

Building Learning Objects from Electronic Documents

Mikel LARRAÑAGA, Jon A. ELORRIAGA, Ana ARRUARTE

Department of Languages and Information Systems,

University of the Basque Country

mikel.larranaga@ehu.es

Abstract: Knowledge reuse has become a main research matter in the Education Technology area. Learning Objects provide a means to promote knowledge reuse, but their success is conditioned by the usefulness of their associated metadata. The annotation of metadata is a hard task that should be eased. This paper presents an approach to semi automatically generate and annotate Learning Objects from electronic documents.

Keywords: Computer Supported Learning Systems, Semi Automatic Domain Acquisition, Didactic Resources, Learning Objects, Metadata, Ontologies

Introduction

Last years one important focus of research in the development of Computer Supported Learning Systems (CSLSs) has been *knowledge reuse*. Learning Objects (LOs) provide a means to facilitate knowledge reuse; LOs are defined as “*reusable pieces of educational material intended to be strung together to form larger educational units such as activities, lessons or whole courses*” [1]. Sharing and reusing LOs enables more effective and efficient creation of learning material [2] due to their potential generativity, adaptability and scalability [3].

A crucial issue in order to promote the use and reuse of LOs is the possibility to retrieve the appropriate LO from a large set. This selection is highly influenced by the appropriateness of the associated metadata. Metadata generation tools are, therefore, essential in order to promote LO sharing and reusing. Since part of the information described in the metadata can be gathered from the own LOs, semi automatic metadata generators should be developed [4].

This paper focuses on the generation of new LOs, including the corresponding metadata, from electronic documents. The work here presented is part of a wider project which main goal is the semi automatic generation of the Domain Module for CSLSs from electronic documents. The paper starts with a brief description of LOM metadata elements and a classification of metadata elements considering the way the metadata can be generated. Following, the LO generation process is described. Finally some conclusions are pointed out.

1. Metadata and Classification of Metadata Elements

The metadata of a LO contains information including details about its content and usage. Thus, the LO annotation is a life-long process in which the information is gathered from the content of the DR, the context (e.g., information from the LO repository), the structure and the usage. Initial metadata can be gathered from the content of the DR. LOM [5] specifies

77 elements that can be used to describe a LO. Next, these elements are classified considering two aspects: (1) how the values of the elements can be propagated from an aggregate LO to its components and vice versa, and (2) which values can be automatically gathered, which ones require to be supervised and which have to be manually specified.

1.1 Classification of the Metadata Elements Considering Information Propagation

When generating the metadata for a LO, mainly if the LO is an aggregation of other LOs, the information of some elements can be propagated from the components to the aggregated LO. According to Cardinaels [6], metadata elements can be classified in the following five types considering information propagation among the aggregated LO and its components:

- **Not-propagated metadata:** Elements which values are not propagated from the aggregate to the components or vice versa. These elements are listed in Table 1.

Table 1. Not-propagated metadata elements.

1.1 General - Identifier	4.3 Technical – Location
1.2 General – Title	5.3 Educational – Interactivity Level
2.1 Life Cycle – Version	5.4 Educational – Semantic Density
2.2. Life Cycle – Status	5.8 Educational – Difficulty
2.3 Life Cycle – Contribute	5.9 Educational – Typical Learning Time
3.1 Meta Metadata – Identifier	8 Annotation
3.2 Meta Metadata – Contribute	

- **Accumulated metadata:** Elements which values on the aggregate are the union of the values of the components (Table 2).

Table 2. Accumulated metadata elements.

1.3 General – Language	4.5 Technical – Installation Remarks
1.4 General – Description	4.6 Technical – Other Platform Requirements
1.5 General – Keyword	5.2 Educational – Learning Resource Type
1.6 General – Coverage	5.10 Educational – Description
4.1 Technical – Format	7 Relation
4.4 Technical – Requirement	9 Classification

- **Additive metadata:** Elements for which the values on the aggregate are the addition of the values on the components (Table 3).

Table 3. Additive metadata elements.

4.2 Technical – Size	4.7 Technical - Duration
----------------------	--------------------------

- **Inherited metadata:** Elements which values propagate from the aggregate to the components (Table 4).

Table 4. Inherited metadata elements.

3.3 Meta-metadata – Metadata Schema	5.7 Educational – Typical Age Range
3.4 Meta-metadata – Language	5.11 Educational – Language
5.1 Educational – Interactivity Type	6.1 Rights – Cost
5.5 Educational – Intended End User Role	6.2 Rights – Copyright
5.6 Educational – Context	6.3 Rights - Description

- **Special metadata:** Elements for which the values of the components is less than or equal to the value of the aggregate (Table 5). E.g., the *aggregation level* of the composite LO is always greater than the *aggregation level* of its components.

Table 5. Special Metadata Elements.

1.7 General – Structure	1.8 General – Aggregation Level
-------------------------	---------------------------------

This classification is essential to determine how the values can be gathered and even propagated from the components to the aggregate or vice versa and, thus, facilitate the automatic generation of metadata. Next, the above elements are classified considering which ones can be automatically set, which ones require human supervision and which must be completely set by human.

1.2 Classification of the Metadata Elements Considering the Level of Human Intervention

Some metadata can be automatically generated from the LO itself (e.g. *Keyword*), i.e., its content, the LO repositories it is stored in or even the Learning Management System that uses the LO [7]. Other metadata elements (e.g., *annotation*) relate to the use of the LO while others describe pedagogical aspects. This paper focuses on the generation of LOs from electronic documents and, therefore, the information that can be gathered from the document, as well as the information that must be provided by the instructional designer, has been identified. The metadata elements can be divided into the following groups:

- **Automatic gathered metadata elements:** Information that can be acquired automatically (see Table 6). Most of these elements (e.g., *1.3 General - Language*, *1.5 General - Keyword*) can be automatically gathered using NLP or text mining techniques. Other elements (e.g., *4.5 Technical - Installation Remarks*) have fixed values; in the approach here presented, all the generated LOs are gathered from electronic documents and, therefore, have similar characteristics.

Table 6. Automatic gathered metadata elements.

1.3 General – Language	4.5 Technical – Installation Remarks
1.5 General – Keyword	4.6 Technical – Other Platform Requirements
1.7 General – Structure	5.1 Educational – Interactivity Type
1.8 General – Aggregation Level	5.2 Educational – Learning Resource Type
3 Metadata	5.3 Educational – Interactivity Level
4.1 Technical – Format	5.5 Educational – Intended End User Role
4.2 Technical – Size	5.11 Educational – Language
4.3 Technical – Location	9.1 Classification – Purpose
4.4 Technical – Requirement	9.4 Classification – Keyword

- **Supervised metadata elements:** Information that may require the supervision of the instructional designer. The values of these elements (Table 7) may be inferred or generated automatically, but require the supervision of the instructional designer.

Table 7. Supervised metadata elements.

1.1 General – Identifier	7.1 Relation – Kind
1.2 General - Title	7.2.1 Relation – Resource – Identifier
2 Life Cycle	9.2 Classification – Taxon Path
5.4 Educational – Semantic Density	

- **Manually specified metadata elements:** Information that must be completely filled by the instructional designer (Table 8).

Table 8. Manually specified metadata elements.

1.4 General – Description	5.9 Educational – Typical Learning Time
1.6 General - Coverage	5.10 Educational – Description
4.7 Technical – Duration	6. Rights
5.6 Educational – Context	9.1 Classification – Purpose
5.7 Educational – Typical Age Range	9.3 Classification - Description
5.8 Educational – Difficulty	

As it can be observed in the above tables, most of the metadata elements can be gathered automatically or semi automatically. Only the metadata elements listed in Table 8 have to be specified manually. However, all this manually specified elements are *inherited metadata elements* (i.e., the information is propagated from the whole LO to its components) or *accumulated metadata element* (i.e., the information of the whole LO is the union of the information of its components). This can be exploited in order to minimize the information that the user has to provide while annotating the LO. The 11 elements listed in Table 8 can be asked to the user once and later use them to set the initial values for all the generated LOs.

2. Semi Automatic Generation of LOs from Electronic Documents

The work presented in this paper is part of a project that aims to facilitate the development of the Domain Module for CSLSs from documents. The only domain knowledge used in the process is the information contained by the analysed document, so the approach is domain independent. The process of semi automatically acquiring the Domain Module from electronic documents entails the acquisition of the domain ontology and the generation of didactic material (i.e., LOs for every topic in the domain) [8]. In the approach here presented, the LOs are built by identifying the text fragments that correspond to DRs (e.g., definitions, exercises, etc.) on the documents. The following procedure is carried out:

1. **Manual Metadata Specification:** The values for the metadata elements that have to be specified by the user are asked in a set of form-based windows. The specified values will be stored and used later while annotating the generated LOs. As all these elements are *inherited metadata elements* or *accumulated metadata elements*, the information will be used for all the generated LOs.
2. **Domain Ontology Building:** The Domain Ontology contains the domain topics and the pedagogical relationships among them. It is gathered from the electronic document using Natural Language Processing (NLP) techniques and Heuristic Reasoning [9]. The instructional designer can supervise and modify the obtained domain ontology with a graphic concept map-based tool, Elkar-Dom [10].
3. **Generation of LOs:** This is an ontology-driven process. The domain ontology, which has been gathered in the previous step, a didactic ontology [11; 12] and a grammar – which defines the most frequently used syntactic structures – are used to identify, classify and annotate fragments that correspond to DRs [13]. The information obtained in this phase is combined with the information provided in the first step to annotate the LOs before storing them in the LO repository (LOR).
4. **Integration of the Domain Ontology and LOs to Build the Domain Module:** Finally, every topic of the domain is linked with its corresponding LOs. The resulting Domain Module is presented to the user to be supervised using Elkar-Dom. The user can modify it and, even, add new LOs retrieving them from the LOR.

3. Conclusions

Knowledge reuse has become one of the main concerns on the research agenda of the Educational Technology area. Learning Objects (LOs) are reusable pieces of didactic material that provide a means to facilitate knowledge reuse on the development of Computer Supported Learning Systems (CSLSs). The easiness to retrieve the appropriate LOs from LO repositories (LORs) is essential to promote the use and reuse of LOs. The accurate retrieval strongly depends on the quality of the metadata that describes the LO.

In this paper a method for semi automatically generating LOs from documents has been described. The method relies on the use of ontologies and NLP techniques. In order to automate the annotation of LOs, the metadata elements have been analysed and classified considering information generation and propagation issues. The conclusion is that most metadata elements can be filled automatically or semiautomatically from the original document. Only a few metadata elements must be manually specified, and in most cases, this information can be propagated to other LOs generated from the electronic document.

Currently, the whole LO generation process is being carried out with Basque language written textbooks in the domain of Nature Sciences provided by the Gipuzkoako Ikastolen Elkartea (GIE – <http://www.ikastola.net/>). GIE is one of the main publishers for Basque-medium education at all educational levels and in all subjects in compulsory education at the Basque Country. Textbooks authors are currently being trained in the use of Elkar-Dom in order to facilitate the evaluation and supervision of the automatically obtained LOs.

Acknowledgements

This work is supported by the Univ. of the Basque Country (UE06/19), the MEC (TIN2006-14968-C02-01) and the Gipuzkoa Council in an EU Program. Also, we want to thank the collaboration of the Gipuzkoako Ikastolen Elkartea.

References

- [1] Brooks, C., Cooke, J., & Vassileva, J. (2003). Versioning of learning objects. Proceedings of 3rd IEEE International Conference on Advanced Learning Technologies, Athens, Greece.
- [2] Wiley, D. A. (2000). Learning Object Design and Sequencing Theory. Brigham Young University.
- [3] Gibbons, A. S., Nelson, J., & Richards, R. (2000). The Nature and Origin of Instructional Objects. In D. A. Wiley (Ed.), *The Instructional Use of Learning Objects*.
- [4] Duval, E., & Hodgins, H. W. (2003). A LOM Research Agenda. Proceedings of Twelfth International Conference on World Wide Web.
- [5] LTSC. (2001). IEEE P1484.12. from Learning Object Metadata Working Group homepage [On-line]. <http://ltsc.ieee.org/wg12/>
- [6] Cardinaels, K. (2007). A Dynamic Learning Object Life Cycle and its Implications for Automatic Metadata Generation. Faculteit Ingenieurswetenschappen, Katholieke Universiteit Leuven.
- [7] Cardinaels, K., Meire, M., & Duval, E. (2005). Automating Metadata Generation: the Simple Indexing Interface. Proceedings of 14th International Conference on World Wide Web (WWW2005), Chibo, Japan.
- [8] Larrañaga, M. (2002). Enhancing ITS building process with semi-automatic domain acquisition using ontologies and NLP techniques. Proceedings of Young Researches track of the Intelligent Tutoring Systems (ITS 2002), Biarritz (France).
- [9] Larrañaga, M., Rueda, U., Elorriaga, J. A., & Arruarte, A. (2004). Acquisition of the Domain Structure from Document Indexes Using Heuristic Reasoning. Proceedings of 7th International Conference on Intelligent Tutoring Systems (ITS 2004), Maceió, Alagoas, Brazil.
- [10] Larrañaga, M., Niebla, I., Rueda, U., Elorriaga, J. A., & Arruarte, A. (2007). Towards Collaborative Domain Module Authoring. Proceedings of IEEE International Conference on Advanced Learning Technologies 2007. (ICALT 2007).
- [11] Leidig, T. (2001). L3–Towards an Open Learning Environment. *ACM Journal of Educational Resources in Computing*, 1(1), 5-11.
- [12] Meder, N. (2000). Didaktische ontologien. Proceedings of Globalisierung und Wissensorganisation: Neue Aspekte für Wissen, Wissenschaft und Informationssysteme.
- [13] Larrañaga, M., Elorriaga, J. A., & Arruarte, A. (2008). Semi Automatic Generation of Didactic Resources from Existing Documents. Proceedings of 9th International Conference on Intelligent Tutoring Systems (ITS 2008), Montreal, Canada.