

Ikaste-domeinuaren sorkuntza erdiautomatikoa

Mikel Larrañaga, Angel Conde, Iñaki Calvo,
Ana Arruarte, Jon A. Elorriaga*

Lengoaia eta Sistema Informatikoak Saila. Informatika Fakultatea.
Euskal Herriko Unibertsitatea (UPV/EHU)

*mikel.larranaga@ehu.es

Jasoa: 2013-07-24

Onartua: 2013-10-10

Laburpena: Teknologian Oinarritutako Hezkuntzarako Tresnak ezinbesteko bihurtu dira gaur egun. Tresna hauek, beren helburua bete ahal izateko, domeinu-modulua —ikasi beharreko domeinuaren adierazpen pedagogikoa— behar dute. Domeinu-modulua sortzeak oso lan handia eskatzen du, ordea. Artikulu honetan, DOM-Sortze, domeinu-modulua testuliburu elektroniketatik erauzteko helburuarekin garatutako tresna aurkezten da.

Hitz gakoak: Teknologian Oinarritutako Hezkuntzarako Tresnak, domeinu-modulua, DOM-Sortze.

Abstract: Nowadays, Technology Supported Learning Systems have become essential. These tools require the Domain Module, i.e., the pedagogical representation of the learning domain, to be defined in order to be effective. However, building the Domain Module is a time and effort consuming task. This paper presents DOM-Sortze, a tool for the acquisition of the Domain Module from electronic textbooks.

Keywords: Technology Supported Learning Systems, domain module, DOM-Sortze.

1. SARRERA

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 Moodle¹ edo Blackboard² moduko ikasketa kudeatzeko sistemak, ezinbesteko bihurtu dira hainbat hezkuntza erakundetan [1]. Are

¹ <http://moodle.org>

² <http://www.blackboard.com>

gehiago, Teknologian Oinarritutako Hezkuntzarako Tresnen erabilerak ikasleen motibazioan ez ezik ikasketaren emaitzetan ere eragin dezake [2].

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 [3]. Tutore Adimendunek, adibidez, osagai hori erabiltzen dute ikasleen ezagutza neurtzeko eta nola jokatu erabakitzeko. Domeinu-modulua egokia ez bada, ikasketa-prozesu eraginkorra burutzea ezinezkoa izango da [3].

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 liburuen egileek ere arazo horri aurre egin behar izaten diote liburuaren edukia egituratu eta antolatuzeko orduan, alegia, topikoak eta hezkuntzarako baliabide egokiak (definizioak, adibideak, ariketak etab.) identifikatu behar dituzte. Ezingo ote genituzke aurretik probetxugarriak suertatu diren liburuak berrerabili domeinu-modulua automatikoki sortzeko?

Azkeneko urte hauetan hezkuntzarako informatikan berrerabilpena bultzatzeko saiakerak burutu dira, hori ahalbidetzen duten estandarrak sortuz [4] eta hezkuntzarako eduki berrerabilgarriak —Ikaste Objektuak— garatuz. Ikaste Objektuen Biltegiak —ARIADNE [5,6] edo Merlot [7]— eta GLOBE³ 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. *DOM-Sortze* domeinuarekiko independentea izateko asmoz garatu da. 2. atalean, domeinu-modulua garatzeko proposatutako prozesua deskribatzen da. 3. atalean, liburuari eragiten zaion aurreprozesua azaltzen da. 4. atalean Ikaste Domeinuaren Ontologiaren erauzketa aurkezten da. 5. atalean, berriz, ikaste-objektuen erauzketa deskribatzen da. Erabiltzaileek burutu dezaketen prozesuaren gainbegiraketa 6. atalean azaltzen da. Ondoren, *DOM-Sortze* balidatzeko egindako esperimentuak aipatzen dira. Eta bukatzeko, ondorioak eta etorkizuneko lerroak aurkezten dira.

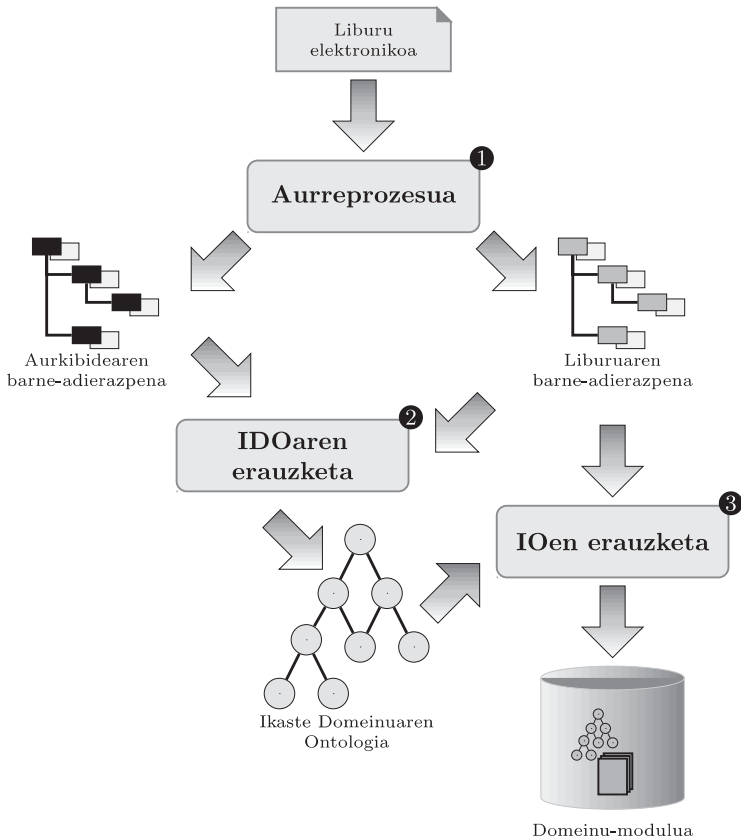
2. DOMEINU-MODULUAREN GARAPENA

Hemen aurkeztutako proposamenean, Adimen Artifizialeko metodoak eta teknikak (adib. Hizkuntzaren Prozesamendua) erabiltzen dituen prozesu erdiautomatikoa baliatu dugu domeinu-modulua garatzeko. Domeinu-moduluak

³ <http://www.globe-info.org/>

bi maila ezberdinetako ezagutza jasotzen du: Ikaste Domeinuaren Ontologia (IDO) eta Ikaste Objektuen (IO) bilduma. Proposamen honetan, domeinu-modulua eraikitzeko hiru urrats egin behar dira (ikus 1. irudia):

1. **Liburuaren aurreprozesua:** Hasteko, aukeratutako liburua informazio-erazketerako egokitu behar da. Bai egokitutako liburua bai liburuaren aurkibidea erabiliko dira jarraian egingo diren erazketa-prozesuetan.
2. **Ikaste Domeinuaren Ontologiaren erazketa:** Urrats honetan, ikasi beharreko topikoak zein beraien arteko erlazio pedagogikoak erazuten dira. Ikaste Domeinuaren Ontologiak Teknologian Oinarritutako Hezkuntzarako Tresnei ikasketa-saioak antolatzea ahalbidetzen die. Horrez gain, ikasleek beren kabuz ikasteko ere erabil dezakete.
3. **Ikaste-objektuen erazketa:** Urrats honetan, ikasteko erabiliko diren edukiak (definizioak, adibideak, ariketak...) erazuten dira.



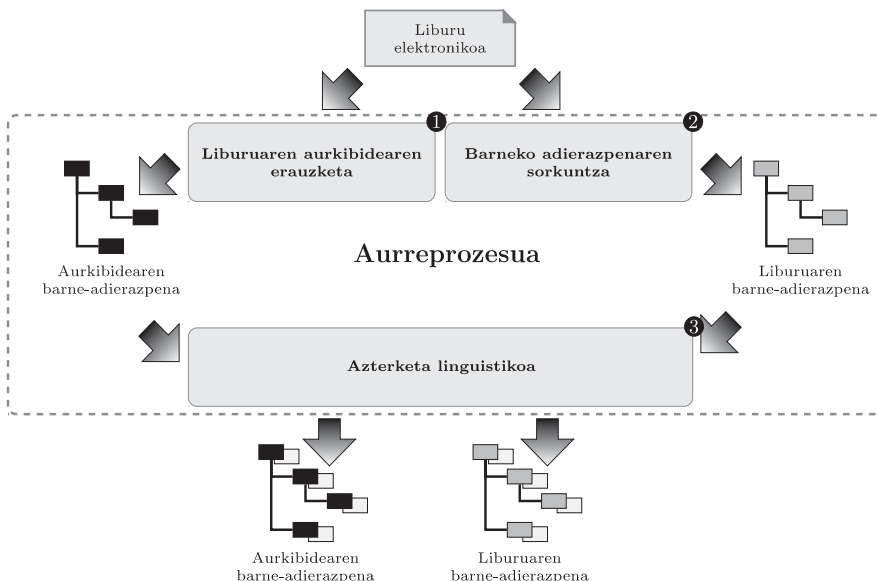
1. irudia. Domeinu-moduluaren sorkuntza-prozesua.

Prozesu erdiautomatiko honen emaitzak —bai Ikaste Domeinuaren Ontologia bai ikaste-objektuak— domeinu-moduluaren egileek berrikustea ezinbestekoa iruditzen zaigu. Berrikuspena elkarlanaren bidez egin ahal izateko *Elkar-DOM* [8] garatu da. Kontzeptu-mapetan oinarritutako tresna da *Elkar-DOM*, domeinu-modulua berrikustea ahalbidetzen duena. Tresna honen bidez emaitzak zuzentzeko edo/eta egileen nahierara egokitzeko aukera dago.

Jarraian, aipatutako urrats bakoitza xehetasun gehiagorekin azaltzen da.

3. LIBURUEN AURREPROZESUA

Liburuaren aurreprozesua jatorrizko dokumentutik abiatzen da eta domeinu-moduluaren erauzketarako egokia den adierazpen estandarra sortzen du (ikus 2. irudia). Liburu elektronikoak formatu ezberdinetan (*pdf*, *doc*, etab.) egon daitezke, eta aurreprozesuak edozein dokumentu mota erauztea ahalbidetzen du. Dokumentu mota horiek guztiak zuhaitz moduko egitura dute; liburuak ataletan banatzen dira, atalak azpiataletan, eta horrela hurrenez hurren. Urrats honetan, zuhaitz moduko egitura hori deskribatzen duen barne-adierazpena erauzten da, ondoren informazio morfosintaktikoarekin aberasten dena.



2. irudia. Liburuaren aurreprozesua.

Liburuen aurkibidea, liburuaren topiko 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.

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

Euskaraz, adibidez, lemei artikulua, numeroa eta deklinabide kasua adierazten duten atzizkiak gehituz osatzen dira hitzak [9]. Hitzen esanahia ulertzeko, ezinbestekoa da azterketa morfosintaktikoa, euskararen kasuan EUSLEM [10] lematizatzaile-etiketatzailea erabiliz egiten dena.

4. IKASTE DOMEINUAREN ONTOLOGIAREN ERAUZKETA

Ontologia-erauzketa —hots, domeinu-ontologiaren edukia hainbat iturburutatik modu automatikoan edo erdiautomatikokoan osatzea— modu ezberdinetan landu da dagoeneko [11]. Proiektu gehienek domeinu-ontologia bat sortzea edo zabaltzea dute helburu. Beste batzuek, ordea, Wordnet [12] moduko ontologia lexikoak aberasteko asmoa dute. Ontologia-erauzketarako ikaste automatikoa eta Hizkuntzaren Prozesamendurako teknikak erabili ohi dira, bai corpusetatik, bai hiztegi elektronikoetatik, bai beste iturburu batzuetatik, ontologiaren edukia lortzeko.

Egitura sintaktikoen pean ezagutza semantikoa dagoelako premisan oinarritzen da ontologia-erauzketa. Adibidez, Text2Onto [13] tresnak Hearst-en patroiak [14] erabiltzen ditu erlazio taxonomikoak erauzteko, eta termino habiaratuak aztertzeke neurriak erabiltzen ditu termino hautagaiak bilatzeko [15]. OntoLT-k [16] —ontologiak erauzteko Protége⁴-rentzako pluginak— *genus et differentia* egiturak baliatzen ditu erlazio taxonomikoak bilatzeko.

Lan honetan, Ikaste Domeinuaren Ontologiak domeinuko topiko nagusiez gain, beraien arteko erlazio pedagogikoak adierazten ditu. Erlazio pedagogikoek *egitura* (*isA* edo *partOf*) eta *ordena* (*prerequisite* edo *next*) adieraz ditzakete. *X isA Y* erlazioak adierazten du *X* topikoa *Y*ren mota zehatz bat dela. *X partOf Y* erlazioak, ordea, *X* topikoa *Y*ren zati bat dela dio, hots, *Y* landuzat emateko ikasi behar diren topikoetako bat dela adie-

⁴ <http://protege.stanford.edu/>

razten du. Halaber, *X prerequisite Y* erlazioak adierazten du *X* ikasten hasi aurretik ikasleak *Y* dagoeneko landuta izan behar duela, eta *X next Y* erlazioak, berriz, *X* landu ondoren *Y* ikastea gomendatzen du.

Ikaste Domeinuaren Ontologiaren erauzketa bi urratsetan gauzatzen da, Hizkuntza Prozesamendua eta arrazoibide heuristikoa erabiltzen dituztenak: aurkibidearen azterketak hasierako ontologiaren bertsioa erazten du, gero dokumentu osoaren azterketarekin aberastuko dena. Prozesu honetan zehar, ontologiaren barne-adierazpena ere erabiltzen da, edukiez gain informazioa eraztean erabilitako heuristikoen informazioa adierazten duena.

4.1. Aurkibidearen azterketa

Hemen aurkeztutako ontologia eraikitzeke 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 aurkibidearen egiturean adierazita daudela kontuan hartuta, Hizkuntzaren Prozesamendurako tresnak eta heuristiko multzo bat erabiliko dira erlazio pedagogikoak erazteko.

Aurkibidearen azterketan zehar bi prozesu egingo dira:

1. **Oinarrizko azterketa:** Ataza honetan aurkibidearen barne-adierazpenetik erazten 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 baten eta bere ume bakoitzaren artean. Gainera, aurkibidearen antolaketak liburua irakurtzeko gomendatutako ordena adierazten du eta *ordena* erlazio pedagogikoak erazteko balio du.
2. **Azterketa heuristikoa:** Oinarrizko azterketan jasotako emaitzak fintzen dira ataza honetan, eta erlazio berri batzuk —*prerequisite* erlazioa batez ere— erazten dira heuristiko sorta batez baliatuz. Azterketa heuristikoa bi urratsetan egikaritzen da. Hasteko, *isA* eta *partOf* erlazioak erazteko erabiltzen diren heuristikoak aplikatzen dira eta, segidan, *next* eta *prerequisite* erlazioak topatzen dituzte-

nak. 1. taulak erakusten du ENH heuristikoak *isA* erlazioa topa dezakeela aurkibide-elementu baten umeez izen bereziak osatzen dituztenean. Izen bereziak topiko baten adibide zehatzak adierazteko erabiltzen dira. Erauzitako erlazioetan, motaz gain, erlaziook topatzean erabilitako heuristikoak eta horien ziurtasun mailak ere gordezten dira.

1. taula. *isA* erlazioa erauzteko ENH heuristikoak erabil daitekeen aurkibide-zati bat.

3.7. Laguntzaile Pertsonal Digitalak (PDA-ak)

3.7.1 **Palm OS**

3.7.2 **Windows CE**

4.2. Liburu osoaren azterketa

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

4.2.1. 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 konbinatzen dituzten termino-erauzketarako prozedura hibridoak nagusitu dira. Lan askotan, lehendabizi patroiak erabiltzen dituzte termino hautagaiak topatzeko eta, neurri estatistikoak erabiliz, hautagai onenen sailkapen bat egiten dute [17].

DOM-Sortze-n, topiko-erauzketa Erauzterm [18] euskararako termino-erauzlearen bidez burutzen da. Erauzterm hitz bakarreko zein hitz anitzeko terminoak erauzteko gai da.

4.2.2. Topikoen arteko erlazio berrien erauzketa

Proposamen honetan, erlazioen erauzketa Hizkuntzaren Prozesamendurako tekniketara oinarritzen da. Prozesu honek liburu osotik topikoen arteko erlazio berriak erauztea ahalbidetzen du patroiak erabiliz. Patroi bidez egitura sintaktikoetan erlazio pedagogikoak topatzea da helburua. Erlazio horiek topatu ahal izateko, ontologiako topikoen agerraldiak etiketatzen dira liburuaren barne-adierazpenean, bertan baitago informazio morfosintaktikoa. Topatutako topikoen habiaratze 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 [19] jarraitzen du, sintaxia aztertzeko modu zabalduenetako bat. 2. taulan ikus daitezkeen moduko erregelak daude definituta gramatikan.

2. taula. *isA* erlazioak topatzeko patroia batzuk.

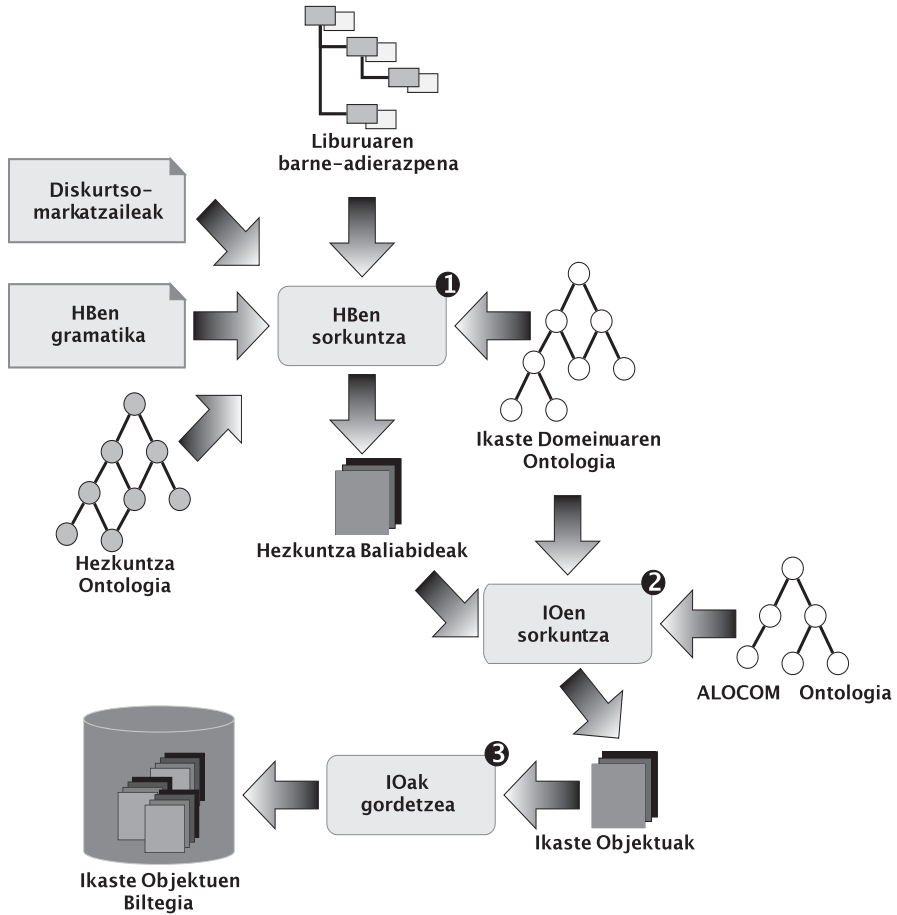
Patroia	<u>@Topic@Topic</u> [<i>det</i>] IZAN
Adibidea	<u>Lurra planeta</u> <i>bat da</i> .
Patroia	<u>@Topic izeneko @Topic</u>
Adibidea	<u>Esne bidea izeneko galaxiak</u> 100 mila milioi izar dituela uste dute zientzialariek.

5. IKASTE-OBJEKTUEN ERAUZKETA

Liburuan dauden hezkuntzarako baliabideak, hots, ontologiako topiko bati edo gehiagori lotuta dauden testu zatiak eta helburu pedagogikoa daukatenak (adib. *definizioak*, *adibideak* edo *ariketak*) lortzean datza ikaste-objektuaren erauzketa. Horretarako, Hizkuntzaren Prozesamendurako teknikak eta ontologiak erabiltzen dira. Domeinuarekiko independentea izan nahi duen hurbilpena izateko asmoz sortu denez, domeinuari lotutako informazio bakarra erabiltzen da, alegia, Ikaste Domeinuaren Ontologiak, daukana, aurreko fasean jaso.

Hemendik aurrera, Hezkuntza Baliabide (HB) terminoa erabiltzen denean, ikasteko asmoz erabiliko den liburu zati bati dagokio, adibidez *ariketeta* edo *definizio* bati. Ikaste-objektua, aldiz, hezkuntzarako baliabide berrerabilgarria da, bilaketa eta erabilera errazten duten metadatuekin aberastua. Ikaste-objektuak erauzteko *ErauzOnt* tresna [20] erabiltzen da, *DOM-Sortze*-ren zati bat dena.

Ikaste-objektuak sortzeko prozedura orokorra hiru atazatan burutzen da (ikus 3. irudia). Lehendabizi, liburutik Hezkuntza Baliabideak erauzten dira. Ondoren, baliabide horiek metadatuekin aberasten dira ikaste-objektuak sortzeko, eta, bukaeran, Ikaste Objektuen Biltegian gordetzen dira, berriro ere erabili ahal izateko.



3. irudia. Ikaste-objektuak (IO) liburutik sortzeko prozedura.

Hezkuntza Baliabideen gramatikak (HB gramatikak) *definizioak, ariketak, teoriak etab.* bilatzeko erregelak definitzen ditu eta testuko Hezkuntza Baliabideak topatzea ahalbidetzen du. 3. taulan *definizioak* topatzeko erabiltzen den adibide bat aurkezten da.

3. taula. Definizioak topatzeko patroia bat.

Patroia	@Topic@Topic [det] IZAN
Adibidea	Lurra planeta bat da.

Topatutako baliabideak bi modutan aberasten dira: alde batetik, jarraian dauden bi baliabide beste berri bat osatzeko batzen dira baldin topiko aldetik eta baliabide moten aldetik antzekoak badira [21,22,23]; 4. taulan erausten dira osatzeko hautagai diren bi baliabide jarraituak. Bestalde, HBen aberaste-prozesuan diskurtso-markatzaileak erabiltzen dira testuaren koherentzia mantentzeko.

4. taula. Elkartu daitezkeen bi baliabide.

Adibidea	
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	<u>Lurra Planeta bat da.</u>

Ondoren, ikaste-objektuak sortzen dira Hezkuntza Baliabideetatik abiatuta. Horretarako, metadatuak modu automatikoan eransten zaizkie. Alde batetik, SAmGI [24] etiketatzailea erabiltzen da hasierako metadatuak lortzeko. Bestalde, gako-hitzak Ikaste Domeinuaren Ontologia erabiliz aberasten dira topikoen arteko erlazioak kontuan har ditzaten. SAmGI-k testu batean *Artizarra*, *Marte* eta *Lurra* aipatzen dituen testuan ez luke jakingo *planetei* buruz ari denik. Ontologia erabilita, aldiz, horrelako ezagutza kontuan hartzen da metadatuak hobetzeko.

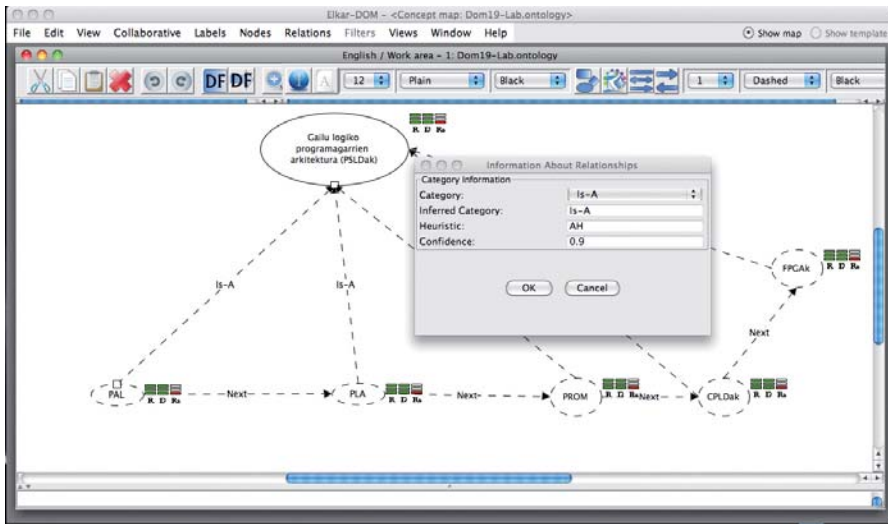
Bukaeran, sortutako ikaste-objektu guztiak Ikaste Objektuen Biltegian gordetzen dira. Erabilitako biltegia ARIADNERen teknologian [5,6] oinarritzen da.

6. ELKARLANAREN BIDEZKO GAINBEGIRATZE-PROZESUA

Aurreko ataletan domeinu-modulua sortzeko egiten diren erauzketa-prozesuak deskribatu dira. Prozesu horiek automatikoak direnez, erabiltzaileen gainbegiratzea ezinbestekoa da domeinu-moduluak izan ditzakeen akatsak zuzendu ahal izateko edo domeinu-modulua beren helburuetara egokitzeko. Gainera, askotan irakasle batek baino gehiagok parte hartzen dute irakasgai bat prestatzerakoan. Beraz, irakasle horiei elkarlana erraztu behar zaie modu automatikoan erauzitako domeinu-modulua egokitzeko. Helburu horrekin, *Elkar-DOM* [8] garatu da. Kontzeptu-

mapetan oinarritutako tresna da *Elkar-DOM*, domeinu-modulua berrikus-tea ahalbidetzen duena. *Elkar-DOM*-en bidez, gainbegiratzea bi mailatan egin daiteke: bata, Ikaste Domeinuaren Ontologia egokitzeko, eta bestea, topiko bakoitzari dagozkion ikaste-objektuak esleitzeko. Berrikuspen-prozesu honetan hainbat irakaslek parte har dezakete, beren irakasgaiekin egin izan duten moduan.

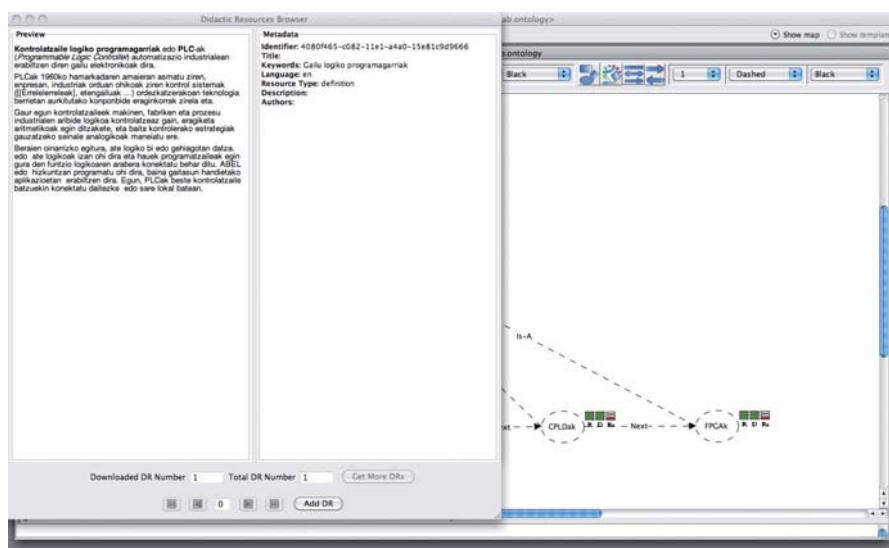
Elkar-DOM-ek Ikaste Domeinuaren Ontologia kontzeptu-mapa baten bidez grafikoki adierazten du (ikus 4. irudia). Adabegiek domeinuko topikoak adierazten dituzte. Geziek, berriz, erlazio pedagogikoak erakusten dituzte. Marra jarraituak elementua —topikoa zein erlazioa— dagoeneko berrikusita dagoela adierazten du. Marra etenak, berriz, baliabidea berrikusteke dagoela adierazten du. Erabiltzaileak topiko edo erlazio baten informazioa ikus dezake dagokion baliabidean klik bikoitza eginez. Erakusten den leihoan, informazioa erazteko erabilitako heuristikoa eta erabakiaren ziurtasun maila ere adierazten dira. Erabiltzaileak erabakia zuzentzeko aukera du. Erabakia baieztatzen edo zuzentzen denean, erabilitako heuristikokoak daukan fidagarritasuna egokitzen da.



4. irudia. Ikaste Domeinuaren Ontologiaren berrikuspena Elkar-DOM erabiliz.

Ikaste Domeinuaren Ontologia berrikusteaz gain, *Elkar-DOM*-ek topikoei ikaste-objektu egokiak esleitzea ahalbidetzen du. Erabiltzaileak topiko bati dagozkion adabegiaren gainean klik bikoitza egiten duenean, topikoaren informazioa agertzen zaio. Leiho horretan, topikoari lotuta dauden

ikaste-objektuak bilatzeko aukera dauka erabiltzaileak, eta horretarako topatu nahi duen ikaste-objektu mota aukeratu behar du. *Elkar-DOM*-ek bilaketa Ikaste Objektu Biltegian gauzatu ondoren, baldintzak betetzen dituzten ikaste-objektu guztiak jasotzen ditu eta erabiltzaileari hautatzeko aukera ematen dio (ikus 5. irudia). Erabiltzaileak ikaste-objektu aproposak aukera ditzan, metadatuak zein ikaste-objektuaren edukia erakusten dizkio *Elkar-DOM*-ek.



5. irudia. Ikaste-objektuen bilaketa *Elkar-DOM* erabiliz.

7. EBALUAZIOA

DOM-Sortze-k domeinu-modulua sortzeko ekar zezakeen onura aztertze-
tze, datu errealekin probatu zen. Ebaluazioa Informazio Erauzketa era-
biltzen den urre-patroi erudian oinarritu zen. Horretarako, adituek zein
baliabideen egileek erudizat hartuko zen informazioa zehaztu zuten. *DOM-
Sortze*-ren jokabidea aztertze hiru neurri erabili ziren:

- **Estaldura:** topatutako elementu garrantzitsuen portzentajea.
- **Doitasuna:** topatutako elementuen artean, garrantzitsuak direnen portzentajea.
- **F1 neurria:** estaldura eta doitasuna neurrien batezbesteko harmo-
nikoa.

Ebaluazioa hiru urratsetan egin zen: lehenengoan, Ikaste Domeinuaren Ontologiaren erauzketa aztertu zen; bigarrenean, ikaste-objektuen erauzketa; azkenik, prozesu osoa. Jarraian, urrats bakoitza deskribatzen da.

7.1. Ikaste Domeinuaren Ontologiaren erauzketa aurkibideetatik

Aurkibideak dira *DOM-Sortze*-k Ikaste Domeinuaren erauzketarako erabiltzen dituen baliabide funtsezkoak, bertatik eskura baitaitezke topiko eta erlazio pedagogiko nagusiak. Gauzak horrela, Euskal Herriko Unibertsitatean eskaintzen ziren 150 irakasgaitako aurkibideak erabili ziren urrats honetan, gizarte-zientziak, ekonomia edo ingeniarietza arloetakoak [25].

Irakasleen laguntzaz, irakasgai bakoitzari zegokion Ikaste Domeinuaren Ontologia sortu zen eskuz, aurkibidean azaltzen ziren topikoetara mugatuz. Ikaste Domeinuaren Ontologia hauek osatu zuten urre-patroia eta *DOM-Sortze*-ren bidez lortutako Ikaste Domeinuaren Ontologiak haiekin parekatu ziren erauzketa ebaluatzeko. Muga hau izanda, erlazioen erauzketa baino ez zen aztertu.

5. taula. Ikaste Domeinuaren Ontologiaren erauzketako emaitzak.

	PartOf	IsA	Next	Prerequisite	Guztira
Estaldura (%)	99,15	79,38	99,86	92,95	98,15
Doitasuna (%)	98,19	89,27	99,61	97,75	98,36
F1 neurria (%)	98,67	84,04	99,74	95,29	98,26

5. taulan ikus daitekeenez, aurkibideak Ikaste Domeinuaren Ontologia erauzketako baliabide aproposak dira. Urre-patroia osatzen zuten Ikaste Domeinuaren Ontologiak aurkibidean aipatzen ziren topikoetara mugatuta zeudenez, proban lortutako estaldura ez da adierazgarria; doitasunak, ordea, erlazioak erauzketako prozesuaren fidagarritasun nabarmena adierazten du.

Arloen arteko konparaketa ere egin zen, eta emaitzak oso antzekoak izan ziren beraien artean, *IsA* erlazioaren kasuan salbu. Erlazio hori erauzketako, askotan domeinu-ezagutza sakona ezinbestekoa da.

Orokorrean domeinu guztietan emaitza onak jaso zirela kontuan izanda, Ikaste Domeinuaren Ontologiaren erauzketa domeinuarekiko independentea dela esan daiteke.

7.2. Ikaste-objektuen erauzketa

Ikaste-objektuen erauzketaren ebaluazioa egiteko Euskal Herriko Ikastolak (EHI)⁵ elkarteak hornitutako lau testuliburu erabili ziren [23]. Lau liburua, Geologia eta Biologiari buruzkoak, Derrigorrezko Bigarren Hezkuntzan erabiltzen dira. Adituek urre-patroia osatzeko topatu beharreko ikaste-objektuak zehaztu zituzten. Urre-patroiak *definizioak*, *fenomeno-deskribapenak*, *adibideak*, *ariketak* eta *ikaste-objektu osatuak* (baliabide mota ezberdinak elkartuz sortutakoak) zeuzkan (ikus 6. taula).

6. taula. Ikaste-objektuen erauzketarako urre-patroiaren datuak.

	Def.	Fen.	Adib.	Arik.	Osat.	Guztira
Kopurua	67	12	8	105	37	229

DOM-Sortze gai izan zen urre-patroia osatzen zuten ikaste-objektuetatik % 70,31 erauzteko, 7. taulan ikus daitekeenez. Egindako azterketan, erauzitako ikaste-objektuetatik % 91,88 egokiak zirela egiaztatu zen.

7. taula. Ikaste-objektuen erauzketaren emaitzak.

	Def.	Fen.	Adib.	Arik.	Osat.	Guztira
Estaldua (%)	59,70	50,00	87,50	81,90	59,46	70,31
Doitasuna (%)	91,14	96,30	100,00	88,55	97,55	91,88
F1 neurria (%)	72,14	65,82	93,33	85,10	73,97	79,66

7.3. Prozesu osoa

Behin Ikaste Domeinuaren Ontologiaren erauzketa eta ikaste-objektuen erauzketa, bakoitza bere aldetik, ebaluatuta, prozesu osoaren ebaluazioari ekin zitzaion. Horretarako EHIk hornitutako liburu berri bat erabili zen [26]. Liburua, Bigarren Hezkuntzakoa, astronomiari buruzkoa zen. Ikaste Domeinuaren Ontologiak (8. taulan deskribatua) eta ikaste-objektuen multzoek (9. taulan adieraziak) osatzen zuten ebaluazioa egiteko sortu zen urre-patroia.

8. taulan ikus daitekeenez, Ikaste Domeinuaren Ontologia 83 topikok eta 135 erlaziok osatzen zuten.

⁵ <http://www.ikastola.net>

8. taula. Urre-patroia – Ikaste Domeinuaren Ontologia.

	Kopurua
Topikoak	83
Erlazioak	135
PartOf	43
IsA	66
Next	13
Prerequisite	13

9. taulan ikus daitekeenez, ikaste-objektuen multzorako adituek *definizioak, teoriak, adibideak, ariketak* eta *ikaste-objektu osatuak* aukeratu zituzten.

9. taula. Urre-patroia – Ikaste-objektuak

	Def.	Teor.	Adib.	Arik.	Osat.	Guztira
Kopurua	27	6	21	6	6	66

10. taulak Ikaste Domeinuaren Ontologiaren erauzketa-prozesuaren emaitzak laburbiltzen ditu. Ikus daitekeenez, topikoen erauzketa arrakastatsua izan zen, urre-patroia osatzen zuten topikoetatik % 87,95 automatikoki topatu baitziren. Erlazioen erauzketan emaitza xumeagoak lortu ziren, % 40,74 soilik erauztea lortu baitzen. Horren arrazoi nagusia azterketarako erabili zen liburuaren aurkibide sinplean egon daiteke. 7.1 atalean deskribatutako proban, heuristikoeak oso doitasun handia erakutsi zuten. 10. taulan ikus daitekeenez,

10. taula. Ikaste Domeinuaren Ontologiaren erauzketaren emaitzak.

		Guztira
Topikoak	Estaldura (%)	87,95
	Doitasuna (%)	17,48
	F1 neurria (%)	29,16
Erlazioak	Estaldura (%)	40,74
	Doitasuna (%)	72,50
	F1 neurria (%)	52,17

erlazioak erazteko erabili ziren erregelak edo patroiek doitasun ona izan zuten. Beraz, aurkibide zehatzago bat izanda emaitza hobekak lortuko liratekeela esan dezakegu.

Ikaste-objektuen erazketari dagokionez, orokorrean 7.2 atalean deskribatutako esperimentuan bezala, emaitza onak lortu ziren (ikus 11. taula).

11. taula. Ikaste-objektuen erazketaren emaitzak.

	Def.	Teor.	Adib.	Arik.	Osat.	Guztira
Estaldura (%)	74,04	33,00	95,24	50,00	100,00	77,27
Doitasuna (%)	78,79	100,00	93,33	71,43	86,21	84,50
F1 neurria (%)	76,36	50,00	94,28	58,82	92,59	80,73

8. ONDORIOAK ETA ETORKIZUNERAKO LANA

Teknologian Oinarritutako Hezkuntzarako Tresnak ezinbesteko bihurtu dira gaur egun. Tresna horiek, beren helburua bete dezaten, *domeinu-modulua* —ikasi beharreko domeinuaren adierazpen pedagogikoa— behar dute. *Domeinu-modulua* definitzeak oso lan handia eskatzen du, ordea. Artikulu honetan, *DOM-Sortze*, *domeinu-modulua* testuliburu elektronikoetatik era erdiautomatikoan erazteko tresna aurkeztu da, domeinuarekiko independentea dena. *DOM-Sortze* ontologietan, arrazonamendu heuristikotan eta Hizkuntzaren Prozesamendurako tekniketari oinarritzen da. Nahiz eta orain arte euskararen gainean bakarrik erabili ahal izan den, *DOM-Sortze* hizkuntzarekiko independentea da, hizkuntzari lotuta dauden ezaugarriak (heuristikoak, gramatikak eta diskurtso markatzaileak) baliabideak baitira. Hizkuntza berri bat onar dezan, *DOM-Sortze* hizkuntza horretarako behar diren prozesatzaile batez hornitu beharko litzateke, eta dagozkion heuristikoak, gramatikak eta diskurtso-markatzaileak definitu. Dagoeneko, *DOM-Sortze* ingelesa tratatu ahal izateko hornitu da.

DOM-Sortze euskaraz idatzitako hainbat dokumentu eta aurkibiderek in probatu da emaitza onak lortuz. Probetarako erabilitako baliabideak arlo ezberdinen ingurukoak izanda, eta guztiekin emaitza positiboak lortu direnez, *DOM-Sortze* domeinuarekiko independentea dela ere esan daiteke.

Etorkizunean, *DOM-Sortze* domeinu-modulu eleaniztunak sortzeko aberastuko da. Horretarako, elkarlana errazten duen *Elkar-DOM* tresna hobetzea ere ezinbestekoa izango da.

9. ESKER ONAK

Lan honek EHUKo UFI11/45 Prestakuntza eta Ikerkuntza Unitatearen eta Eusko Jaurlaritzako GIC12/79 ikerkuntza taldeen babesa jaso du.

10. BIBLIOGRAFIA

- [1] PARSAD B. eta LEWIS L., 2008, *Distance Education at Degree-Granting Postsecondary Institutions: 2006--07*.
- [2] CHEN P.-S. D., LAMBERT A. D. eta GUIDRY K. R., 2010, «Engaging on-line learners: The impact of Web-based learning technology on college student engagement», *Computers & Education*, **54**(4), 1222-1232.
- [3] ANDERSON J. R., 1988, *The Expert Module*, Foundations of Intelligent Tutoring Systems, Lawrence Erlbaum Associates, Inc., 21-54.
- [4] LTSC, 2001, *1484.12.1 IEEE LTSC Draft Standard for Learning Object Metadata*.
- [5] DUVAL E., FORTE, E., CARDINAELS K., VERHOEVEN B., DURM R. V., HENDRIKX K., FORTE M. W., EBEL N., MACOWICZ M., WARKENTYNE K. eta HAENNI F., 2001, «The ARIADNE Knowledge Pool System», *Communications of the ACM*, **44**(5), 72-78.
- [6] TERNIER S., VERBERT K., PARRA G., VANDEPUTTE B., KLERKX J., DUVAL E., ORDONEZ V. eta OCHOA X., 2009, «The Ariadne Infrastructure for Managing and Storing Metadata», *IEEE Internet Computing*, **13**(4), 18-25.
- [7] CAFOLLA R., 2006, «Project Merlot: Bringing Peer Review to Web-based Educational Resources», *Journal of Technology and Teacher Education*, **14**(2), 313-323.
- [8] LARRAÑAGA M., NIEBLA I., RUEDA U., ELORRIAGA J. A. eta ARRUARTE A., 2007, «Towards Collaborative Domain Module Authoring», *Proceedings of the 7th IEEE International Conference on Advanced Learning Technologies*, ICALT 2007, IEEE Computer Society, 814-818.
- [9] ADURIZ I., AGIRRE E., ALDEZABAL I., ALEGRIA I., ANSA O., ARREGI X., ARRIOLA J. M., ARTOLA X., DE ILARRAZA A. D., EZEIZA N., GOJENOLA K., MARITXALAR A., MARITXALAR M., ORONOZ, M., SARASOLA K., SOROA A., URIZAR R. eta URKIA M., 1998, «A Framework for the Automatic Processing of Basque», *Proceedings of the 1st International Conference on Language Resources and Evaluation (LREC 1998)*.
- [10] ADURIZ I., ALDEZABAL I., ALEGRIA I., ARTOLA X., EZEIZA N. eta URIZAR R., 1996, «EUSLEM: A Lemmatiser / Tagger for Basque», *Proceedings of the 7th EURALEX International Congress on Lexicography*, EURALEX 1996, 17-26.

- [11] BUITELAAR P., CIMIANO P. eta MAGNINI B., 2005, «Ontology Learning from Text: An Overview», *Ontology Learning from Text: Methods, Applications and Evaluation*, IOS Press, 3-12.
- [12] FELLBAUM C., 1998, *WordNet: An Electronic Lexical Database*, MIT Press.
- [13] CIMIANO P., eta VÖLKER J., 2005, «Text2Onto - A Framework for Ontology Learning and Data-driven Change Discovery», *10th International Conference on Applications of Natural Language to Information Systems, NLDB 2005*, Springer, 227-238.
- [14] HEARST M. A., 1992, «Automatic Acquisition of Hyponyms from Large Text Corpora», *Proceedings of the 14th Conference on Computational Linguistics, COLING-92*, Association for Computational Linguistics, 539-545.
- [15] FRANTZI K. T., ANANIADOU S. eta ICHI TSUJII J., 1998, «The C-value/NC-value Method of Automatic Recognition for Multi-Word Terms», *Proceedings of the 2nd European Conference on Research and Advanced Technology for Digital Libraries, ECDL '98*, Springer, 585-604.
- [16] BUITELAAR P., OLEJNIK D. eta SINTEK M., 2004, «A Protégé Plug-In for Ontology Extraction from Text Based on Linguistic Analysis», *The Semantic Web: Research and Applications, First European Semantic Web Symposium, ESWS 2004*, Springer, 31-44.
- [17] JUSTESON J. S., eta KATZ S. M., 1995, «Technical Terminology: Some Linguistic Properties and an Algorithm for Identification of Terms in Text», *Natural Language Engineering*, **1**(1), 9-27.
- [18] ALEGRIA I., GURRUTXAGA A., LIZASO P., SARALEGI X., UGARTETXEA S. eta URIZAR R., 2004, «An Xml-Based Term Extraction Tool for Basque», *Proceedings of the 4rd International Conference on Language Resources and Evaluation (LREC 2004)*.
- [19] KARLSSON F., VOUTILAINEN A. eta HEIKKILA J., 1995, *Constraint Grammar: Language-independent System for Parsing Unrestricted Text*, Mouton de Gruyter.
- [20] LARRAÑAGA M., CALVO I., ELORRIAGA J. A., ARRUARTE A., VERBERT K. eta DUVAL E., 2011, «ErauzOnt: A Framework for Gathering Learning Objects from Electronic Documents», *Proceedings of the 11th IEEE International Conference on Advanced Learning Technologies, ICALT 2011*, IEEE Computer Society, 656-658.
- [21] LARRAÑAGA M., ELORRIAGA J. A. eta ARRUARTE A., 2008, «Semi Automatic Generation of Didactic Resources from Existing Documents», *Proceedings of the 9th International Conference on Intelligent Tutoring Systems, ITS 2008*, Springer, 728-730.
- [22] LARRAÑAGA M., ELORRIAGA J. A. eta ARRUARTE A., 2008, «A Heuristic NLP Based Approach for Getting Didactic Resources from Electronic Documents», *Times of Convergence. Technologies Across Learning Contexts, Proceedings of Third European Conference on Technology Enhanced Learning, EC-TEL 2008*, Springer, 197-202.

- [23] LARRAÑAGA M., CONDE A., CALVO I., ARRUARTE A. eta ELORRIAGA J. A., 2012, «Evaluating the Automatic Extraction of Learning Objects from Electronic Textbooks using ErauzOnt», *Proceedings of 11th International Conference on Intelligent Tutoring System*, ITS 2012, Springer, 655-656.
- [24] MEIRE M., OCHOA X. eta DUVAL E., 2007, «SAmgI: Automatic Metadata Generation v2.0», *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007*, ED-MEDIA 2007, AACE, 1195-1204.
- [25] LARRAÑAGA M., RUEDA U., ELORRIAGA J. A. eta ARRUARTE A., 2004, «Acquisition of the Domain Structure from Document Indexes Using Heuristic Reasoning», *Proceedings of the 7th International Conference on Intelligent Tutoring Systems*, ITS 2004, Springer, 175-186.
- [26] LARRAÑAGA M., CONDE A., CALVO I., ELORRIAGA J. A. eta ARRUARTE A., Argitaratze-prozesuan, «Automatic Generation of the Domain Module from Electronic Textbooks. Method & Validation», *IEEE Transactions on Knowledge and Data Engineering*. <http://doi.ieeecomputersociety.org/10.1109/TKDE.2013.36>