Ontology-based Learning Content Repurposing: the ALOCoM Framework

Katrien Verbert[1], Jelena Jovanović[2], Erik Duval[1], Dragan Gašević[2] , Michael Meire[1]

[1]Dept. Computerwetenschappen, Katholieke Universiteit Leuven
Celestijnenlaan 200A, B-3001 Leuven, Belgium
{Katrien.Verbert, Erik.Duval, Michael.Meire}@cs.kuleuven.ac.be

[2]FON-School of Business Administration, University of Belgrade
POB 52, Jove Ilica 154, Belgrade, Serbia and Montenegro
jeljov@fon.bg.ac.yu, gasevic@yahoo.com

Abstract: This paper reports on the development of a framework that enables on-the-fly composition of learning object components, more specifically components of slide presentations. In earlier work, we developed an ontology that explicitly defines the structure of learning objects, formally specifying learning object component types and relationships between these components. In this paper, we present a framework that disaggregates slide presentations into this ontology format and reassembles their components (definitions, references …) into new slide presentations.

Introduction

Issues concerning learning object (LO) re-use and repurposing are currently among the most important research topics in the learning technology community (Duval & Hodgins 2003). In many cases, we need to repurpose specific parts of a LO instead of the LO in its entirety. In that case, a definition, example or illustration is repurposed by copy and paste in new and different LOs. This approach is non-scalable in terms of maintenance, since each time a component is copied, a new place is created that needs to be maintained. It is possible to re-use learning objects in a much more sophisticated way if we can access the components of a learning object and repurpose them on-the-fly. This requires a more innovative and flexible underlying model for learning object components (Duval & Hodgins 2003). In earlier work, we developed an ontology that is an abstract learning object content model (ALOCoM), defining a framework for learning objects and their components (Verbert & Duval 2004). The ontology provides an explicit definition of the LO content structure, formally specifying both LO component types and relationships between those components. In this paper, we propose a framework that uses the developed ontology for composing and decomposing slide presentations.

In the next section, we briefly outline the ALOCoM ontology. In section 3, we present the transformation framework and section 4 illustrates a scenario applying the framework. Conclusions and remarks on future work conclude this paper.

The ALOCoM Ontology

In earlier work, we developed the ALOCoM ontology as a generic abstract learning object content model for learning objects and their components (Verbert & Duval 2004). The ontology distinguishes between content fragments (CFs), content objects (COs) and learning objects (LOs). CFs are learning content elements in their most basic form, like text, audio and video. These elements can be regarded as raw digital resources and are uncombined with other elements. COs aggregate CFs and add navigation. Navigation elements enable structuring of content fragments in a content object. Besides CFs, COs also include other COs. Finally, LOs aggregate COs and add a learning objective. Further, we defined content types for each of these components. We introduced CF types such as image, text, audio and video. For defining CO types, we investigated existing Information Architectures, like the Information Block Architecture (Horn 1998) developed by Dr. Horn and the IBM Darwin Information Typing

Architecture (Priestley 2001). These architectures define information types (e.g. concept, principle, task) and their building blocks (e.g. example, definition, analogy). As a starting point, we defined the CO types and their structure using DITA concepts, since DITA is a recent architecture with rich documentation and online support (Priestley 2001). Besides CF and CO types, the ontology identifies LO types. For now, only a slide presentation LO type is defined. Finally, the ontology defines the relationships between the LO components. Aggregation and navigation relations are specified. Aggregation relationships between components are represented in the form of a “hasPart” and its inverse “isPartOf” ontology properties. Each composite content unit (i.e. CO and LO) can have an arbitrary number of “hasPart” properties relating the content unit to its components. Navigational relationships were initially specified in the form of “previous” and “next” properties. However, that solution is inappropriate since it mixes content and navigation and hence degrades the level of content reusability. Therefore, we provide an alternative solution in the form of an ordering property attached to each CO and LO. The range of this property is a list that defines the order of components in a composite content unit (i.e. CO or LO). A composite content unit can have an arbitrary number of ordering properties, each one defining a specific learning path. For more information about the ontology, see (Jovanović et al. 2005)

The ALOCoM Framework

Our main focus is on the development of a tool for extracting/transforming LO content into ontology-aware content (we call these tools disaggregators) as well as for repurposing ontology-aware content in real-world applications (aggregators). The proposed ALOCoM framework supports both the process of aggregating and the process of disaggregating LOs. The framework maps different tool specific formats into ALOCoM ontology-aware content and vice versa, ontology aware content into tool specific content. For now, the framework supports slide presentations. Figure 2 illustrates the framework. The colors of the arrows symbolically differentiate transformations that we have implemented so far (green) from those that we need to develop to make the framework fully operational (red). For example, we have developed a transformation from OpenOffice.org slides into a SCORM Content Package (hence green arrow ending), but the reverse conversion is not yet supported (red color on the opposite side).

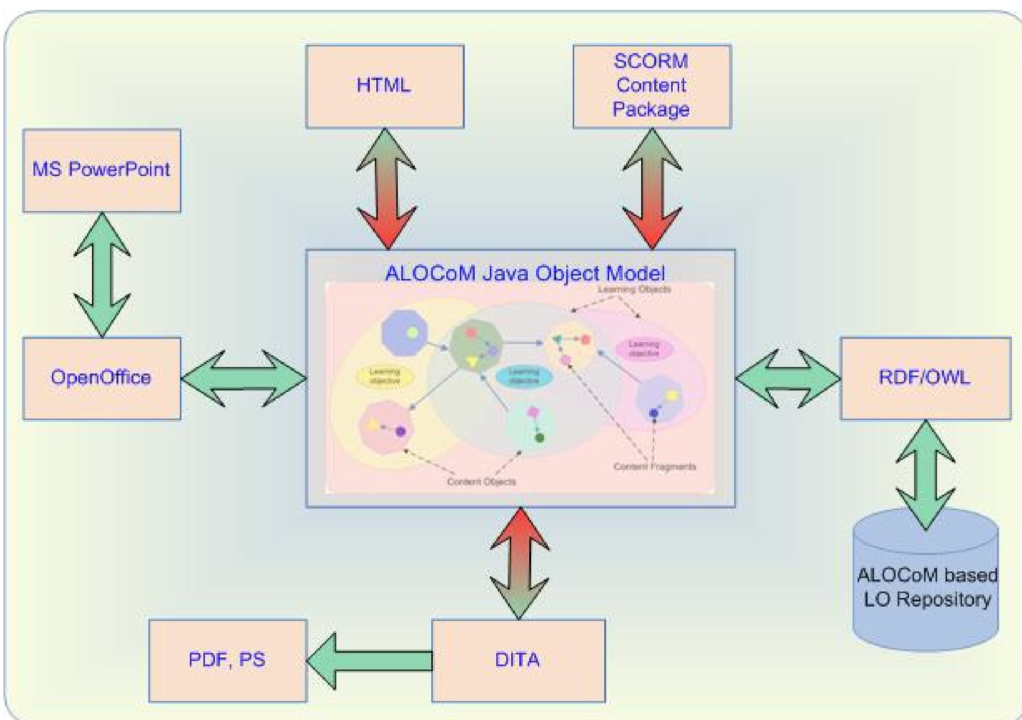

Figure 1: The ALOCoM framework

Since the most popular tools for slide presentation authoring are MS PowerPoint and OpenOffice.org, the proposed framework focuses for now on slides presentations authored using these tools. The process of content disaggregation performed inside the framework has an OpenOffice.org (OO) slide presentation as its input. If a slide presentation is in the MS PowerPoint format, it is first transformed into the OO format. A slide presentation in the OO format is parsed, decomposed into its components (according to the ALOCoM ontology) and transformed into an ALOCoM Java Object Model. In the next step, we use Jena – a HP Java-based framework for the Semantic Web (http://jena.sourceforge.net/) - to generate RDF/OWL instances out of the ALOCoM model. These instances are stored in an ALOCoM compliant LO Repository (LOR).

We use the RDQL query language to search the database and retrieve components of a certain type (e.g. definition, example). These components are then reassembled to create a new LO. Export functions to OpenOffice.org, SCORM, HTML, PDF and PS are provided as illustrated in figure 1. Finally, the generated OpenOffice.org slide presentation can be exported to the MS PowerPoint format.

For more information about transforming OO presentations into ALOCoM, we refer to (Verbert et al. 2005). The rest of this section elaborates on the ALOCoM LOR and the generation of new LOs out of components.

Ontology-based LOR

The ALOCoM framework uses Jena to create RDF/OWL instances. Jena offers persistent storage models, which are continually and transparently persisted to a backing store. Persistent models can be maintained in the file system, or in a relational database. The database engines currently supported are PostgreSQL, Oracle, and MySQL (McCarty 2004). Currently, we have set up the ontology-based LOR using MySQL server.

We query the LOR using RDQL, a query language for RDF. While not yet a formal standard, RDQL is widely implemented by RDF frameworks. RDQL allows complex queries to be expressed concisely, with a query engine performing the hard work of accessing the data model (McCarty 2004).

Generating new LOs out of components

LOs disaggregated in the ALOCoM format provide us with a flexible solution for repurposing LO components. LO components at different levels of granularity are available (CF, CO, LO). For instance, we can retrieve complete slide presentations at the LO level, definitions and examples at the CO level or just text fragments or images at the CF level. These components need to be reassembled in new LOs. Currently, all selected components are assembled in a new slide presentation. The framework supports MS PowerPoint, OpenOffice.org, HTML, PDF, PS and SCORM output formats. For more information about the export functions to SCORM, PDF, PS and HTML, we refer to (Verbert et al. 2005). We will illustrate the generation of a new MS PowerPoint slide presentation in the rest of this section.

Export of an ALOCoM slide presentation to the MS PowerPoint format proceeds in two steps. In the first step, the ALOCoM presentation is exported to an OpenOffice.org presentation. To generate OpenOffice.org slide presentations, we use the OpenOffice.org Application Programming Interface (API). This API is a comprehensive specification that describes the programmable features of OpenOffice.org (http://api.openoffice.org). Impress is the OO application we use to generate OO slide presentations. This application is very presentation oriented, in the sense that every represented component is a kind of shape with presentation properties. Structure related information in ALOCoM components is mapped to these presentation elements in the OO format. For instance, the title and body of an ALOCoM slide are mapped to two different rectangles in an OO slide. Since we do not keep track of presentation related information, we use default presentation styles for the title of a slide, list items and others. In the second step, the OpenOffice.org presentation is exported to a MS PowerPoint presentation, again using programming features of the OpenOffice.org API.

Scenario

We use the framework to repurpose existing LO components in new meaningful LOs. Currently, we enable uploading of both MS PowerPoint and OpenOffice.org slide presentations. All components of these presentations are available for repurposing.

A typical usage scenario of the proposed framework goes as follows: Suppose an author is creating a slide presentation on differential equations. He/she wants to start with a definition, followed by three examples. The author enters “differential equations” as keywords and selects “definition” and “example” as types of components that he/she is interested in. The system then searches the LOR and retrieves all components of the selected types dealing with the selected topic. The author chooses the most relevant components from the set prepared for him/her. Furthermore, the author wants to include a reference to a book (s)he wants to recommend and an image of the book. Again the author searches the LOR and selects the component he/she wants to repurpose from the set of retrieved components. The author is free to choose the presentation form of the generated content assembly among MS

PowerPoint, OpenOffice.org, HTML and PDF formats. The author enhances the automatically generated slide presentation with some additional information on the topic and the presentation is ready for in-class use.

Conclusions

In this paper, we presented the ALOCoM framework as a solution for repurposing LO components. The framework enables repurposing of meaningful LO components (e.g. definition, example, reference) in existing slide presentations. Furthermore, these components are automatically reassembled into new LOs and launched in the authoring tool the author is using. Next steps will extend the framework to support MS Word and OpenOffice.org Text Documents. This extension needs to incorporate some additional component types that are not used in slide presentations, for instance different levels of headings.

Currently, we can search for component types and keywords in the ALOCoM LOR. We need more metadata for each of the ALOCoM components. To that extent, we are planning to make use of a framework for automatic metadata generation (http://ariadne.cs.kuleuven.ac.be/amg), which is made available as a web service (IndexingService). The advantage of using this framework is that it contains components that generate rich metadata instances, including for instance the language of a LO (component). The result of using this framework will enable an improved search and retrieval of relevant LO components.

References

Duval, E. and Hodgins, W. (2003). "A LOM research agenda," *In Proceedings of the twelfth international conference on World Wide Web*, Budapest, Hungary, 2003, pp. 1-9.

Horn, R. E. Structured writing as a paradigm. In Instructional Development: the State of the Art. Englewood Cliffs, N.J., 1998.

Jovanović, J., Gašević, D., Verbert, K. and Duval, E. (2005). Ontology of learning object content structure (submitted)

McCarty P. (2004) Use RDF models in your Java applications with the Jena Semantic Web Framework, DeveloperWorks, http://www-128.ibm.com/developerworks/java/library/j-jena/

Priestley, M. (2001). "DITA XML: a reuse by reference architecture for technical documentation," *In Proceedings of the 19th annual international conference on Computer documentation*, Sante Fe, USA, 2001, pp. 152-156.

Verbert, K. and Duval, E. (2004). "Towards a global architecture for learning objects: a comparative analysis of learning object content models," *In Proceedings of the 16th ED-MEDIA 2004 World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Lugano, Switzerland, 2004, pp. 202-209.

Verbert, K., Jovanović, J., Gašević, D., Duval, E. and Meire, M. (2005). "Towards a global component architecture for learning objects: a slide presentation framework." (submitted).

Acknowledgements

We gratefully acknowledge the financial support of the K.U.Leuven research council through the BALO project, and of the European Commission through the ProLearn Network of excellence, which has facilitated the collaboration between the Computer Science Department of the Katholieke Universiteit Leuven and the GOOD OLD AI group at the University of Belgrade.