TECHNOLOGY DEVELOPMENT OF MODULE TEXT

B.Sivaranjani
Assistant Professor,
Dr. N.G.P Arts and Science College,Coimbatore.

S.Meena Selvaraj
M.Phil Scholar,
Dr. N.G.P Arts and Science College,Coimbatore.

ABSTRACT

Technology-supported learning systems have proved to be helpful in many learning situations. These systems require an appropriate representation of the knowledge to be learned, the Domain Module. The authoring of the Domain Module is cost and labor intensive, but its development cost might be lightened by profiting from semiautomatic Domain Module authoring techniques and promoting knowledge reuse.

DOM-Sortze is a system that uses natural language processing techniques, heuristic reasoning, and on tologies for the semiautomatic construction of the Domain Module from electronic textbooks.

KEYWORDS:

- Heuristics
- Lingusitic
- Pedagogical
- Ontology

1.INTRODUCTION

The revolution of information and communication technologies (ICTs) has affected education, providing means to enhance both the teaching and learning Processes.

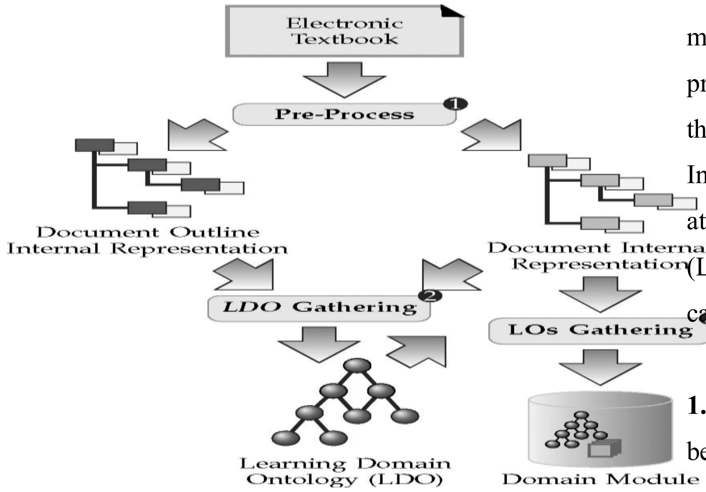


Fig.1 Domain Module building process

Nowadays, technology-supported learning systems (TSLSs), such as intelligent tutoring systems (ITSs), adaptive hypermedia systems (AHSs), and, especially, teaching management systems (LMSs) such as Moodle1 or Blackboard, 2 are being widely used in many academic institutions and becoming essential for education . Furthermore, a positive relationship between the use of web-based learning technology and student engagement and desirable learning outcomes has been observed . To be effective, TSLSs require an appropriate Domain Module, i.e., the pedagogical representation of the domain to be learned. 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. The Domain Module enables either the students to learn by themselves, in the case of exploratory learning systems, or to guide students through the learning process in instructive TSLS . An incomplete Domain Module may result in a system that is only able to provide part of the instruction required in the domain

- http://moodle.org
- http://www.blackboard.com

2.BUILDING THE DOMAIN MODULE

The approach here presented uses artificial intelligence methods and techniques such as natural language processing (NLP) and heuristic reasoning to achieve the semiautomatic generation of the Domain Module. In this work, the Domain Module encodes knowledge at two different levels, the learning domain ontology (LDO) and the set of LOs. The following steps are carried out to develop the Domain Module (see fig.1)

1. Textbook preprocessing. First, the document must be prepared for the subsequent knowledge acquisition processes. This process is described in Section 3, and the outcomes are then used to gather the two levels of knowledge encoded in the Domain Module.

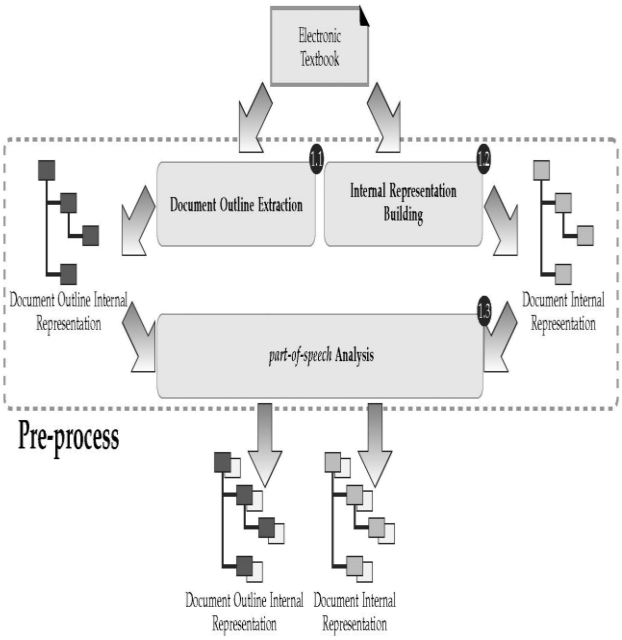
2. LDO gathering.

At this phase, the domain topics to be mastered, as well as the pedagogical relationships among them are identified and represented in the LDO. The LDO will allow either the TSLS to plan the learning session or the students to guide themselves during the learning process.

3 TEXTBOOK PREPROCESSING

In this phase, the system prepares the electronic document and gathers a standardized representation of it, to later run the knowledge acquisition processes (see Fig. 2). As electronic documents are available in many different formats, such as pdf, rtf, doc, or odf, a preprocess is carried out first to prepare the document.

The content of electronic documents is organized using a hierarchical structure; documents contain chapters, which in turn are divided into sections, and so on. A tree-like internal representation of the document is built, so that the rest of the task can be performed with no dependence 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 of the document, can also be numbered or



indented in different ways showing its structure. Thus, homogenized internal representation of the outline is also gathered in the preprocess. The obtained internal representations for the outline and the document body are then linguistically analyzed to enhance them with the part-of-speech information that will be used in the following steps.

Linguistic analysis is essential, especially for agglutinative languages such as Basque, where most words are formed by joining morphemes together.

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 . As prepositional functions are realized by case suffixes inside word forms, Basque presents a relatively high power to generate inflected word forms, which makes morph syntactic analysis very important to be able to extract information from text fragments.

The linguistic analysis is carried out by a web service, which uses EUSLEM , a lemmatize/tagger for the Basque language.

4. GATHERING THE LDO

Ontology learning, i.e., gathering domain ontologism from different resources in an automatic or semiautomatic way has been addressed in many works [7], [8]. Most of these projects aim at building or extending a domain ontology or populating lexical ontologism such as WorldNet .

Ontology learning usually combines machine learning and NLP techniques to build domain ontologism 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 sources of information for ontology learning

. In the approach here presented, the LDO contains the main domain topics and the pedagogical relationships among them. Pedagogical relationships can be structural—is A and part Of—or sequential—prerequisite and next. The X iIs A___!Y relationship declares that the topic X is a particular kind of Y . The X part Of_____!Y describes that X is part of Y , i.e.,X is one of the topics to learn to fully master Y . The Y prerequisite __________! X relationship states that a topic X must be mastered before attempting to learn topic Y, whileX next____!Y expresses that it is recommended to learn topic Y right after mastering topic X. Ontology learning relies on the assumption that there issemantic knowledge underlying syntactic structures.

For example, Text2Onto uses Hearst's patterns togather taxonomic relationships, and nested term-based methods to identify the set of candidate domain topics.OntoLT , a Prote´ge´3 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).

The LDO describes a certain domain for learning purposes; it is an application ontology according to Guarino's classification, and it also describes 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 document body analysis, which enhances the ontology with new topics and relationships.

During the LDO gathering process, an internal representation is used; in this representation, besides the learning topics and the relationships, information about the gathering process itself—used heuristics, confidence on the heuristics, and so on—is also included. Once the LDO has-been gathered and reviewed by teachers or instructional designers using Elkar -DOM , the ontology is represented in OWL .

4.1 OUTLINE ANALYSIS

Document outlines are the main sources of information for acquiring the LDO in a semiautomatic way as they are usually well structured and contain the main topics of the domain. Besides, they are considerably summarized, and therefore meaningful information can be extracted with a low-cost process. The reason is that authors of textbooks have previously analyzed the domain and decided how to organize the content according to pedagogical principles.

Provided that the organization of the textbook is reflected in the outline, NLP techniques and a collection of heuristics are used to infer the implicit pedagogical relationships.

The outline analysis is composed of two phases: Basic analysis. In this task the main topics of the domain and the relationships among these topics are mined from the homogenized 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 sub items.

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 both categorize the relationships identified in the previous step and also mine new ones, mainly prerequisite relationships. The identified relationships are labeled with the inferred kinds, the heuristic used, and the confidence on the inferred information.

The heuristics entail the condition to be matched, the empirically gathered confidence on the heuristic (percent of times it has correctly identified a relationship) and the post condition, i.e., the relationships that are recognized.

The heuristic analysis is carried out in two steps: first, the heuristics for identifying structural relationships are applied to elicit is A and part of relationships. Then, the heuristics for sequential relationships are used to identify next and prerequisite relationships.

The heuristic analysis works on the assumption that each outline item reflects a unique domain topic. However, there are some common items such as “Introduction,” “Summary,” or “Conclusions” that can appear more than once in the outline. To solve these situations, a graphical tool was developed to allow the Domain Module authors.

To modify the outline so that it conforms with the assumptions of the heuristic analysis. A set of heuristics must be defined to perform this analysis. Some of the heuristics are language dependent, as they rely on syntactic structures that may vary depending on the language they are defined for. As mentioned above, this work has been applied on documents in the Basque language. The set of heuristics was identified on a set of 150 outlines of different subjects at the University of the Basque Country (UPV/EHU). Their study allowed the identification of the set of heuristics and their confidence level .

The following procedure was carried out to identify the heuristics:

1. A small set of outlines related to computer science was analyzed to detect some patterns that might help in the classification of relationships.
2. These heuristics were tested on a wide set of outlines related to different domains.
3. The relationships identified by the heuristics were contrasted with the real ones, i.e., manually labeled relationships.
4. After analyzing the results, paying special attention to the detected lacks in the heuristics, some new heuristics were defined.
5. The final set of heuristics was tested on a set of 150 outlines of different subjects of different domains (e.g., economics, psychology, or computer engineering) taught at the University of the Basque Country obtaining good results (98.36 percent precision and 98.15 percent recall).

4.1.1 Heuristics for Structural Relationships

The heuristics for structural relationships allow identifying the kind of relationship between an item of the outline 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 analyzed outlines. The analysis of the outlines also showed that the most common structural relation is the part of relationship. Therefore, by default, the structural relationships are labeled as part of. In addition, some homogeneous structures were observed in the identified isA relationships.

A set of group heuristics, i.e., heuristics that check if the outline item or all its subitems meet a particular condition (see section Group Structural Heuristics), that allow recognizing such isA relationships were defined. Individual heuristics, i.e., heuristics that check whether a particular subitem meets a condition, were defined for identifying structural relationships in outline items with heterogeneous subitems (see section Individual structural heuristics).

The following process is applied for the identification of the structural relationships:

1) If a group heuristic triggers, then isA relationships are defined between the outline item and all its sub items.

2) Otherwise, for every subtitem an individual heuristic that matches is looked 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 is A relationship using.

Individual structural heuristics. These heuristics check if an individual subitem meets a condition. The condition may also involve the general item. The empirical analysis showed that different heuristics of this kind can trigger together in the same group of subitems. Table 1 shows some outline fragments in which the heuristics for recognizing the isA relationships can be applied.

1. Multiword heuristic (MWH). Multiword terms contain information to infer the isA relation. 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. This pattern can be found in several ways: noun þ adjective, noun þ noun phrase, and so on. 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.

2. Entity name heuristic (ENH).

Entity names are used to identify examples of a particular entity. When the sub items contain entity names, the relation between the item and the subitems can be considered as the is A relationship. In Table 1, Palm Os corresponds to an entity name, which is a particular instance of Operative Systems of Personal Digital Assistants.

3. **Acronyms heuristic (AH).**

The authors also use acronyms to refer to domain topics which have long and frequently used names. When the sub items contain only acronyms, the structural relation is likely to be the isA relation. In Table 1, the XUL acronym represents the names of some languages for designing graphical interfaces, thus, there is an implicit I sA relation between the item and its subitems.

4. Head of the phrase þ multiword heuristic (He-MWH).

This heuristic checks if the subitem of an outline item forms a multiword term from the head of the phrase of the outline item. Table 1 shows an application example, as agente (agent) is used to refer to Agente laguntzaileak (Auxiliar Agents) in a specific context, and therefore the isA relationship can be inferred.

5. Acronyms þ multiword heuristic (AþMWH).

This heuristic allows identifying isA relationships if the sub items form a multiword from the acronym of the upper level outline item.

6. 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 relation between the general item and its subitems is reinforced by this heuristic (see Table 2).

Group structural heuristics. While individual structural heuristics test if an outline item and a particular subitem match a certain condition to determine the kind of structural relationship, the 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:

1. Keyword heuristic (KH). This heuristic relies on a set of keywords—for example, adibideak (examples), elementuak (elements), motak (kinds), or tresnak (tools)—to identify isA relationships 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 is A relationship is defined between the outline item and all its subitems. The set of keywords was identified on the set of outlines analyzed to define the heuristics. However, they are stored in a configuration file and can be easily modified if new keywords are found.

2. Common head þ multiword heuristic (CHeþMWH).

This heuristic checks whether all the subitems of an outline item share a homogeneous structure. In particular, this heuristic examines if all the subitems form a multiword term and share the head of the noun phrase (see Table 3). This heuristic also identifies isA relationships among the outline item and all its subitems.

4.1.2 Heuristics for Sequential Relationships

Two kinds of sequential relationships can be found. The next relation, which states that a learning topic should be learned just after another one, appears between items at the same level, i.e., subitems of the same general item. A prerequisite relation between two domain topics states that a domain topic must be mastered before trying to learn the other topic. 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. Therefore, by default, any sequential relation is translated into next.

All the heuristics for sequential relationships check if aparticular outline item and a previous one meet a condition to recognize a sequential relationship. 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, not necessarily at the same level Head of the phraseþReference heuristic (HeþRH). References to the head of a previous outline item allow recognizing prerequisite relationships.

Acronym þ reference heuristic (AþRH).

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.possessive genitive based on the head of the previous outline item at the same nesting level.

Acronym þ possessive genitive heuristic (AþPGH2).

This heuristic is triggered by outline items formed by apossessive genitive based on the acronym of the previous outline item at the same nesting level.

4.2 Document Body Analysis

In this stage, the LDO is enhanced with new topics andrelationships gathered from the document body. 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 (see 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 document body is analyzed to get such new topics. In the last few years, the use of hybrid methods that combine NLP techniques and statistical methods has prevailed in term extraction. Many approaches use a set of patterns such as 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 those that are most appropriate .

In DOM-Sortze, term extraction is carried out using Erauzterm , 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 the domain relatedness.

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 recognize 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 topics 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 relating 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 analyzing and disambiguation systems, has been used to develop and apply the grammar on the documents.

The grammar for identifying pedagogical relationships entails rules for recognizing structural relationships—isA and partOf—and the prerequisite sequential relationship.

The rules have been defined after an empirical analysis of a set of textbooks corresponding to primary school. 13 rules for the isA relationships, 6 for partOf, and 1 for prerequisite are defined in the grammar.

These rules include, among others, the equivalent for Hearst's patterns for the Basque language. Hearst's patterns have been used in many ontology learning approaches to identify taxonomic relationships (hypernyms/hyponyms) from syntactic structures like NP0 such as NP1, NP2,. . . (and|or) NPn, where NPi refers to a noun phrase corresponding to a term or topic.

Table 5 shows two examples in which isA relationships can be identified when two ontology topics (@Topic) are related through the izan (to be) verb.

When two domain topics are related by the izeneko (referred to as) expression, an isA relationship between those topics can be inferred which identifies part of relationships. 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, part of relationships can be defined among unibertsoa (universe) and galaxiak (galaxies).

5 GATHERING LOS FROM DOCUMENTS

The generation of LOs for the domain topics is achieved by identifying and gathering DRs, i.e., consistent fragments of the document related to one or more topics with a particular educational purpose. The identification and extraction of these pieces is carried out in an ontology-driven process that also uses NLP techniques.

As the LO generating approach presented in this work aims to be domainindependent, the only domain-specific knowledge used is the LDO that has been gathered from the electronic textbook in the previous phase.

From now on, a DR will refer to a piece of the document meant to be used in the learningsessions (e.g., definition, exercise, . . .) while a LO refers to a reusable DR enriched with metadata.

The LO generation process here described is carried out by ErauzOnt [23], which is part of the DOM-Sortze framework.

Fig. 3 describes the process for gathering the LOs from the electronic document, which entails the following tasks: generating DRs from the document, annotating the DRs to

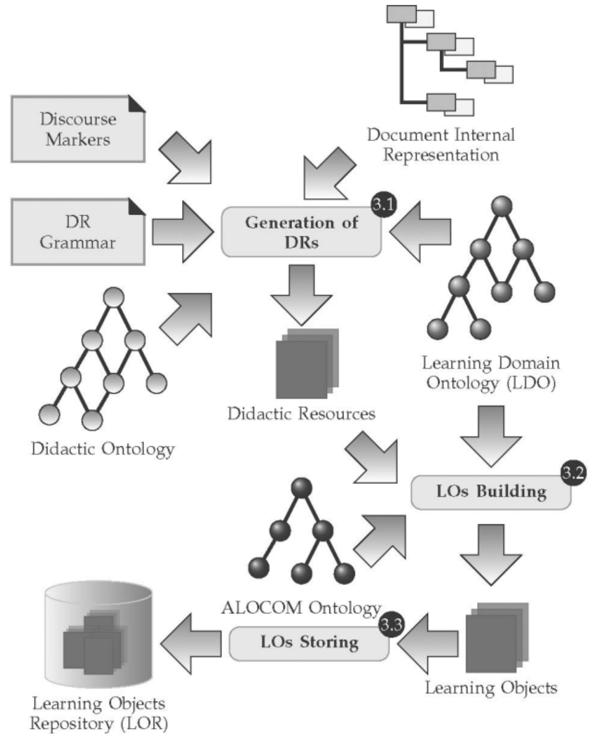


Fig.3 Generation of Los documents

become LOs, and, finally, storing the generated LOs in a LOR for further use.

The LDO, a DR grammar, discourse markers and a didactic ontology [24] are used to gather DRs from the internal representation of the electronic textbook with the part-of-speech information. DR Generation is described in depth in Section 5.1.

The LDO, and the ALOCOM ontology [25] are used to build the LOs fro the gathered DRs (see Section 5.2), and, finally, the LOs are stored in the LOR to facilitate their use and reuse (see Section 5.3).

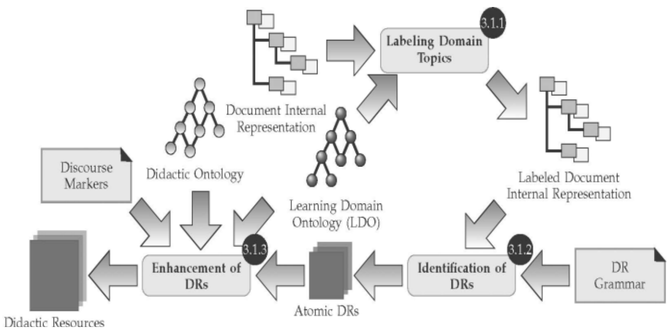
5.1 Generation of the DRs.

The identification of the DRs is carried out by identifying relevant text fragments that correspond to definitions, examples, facts, theories, principle statements, and problem statements for the LDO topics, as shown in Fig. 4.

First, the appearances of the LDO topics are labeled in the document internal representation with the part-of-speech information.Next, the DR grammar is used to find text fragments that might contain appropriate resources. The DR grammar includes a set of rules that define the different patterns or syntactic structures that have been found in a textbook for primary school and have been tested on a set of textbooks for primary school.

These patterns are the most common syntactic structures observed in several topic definitions, and so on. Similar patterns are used for English 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 a low precision. The precision of the grammarrules is used to determine the confidence on these rules. The identified DRs contain the sentence that triggered the rule for the corresponding DR and all the sentences that follow which refer to the same topic(s). Every DR is labeled with the domain topics referred and with the rules of



The DR that identified it. This information is used later in the LO annotation process.

Given that the DRs identified by the DR grammar are usually quite simple, they are enhanced in two ways to make them more accurate. On the one hand, consecutive DRs are combined if they are similer, to which end similarity measures have been defined. shows two consecutive atomic definitions that might be combined to get a more comprehensive DR, as the second definition reinforces the description provided by the first one.

On the other hand, and to keep the cohesion of the DRs, previous fragments are added to each DR if it contains references to previous DRs or sentences. 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 similar DRs) and the similarity rates.

Besides, the referred topics and the DR grammar rules used to identify the DR are also kept in every DR.

CONCLUSION AND FUTURE WORK

This paper has presented DOM-Sortze, a system for the semiautomatic generation of the Domain Module from electronic textbooks. The system employs NLP techniques, heuristic reasoning, and ontologies for the knowledge acquisition processes.

DOM-Sortze has been tested using an electronic textbook and comparing the automatically generated elements with the Domain Module manually developed by instructional designers. The aim was to evaluate how DOM-Sortze contributes to Domain Module authoring.

The electronic document used for the experiment was one of the books, written in the Basque language, used in the Nature Sciences subject in the first course of mandatory secondary education. 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 DOM-Sortze, the Domain Module entails an LDO, that contains the main domain topics and the pedagogical relationships among the topics, and the learning objects(LOs) that are used to enable mastering each domain topic. 87.95 percent of the domain topics and 40.74 percent of the pedagogical relationships were automatically gathered with 17.48 and 72.50 percent precision, respectively.

DOM-Sortze was able to automatically gather 77.27 percent of the LOsfor the Domain Module with 84.50 percent precision. The automatically gathered LDO can be supervised and enhanced using Elkar-DOM [4].

Currently, DOM-Sortze is being enhanced to support new languages such as English. In fact, the acquisition of LOs has already been adapted and tested on a textbook on Object Oriented Programming obtaining similar results to those presented in this work.

The analyzed book has 67 pages and 29,300 words. The DR grammar for English achieved 80.09

percent accuracy versus the 70 percent achieved for Basque. The biggest differences were observed in the rules for definitions and problem-statements.

The rules for definitions performed better in English, probably because the sentences employed in the book are shorter, and therefore less complex. The performance on problem-statements was worse for English. The identification of problem-statements in Basque is facilitated by an auxiliary verb used for imperative cases.

However, identifying imperative cases in English is harder. 75.93 percent of the LOs were gathered with 86.79 percent precision. Future work on DOM-Sortze comprises improving the generation of the LDO. It is planned to enhance the grammar for identifying pedagogical relationships to increase the recall of the relationships. Alternative ways to gather prerequisite relationships, which have a very poor recall, will be also tested. Besides, attributes of the domain topics, such as the domain relevance or the difficulty, which might be estimated using term hood measures (e.g., TFIDF are aimed to be automatically gathered.

Although DOM-Sortze is currently able to process images in the electronic document, it only considers their position in the text, and not where the image is referenced, and therefore useful. Thus, the treatment of images must be improved.

DOM-Sortze is being enhanced to support multilingual Domain Module generation. The LDO ontology supports the multilingual representation of the domain topics, and machine translation might be used to get approximate translations of the gathered LOs, used for searching and retrieving from the LOR or webpages.

Besides, machine learning methods are planned to be used to infer new rules that might improve the identification of pedagogical relationships or the DRs in the electronic textbooks.

ACKNOWLEDGEMENTS

This work was supported by the Spanish Ministry of Education (TIN2009-14380), the Basque Government (IT421-10), and the University of the Basque Country, under grant UFI11/45

REFERENCES

[1] B. Parsad and L. Lewis, “Distance Education at Degree-Granting Postsecondary Institutions: 2006-07,” technical report, Nat’l Center for Education Statistics, Inst. of Education Sciences, US Department of Education, 2008.

[2] P.-S.D. Chen, A.D. Lambert, and K.R. Guidry, “Engaging Online Learners: The Impact of Web-Based Learning Technology on College Student Engagement,” Computers and Education, vol. 54, no. 4, pp. 1222-1232, May 2010.

[3] M. Larran˜ aga, A. Conde, I. Calvo, A. Arruarte, and J.A.“Evaluating the Automatic Extraction of Learning Objects fromElectronic Textbooks Using Erauzont,” Proc. 11th Int’l Conf.Intelligent Tutoring Systems (ITS ’12), pp. 655-656, 2012.

[4] K. Verbert, “An Architecture and Framework for Flexible Reuseof Learning Object Components,” PhD dissertation, Faculteit Ingenieurswetenschappen, Katholieke Univ. Leuven, Feb. 2008.

[5] B. Liu, C.W. Chin, and H.T. Ng, “Mining Topic-Specific Conceptsand Definitions on the Web,” Proc. 12th Int’l Conf. World Wide Web (WWW), pp. 251-260, 2003.

[6] E. Agirre, O.L. de Lacalle, and A. Soroa, “Knowledge-BasedWSD and Specific Domains: Performing Better Than Generic Supervised WSD,” Proc. 21st Int’l Joint Conf. Artifical Intelligence(IJCAI ’09), pp. 1501-1506, 2009.

[7] T. Hughes and D. Ramage, “Lexical Semantic Relatedness withRandom Graph Walks,” Proc. EMNLP-CONLL-2007, pp. 581-589,2007.