.w3.org/2006/12/owl2-xml#" >
    <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
    <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
    <!ENTITY lD "http://lsi.vc.ehu.es/2008/10/DidacticResources.owl#" >
```

```
]>

<rdf:RDF xmlns="http://lsi.vc.ehu.es/balTaxonomia#"
      xml:base="http://lsi.vc.ehu.es/balTaxonomia"
      xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
      xmlns:owl2xml="http://www.w3.org/2006/12/owl2-xml#"
      xmlns:balTaxonomia="http://lsi.vc.ehu.es/balTaxonomia#"
      xmlns:lD="http://lsi.vc.ehu.es/2008/10/DidacticResources.owl#"
      xmlns:owl="http://www.w3.org/2002/07/owl#"
      xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
    <owl:Ontology rdf:about="">
        <owl:imports rdf:resource="http://lsi.vc.ehu.es/2008/10/
            DidacticResources.owl"/>
    </owl:Ontology>

    <!--
    ///////////////////
    // Object Properties
    ///////////////////
     -->

    <!-- http://lsi.vc.ehu.es/2008/10/DidacticResources.owl#isA -->
    <owl:ObjectProperty rdf:about="&lD;isA"/>

    <!--
    ///////////////////
    // Classes
    ///////////////////
     -->

    <!-- http://lsi.vc.ehu.es/2008/10/DidacticResources.owl#
        DidacticResource -->
    <owl:Class rdf:about="&lD;DidacticResource"/>

    <!--
    ////////////////////
    // Individuals
    ////////////////////
     -->

    <!-- http://lsi.vc.ehu.es/balTaxonomia#T0 -->
    <lD:DidacticResource rdf:about="#T0">
        <rdfs:label xml:lang="en">Knowledge</rdfs:label>
    </lD:DidacticResource>
```

```
<!-- http://lsi.vc.ehu.es/balTaxonomia#T1 -->
<lD:DidacticResource rdf:about="#T1">
    <rdfs:label xml:lang="en">Explanation</rdfs:label>
    <lD:isA rdf:resource="#T0"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T10 -->
<lD:DidacticResource rdf:about="#T10">
    <rdfs:label xml:lang="en">Assumtion</rdfs:label>
    <lD:isA rdf:resource="#T1"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T11 -->
<lD:DidacticResource rdf:about="#T11">
    <rdfs:label xml:lang="en">Hyphotesis</rdfs:label>
    <lD:isA rdf:resource="#T10"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T12 -->
<lD:DidacticResource rdf:about="#T12">
    <rdfs:label xml:lang="en">Argumentation</rdfs:label>
    <lD:isA rdf:resource="#T1"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T13 -->
<lD:DidacticResource rdf:about="#T13">
    <rdfs:label xml:lang="en">Conclusion</rdfs:label>
    <lD:isA rdf:resource="#T12"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T14 -->
<lD:DidacticResource rdf:about="#T14">
    <rdfs:label xml:lang="en">Proof</rdfs:label>
    <lD:isA rdf:resource="#T12"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T15 -->
<lD:DidacticResource rdf:about="#T15">
    <rdfs:label xml:lang="en">Reflection</rdfs:label>
    <lD:isA rdf:resource="#T12"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T16 -->
<lD:DidacticResource rdf:about="#T16">
    <rdfs:label xml:lang="en">Interpretation</rdfs:label>
    <lD:isA rdf:resource="#T1"/>
```

```
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T17 -->
<lD:DidacticResource rdf:about="#T17">
    <rdfs:label xml:lang="en">Orientation</rdfs:label>
    <lD:isA rdf:resource="#T1"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T18 -->
<lD:DidacticResource rdf:about="#T18">
    <rdfs:label xml:lang="en">History</rdfs:label>
    <lD:isA rdf:resource="#T17"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T19 -->
<lD:DidacticResource rdf:about="#T19">
    <rdfs:label xml:lang="en">Fact</rdfs:label>
    <lD:isA rdf:resource="#T17"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T2 -->
<lD:DidacticResource rdf:about="#T2">
    <rdfs:label xml:lang="en">Example</rdfs:label>
    <lD:isA rdf:resource="#T1"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T20 -->
<lD:DidacticResource rdf:about="#T20">
    <rdfs:label xml:lang="en">Scenario</rdfs:label>
    <lD:isA rdf:resource="#T17"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T21 -->
<lD:DidacticResource rdf:about="#T21">
    <rdfs:label xml:lang="en"
        >HypotheticalSituation</rdfs:label>
    <lD:isA rdf:resource="#T20"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T22 -->
<lD:DidacticResource rdf:about="#T22">
    <rdfs:label xml:lang="en">VirtualWorld</rdfs:label>
    <lD:isA rdf:resource="#T20"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T23 -->
```

```
<lD:DidacticResource rdf:about="#T23">
    <rdfs:label xml:lang="en">StoryProblem</rdfs:label>
    <lD:isA rdf:resource="#T20"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T24 -->
<lD:DidacticResource rdf:about="#T24">
    <rdfs:label xml:lang="en">Summary</rdfs:label>
    <lD:isA rdf:resource="#T17"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T25 -->
<lD:DidacticResource rdf:about="#T25">
    <rdfs:label xml:lang="en">Overview</rdfs:label>
    <lD:isA rdf:resource="#T17"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T26 -->
<lD:DidacticResource rdf:about="#T26">
    <rdfs:label xml:lang="en">Reference</rdfs:label>
    <lD:isA rdf:resource="#T0"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T27 -->
<lD:DidacticResource rdf:about="#T27">
    <rdfs:label xml:lang="en">CrossReference</rdfs:label>
    <lD:isA rdf:resource="#T26"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T28 -->
<lD:DidacticResource rdf:about="#T28">
    <rdfs:label xml:lang="en"
        >GlossaryReferece</rdfs:label>
    <lD:isA rdf:resource="#T27"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T29 -->
<lD:DidacticResource rdf:about="#T29">
    <rdfs:label xml:lang="en">AnnexReference</rdfs:label>
    <lD:isA rdf:resource="#T27"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T3 -->
<lD:DidacticResource rdf:about="#T3">
    <rdfs:label xml:lang="en">CounterExample</rdfs:label>
    <lD:isA rdf:resource="#T2"/>
```

```
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T30 -->
<lD:DidacticResource rdf:about="#T30">
    <rdfs:label xml:lang="en"
        >ArchiveReference</rdfs:label>
    <lD:isA rdf:resource="#T26"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T31 -->
<lD:DidacticResource rdf:about="#T31">
    <rdfs:label xml:lang="en"
        >DocumentReference</rdfs:label>
    <lD:isA rdf:resource="#T26"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T32 -->
<lD:DidacticResource rdf:about="#T32">
    <rdfs:label xml:lang="en"
        >HandbookReference</rdfs:label>
    <lD:isA rdf:resource="#T31"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T33 -->
<lD:DidacticResource rdf:about="#T33">
    <rdfs:label xml:lang="en"
        >LexicalReference</rdfs:label>
    <lD:isA rdf:resource="#T31"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T34 -->
<lD:DidacticResource rdf:about="#T34">
    <rdfs:label xml:lang="en"
        >ProtocolReference</rdfs:label>
    <lD:isA rdf:resource="#T31"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T35 -->
<lD:DidacticResource rdf:about="#T35">
    <rdfs:label xml:lang="en"
        >StatisticReference</rdfs:label>
    <lD:isA rdf:resource="#T31"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T36 -->
<lD:DidacticResource rdf:about="#T36">
```

```
    <rdfs:label xml:lang="en">ReportReference</rdfs:label>
    <lD:isA rdf:resource="#T31"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T37 -->
<lD:DidacticResource rdf:about="#T37">
    <rdfs:label xml:lang="en">OnlineReference</rdfs:label>
    <lD:isA rdf:resource="#T26"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T38 -->
<lD:DidacticResource rdf:about="#T38">
    <rdfs:label xml:lang="en">Action</rdfs:label>
    <lD:isA rdf:resource="#T0"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T39 -->
<lD:DidacticResource rdf:about="#T39">
    <rdfs:label xml:lang="en">Rule</rdfs:label>
    <lD:isA rdf:resource="#T38"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T4 -->
<lD:DidacticResource rdf:about="#T4">
    <rdfs:label xml:lang="en">Declaration</rdfs:label>
    <lD:isA rdf:resource="#T1"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T40 -->
<lD:DidacticResource rdf:about="#T40">
    <rdfs:label xml:lang="en">Procedure</rdfs:label>
    <lD:isA rdf:resource="#T38"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T41 -->
<lD:DidacticResource rdf:about="#T41">
    <rdfs:label xml:lang="en"
        >OperatingDirections</rdfs:label>
    <lD:isA rdf:resource="#T40"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T42 -->
<lD:DidacticResource rdf:about="#T42">
    <rdfs:label xml:lang="en"
        >SocialDirections</rdfs:label>
    <lD:isA rdf:resource="#T40"/>
```

```
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T43 -->
<lD:DidacticResource rdf:about="#T43">
    <rdfs:label xml:lang="en">CheckList</rdfs:label>
    <lD:isA rdf:resource="#T40"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T44 -->
<lD:DidacticResource rdf:about="#T44">
    <rdfs:label xml:lang="en"
        >AdministrativeDirections</rdfs:label>
    <lD:isA rdf:resource="#T40"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T45 -->
<lD:DidacticResource rdf:about="#T45">
    <rdfs:label xml:lang="en">Strategy</rdfs:label>
    <lD:isA rdf:resource="#T38"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T46 -->
<lD:DidacticResource rdf:about="#T46">
    <rdfs:label xml:lang="en">Legal</rdfs:label>
    <lD:isA rdf:resource="#T38"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T47 -->
<lD:DidacticResource rdf:about="#T47">
    <rdfs:label xml:lang="en">LawComment</rdfs:label>
    <lD:isA rdf:resource="#T46"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T48 -->
<lD:DidacticResource rdf:about="#T48">
    <rdfs:label xml:lang="en">Law</rdfs:label>
    <lD:isA rdf:resource="#T46"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T49 -->
<lD:DidacticResource rdf:about="#T49">
    <rdfs:label xml:lang="en">Degree</rdfs:label>
    <lD:isA rdf:resource="#T46"/>
</lD:DidacticResource>

<!-- http://lsi.vc.ehu.es/balTaxonomia#T5 -->
```

```
    <lD:DidacticResource rdf:about="#T5">
        <rdfs:label xml:lang="en">Definition</rdfs:label>
        <lD:isA rdf:resource="#T4"/>
    </lD:DidacticResource>

    <!-- http://lsi.vc.ehu.es/balTaxonomia#T50 -->
    <lD:DidacticResource rdf:about="#T50">
        <rdfs:label xml:lang="en">Principle</rdfs:label>
        <lD:isA rdf:resource="#T38"/>
    </lD:DidacticResource>

    <!-- http://lsi.vc.ehu.es/balTaxonomia#T51 -->
    <lD:DidacticResource rdf:about="#T51">
        <rdfs:label xml:lang="en">Exercise</rdfs:label>
        <lD:isA rdf:resource="#T0"/>
    </lD:DidacticResource>

    <!-- http://lsi.vc.ehu.es/balTaxonomia#T6 -->
    <lD:DidacticResource rdf:about="#T6">
        <rdfs:label xml:lang="en"
            >FormulaDefinition</rdfs:label>
        <lD:isA rdf:resource="#T5"/>
    </lD:DidacticResource>

    <!-- http://lsi.vc.ehu.es/balTaxonomia#T7 -->
    <lD:DidacticResource rdf:about="#T7">
        <rdfs:label xml:lang="en">TermDefinition</rdfs:label>
        <lD:isA rdf:resource="#T5"/>
    </lD:DidacticResource>

    <!-- http://lsi.vc.ehu.es/balTaxonomia#T8 -->
    <lD:DidacticResource rdf:about="#T8">
        <rdfs:label xml:lang="en">Description</rdfs:label>
        <lD:isA rdf:resource="#T4"/>
    </lD:DidacticResource>

    <!-- http://lsi.vc.ehu.es/balTaxonomia#T9 -->
    <lD:DidacticResource rdf:about="#T9">
        <rdfs:label xml:lang="en">Theorem</rdfs:label>
        <lD:isA rdf:resource="#T4"/>
    </lD:DidacticResource>
</rdf:RDF>
```

The Didactic Ontology is built on the ontology shown in Listing C.2, which defines the *DidacticResource* class, and *isA* relationships.

Listing C.2 – Didactic Resources in OWL

```xml
<?xml version="1.0"?>


<!DOCTYPE rdf:RDF [
    <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
    <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
    <!ENTITY owl2xml "http://www.w3.org/2006/12/owl2-xml#" >
    <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
    <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
    <!ENTITY DidacticResources "http://lsi.vc.ehu.es/2008/10/
        DidacticResources.owl#" >
]>


<rdf:RDF xmlns="http://lsi.vc.ehu.es/2008/10/DidacticResources.owl#"
     xml:base="http://lsi.vc.ehu.es/2008/10/DidacticResources.owl"
     xmlns:DidacticResources="http://lsi.vc.ehu.es/2008/10/
         DidacticResources.owl#"
     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
     xmlns:owl2xml="http://www.w3.org/2006/12/owl2-xml#"
     xmlns:owl="http://www.w3.org/2002/07/owl#"
     xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

    <!-- http://lsi.vc.ehu.es/2008/10/DidacticResources.owl#isA -->
    <owl:ObjectProperty rdf:about="#isA"/>

    <!--
    /////////////////
    // Classes
    /////////////////
     -->
    <!-- http://lsi.vc.ehu.es/2008/10/DidacticResources.owl#
        DidacticResource -->
    <owl:Class rdf:about="#DidacticResource">
        <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    </owl:Class>

    <!-- http://www.w3.org/2002/07/owl#Thing -->
    <owl:Class rdf:about="&owl;Thing"/>
</rdf:RDF>
```

C.2 ALOCOM Ontology

The ALOCOM ontology (Verbert *et al.*, 2005; Verbert, 2008), which represents a content model for LOs is used as vocabulary for describing the kind of LOs elicited from the electronic textbooks. This ontology is shown in Listing C.3.

Listing C.3 – ALOCOM Ontology Represented in OWL

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [
    <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
    <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
    <!ENTITY owl2xml "http://www.w3.org/2006/12/owl2-xml#" >
    <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
    <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
    <!ENTITY Ontology1185016977 "http://www.owl-ontologies.com/
        Ontology1185016977.owl#" >
]>
<rdf:RDF xmlns="http://www.owl-ontologies.com/Ontology1185016977.owl#"
     xml:base="http://www.owl-ontologies.com/Ontology1185016977.owl"
     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
     xmlns:owl2xml="http://www.w3.org/2006/12/owl2-xml#"
     xmlns:owl="http://www.w3.org/2002/07/owl#"
     xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

    <!--
    //////////////
    // Classes
    //////////////
     -->

    <!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
        Additional_Resources -->
    <owl:Class rdf:about="#Additional_Resources">
        <rdfs:subClassOf rdf:resource="#Content_Object"/>
    </owl:Class>

    <!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Analogy
        -->
    <owl:Class rdf:about="#Analogy">
        <rdfs:subClassOf rdf:resource="#Content_Object"/>
    </owl:Class>
```

```
<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Animation
    -->
<owl:Class rdf:about="#Animation">
    <rdfs:subClassOf rdf:resource="#Content_Fragment"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Audio -->
<owl:Class rdf:about="#Audio">
    <rdfs:subClassOf rdf:resource="#Content_Fragment"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Chapter
    -->
<owl:Class rdf:about="#Chapter">
    <rdfs:subClassOf rdf:resource="#Larger_Objective_LO"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Chart -->
<owl:Class rdf:about="#Chart">
    <rdfs:subClassOf rdf:resource="#Graphic"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Concept
    -->
<owl:Class rdf:about="#Concept">
    <rdfs:subClassOf rdf:resource="#Single_Objective_LO"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Content_Fragment -->
<owl:Class rdf:about="#Content_Fragment">
    <rdfs:subClassOf rdf:resource="#LO_Component"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Content_Object -->
<owl:Class rdf:about="#Content_Object">
    <rdfs:subClassOf rdf:resource="#LO_Component"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Course --
    >
<owl:Class rdf:about="#Course">
    <rdfs:subClassOf rdf:resource="#LO_Aggregation"/>
</owl:Class>
```

```xml
<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Curriculum -->
<owl:Class rdf:about="#Curriculum">
    <rdfs:subClassOf rdf:resource="#LO_Aggregation"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Definition -->
<owl:Class rdf:about="#Definition">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Demonstration -->
<owl:Class rdf:about="#Demonstration">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Diagram
    -->
<owl:Class rdf:about="#Diagram">
    <rdfs:subClassOf rdf:resource="#Graphic"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Example
    -->
<owl:Class rdf:about="#Example">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Excursion
    -->
<owl:Class rdf:about="#Excursion">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Exercise
    -->
<owl:Class rdf:about="#Exercise">
    <rdfs:subClassOf rdf:resource="#Interactivity"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Experiment -->
<owl:Class rdf:about="#Experiment">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
```

```
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Explanation -->
<owl:Class rdf:about="#Explanation">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Fact -->
<owl:Class rdf:about="#Fact">
    <rdfs:subClassOf rdf:resource="#Single_Objective_LO"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Glossary
    -->
<owl:Class rdf:about="#Glossary">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Graph -->
<owl:Class rdf:about="#Graph">
    <rdfs:subClassOf rdf:resource="#Graphic"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Graphic
    -->
<owl:Class rdf:about="#Graphic">
    <rdfs:subClassOf rdf:resource="#Content_Fragment"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Guideline
    -->
<owl:Class rdf:about="#Guideline">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Illustration -->
<owl:Class rdf:about="#Illustration">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Importance -->
<owl:Class rdf:about="#Importance">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
```

```
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Interactivity -->
<owl:Class rdf:about="#Interactivity">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Introduction -->
<owl:Class rdf:about="#Introduction">
    <rdfs:subClassOf rdf:resource="#Explanation"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    LO_Aggregation -->
<owl:Class rdf:about="#LO_Aggregation">
    <rdfs:subClassOf rdf:resource="#Learning_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    LO_Component -->
<owl:Class rdf:about="#LO_Component"/>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Larger_Objective_LO -->
<owl:Class rdf:about="#Larger_Objective_LO">
    <rdfs:subClassOf rdf:resource="#Learning_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Learning_Object -->
<owl:Class rdf:about="#Learning_Object"/>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Lesson --
    >
<owl:Class rdf:about="#Lesson">
    <rdfs:subClassOf rdf:resource="#Larger_Objective_LO"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Literature -->
<owl:Class rdf:about="#Literature">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>
```

```
<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Map -->
<owl:Class rdf:about="#Map">
    <rdfs:subClassOf rdf:resource="#Graphic"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Module --
   >
<owl:Class rdf:about="#Module">
    <rdfs:subClassOf rdf:resource="#LO_Aggregation"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Motivation -->
<owl:Class rdf:about="#Motivation">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Next_Steps -->
<owl:Class rdf:about="#Next_Steps">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Non-
    Example -->
<owl:Class rdf:about="#Non-Example">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Objective
    -->
<owl:Class rdf:about="#Objective">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Open_Question -->
<owl:Class rdf:about="#Open_Question">
    <rdfs:subClassOf rdf:resource="#Interactivity"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Outline
   -->
<owl:Class rdf:about="#Outline">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>
```

```
<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Overview
    -->
<owl:Class rdf:about="#Overview">
    <rdfs:subClassOf rdf:resource="#Explanation"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Paragraph
    -->
<owl:Class rdf:about="#Paragraph">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Photograph -->
<owl:Class rdf:about="#Photograph">
    <rdfs:subClassOf rdf:resource="#Graphic"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Pictograph -->
<owl:Class rdf:about="#Pictograph">
    <rdfs:subClassOf rdf:resource="#Graphic"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Prerequisites -->
<owl:Class rdf:about="#Prerequisites">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Principle
    -->
<owl:Class rdf:about="#Principle">
    <rdfs:subClassOf rdf:resource="#Single_Objective_LO"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Principle_Statement -->
<owl:Class rdf:about="#Principle_Statement">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Problem_Statement -->
<owl:Class rdf:about="#Problem_Statement">
```

```
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Procedure
     -->
<owl:Class rdf:about="#Procedure">
    <rdfs:subClassOf rdf:resource="#Single_Objective_LO"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Process
    -->
<owl:Class rdf:about="#Process">
    <rdfs:subClassOf rdf:resource="#Single_Objective_LO"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Questionnaire -->
<owl:Class rdf:about="#Questionnaire">
    <rdfs:subClassOf rdf:resource="#Interactivity"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Reference
     -->
<owl:Class rdf:about="#Reference">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Remark --
    >
<owl:Class rdf:about="#Remark">
    <rdfs:subClassOf rdf:resource="#Explanation"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Review --
    >
<owl:Class rdf:about="#Review">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Scenario
    -->
<owl:Class rdf:about="#Scenario">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>
```

```
<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Self-
    assessment -->
<owl:Class rdf:about="#Self-assessment">
    <rdfs:subClassOf rdf:resource="#Interactivity"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Simulation -->
<owl:Class rdf:about="#Simulation">
    <rdfs:subClassOf rdf:resource="#Interactivity"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#
    Single_Objective_LO -->
<owl:Class rdf:about="#Single_Objective_LO">
    <rdfs:subClassOf rdf:resource="#Learning_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Story -->
<owl:Class rdf:about="#Story">
    <rdfs:subClassOf rdf:resource="#LO_Aggregation"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Summary
    -->
<owl:Class rdf:about="#Summary">
    <rdfs:subClassOf rdf:resource="#Explanation"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Symbol --
    >
<owl:Class rdf:about="#Symbol">
    <rdfs:subClassOf rdf:resource="#Graphic"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Table -->
<owl:Class rdf:about="#Table">
    <rdfs:subClassOf rdf:resource="#Content_Object"/>
</owl:Class>

<!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Text -->
<owl:Class rdf:about="#Text">
    <rdfs:subClassOf rdf:resource="#Content_Fragment"/>
</owl:Class>
```

```
    <!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Theorem
        -->
    <owl:Class rdf:about="#Theorem">
        <rdfs:subClassOf rdf:resource="#Content_Object"/>
    </owl:Class>

    <!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Timeline
        -->
    <owl:Class rdf:about="#Timeline">
        <rdfs:subClassOf rdf:resource="#Graphic"/>
    </owl:Class>

    <!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Unit -->
    <owl:Class rdf:about="#Unit">
        <rdfs:subClassOf rdf:resource="#LO_Aggregation"/>
    </owl:Class>

    <!-- http://www.owl-ontologies.com/Ontology1185016977.owl#Video -->
    <owl:Class rdf:about="#Video">
        <rdfs:subClassOf rdf:resource="#Content_Fragment"/>
    </owl:Class>
</rdf:RDF>
```

C.3 Learning Domain Ontology

The LDOs are represented in *DOM-Sortze* using two different formalisms. The XML formalism described in Section C.3.1 is used during the construction of the LDO. Once the LDO has been supervised, the LDO is represented as shown in Section C.3.2.

C.3.1 XML Representation of the Learning Domain Ontology during the Acquisition of the Learning Domain Ontology

The XML scheme shown in Listing C.4 defines the primitives used to describe the LDO. It contains the elements needed to define the topics, the relationships and the information about the heuristics used to elicit the ontology components.

Listing C.4 – XML Scheme for the Learning Domain Ontology

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
    <xs:element name="InferredOntology" type="OntologyType"/>
    <xs:complexType name="OntologyType">
        <xs:all>
            <xs:element name="TopicSet" type="TopicSetType"/>
            <xs:element name="RelationSet" type="RelationSetType"
               maxOccurs="1" minOccurs="0"/>
        </xs:all>
    </xs:complexType>
    <xs:complexType name="TopicSetType">
        <xs:sequence>
            <xs:element name="Topic" type="TopicType" maxOccurs="
               unbounded"/>
        </xs:sequence>
    </xs:complexType>
    <xs:complexType name="TopicType">
        <xs:all>
            <xs:element name="ItemId" type="xs:string"/>
            <xs:element name="ItemContent" type="xs:string"/>
            <xs:element name="Category" type="TopicCategoryType"
               maxOccurs="1" minOccurs="0"/>
            <xs:element name="Relevance" type="RelevanceType" maxOccurs
               ="1" minOccurs="0"/>
            <xs:element name="Difficulty" type="DifficultyType"
               maxOccurs="1" minOccurs="0"/>
```

```
            <xs:element name="DRs" type="DRsType" maxOccurs="1"
                minOccurs="1"/>
        </xs:all>
    </xs:complexType>
    <xs:complexType name="DRsType">
        <xs:sequence>
            <xs:element name="DR" type="DRType" minOccurs="0" maxOccurs
                ="unbounded"> </xs:element>
        </xs:sequence>
    </xs:complexType>
    <xs:complexType name="RelationSetType">
        <xs:sequence>
            <xs:element name="Relation" type="RelationType" maxOccurs="
                unbounded"/>
        </xs:sequence>
    </xs:complexType>
    <xs:complexType name="RelationType">
        <xs:all>
            <xs:element name="RelationID" type="xs:string"/>
            <xs:element name="Source" type="xs:string"/>
            <xs:element name="Target" type="xs:string"/>
            <xs:element name="Category" type="RelationCategoryType"/>
        </xs:all>
    </xs:complexType>
    <xs:complexType name="TopicCategoryType">
        <xs:sequence>
            <xs:element name="Category" type="TCategoryType" minOccurs=
                "1" maxOccurs="1"/>
            <xs:element name="InferredCategory" type="TCategoryType"
                minOccurs="0"/>
            <xs:element name="InferredBy" type="InferredByType"
                minOccurs="0"/>
            <xs:element name="Confidence" type="xs:float" minOccurs="0"
                />
        </xs:sequence>
    </xs:complexType>
    <xs:complexType name="RelevanceType">
        <xs:sequence>
            <xs:element name="Relevance" type="xs:string"/>
            <xs:element name="InferredRelevance" minOccurs="0"/>
            <xs:element name="InferredBy" type="InferredByType"
                minOccurs="0"/>
            <xs:element name="Confidence" type="xs:float" minOccurs="0"
                />
        </xs:sequence>
    </xs:complexType>
```

```
<xs:complexType name="DifficultyType">
    <xs:sequence>
        <xs:element name="Difficulty" type="xs:string"/>
        <xs:element name="InferredDifficulty" minOccurs="0"/>
        <xs:element name="InferredBy" type="InferredByType"
            minOccurs="0"/>
        <xs:element name="Confidence" type="xs:float" minOccurs="0"
            />
    </xs:sequence>
</xs:complexType>
<xs:complexType name="RelationCategoryType">
    <xs:sequence>
        <xs:element name="Category" type="RCategoryType"/>
        <xs:element name="InferredCategory" type="RCategoryType"
            minOccurs="0"/>
        <xs:element name="InferredBy" type="InferredByType"
            minOccurs="0"/>
        <xs:element name="Confidence" type="xs:float" minOccurs="0"
            />
    </xs:sequence>
</xs:complexType>
<xs:complexType name="InferredByType">
    <xs:choice>
        <xs:element name="UsedHeuristic" type="HeuristicNameType"/>
        <xs:element name="CombinedValue" type="HeuristicNameType"/>
    </xs:choice>
</xs:complexType>
<xs:complexType name="HeuristicNameType">
    <xs:sequence>
        <xs:element name="HeuristicName" type="xs:string" maxOccurs
            ="1"/>
    </xs:sequence>
</xs:complexType>
<xs:complexType name="DRType">
    <xs:all>
        <xs:element name="DRID" type="xs:string"/>
        <xs:element name="DRTYPE" type="xs:string"/>
        <xs:element name="DRPATH" type="xs:string"/>
    </xs:all>
</xs:complexType>
<xs:simpleType name="TCategoryType">
    <xs:restriction base="xs:string">
        <xs:enumeration value="Concept"/>
        <xs:enumeration value="Procedure"/>
        <xs:enumeration value="Principle"/>
        <xs:enumeration value="Fact"/>
```

```
        </xs:restriction>
    </xs:simpleType>
    <xs:simpleType name="RCategoryType">
        <xs:restriction base="xs:string">
            <xs:enumeration value="Is-A"/>
            <xs:enumeration value="Part-of"/>
            <xs:enumeration value="Prerequisite"/>
            <xs:enumeration value="Next"/>
            <xs:enumeration value="Same-That"/>
            <xs:enumeration value="InstanceOf"/>
        </xs:restriction>
    </xs:simpleType>
</xs:schema>
```

Next, an example of an LDO described using this formalism is presented (Listing C.5).

Listing C.5 – Example of a Learning Domain Ontology Represented in XML

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!-- Generated by GCMT -->
<InferredOntology xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:noNamespaceSchemaLocation="file:/Users/mikel/Documents/
            Tesia/Heurist/Emaitzak.xsd">
        <TopicSet>
                <Topic>
                        <ItemId>T11</ItemId>
                        <ItemContent>PLDen programazio ekipoak.</
                            ItemContent>
                        <Category>
                                <Category>Concept</Category>
                        </Category>
                        <Relevance>
                                <Relevance>HIGH</Relevance>
                        </Relevance>
                        <Difficulty>
                                <Difficulty>HIGH</Difficulty>
                        </Difficulty>
                        <DRs/>
                </Topic>
                <Topic>
                        <ItemId>T71</ItemId>
                        <ItemContent>Arrisku logikoen eliminazioa</
                            ItemContent>
                        <Category>
```

```
                <Category>Concept</Category>
            </Category>
            <Relevance>
                <Relevance>HIGH</Relevance>
            </Relevance>
            <Difficulty>
                <Difficulty>HIGH</Difficulty>
            </Difficulty>
            <DRs/>
        </Topic>
        <Topic>
            <ItemId>T24</ItemId>
            <ItemContent>CPLDak</ItemContent>
            <Category>
                <Category>Concept</Category>
            </Category>
            <Relevance>
                <Relevance>HIGH</Relevance>
            </Relevance>
            <Difficulty>
                <Difficulty>HIGH</Difficulty>
            </Difficulty>
            <DRs/>
        </Topic>
        <Topic>
            <ItemId>T3</ItemId>
            <ItemContent>Softwareko tresnak PLDak garatzeko
                . Hardwareko deskripzio
                lengoaiak</ItemContent>
            <Category>
                <Category>Concept</Category>
            </Category>
            <Relevance>
                <Relevance>HIGH</Relevance>
            </Relevance>
            <Difficulty>
                <Difficulty>HIGH</Difficulty>
            </Difficulty>
            <DRs/>
        </Topic>
        <Topic>
            <ItemId>T1</ItemId>
            <ItemContent>Zirkuitu integratuen sailkapena</
                ItemContent>
            <Category>
                <Category>Concept</Category>
```

```xml
                </Category>
                <Relevance>
                        <Relevance>HIGH</Relevance>
                </Relevance>
                <Difficulty>
                        <Difficulty>HIGH</Difficulty>
                </Difficulty>
                <DRs/>
        </Topic>
        <Topic>
                <ItemId>T10</ItemId>
                <ItemContent>PLDen Programazioa.</ItemContent>
                <Category>
                        <Category>Concept</Category>
                </Category>
                <Relevance>
                        <Relevance>HIGH</Relevance>
                </Relevance>
                <Difficulty>
                        <Difficulty>HIGH</Difficulty>
                </Difficulty>
                <DRs/>
        </Topic>
        <Topic>
                <ItemId>T81</ItemId>
                <ItemContent>Mealy eta Moore estatu-makinen
                    diseinua</ItemContent>
                <Category>
                        <Category>Concept</Category>
                </Category>
                <Relevance>
                        <Relevance>HIGH</Relevance>
                </Relevance>
                <Difficulty>
                        <Difficulty>HIGH</Difficulty>
                </Difficulty>
                <DRs/>
        </Topic>
        <Topic>
                <ItemId>T25</ItemId>
                <ItemContent>FPGAk</ItemContent>
                <Category>
                        <Category>Concept</Category>
                </Category>
                <Relevance>
                        <Relevance>HIGH</Relevance>
```

```
            </Relevance>
            <Difficulty>
                <Difficulty>HIGH</Difficulty>
            </Difficulty>
            <DRs/>
        </Topic>
        <Topic>
            <ItemId>T51</ItemId>
            <ItemContent>JEDEC</ItemContent>
            <Category>
                <Category>Concept</Category>
            </Category>
            <Relevance>
                <Relevance>HIGH</Relevance>
            </Relevance>
            <Difficulty>
                <Difficulty>HIGH</Difficulty>
            </Difficulty>
            <DRs/>
        </Topic>
        <Topic>
            <ItemId>T5</ItemId>
            <ItemContent>PLDak garatzeko fitxategi motak.
              Irteera formatuak:</ItemContent>
            <Category>
                <Category>Concept</Category>
            </Category>
            <Relevance>
                <Relevance>HIGH</Relevance>
            </Relevance>
            <Difficulty>
                <Difficulty>HIGH</Difficulty>
            </Difficulty>
            <DRs/>
        </Topic>
        <Topic>
            <ItemId>T52</ItemId>
            <ItemContent>Intel hexadecimal</ItemContent>
            <Category>
                <Category>Concept</Category>
            </Category>
            <Relevance>
                <Relevance>HIGH</Relevance>
            </Relevance>
            <Difficulty>
                <Difficulty>HIGH</Difficulty>
```

```
                </Difficulty>
                <DRs/>
            </Topic>
            <Topic>
                <ItemId>T61</ItemId>
                <ItemContent>Faktorizazioa</ItemContent>
                <Category>
                    <Category>Concept</Category>
                </Category>
                <Relevance>
                    <Relevance>HIGH</Relevance>
                </Relevance>
                <Difficulty>
                    <Difficulty>HIGH</Difficulty>
                </Difficulty>
                <DRs/>
            </Topic>
            <Topic>
                <ItemId>T4</ItemId>
                <ItemContent>PDLak programatzeko lengoaiaren
                  elementuak</ItemContent>
                <Category>
                    <Category>Concept</Category>
                </Category>
                <Relevance>
                    <Relevance>HIGH</Relevance>
                </Relevance>
                <Difficulty>
                    <Difficulty>HIGH</Difficulty>
                </Difficulty>
                <DRs/>
            </Topic>
            <Topic>
                <ItemId>T0</ItemId>
                <ItemContent>Logika programagarria</ItemContent
                  >
                <Category>
                    <Category>Concept</Category>
                </Category>
                <Relevance>
                    <Relevance>HIGH</Relevance>
                </Relevance>
                <Difficulty>
                    <Difficulty>HIGH</Difficulty>
                </Difficulty>
                <DRs/>
```

```
    </Topic>
    <Topic>
        <ItemId>T101</ItemId>
        <ItemContent>Programatutako PLDen egiaztapena.<
            /ItemContent>
        <Category>
            <Category>Concept</Category>
        </Category>
        <Relevance>
            <Relevance>HIGH</Relevance>
        </Relevance>
        <Difficulty>
            <Difficulty>HIGH</Difficulty>
        </Difficulty>
        <DRs/>
    </Topic>
    <Topic>
        <ItemId>T2</ItemId>
        <ItemContent>Gailu logiko programagarrien
            arkitektura (PSLDak)</ItemContent>
        <Category>
            <Category>Procedure</Category>
        </Category>
        <Relevance>
            <Relevance>HIGH</Relevance>
        </Relevance>
        <Difficulty>
            <Difficulty>HIGH</Difficulty>
        </Difficulty>
        <DRs/>
    </Topic>
    <Topic>
        <ItemId>T9</ItemId>
        <ItemContent>Diseinuen simulazioa</ItemContent>
        <Category>
            <Category>Concept</Category>
        </Category>
        <Relevance>
            <Relevance>HIGH</Relevance>
        </Relevance>
        <Difficulty>
            <Difficulty>HIGH</Difficulty>
        </Difficulty>
        <DRs/>
    </Topic>
    <Topic>
```

```
            <ItemId>T8</ItemId>
            <ItemContent>Estatu−makinen diseinua.</
                ItemContent>
            <Category>
                    <Category>Concept</Category>
            </Category>
            <Relevance>
                    <Relevance>HIGH</Relevance>
            </Relevance>
            <Difficulty>
                    <Difficulty>HIGH</Difficulty>
            </Difficulty>
            <DRs/>
    </Topic>
    <Topic>
            <ItemId>T21</ItemId>
            <ItemContent>PAL</ItemContent>
            <Category>
                    <Category>Concept</Category>
            </Category>
            <Relevance>
                    <Relevance>HIGH</Relevance>
            </Relevance>
            <Difficulty>
                    <Difficulty>HIGH</Difficulty>
            </Difficulty>
            <DRs/>
    </Topic>
    <Topic>
            <ItemId>T22</ItemId>
            <ItemContent>PLA</ItemContent>
            <Category>
                    <Category>Concept</Category>
            </Category>
            <Relevance>
                    <Relevance>HIGH</Relevance>
            </Relevance>
            <Difficulty>
                    <Difficulty>HIGH</Difficulty>
            </Difficulty>
            <DRs/>
    </Topic>
    <Topic>
            <ItemId>T7</ItemId>
            <ItemContent>Sekuentzia−zirkuituen diseinua,
                erlojurik gabe.</ItemContent>
```

```
        <Category>
            <Category>Concept</Category>
        </Category>
        <Relevance>
            <Relevance>HIGH</Relevance>
        </Relevance>
        <Difficulty>
            <Difficulty>HIGH</Difficulty>
        </Difficulty>
        <DRs/>
    </Topic>
    <Topic>
        <ItemId>T23</ItemId>
        <ItemContent>PROM</ItemContent>
        <Category>
            <Category>Concept</Category>
        </Category>
        <Relevance>
            <Relevance>HIGH</Relevance>
        </Relevance>
        <Difficulty>
            <Difficulty>HIGH</Difficulty>
        </Difficulty>
        <DRs/>
    </Topic>
    <Topic>
        <ItemId>T111</ItemId>
        <ItemContent>Gailu logiko programagarriak in
          system</ItemContent>
        <Category>
            <Category>Concept</Category>
        </Category>
        <Relevance>
            <Relevance>HIGH</Relevance>
        </Relevance>
        <Difficulty>
            <Difficulty>HIGH</Difficulty>
        </Difficulty>
        <DRs/>
    </Topic>
    <Topic>
        <ItemId>T6</ItemId>
        <ItemContent>Sekuentzia-zirkuitu sinkronoen
          diseinua.</ItemContent>
        <Category>
            <Category>Concept</Category>
```

```
                </Category>
                <Relevance>
                        <Relevance>HIGH</Relevance>
                </Relevance>
                <Difficulty>
                        <Difficulty>HIGH</Difficulty>
                </Difficulty>
                <DRs/>
        </Topic>
        <Topic>
                <ItemId>T102</ItemId>
                <ItemContent>Test  Bektoreak</ItemContent>
                <Category>
                        <Category>Concept</Category>
                </Category>
                <Relevance>
                        <Relevance>HIGH</Relevance>
                </Relevance>
                <Difficulty>
                        <Difficulty>HIGH</Difficulty>
                </Difficulty>
                <DRs/>
        </Topic>
</TopicSet>
<RelationSet>
        <Relation>
                <RelationID>IS-A16</RelationID>
                <Target>T0</Target>
                <Source>T5</Source>
                <Category>
                        <Category>Part-of</Category>
                        <InferredCategory>Part-of</
                            InferredCategory>
                        <InferredBy>
                                <UsedHeuristic>
                                        <HeuristicName>
                                            Structural-default</
                                            HeuristicName>
                                </UsedHeuristic>
                        </InferredBy>
                        <Confidence>0.77</Confidence>
                </Category>
        </Relation>
        <Relation>
                <RelationID>IS-A44</RelationID>
                <Target>T4</Target>
```

```
            <Source>T3</Source>
            <Category>
                <Category>Next</Category>
                <InferredCategory>Next</
                    InferredCategory>
                <InferredBy>
                    <UsedHeuristic>
                        <HeuristicName>
                            Sequential-default</
                            HeuristicName>
                    </UsedHeuristic>
                </InferredBy>
                <Confidence>0.76</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A41</RelationID>
            <Target>T0</Target>
            <Source>T4</Source>
            <Category>
                <Category>Part-of</Category>
                <InferredCategory>Part-of</
                    InferredCategory>
                <InferredBy>
                    <UsedHeuristic>
                        <HeuristicName>
                            Structural-default</
                            HeuristicName>
                    </UsedHeuristic>
                </InferredBy>
                <Confidence>0.77</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A35</RelationID>
            <Target>T10</Target>
            <Source>T9</Source>
            <Category>
                <Category>Next</Category>
                <InferredCategory>Part-of</
                    InferredCategory>
                <InferredBy>
                    <UsedHeuristic>
                        <HeuristicName>
                            Sequential-default</
                            HeuristicName>
```

```
                    </UsedHeuristic>
                </InferredBy>
                <Confidence>0.76</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A34</RelationID>
            <Target>T2</Target>
            <Source>T23</Source>
            <Category>
                <Category>Is-A</Category>
                <InferredCategory>Is-A</
                  InferredCategory>
                <InferredBy>
                    <UsedHeuristic>
                        <HeuristicName>AH</
                          HeuristicName>
                    </UsedHeuristic>
                </InferredBy>
                <Confidence>1</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A43</RelationID>
            <Target>T7</Target>
            <Source>T71</Source>
            <Category>
                <Category>Part-of</Category>
                <InferredCategory>Part-of</
                  InferredCategory>
                <InferredBy>
                    <UsedHeuristic>
                        <HeuristicName>
                          Structural-default</
                          HeuristicName>
                    </UsedHeuristic>
                </InferredBy>
                <Confidence>0.77</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A6</RelationID>
            <Target>T8</Target>
            <Source>T81</Source>
            <Category>
                <Category>Is-A</Category>
```

```
                <InferredCategory>Is-A</
                   InferredCategory>
                <InferredBy>
                        <UsedHeuristic>
                                <HeuristicName>MWH</
                                   HeuristicName>
                        </UsedHeuristic>
                </InferredBy>
                <Confidence>0.92</Confidence>
        </Category>
</Relation>
<Relation>
        <RelationID>IS-A8</RelationID>
        <Target>T22</Target>
        <Source>T21</Source>
        <Category>
                <Category>Next</Category>
                <InferredCategory>Next</
                   InferredCategory>
                <InferredBy>
                        <UsedHeuristic>
                                <HeuristicName>
                                   Sequential-default</
                                   HeuristicName>
                        </UsedHeuristic>
                </InferredBy>
                <Confidence>0.76</Confidence>
        </Category>
</Relation>
<Relation>
        <RelationID>IS-A23</RelationID>
        <Target>T102</Target>
        <Source>T101</Source>
        <Category>
                <Category>Next</Category>
                <InferredCategory>Next</
                   InferredCategory>
                <InferredBy>
                        <UsedHeuristic>
                                <HeuristicName>
                                   Sequential-default</
                                   HeuristicName>
                        </UsedHeuristic>
                </InferredBy>
                <Confidence>0.76</Confidence>
        </Category>
```

```
</Relation>
<Relation>
    <RelationID>IS-A26</RelationID>
    <Target>T0</Target>
    <Source>T10</Source>
    <Category>
        <Category>Part-of</Category>
        <InferredCategory>Part-of</
            InferredCategory>
        <InferredBy>
            <UsedHeuristic>
                <HeuristicName>
                    Structural-default</
                    HeuristicName>
            </UsedHeuristic>
        </InferredBy>
        <Confidence>0.77</Confidence>
    </Category>
</Relation>
<Relation>
    <RelationID>IS-A12</RelationID>
    <Target>T24</Target>
    <Source>T23</Source>
    <Category>
        <Category>Next</Category>
        <InferredCategory>Next</
            InferredCategory>
        <InferredBy>
            <UsedHeuristic>
                <HeuristicName>
                    Sequential-default</
                    HeuristicName>
            </UsedHeuristic>
        </InferredBy>
        <Confidence>0.76</Confidence>
    </Category>
</Relation>
<Relation>
    <RelationID>IS-A5</RelationID>
    <Target>T10</Target>
    <Source>T101</Source>
    <Category>
        <Category>Part-of</Category>
        <InferredCategory>Part-of</
            InferredCategory>
        <InferredBy>
```

```
                        <UsedHeuristic>
                            <HeuristicName>
                              Structural-default</
                              HeuristicName>
                        </UsedHeuristic>
                    </InferredBy>
                    <Confidence>0.77</Confidence>
                </Category>
            </Relation>
            <Relation>
                <RelationID>IS-A37</RelationID>
                <Target>T6</Target>
                <Source>T61</Source>
                <Category>
                    <Category>Part-of</Category>
                    <InferredCategory>Part-of</
                      InferredCategory>
                    <InferredBy>
                        <UsedHeuristic>
                            <HeuristicName>
                              Structural-default</
                              HeuristicName>
                        </UsedHeuristic>
                    </InferredBy>
                    <Confidence>0.77</Confidence>
                </Category>
            </Relation>
            <Relation>
                <RelationID>IS-A19</RelationID>
                <Target>T2</Target>
                <Source>T5</Source>
                <Category>
                    <Category>Prerequisite</Category>
                    <InferredCategory>Prerequisite</
                      InferredCategory>
                    <InferredBy>
                        <UsedHeuristic>
                            <HeuristicName>RH</
                              HeuristicName>
                        </UsedHeuristic>
                    </InferredBy>
                    <Confidence>0.93</Confidence>
                </Category>
            </Relation>
            <Relation>
                <RelationID>IS-A27</RelationID>
```

```
            <Target>T52</Target>
            <Source>T51</Source>
            <Category>
                <Category>Next</Category>
                <InferredCategory>Next</
                    InferredCategory>
                <InferredBy>
                    <UsedHeuristic>
                        <HeuristicName>
                            Sequential-default</
                            HeuristicName>
                    </UsedHeuristic>
                </InferredBy>
                <Confidence>0.76</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A9</RelationID>
            <Target>T0</Target>
            <Source>T6</Source>
            <Category>
                <Category>Part-of</Category>
                <InferredCategory>Part-of</
                    InferredCategory>
                <InferredBy>
                    <UsedHeuristic>
                        <HeuristicName>
                            Structural-default</
                            HeuristicName>
                    </UsedHeuristic>
                </InferredBy>
                <Confidence>0.77</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A17</RelationID>
            <Target>T2</Target>
            <Source>T22</Source>
            <Category>
                <Category>Is-A</Category>
                <InferredCategory>Is-A</
                    InferredCategory>
                <InferredBy>
                    <UsedHeuristic>
                        <HeuristicName>AH</
                            HeuristicName>
```

```
                </UsedHeuristic>
            </InferredBy>
            <Confidence>1</Confidence>
        </Category>
    </Relation>
    <Relation>
        <RelationID>IS-A11</RelationID>
        <Target>T5</Target>
        <Source>T51</Source>
        <Category>
            <Category>Is-A</Category>
            <InferredCategory>Is-A</
               InferredCategory>
            <InferredBy>
                <UsedHeuristic>
                    <HeuristicName>AH</
                       HeuristicName>
                </UsedHeuristic>
            </InferredBy>
            <Confidence>1</Confidence>
        </Category>
    </Relation>
    <Relation>
        <RelationID>IS-A39</RelationID>
        <Target>T0</Target>
        <Source>T3</Source>
        <Category>
            <Category>Part-of</Category>
            <InferredCategory>Part-of</
               InferredCategory>
            <InferredBy>
                <UsedHeuristic>
                    <HeuristicName>
                       Structural-default</
                       HeuristicName>
                </UsedHeuristic>
            </InferredBy>
            <Confidence>0.77</Confidence>
        </Category>
    </Relation>
    <Relation>
        <RelationID>IS-A15</RelationID>
        <Target>T7</Target>
        <Source>T6</Source>
        <Category>
            <Category>Next</Category>
```

```
                <InferredCategory>Next</
                  InferredCategory>
                <InferredBy>
                    <UsedHeuristic>
                        <HeuristicName>
                          Sequential-default</
                          HeuristicName>
                    </UsedHeuristic>
                </InferredBy>
                <Confidence>0.76</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A40</RelationID>
            <Target>T0</Target>
            <Source>T7</Source>
            <Category>
                <Category>Part-of</Category>
                <InferredCategory>Part-of</
                  InferredCategory>
                <InferredBy>
                    <UsedHeuristic>
                        <HeuristicName>
                          Structural-default</
                          HeuristicName>
                    </UsedHeuristic>
                </InferredBy>
                <Confidence>0.77</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A7</RelationID>
            <Target>T8</Target>
            <Source>T7</Source>
            <Category>
                <Category>Next</Category>
                <InferredCategory>Next</
                  InferredCategory>
                <InferredBy>
                    <UsedHeuristic>
                        <HeuristicName>
                          Sequential-default</
                          HeuristicName>
                    </UsedHeuristic>
                </InferredBy>
                <Confidence>0.76</Confidence>
```

```
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A36</RelationID>
            <Target>T2</Target>
            <Source>T21</Source>
            <Category>
                <Category>Is-A</Category>
                <InferredCategory>Is-A</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
                    <HeuristicName>AH</
                        HeuristicName>
                </UsedHeuristic>
                </InferredBy>
                <Confidence>1</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A33</RelationID>
            <Target>T0</Target>
            <Source>T11</Source>
            <Category>
                <Category>Part-of</Category>
                <InferredCategory>Part-of</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
                    <HeuristicName>Structural-
                        default</HeuristicName>
                </UsedHeuristic>
                </InferredBy>
                <Confidence>0.77</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A31</RelationID>
            <Target>T23</Target>
            <Source>T22</Source>
            <Category>
                <Category>Next</Category>
                <InferredCategory>Next</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
                    <HeuristicName>Sequential-
                        default</HeuristicName>
                </UsedHeuristic>
```

```
                </InferredBy>
                <Confidence>0.76</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A24</RelationID>
            <Target>T5</Target>
            <Source>T4</Source>
            <Category>
                <Category>Next</Category>
                <InferredCategory>Next</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
                        <HeuristicName>Sequential-
                            default</HeuristicName>
                </UsedHeuristic>
                </InferredBy>
                <Confidence>0.76</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A13</RelationID>
            <Target>T11</Target>
            <Source>T111</Source>
            <Category>
                <Category>Part-of</Category>
                <InferredCategory>Part-of</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
                        <HeuristicName>Structural-
                            default</HeuristicName>
                </UsedHeuristic>
                </InferredBy>
                <Confidence>0.77</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A14</RelationID>
            <Target>T10</Target>
            <Source>T2</Source>
            <Category>
                <Category>Prerequisite</Category>
                <InferredCategory>Prerequisite</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
```

```
                        <HeuristicName>PGH2</
                            HeuristicName>
                </UsedHeuristic>
                </InferredBy>
                <Confidence>0.96</Confidence>
        </Category>
</Relation>
<Relation>
        <RelationID>IS-A28</RelationID>
        <Target>T10</Target>
        <Source>T102</Source>
        <Category>
                <Category>Part-of</Category>
                <InferredCategory>Part-of</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
                        <HeuristicName>Structural-
                            default</HeuristicName>
                </UsedHeuristic>
                </InferredBy>
                <Confidence>0.77</Confidence>
        </Category>
</Relation>
<Relation>
        <RelationID>IS-A32</RelationID>
        <Target>T5</Target>
        <Source>T52</Source>
        <Category>
                <Category>Is-A</Category>
                <InferredCategory>Is-A</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
                        <HeuristicName>EH</
                            HeuristicName>
                </UsedHeuristic>
                </InferredBy>
                <Confidence>1</Confidence>
        </Category>
</Relation>
<Relation>
        <RelationID>IS-A2</RelationID>
        <Target>T2</Target>
        <Source>T1</Source>
        <Category>
                <Category>Next</Category>
```

```
                <InferredCategory>Next</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
                        <HeuristicName>Sequential−
                            default</HeuristicName>
                </UsedHeuristic>
                </InferredBy>
                <Confidence>0.76</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS−A3</RelationID>
            <Target>T9</Target>
            <Source>T8</Source>
            <Category>
                <Category>Next</Category>
                <InferredCategory>Next</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
                        <HeuristicName>Sequential−
                            default</HeuristicName>
                </UsedHeuristic>
                </InferredBy>
                <Confidence>0.76</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS−A20</RelationID>
            <Target>T2</Target>
            <Source>T24</Source>
            <Category>
                <Category>Is−A</Category>
                <InferredCategory>Is−A</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
                        <HeuristicName>AH</
                            HeuristicName>
                </UsedHeuristic>
                </InferredBy>
                <Confidence>1</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS−A1</RelationID>
            <Target>T2</Target>
            <Source>T25</Source>
```

```xml
            <Category>
                  <Category>Is-A</Category>
                  <InferredCategory>Is-A</
                     InferredCategory>
                  <InferredBy><UsedHeuristic>
                        <HeuristicName>AH</
                           HeuristicName>
                  </UsedHeuristic>
                  </InferredBy>
                  <Confidence>1</Confidence>
            </Category>
      </Relation>
      <Relation>
            <RelationID>IS-A21</RelationID>
            <Target>T6</Target>
            <Source>T5</Source>
            <Category>
                  <Category>Next</Category>
                  <InferredCategory>Next</
                     InferredCategory>
                  <InferredBy><UsedHeuristic>
                        <HeuristicName>Sequential-
                           default</HeuristicName>
                  </UsedHeuristic>
                  </InferredBy>
                  <Confidence>0.76</Confidence>
            </Category>
      </Relation>
      <Relation>
            <RelationID>IS-A29</RelationID>
            <Target>T2</Target>
            <Source>T4</Source>
            <Category>
                  <Category>Prerequisite</Category>
                  <InferredCategory>Prerequisite</
                     InferredCategory>
                  <InferredBy><UsedHeuristic>
                        <HeuristicName>RH</
                           HeuristicName>
                  </UsedHeuristic>
                  </InferredBy>
                  <Confidence>0.93</Confidence>
            </Category>
      </Relation>
      <Relation>
            <RelationID>IS-A30</RelationID>
```

```
            <Target>T0</Target>
            <Source>T9</Source>
            <Category>
                <Category>Part-of</Category>
                <InferredCategory>Part-of</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
                    <HeuristicName>Structural-
                        default</HeuristicName>
                </UsedHeuristic>
                </InferredBy>
                <Confidence>0.77</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A10</RelationID>
            <Target>T2</Target>
            <Source>T101</Source>
            <Category>
                <Category>Prerequisite</Category>
                <InferredCategory>Prerequisite</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
                    <HeuristicName>PGH2</
                        HeuristicName>
                </UsedHeuristic>
                </InferredBy>
                <Confidence>0.96</Confidence>
            </Category>
        </Relation>
        <Relation>
            <RelationID>IS-A42</RelationID>
            <Target>T2</Target>
            <Source>T3</Source>
            <Category>
                <Category>Prerequisite</Category>
                <InferredCategory>Prerequisite</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
                    <HeuristicName>RH</
                        HeuristicName>
                </UsedHeuristic>
                </InferredBy>
                <Confidence>0.93</Confidence>
            </Category>
        </Relation>
```

```
<Relation>
        <RelationID>IS-A38</RelationID>
        <Target>T0</Target>
        <Source>T8</Source>
        <Category>
                <Category>Part-of</Category>
                <InferredCategory>Part-of</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
                        <HeuristicName>Structural-
                            default</HeuristicName>
                </UsedHeuristic>
                </InferredBy>
                <Confidence>0.77</Confidence>
        </Category>
</Relation>
<Relation>
        <RelationID>IS-A4</RelationID>
        <Target>T25</Target>
        <Source>T24</Source>
        <Category>
                <Category>Next</Category>
                <InferredCategory>Next</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
                        <HeuristicName>Sequential-
                            default</HeuristicName>
                </UsedHeuristic>
                </InferredBy>
                <Confidence>0.76</Confidence>
        </Category>
</Relation>
<Relation>
        <RelationID>IS-A18</RelationID>
        <Target>T0</Target>
        <Source>T1</Source>
        <Category>
                <Category>Part-of</Category>
                <InferredCategory>Part-of</
                    InferredCategory>
                <InferredBy><UsedHeuristic>
                        <HeuristicName>Structural-
                            default</HeuristicName>
                </UsedHeuristic>
                </InferredBy>
                <Confidence>0.77</Confidence>
```

```
                </Category>
            </Relation>
            <Relation>
                <RelationID>IS−A25</RelationID>
                <Target>T0</Target>
                <Source>T2</Source>
                <Category>
                    <Category>Part−of</Category>
                    <InferredCategory>Part−of</
                        InferredCategory>
                    <InferredBy><UsedHeuristic>
                        <HeuristicName>Structural−
                            default</HeuristicName>
                    </UsedHeuristic>
                    </InferredBy>
                    <Confidence>0.77</Confidence>
                </Category>
            </Relation>
            <Relation>
                <RelationID>IS−A22</RelationID>
                <Target>T10</Target>
                <Source>T11</Source>
                <Category>
                    <Category>Prerequisite</Category>
                    <InferredCategory>Prerequisite</
                        InferredCategory>
                    <InferredBy><UsedHeuristic>
                        <HeuristicName>RH</
                            HeuristicName>
                    </UsedHeuristic>
                    </InferredBy>
                    <Confidence>0.93</Confidence>
                </Category>
            </Relation>
        </RelationSet>
</InferredOntology>
```

C.3.2 OWL-based Representation of the Learning Domain Ontology

Finally, the LDO is represented in OWL, which might be understood by a wider set of systems, using the formalism presented in Listing C.6.

Listing C.6 – OWL Ontology for Describing the Learning Domain Ontology

```
<?xml version="1.0"?>

<!DOCTYPE rdf:RDF [
    <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
    <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
    <!ENTITY owl2xml "http://www.w3.org/2006/12/owl2-xml#" >
    <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
    <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
    <!ENTITY LearningDomain "http://lsi.vc.ehu.es/2008/10/
        LearningDomain.owl#" >
]>

<rdf:RDF xmlns="http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#"
     xml:base="http://lsi.vc.ehu.es/2008/10/LearningDomain.owl"
     xmlns:LearningDomain="http://lsi.vc.ehu.es/2008/10/LearningDomain.
         owl#"
     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
     xmlns:owl2xml="http://www.w3.org/2006/12/owl2-xml#"
     xmlns:owl="http://www.w3.org/2002/07/owl#"
     xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
    <owl:Ontology rdf:about="">
        <rdfs:comment rdf:datatype="&xsd;string"
            >Base Ontology for describing the Domain Module for
                Technology Supported Learning Systems</rdfs:comment>
    </owl:Ontology>

    <!--
    ///////////////////
    // Object Properties
    ///////////////////
     -->

    <!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#hasDR -->
    <owl:ObjectProperty rdf:about="#hasDR">
        <rdfs:range rdf:resource="#DidacticResource"/>
        <rdfs:domain rdf:resource="#DomainTopic"/>
```

```
</owl:ObjectProperty>

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#
    hasDifficultyLevel -->
<owl:ObjectProperty rdf:about="#hasDifficultyLevel"/>

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#hasRelevance
    -->
<owl:ObjectProperty rdf:about="#hasRelevance"/>

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#isA -->
<owl:ObjectProperty rdf:about="#isA"/>

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#next -->
<owl:ObjectProperty rdf:about="#next"/>

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#partOf -->
<owl:ObjectProperty rdf:about="#partOf"/>

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#prerequisite
    -->
<owl:ObjectProperty rdf:about="#prerequisite"/>

<!--
/////////////////
// Data properties
/////////////////
 -->

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#drID -->
<owl:DatatypeProperty rdf:about="#drID">
    <rdfs:domain rdf:resource="#DidacticResource"/>
    <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#drKind -->
<owl:DatatypeProperty rdf:about="#drKind">
    <rdfs:domain rdf:resource="#DidacticResource"/>
    <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#drPath -->
<owl:DatatypeProperty rdf:about="#drPath">
    <rdfs:domain rdf:resource="#DidacticResource"/>
    <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>
```

```
<!--
///////////
// Classes
///////////
 -->
<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#Concept -->
<owl:Class rdf:about="#Concept">
    <rdfs:subClassOf rdf:resource="#DomainTopic"/>
</owl:Class>

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#
    DidacticResource -->
<owl:Class rdf:about="#DidacticResource"/>

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#
    DificultyRelevanceLevel -->
<owl:Class rdf:about="#DificultyRelevanceLevel">
    <owl:equivalentClass>
        <owl:Class>
            <owl:oneOf rdf:parseType="Collection">
                <rdf:Description rdf:about="#MEDIUM"/>
                <rdf:Description rdf:about="#HIGH"/>
                <rdf:Description rdf:about="#LOW"/>
            </owl:oneOf>
        </owl:Class>
    </owl:equivalentClass>
</owl:Class>

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#DomainTopic --
    >
<owl:Class rdf:about="#DomainTopic">
    <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasDifficultyLevel"/>
            <owl:onClass rdf:resource="#DomainTopic"/>
            <owl:maxQualifiedCardinality rdf:datatype="&xsd;
                nonNegativeInteger">1</owl:maxQualifiedCardinality>
        </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#hasRelevance"/>
            <owl:onClass rdf:resource="#DomainTopic"/>
```

```
                <owl:maxQualifiedCardinality rdf:datatype="&xsd;
                    nonNegativeInteger">1</owl:maxQualifiedCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
    </owl:Class>

    <!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#Fact -->
    <owl:Class rdf:about="#Fact">
        <rdfs:subClassOf rdf:resource="#DomainTopic"/>
    </owl:Class>

    <!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#Principle -->
    <owl:Class rdf:about="#Principle">
        <rdfs:subClassOf rdf:resource="#DomainTopic"/>
    </owl:Class>

    <!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#Procedure -->
    <owl:Class rdf:about="#Procedure">
        <rdfs:subClassOf rdf:resource="#DomainTopic"/>
    </owl:Class>

    <!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#Process -->
    <owl:Class rdf:about="#Process">
        <rdfs:subClassOf rdf:resource="#DomainTopic"/>
    </owl:Class>

    <!-- http://www.w3.org/2002/07/owl#Thing -->
    <owl:Class rdf:about="&owl;Thing"/>

    <!--
    //////////////
    // Individuals
    //////////////
     -->
    <!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#HIGH -->
    <lD:DidacticResource rdf:about="#HIGH"/>

    <!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#LOW -->
    <lD:DidacticResource rdf:about="#LOW"/>

    <!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#MEDIUM -->
    <lD:DidacticResource rdf:about="#MEDIUM"/>
</rdf:RDF>
```

Listing C.7 shows an example of an LDO Represented in OWL.

Listing C.7 – Example of an Learning Domain Ontology Represented in OWL

```xml
<?xml version="1.0"?>


<!DOCTYPE rdf:RDF [
    <!ENTITY Ont "http://lsi.vc.ehu.es/Ont#" >
    <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
    <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
    <!ENTITY owl2xml "http://www.w3.org/2006/12/owl2-xml#" >
    <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
    <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
    <!ENTITY LearningDomain "http://lsi.vc.ehu.es/2008/10/
        LearningDomain.owl#" >
]>

<rdf:RDF xmlns="http://lsi.vc.ehu.es/Ont#"
     xml:base="http://lsi.vc.ehu.es/Ont"
     xmlns:LearningDomain="http://lsi.vc.ehu.es/2008/10/LearningDomain.
         owl#"
     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
     xmlns:owl2xml="http://www.w3.org/2006/12/owl2-xml#"
     xmlns:owl="http://www.w3.org/2002/07/owl#"
     xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:Ont="http://lsi.vc.ehu.es/Ont#">
    <owl:Ontology rdf:about="">
        <rdfs:comment xml:lang="eu">Proba Ontologia</rdfs:comment>
        <owl:imports rdf:resource="http://lsi.vc.ehu.es/2008/10/
            LearningDomain.owl"/>
    </owl:Ontology>

    <!--
    ///////////////////
    // Object Properties
    ///////////////////
     -->

    <!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#
        hasDifficultyLevel -->
    <owl:ObjectProperty rdf:about="&LearningDomain;hasDifficultyLevel"/
        >

    <!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#hasRelevance
        -->
    <owl:ObjectProperty rdf:about="&LearningDomain;hasRelevance"/>
```

```
<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#isA -->
<owl:ObjectProperty rdf:about="&LearningDomain;isA"/>

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#next -->
<owl:ObjectProperty rdf:about="&LearningDomain;next"/>

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#partOf -->
<owl:ObjectProperty rdf:about="&LearningDomain;partOf"/>

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#prerequisite
   -->
<owl:ObjectProperty rdf:about="&LearningDomain;prerequisite"/>

<!--
/////////////
// Classes
/////////////
 -->
<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#Concept -->
<owl:Class rdf:about="&LearningDomain;Concept"/>

<!-- http://www.w3.org/2002/07/owl#Thing -->
<owl:Class rdf:about="&owl;Thing"/>

<!--
//////////////
// Individuals
//////////////
 -->

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#LOW -->
<lD:DidacticResource rdf:about="&LearningDomain;LOW"/>

<!-- http://lsi.vc.ehu.es/2008/10/LearningDomain.owl#MEDIUM -->
<lD:DidacticResource rdf:about="&LearningDomain;MEDIUM"/>

<!-- http://lsi.vc.ehu.es/Ont#T0 -->
<LearningDomain:Concept rdf:about="#T0">
    <rdfs:label xml:lang="eu">Canis Major</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
</LearningDomain:Concept>
```

```
<!-- http://lsi.vc.ehu.es/Ont#T1 -->
<LearningDomain:Concept rdf:about="#T1">
    <rdfs:label xml:lang="eu"
        >Canis Major konstelazio</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <owl:sameAs rdf:resource="#T0"/>
    <LearningDomain:isA rdf:resource="#T42"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T10 -->
<LearningDomain:Concept rdf:about="#T10">
    <rdfs:label xml:lang="eu"
        >Taurus konstelazio</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:isA rdf:resource="#T42"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T11 -->
<LearningDomain:Concept rdf:about="#T11">
    <rdfs:label xml:lang="eu"
        >ilargiaren translazio mugimendua</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <owl:sameAs rdf:resource="#T17"/>
    <LearningDomain:isA rdf:resource="#T52"/>
    <LearningDomain:isA rdf:resource="#T87"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T12 -->
<LearningDomain:Concept rdf:about="#T12">
    <rdf:type rdf:resource="&LearningDomain;Concept"/>
    <rdfs:label xml:lang="eu"
        >behatoki astronomiko</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T83"/>
```

```
    <LearningDomain:next rdf:resource="#T84"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T13 -->
<LearningDomain:Concept rdf:about="#T13">
    <rdfs:label xml:lang="eu">eguzki-eklipse</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:next rdf:resource="#T28"/>
    <LearningDomain:isA rdf:resource="#T46"/>
    <LearningDomain:prerequisite rdf:resource="#T55"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T14 -->
<LearningDomain:Concept rdf:about="#T14">
    <rdfs:label xml:lang="eu"
        >eguzki-eklipse oso</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:isA rdf:resource="#T13"/>
    <LearningDomain:next rdf:resource="#T89"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T15 -->
<LearningDomain:Concept rdf:about="#T15">
    <rdfs:label xml:lang="eu">eguzki-sistema</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T24"/>
    <LearningDomain:partOf rdf:resource="#T81"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T16 -->
<LearningDomain:Concept rdf:about="#T16">
    <rdfs:label xml:lang="eu"
        >eguzki-sistemako astroen mugimendu</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
```

```
        <LearningDomain:prerequisite rdf:resource="#T15"/>
        <LearningDomain:partOf rdf:resource="#T80"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T17 -->
    <LearningDomain:Concept rdf:about="#T17">
        <rdfs:label xml:lang="eu"
            >ilargiaren translazio</rdfs:label>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T18 -->
    <LearningDomain:Concept rdf:about="#T18">
        <rdfs:label xml:lang="eu"
            >eredu geozentriko</rdfs:label>
        <LearningDomain:hasDifficultyLevel rdf:resource="&
            LearningDomain;LOW"/>
        <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
            MEDIUM"/>
        <LearningDomain:next rdf:resource="#T49"/>
        <LearningDomain:isA rdf:resource="#T79"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T19 -->
    <LearningDomain:Concept rdf:about="#T19">
        <rdfs:label xml:lang="eu"
            >errefrakzio teleskopio</rdfs:label>
        <LearningDomain:hasDifficultyLevel rdf:resource="&
            LearningDomain;LOW"/>
        <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
            MEDIUM"/>
        <LearningDomain:isA rdf:resource="#T90"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T20 -->
    <LearningDomain:Concept rdf:about="#T20">
        <rdfs:label xml:lang="eu">teleskopio</rdfs:label>
        <LearningDomain:hasDifficultyLevel rdf:resource="&
            LearningDomain;LOW"/>
        <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
            MEDIUM"/>
        <LearningDomain:isA rdf:resource="#T53"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T21 -->
    <LearningDomain:Concept rdf:about="#T21">
        <rdfs:label xml:lang="eu"
```

```
            >errotazio-periodo</rdfs:label>
        <LearningDomain:hasDifficultyLevel rdf:resource="&
            LearningDomain;LOW"/>
        <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
            MEDIUM"/>
        <LearningDomain:partOf rdf:resource="#T23"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T22 -->
    <LearningDomain:Concept rdf:about="#T22">
        <rdfs:label xml:lang="eu"
            >errotazio mugimenduaren ondorio</rdfs:label>
        <LearningDomain:hasDifficultyLevel rdf:resource="&
            LearningDomain;LOW"/>
        <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
            MEDIUM"/>
        <LearningDomain:prerequisite rdf:resource="#T37"/>
        <LearningDomain:partOf rdf:resource="#T85"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T23 -->
    <LearningDomain:Concept rdf:about="#T23">
        <rdfs:label xml:lang="eu"
            >errotazio mugimendu</rdfs:label>
        <LearningDomain:hasDifficultyLevel rdf:resource="&
            LearningDomain;LOW"/>
        <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
            MEDIUM"/>
        <LearningDomain:isA rdf:resource="#T16"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T24 -->
    <LearningDomain:Concept rdf:about="#T24">
        <rdfs:label xml:lang="eu">esne-bide</rdfs:label>
        <LearningDomain:hasDifficultyLevel rdf:resource="&
            LearningDomain;LOW"/>
        <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
            MEDIUM"/>
        <LearningDomain:isA rdf:resource="#T41"/>
        <LearningDomain:isA rdf:resource="#T60"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T25 -->
    <LearningDomain:Concept rdf:about="#T25">
        <rdfs:label xml:lang="eu">espazio-ontzi</rdfs:label>
```

```
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;LOW"
        />
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:partOf rdf:resource="#T84"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T26 -->
<LearningDomain:Concept rdf:about="#T26">
    <rdfs:label xml:lang="eu">hartz nagusi</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:isA rdf:resource="#T42"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T27 -->
<LearningDomain:Concept rdf:about="#T27">
    <rdfs:label xml:lang="eu"
        >ilargiaren translazioaren ondorio</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;MEDIUM"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T11"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T28 -->
<LearningDomain:Concept rdf:about="#T28">
    <rdfs:label xml:lang="eu">ilargi-eklipse</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:isA rdf:resource="#T46"/>
    <LearningDomain:prerequisite rdf:resource="#T86"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T29 -->
<LearningDomain:Concept rdf:about="#T29">
    <rdfs:label xml:lang="eu"
        >Kontaezin ahala galaxia</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;MEDIUM"/>
```

```
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:isA rdf:resource="#T79"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T3 -->
<LearningDomain:Concept rdf:about="#T3">
    <rdfs:label xml:lang="eu"
        >lurraren translazio mugimendu</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <owl:sameAs rdf:resource="#T4"/>
    <LearningDomain:isA rdf:resource="#T52"/>
    <LearningDomain:isA rdf:resource="#T85"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T30 -->
<LearningDomain:Concept rdf:about="#T30">
    <rdfs:label xml:lang="eu">ilargiaren aldi</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;LOW"
        />
    <LearningDomain:partOf rdf:resource="#T16"/>
    <LearningDomain:isA rdf:resource="#T27"/>
    <LearningDomain:next rdf:resource="#T88"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T31 -->
<LearningDomain:Concept rdf:about="#T31">
    <rdfs:label xml:lang="eu">ipar izar</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:isA rdf:resource="#T43"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T32 -->
<LearningDomain:Concept rdf:about="#T32">
    <rdfs:label xml:lang="eu"
        >irrati-teleskopio</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
```

```
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T12"/>
    <LearningDomain:isA rdf:resource="#T20"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T33 -->
<LearningDomain:Concept rdf:about="#T33">
    <rdfs:label xml:lang="eu"
        >islapen-teleskopio</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:isA rdf:resource="#T90"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T34 -->
<LearningDomain:Concept rdf:about="#T34">
    <rdfs:label xml:lang="eu"
        >itsasoaren maila</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T88"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T35 -->
<LearningDomain:Concept rdf:about="#T35">
    <rdfs:label xml:lang="eu"
        >lurraren translazio periodo</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;LOW"
        />
    <LearningDomain:partOf rdf:resource="#T3"/>
    <LearningDomain:isA rdf:resource="#T51"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T36 -->
<LearningDomain:Concept rdf:about="#T36">
    <rdfs:label xml:lang="eu"
        >lurraren errotazio</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
```

```
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <owl:sameAs rdf:resource="#T37"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T37 -->
<LearningDomain:Concept rdf:about="#T37">
    <rdfs:label xml:lang="eu"
        >lurraren errotazio mugimendu</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:isA rdf:resource="#T23"/>
    <LearningDomain:isA rdf:resource="#T85"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T38 -->
<LearningDomain:Concept rdf:about="#T38">
    <rdfs:label xml:lang="eu">marea bizi</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:next rdf:resource="#T39"/>
    <LearningDomain:isA rdf:resource="#T88"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T39 -->
<LearningDomain:Concept rdf:about="#T39">
    <rdfs:label xml:lang="eu">marea hil</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:isA rdf:resource="#T88"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T4 -->
<LearningDomain:Concept rdf:about="#T4">
    <rdfs:label xml:lang="eu"
        >lurraren transalazio</rdfs:label>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T40 -->
<LearningDomain:Concept rdf:about="#T40">
```

```
    <rdfs:label xml:lang="eu">unibertso</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T81"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T41 -->
<LearningDomain:Concept rdf:about="#T41">
    <rdfs:label xml:lang="eu">galaxia</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T40"/>
    <LearningDomain:partOf rdf:resource="#T81"/>
    <LearningDomain:next rdf:resource="#T82"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T42 -->
<LearningDomain:Concept rdf:about="#T42">
    <rdfs:label xml:lang="eu">konstelazio</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T40"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T43 -->
<LearningDomain:Concept rdf:about="#T43">
    <rdfs:label xml:lang="eu">izar</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T40"/>
    <LearningDomain:partOf rdf:resource="#T41"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T44 -->
<LearningDomain:Concept rdf:about="#T44">
    <rdfs:label xml:lang="eu">planeta</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
```

```
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T40"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T45 -->
<LearningDomain:Concept rdf:about="#T45">
    <rdfs:label xml:lang="eu">satelite</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T40"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T46 -->
<LearningDomain:Concept rdf:about="#T46">
    <rdfs:label xml:lang="eu">eklipse</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T16"/>
    <LearningDomain:isA rdf:resource="#T27"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T47 -->
<LearningDomain:Concept rdf:about="#T47">
    <rdfs:label xml:lang="eu"
         >metodo zientifiko</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T48 -->
<LearningDomain:Concept rdf:about="#T48">
    <rdfs:label xml:lang="eu">ordu-sektore</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;LOW"
        />
    <LearningDomain:isA rdf:resource="#T22"/>
</LearningDomain:Concept>
```

```
<!-- http://lsi.vc.ehu.es/Ont#T49 -->
<LearningDomain:Concept rdf:about="#T49">
    <rdfs:label xml:lang="eu"
        >eredu heliozentriko</rdfs:label>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
       MEDIUM"/>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
       LearningDomain;MEDIUM"/>
    <LearningDomain:next rdf:resource="#T29"/>
    <LearningDomain:isA rdf:resource="#T79"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T5 -->
<LearningDomain:Concept rdf:about="#T5">
    <rdfs:label xml:lang="eu">egun</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
       LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;LOW"
       />
    <LearningDomain:isA rdf:resource="#T22"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T51 -->
<LearningDomain:Concept rdf:about="#T51">
    <rdfs:label xml:lang="eu"
       >translazio-periodo</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
       LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
       MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T52"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T52 -->
<LearningDomain:Concept rdf:about="#T52">
    <rdfs:label xml:lang="eu"
       >translazio mugimendu</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
       LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
       MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T16"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T53 -->
<LearningDomain:Concept rdf:about="#T53">
```

```
    <rdfs:label xml:lang="eu">tresna optiko</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:next rdf:resource="#T12"/>
    <LearningDomain:partOf rdf:resource="#T83"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T54 -->
<LearningDomain:Concept rdf:about="#T54">
    <rdfs:label xml:lang="eu"
        >unitate astronomiko</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:isA rdf:resource="#T64"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T55 -->
<LearningDomain:Concept rdf:about="#T55">
    <rdfs:label xml:lang="eu">Eguzki</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T15"/>
    <LearningDomain:isA rdf:resource="#T43"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T56 -->
<LearningDomain:Concept rdf:about="#T56">
    <rdfs:label xml:lang="eu">Lur</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:isA rdf:resource="#T44"/>
    <LearningDomain:isA rdf:resource="#T76"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T57 -->
<LearningDomain:Concept rdf:about="#T57">
    <rdfs:label xml:lang="eu">Nebulosa</rdfs:label>
    <LearningDomain:partOf rdf:resource="#T40"/>
```

```
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T58 -->
<LearningDomain:Concept rdf:about="#T58">
    <rdfs:label xml:lang="eu">Kometa</rdfs:label>
    <LearningDomain:partOf rdf:resource="#T40"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T59 -->
<LearningDomain:Concept rdf:about="#T59">
    <rdfs:label xml:lang="eu">Asteroide</rdfs:label>
    <LearningDomain:partOf rdf:resource="#T40"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T6 -->
<LearningDomain:Concept rdf:about="#T6">
    <rdfs:label xml:lang="eu"
        >Hubble teleskopio</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;LOW"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T12"/>
    <LearningDomain:isA rdf:resource="#T33"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T60 -->
<LearningDomain:Concept rdf:about="#T60">
    <rdfs:label xml:lang="eu">Kiribil</rdfs:label>
    <LearningDomain:isA rdf:resource="#T41"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T61 -->
<LearningDomain:Concept rdf:about="#T61">
    <rdfs:label xml:lang="eu">Barradun</rdfs:label>
    <LearningDomain:isA rdf:resource="#T41"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T62 -->
<LearningDomain:Concept rdf:about="#T62">
    <rdfs:label xml:lang="eu">Irregular</rdfs:label>
    <LearningDomain:isA rdf:resource="#T41"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T63 -->
<LearningDomain:Concept rdf:about="#T63">
```

```
    <rdfs:label xml:lang="eu">Eliptiko</rdfs:label>
    <LearningDomain:isA rdf:resource="#T41"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T64 -->
<LearningDomain:Concept rdf:about="#T64">
    <rdfs:label xml:lang="eu"
        >Unibertsoko distantzia neurtzeko metodo</rdfs:label>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
       MEDIUM"/>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
       LearningDomain;MEDIUM"/>
    <LearningDomain:prerequisite rdf:resource="#T40"/>
    <LearningDomain:partOf rdf:resource="#T82"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T65 -->
<LearningDomain:Concept rdf:about="#T65">
    <rdfs:label xml:lang="eu">argi-urte</rdfs:label>
    <LearningDomain:isA rdf:resource="#T64"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T66 -->
<LearningDomain:Concept rdf:about="#T66">
    <rdfs:label xml:lang="eu">u.a.</rdfs:label>
    <owl:sameAs rdf:resource="#T54"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T67 -->
<LearningDomain:Concept rdf:about="#T67">
    <rdfs:label xml:lang="eu"
        >Proxima Centauri</rdfs:label>
    <LearningDomain:isA rdf:resource="#T43"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T68 -->
<LearningDomain:Concept rdf:about="#T68">
    <rdfs:label xml:lang="eu">Merkurio</rdfs:label>
    <LearningDomain:isA rdf:resource="#T44"/>
    <LearningDomain:isA rdf:resource="#T76"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T69 -->
<LearningDomain:Concept rdf:about="#T69">
    <rdfs:label xml:lang="eu">Artizar</rdfs:label>
    <LearningDomain:isA rdf:resource="#T44"/>
```

```xml
        <LearningDomain:isA rdf:resource="#T77"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T7 -->
    <LearningDomain:Concept rdf:about="#T7">
        <rdfs:label xml:lang="eu">gau</rdfs:label>
        <LearningDomain:hasDifficultyLevel rdf:resource="&
            LearningDomain;LOW"/>
        <LearningDomain:hasRelevance rdf:resource="&LearningDomain;LOW"
            />
        <LearningDomain:isA rdf:resource="#T22"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T70 -->
    <LearningDomain:Concept rdf:about="#T70">
        <rdfs:label xml:lang="eu">Marte</rdfs:label>
        <LearningDomain:isA rdf:resource="#T44"/>
        <LearningDomain:isA rdf:resource="#T77"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T71 -->
    <LearningDomain:Concept rdf:about="#T71">
        <rdfs:label xml:lang="eu">Jupiter</rdfs:label>
        <LearningDomain:isA rdf:resource="#T44"/>
        <LearningDomain:isA rdf:resource="#T77"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T72 -->
    <LearningDomain:Concept rdf:about="#T72">
        <rdfs:label xml:lang="eu">Saturno</rdfs:label>
        <LearningDomain:isA rdf:resource="#T44"/>
        <LearningDomain:isA rdf:resource="#T77"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T73 -->
    <LearningDomain:Concept rdf:about="#T73">
        <rdfs:label xml:lang="eu">Urano</rdfs:label>
        <LearningDomain:isA rdf:resource="#T44"/>
        <LearningDomain:isA rdf:resource="#T77"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T74 -->
    <LearningDomain:Concept rdf:about="#T74">
        <rdfs:label xml:lang="eu">Neptuno</rdfs:label>
        <LearningDomain:isA rdf:resource="#T44"/>
        <LearningDomain:isA rdf:resource="#T77"/>
```

```
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T75 -->
<LearningDomain:Concept rdf:about="#T75">
    <rdfs:label xml:lang="eu">Pluton</rdfs:label>
    <LearningDomain:isA rdf:resource="#T44"/>
    <LearningDomain:isA rdf:resource="#T77"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T76 -->
<LearningDomain:Concept rdf:about="#T76">
    <rdfs:label xml:lang="eu">Barneko planeta</rdfs:label>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T15"/>
    <LearningDomain:isA rdf:resource="#T44"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T77 -->
<LearningDomain:Concept rdf:about="#T77">
    <rdfs:label xml:lang="eu">Kanpoko planeta</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;MEDIUM"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T15"/>
    <LearningDomain:isA rdf:resource="#T44"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T78 -->
<LearningDomain:Concept rdf:about="#T78">
    <rdfs:label xml:lang="eu">Kosmologi</rdfs:label>
    <LearningDomain:partOf rdf:resource="#T81"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T79 -->
<LearningDomain:Concept rdf:about="#T79">
    <rdfs:label xml:lang="eu"
        >Unibertsori buruzko hainbat ideia</rdfs:label>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;MEDIUM"/>
    <LearningDomain:prerequisite rdf:resource="#T40"/>
```

```
    <LearningDomain:partOf rdf:resource="#T78"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T8 -->
<LearningDomain:Concept rdf:about="#T8">
    <rdfs:label xml:lang="eu"
        >translazio mugimenduaren ondorio</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;MEDIUM"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:prerequisite rdf:resource="#T3"/>
    <LearningDomain:partOf rdf:resource="#T85"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T80 -->
<LearningDomain:Concept rdf:about="#T80">
    <rdfs:label xml:lang="eu">Izarrei begira</rdfs:label>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T81 -->
<LearningDomain:Concept rdf:about="#T81">
    <rdfs:label xml:lang="eu"
        >Unibertsoa eta eguzki-sistema</rdfs:label>
    <LearningDomain:partOf rdf:resource="#T80"/>
    <LearningDomain:next rdf:resource="#T83"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T82 -->
<LearningDomain:Concept rdf:about="#T82">
    <rdfs:label xml:lang="eu"
        >Distantziak Unibertsoan</rdfs:label>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;MEDIUM"/>
    <LearningDomain:prerequisite rdf:resource="#T40"/>
    <LearningDomain:next rdf:resource="#T78"/>
    <LearningDomain:partOf rdf:resource="#T81"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T83 -->
<LearningDomain:Concept rdf:about="#T83">
    <rdfs:label xml:lang="eu"
        >Unibertsoa ezagutzeko teknologia</rdfs:label>
    <LearningDomain:prerequisite rdf:resource="#T40"/>
```

```
    <LearningDomain:partOf rdf:resource="#T80"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T84 -->
<LearningDomain:Concept rdf:about="#T84">
    <rdfs:label xml:lang="eu">Astronautika</rdfs:label>
    <LearningDomain:partOf rdf:resource="#T83"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T85 -->
<LearningDomain:Concept rdf:about="#T85">
    <rdfs:label xml:lang="eu"
        >Lurraren mugimendu</rdfs:label>
    <LearningDomain:partOf rdf:resource="#T16"/>
    <LearningDomain:prerequisite rdf:resource="#T56"/>
    <LearningDomain:next rdf:resource="#T87"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T86 -->
<LearningDomain:Concept rdf:about="#T86">
    <rdfs:label xml:lang="eu">Ilargi</rdfs:label>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;MEDIUM"/>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T15"/>
    <LearningDomain:isA rdf:resource="#T45"/>
    <LearningDomain:prerequisite rdf:resource="#T56"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T87 -->
<LearningDomain:Concept rdf:about="#T87">
    <rdfs:label xml:lang="eu"
        >Ilargiaren mugimendu</rdfs:label>
    <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
        MEDIUM"/>
    <LearningDomain:hasDifficultyLevel rdf:resource="&
        LearningDomain;MEDIUM"/>
    <LearningDomain:partOf rdf:resource="#T16"/>
    <LearningDomain:next rdf:resource="#T30"/>
    <LearningDomain:prerequisite rdf:resource="#T86"/>
</LearningDomain:Concept>

<!-- http://lsi.vc.ehu.es/Ont#T88 -->
<LearningDomain:Concept rdf:about="#T88">
    <rdfs:label xml:lang="eu">Marea</rdfs:label>
```

```
        <LearningDomain:partOf rdf:resource="#T16"/>
        <LearningDomain:isA rdf:resource="#T27"/>
        <LearningDomain:prerequisite rdf:resource="#T37"/>
        <LearningDomain:next rdf:resource="#T46"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T89 -->
    <LearningDomain:Concept rdf:about="#T89">
        <rdfs:label xml:lang="eu"
            >Eguzki-eklipse partzial</rdfs:label>
        <LearningDomain:isA rdf:resource="#T13"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T9 -->
    <LearningDomain:Concept rdf:about="#T9">
        <rdfs:label xml:lang="eu">Sirius izar</rdfs:label>
        <LearningDomain:hasDifficultyLevel rdf:resource="&
            LearningDomain;LOW"/>
        <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
            MEDIUM"/>
        <LearningDomain:isA rdf:resource="#T43"/>
    </LearningDomain:Concept>

    <!-- http://lsi.vc.ehu.es/Ont#T90 -->
    <LearningDomain:Concept rdf:about="#T90">
        <rdfs:label xml:lang="eu"
            >Teleskopio optiko</rdfs:label>
        <LearningDomain:hasDifficultyLevel rdf:resource="&
            LearningDomain;LOW"/>
        <LearningDomain:hasRelevance rdf:resource="&LearningDomain;
            MEDIUM"/>
        <LearningDomain:isA rdf:resource="#T20"/>
        <LearningDomain:isA rdf:resource="#T53"/>
    </LearningDomain:Concept>
</rdf:RDF>
```