Ontology-based Learning Content Repurposing

Katrien Verbert
Dept. Computerwetenschappen,
Katholieke Universiteit Leuven
Celestijnenlaan 200A,
B-3001 Leuven, Belgium
Katrien.verbert@cs.kuleuven.ac.be

Dragan Gašević, Jelena Jovanović
FON-School of Business Administration,
University of Belgrade
POB 52, Jove Ilića 154, Belgrade,
Serbia and Montenegro
gasevic@yahoo.com, jeljov@fon.bg.ac.yu
http://goodoldai.org.yu

Erik Duval
Dept. Computerwetenschappen,
Katholieke Universiteit Leuven
Celestijnenlaan 200A,
B-3001 Leuven, Belgium
Erik.duval@cs.kuleuven.ac.be

ABSTRACT
This paper investigates basic research issues that need to be addressed for developing an architecture that enables repurposing of learning objects in a flexible way. Currently, there are a number of Learning Object Content Models (e.g. the SCORM Content Aggregation Model) that define learning objects and their components in a more or less precise way. However, these models do not allow repurposing of fine-grained components (sentences, images…). We developed an ontology-based solution for content repurposing. The ontology is a solid basis for an architecture that will enable on-the-fly access to learning object components and that will facilitate repurposing these components.

Categories and Subject Descriptors
K.3 [**Computing Milieux**]: Computers and Education; H.3.m [**Information Storage and Retrieval**]: Miscellaneous

General Terms: Learning objects

Keywords: Content models, metadata, ontologies, repurposing

1. INTRODUCTION
Learning objects (LOs) and their reusability are one of the most important current research topics in the learning technology community [1]. Reusability of LOs is conventionally regarded as the use of entire LOs in different contexts. The Learning Object Metadata (LOM) standard [2] provides a set of metadata elements for describing LOs: this facilitates finding relevant LOs. However, in many cases we need to reuse specific parts of a LO, rather then the LO as a whole. In such situations, current practice is to copy & paste in order to reuse specifically those parts of a document (e.g. a definition, an example or an illustration) that are relevant. However, this can be rather tedious and time-consuming. More important, such an approach is non-scalable in terms of maintenance, since each time you copy a content unit, a new place is created that needs to be maintained. Our goal is to release authors from the task of reusing parts of LOs manually, by automating that process as much as possible. Therefore, we need a LO content format that includes an explicit definition of the structure of the LO. We developed an ontology that provides an explicit definition of the LO content structure, formally specifying both LO component types and relationships between those components. Furthermore, we need tools for extracting/transforming LO content into this ontology content format (we call those tools disaggregators) as well as tools for repurposing ontology-aware content in real-world applications. This approach will enable not only repurposing of complete LOs, but also the retrieval and repurposing of relevant components.

WWW 2005, May 10-14, 2005, Chiba, Japan.
ACM 1-59593-051-5/05/0005.

In the next section, we briefly outline the ALOCoM ontology. Section 3 illustrates the role of the ontology in the process of authoring learning materials. Section 4 elaborates on tool support and conclusions and remarks on future work conclude this paper.

2. THE ALOCOM ONTOLOGY
We developed a generic content model (ALOCoM) that defines LOs and their components [7]. The model differentiates between Content Fragments (CF), Content Objects (CO), and Learning Objects (LO). CFs are content units in their most basic form, like text, audio and video. Basically, CFs are raw digital resources. They can be further specialized into discrete (graphic, text, image) and continuous (audio, video, simulation and animation) elements. COs aggregate CFs and add navigation. Navigation elements should enable proper structuring of CFs within a CO. Besides CFs, a CO can include other COs as well. At the next aggregation level, a LO is defined as a collection of COs with an associated learning objective.

We defined content types for each of these components. We introduced CF types such as images, text, audio and video. For defining CO types, we investigated existing Information Architectures, like the Information Block Architecture [4] developed by Dr. Horn and the IBM Darwin Information Typing Architecture [6]. 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 CO types and their structure using DITA concepts, since DITA is a recent architecture with rich documentation and online support [6]. Besides CF and CO types, the ontology identifies LO types. For now, only a slide presentation type is defined. Finally, the ontology defines the relationships between the LO components. Both aggregation and navigation relations are specified. For more details about the ontology, we refer to [7].

3. THE ROLE OF THE ONTOLOGY IN THE AUTHORING PROCESS
A teacher (or another kind of author) uses an authoring tool to produce learning materials for students. Since our aim is to facilitate the process of learning content authoring, we provide the teacher with the technology that enables him/her to reuse existing components deposed in LORs. In our model, the LOR should provide access not only to complete LOs, but also to smaller components, i.e. COs and CFs as they are defined in ALOCoM. In order to have LORs with ALOCoM ontology-aware content, we transform LOs into a form compliant with the ALOCoM ontology. That is why we need tools we call disaggregators. They take as input LOs in any domain/tool specific format and convert them into an ALOCoM compliant output format. Having learning

content disaggregated to the level of a single image, text fragment or audio/video clip, we enable more flexible content reuse and make the process of composing new learning materials more lightweight and less time consuming. Of course, our intent is not to engage a teacher in the process of searching for and putting together numerous content fragments. Instead he/she must have technical support that automatically searches the LOR for a particular image, example, definition or table and assembles them into a meaningful learning unit. Finally, the teacher can refine and fine-tune the learning material thus generated. Of course, the author can also add new content to the assembly of reused components. Further, this kind of scenario demands for semantic annotations of LOs (both content and metadata, for details see [3]) deposited in LORs. Those annotations should be in accordance with domain ontologies. Therefore, we argue for content structuring according to the ALOCOM ontology as well as for its semantic markup compliant with appropriate domain ontologies.

Suppose a teacher is preparing a course on the Darwin evolution theory. For such a course, (s)he needs textual descriptions, images, some summary tables and possibly some video clips. The conventional approach is to search LORs for domain relevant LOs, browse through obtained files, mark useful components (text, image etc.), copy & paste those components into the LO under construction and finally fine-tune the assembly. In the approach that we propose, the author selects the most relevant components from the automatically determined set prepared for him/her and defines their correct sequence. The author can enhance the generated materials, e.g. giving his/her own overview of the evolution process, provide students with additional readings in the form of footnotes or give an example.

4. TOOL SUPPORT

We need tools for extracting/transforming LO content into ontology-aware content as well as tools for repurposing ontology-aware content in new applications.

4.1. Creating ontology-aware content

Extracting ALOCoM aware content from the content produced by real-world authoring tools (e.g. MS PowerPoint, OpenOffice.org etc.) is a complex task since it must consider a huge variety of source formats and implementation techniques. Most tools use their own (vendor-specific) format for storing learning materials. Although the great majority of them support other formats through an export function, it is often the case that information is lost when exporting tool-specific formats (e.g. MS Word) into other formats (e.g. OpenOffice.org). Our goal is to develop a tool that maps all different tool-specific formats into the ALOCoM format. This task requires knowledge of the internal organization (i.e. structure) of the format we are mapping into the ontology-aware content. For content authoring tools that have their own API (e.g. Microsoft's .Net API for the MS Office tools), we use this API to retrieve both content and structure of their documents. For other formats like HTML and PDF, we can use Java parsers to retrieve the same information. Finally, instances of the ALOCoM ontology are created using Jena – a Java-based framework for the Semantic Web [5].

4.2. Repurposing ontology-aware content

Our goal is to create new LOs out of components in the ALOCoM format. Having LOs disaggregated into the ALOCoM format, we have access to their meaningful components like their definitions and references. These components are reassembled in new LOs. Since authors use their favorite authoring tools to prepare LOs, we need to export the newly created LOs to these tool-specific formats. We use Jena to retrieve the content of ALOCoM components. We export this content to specific formats, for content authoring tools that have their own API we use again their well-defined programming libraries.

5. CONCLUSIONS

In this paper, we have argued for the need of an architecture that enables on-the-fly repurposing of LO components. Such an architecture requires a flexible underlying model for LO components that we provide in the form of the ALOCoM ontology. Currently, the ontology fully covers slide presentations. The ontology is a solid foundation for the development of an ontology-based infrastructure for authoring learning materials, extensively grounded on repurposing of existing components.

The work reported in this paper is still in its rather early stages. We are currently working on the transformation from different tool-specific contents (e.g. OpenOffice.org slides, Word documents) to the ALOCoM-ontology compliant content and vice versa. For now, the framework transforms MS PowerPoint and OpenOffice.org slide presentations. We will add support for new LO types (e.g. reports), and we will refine the ALOCoM ontology to better distinguish between navigation, content and presentation in learning materials, using well-known principles from the Web engineering community. Further, our plans include the enrichment of the ALOCoM components with semantic markup, using concepts from appropriate domain ontologies. Finally, we are considering extending the ALOCoM ontology with additional concepts and properties that would facilitate adaptation of content elements according to the specific preferences and needs of each individual learner.

6. ACKNOWLEDGMENTS

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.

7. REFERENCES

[1] Duval, E. and Hodgins, W., "A LOM research agenda," *In Proceedings of the 12th International World Wide Web Conference.*, Budapest, Hungary, 2003, pp. 1-9.

[2] Duval, E. (2002), editor. 1484.12.1 IEEE Standard for learning Object Metadata. June 2002.

[3] Gašević, D. et al, "Enhancing Learning Object Content on the Semantic Web," *In Proceedings of the 4th IEEE International Conference on Advanced Learning Technologies*, Joensuu, Finland, 2004, pp. 714-716.

[4] Horn, R. E., "Structured writing as a paradigm", In *Instruct. Devel.: the State of the Art.* Englewood Cliffs, N.J, 1998.

[5] McBride, B., "Jena: A Semantic Web Toolkit," *IEEE Inter. Comp.*, Vol. 6, No.6, 2002, pp. 55-59.

[6] Priestley, M., "DITA XML: a reuse by reference architecture for technical documentation," *In Proceedings of the 19th International Conference on Computer Documentation.*, Sante Fe, USA, 2001, pp. 152-156.

[7] Verbert, K. et al, "Towards a Global Component Architecture for Learning Objects: an Ontology Based Approach", *In Proc. of OTM 2004 Workshop on Ontologies, Semantics and E-learning,* Agia Napa, Cyprus, 2004.