[MHRSt06] *Marek Meyer, Tomas Hildebrandt, Christoph Rensing, Ralf Steinmetz;* **Requirements and an architecture for a multimedia content re-purposing framework;** First European Conference on Technology Enhanced Learning, Kreta, Oktober 2006, S.

Requirements and an Architecture for a Multimedia Content Re-Purposing Framework

Marek Meyer[1], Tomas Hildebrandt[2], Christoph Rensing[2], and Ralf Steinmetz[2]

[1] SAP AG, SAP Research CEC Darmstadt, Bleichstr. 8, 64283, Germany,
marek.meyer@sap.com,
[2] KOM Multimedia Communications Lab, Darmstadt University of Technology,
Merckstr. 25, 64832 Darmstadt, Germany,
{hildebrandt, rensing, steinmetz}@kom.tu-darmstadt.de

Abstract. Production of Learning Resources is very expensive. Hence re-use is a common way to lower these costs. In contrary to a re-use "as is" re-purposing means to transform a Learning Resource to suit a new learning or teaching context. We consider re-purposing to be a promising approach to improve the re-use of Learning Resources. As a basis for re-purposing tools, a re-purposing framework shall provide fundamental functionality in order to facilitate the development of re-purposing tools. This paper analyzes the requirements for such a framework, proposes an actual architecture and presents the current state of the implementation.

1 Introduction

E-Learning has gained a major role eduction. Web Based Trainings (WBT) and other kinds of multimedia content are important especially for adult education. For these formats, powerful tools exist for creating Learning Resources. But once a Learning Resource is created and published, the work of these tools is considered to be finished. The challenge in re-using digital Learning Resources is the combination of content, multimedia and didactics.

From a didactical point of view, it would be best to design and produce learning materials completely from scratch, because this approach would lead to a maximum of quality. But the production of e-learning materials is very expensive. Therefore, new ways have to be found to reduce the costs of the authoring process. The common approach to lower production costs is to re-use existing content. The term Learning Object has been introduced to define reusable units of learning materials; re-use here means to use a learning object again without changing it.

A broader definition of re-use is based on the insight that re-use may include modifying the Learning Resource in order to use it (re-purposing) [1]. Re-purposing may consist of different steps, such as modularization and adaptation. A re-purposing tool is a tool that supports an author to transform a Learning Resource to suit a new learning or teaching context. It guides the author to reach that goal and should enable her to pursue the big picture instead of small editing steps.

Developing such a re-purposing tool is a complex task. Therefore, this paper focuses on the development of a re-purposing framework, which facilitates the implementation of re-purposing tools.

This paper is organized as follows: In the second chapter, we analyze the requirements for a re-purposing framework. Related work is discussed in chapter three. Chapter four presents a brief solution approach which is then transformed into an actual architecture and implementation in chapters five and six. Finally, conclusions and an outlook are presented in the seventh chapter.

2 Requirements for a Re-Purposing Framework

As shown in the introduction, there is need for a generic re-purposing framework. The requirements for such a framework are analyzed in this section. The re-purposing framework should not be one application with a fixed functionality and front-end, but a basis on which different applications can be built. The core requirements are:

Format Independence: The intended framework must be format-independent in a way that an application using the framework will be enabled to deal with resources in different formats. Ideally, the application should be format-independent.

Support for Structured Multi-Document Resources: Learning Resources may consist of multiple documents. Usually, there are links between these documents, which turn a set of single documents into a meaningful ensemble. Those links may be hyperlink-like references between content-documents or references in a manifest document. Therefore it is necessary to regard several documents as one logical resource as well as to give the possibility to regard each single document.

Enhanced Support for Content Analysis: An intelligent re-purposing tool, which supports the user to reach her goals, has to know about the details of the Learning Resource contents. It would, for example, be helpful for restructuring tasks if the structure of a course could be shown to the user; possibly enriched with information about the relations between different text blocks (e.g. coherence of text blocks or which didactical roles they play).

Modification of Content Without Unintended Information Loss: Analysis of the content is only the first step of re-purposing. The data has also to be changed by the application. These changes should take place without losing relevant information (e.g. formatting). This is a challenge because the analysis works on an abstract data model.

Extensibility: The re-purposing framework should be designed as an open framework for various re-purposing applications. Hence, extensibility should be considered as one of the core requirements. Extensibility comprises especially new formats, new re-purposing processes and new content analysis methods.

3 Related Work

Authoring tools for Learning Resources mostly support creation and simple editing, but no re-purposing. Furthermore, many authoring tools bring along their

own source format. A created Learning Resource is rendered into a target format for publishing. Those Learning Resources can only be edited by the authoring tool, and only if the original source is available. There are some tools that provide basic adaptation support. But the functionality is always restricted to a small set of document formats and adaptations types. Examples for such tools are SYSTRAN office plug-ins [2] and the slide master feature of Microsoft PowerPoint. However, the approach of using one separate tool for each adaptation type for each format leads to a confusing number of programs. And each of these programs behaves differently. We currently do not know of any editing tool that offers support for arbitrary adaptations for different document formats.

A wide-spread approach for covering several document formats by one tool is to use an abstract intermediate format. A popular approach is to use the XML File Format of OpenOffice as intermediate format and filters for importing and exporting various other document formats [3, 4]. Another project that uses an intermediate format is ALOCoM [5]. ALOCoM uses a proprietary intermediate format in combination with a simple ontology for content elements.

Bayerl et al. have developed methods for semantic analysis of document markup (cf. [6]). They propose to use three levels of semantic document markup: a structural level, a thematic level and a rhetorical level. An implementation exists that uses XML as markup language and Prolog for semantic analysis. However, only one document format is supported: all contents have to be available as DocBook documents.

4 Application Examples

Two example applications may illustrate the use of a re-purposing framework. The first example is an adaptation tool that adapts a Learning Resource to a particular corporate design. The second example is a modularization tool, which decomposes a Learning Resource into several smaller Learning Resources.

A corporate design consists of several layout elements and styles, which form a unique appearance that is recognized by customers. Parts of a corporate design are e.g. logos, particular fonts, text colors and sizes, as well as rules for text alignment and image placement. The difficulty in changing the layout of a Learning Resource in order to fit a new corporate design is to identify layout elements with a particular meaning. For example, headlines have to be formatted different than plain text or image captures. If logos are to be replaced, we have to tell apart logos from all other images. An adaptation tool could benefit from a re-purposing framework by re-using existing functionality and focusing on the primary task.

As a second example we consider a tool that modularizes a Learning Resource into several smaller Learning Resources. Again detailed semantic information about the content elements would be required for performing the task. Consider that a modularization algorithm calculates module boundaries based on the similarity of text passages. Using the re-purposing framework, the modularization tool could first calculate the similarity of text passages and attach

that information to the content elements. The boundaries would be technically realized by propagating the boundaries to the format-specific layer, where finally the Learning Resources are technically decomposed.

5 Re-Purposing Framework

For realizing a re-purposing framework that fulfills the aforementioned requirements, we decided to create representations of the Learning Resource contents on multiple layers, each with an increasing abstraction level. The proposed layers are:

- A physical representation layer which represents the involved files as they exist on disk
- An object-oriented representation layer that represents the content structure as an object tree
- A semantic representation layer that represents the content structure as a semantic graph enriched by semantic information

On the object-oriented representation layer a so-called Object-Oriented Content Representation (OOCR) will represent the structural entities of the documents as a tree of objects. These objects are also the entities which provide methods for modifying the content.

The semantic representation layer will consist of an RDF model called Semantic Content Representation (SCR) that contains the structural information of the OOCR plus additional semantic markup and relations. The SCR also provides the required format abstraction by using a generic Content Ontology. The Content Ontology will provide the necessary concepts for interpretation of the SCR.

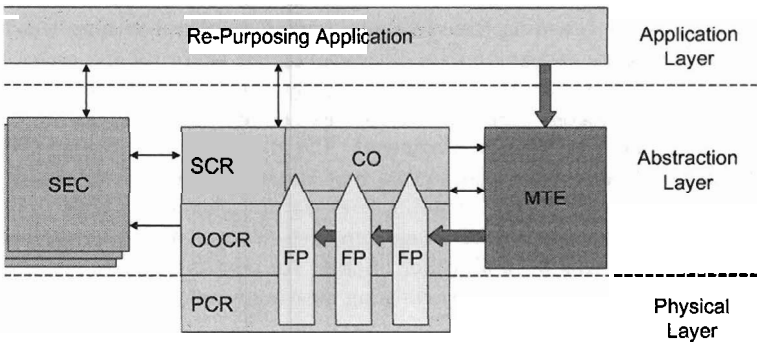

Fig. 1. Components of the Re-Purposing Framework.

However, although an RDF model is very helpful for analysis, it is not suitable for changing the content. If changes were performed directly on the Semantic

Content Representation, it would be necessary to transform this model back into the original format - which would lead to information loss. Therefore, all modifications are propagated by a Modification Transaction Engine (MTE) to the OOCR, where they can be performed format-specific. Content modifications are formulated by re-purposing applications as atomic, format-independent modification commands.

The complete architecture is illustrated in Fig. 1. Re-purposing tools mainly interact with the framework via the SCR.

The ability for handling particular document formats is added to the framework by format plug-ins. A format plug-in contributes format-specific extensions to the framework for handling this document format.

6 Implementation

We have implemented a first version of the presented re-purposing framework. Currently we provide support for two formats: HTML and the Sharable Content Object Reference Model (SCORM) [7]. The framework is combined with a simple SCORM editor that allows to edit and aggregate SCORM-compliant Learning Resources. Framework and SCORM editor together are planned as a basis for a sophisticated re-purposing tool.

The whole application is implemented in Java on top of the Eclipse Rich Client Platform (http://www.eclipse.org/rcp/). It is realized as multiple Eclipse plug-ins, which represent the different components of the architecture. There is one plug-in for the basis system and for each supported content format a separate plug-in. Re-purposing applications are integrated by adding a new Eclipse plug-in. For the Semantic Content Representation, we build an RDF model using the Jena semantic web framework.

Up to now, some simple adaptation processes have been implemented to prove the feasibility of the framework architecture. The results are quite promising, as the adaptation processes work satisfactory for the supported formats right now. We have used three different SCORM courses for testing (consisting of 27 to 95 HTML documents). The results of a performance measurement show that the tested courses can be transformed into OOCR and SCR representations in less that 7 seconds, which is .

7 Conclusion and Outlook

In this paper, we have presented the requirements for a re-purposing framework and our architecture for such a framework. The approach of that framework is to separate a re-purposing application as much as possible from the handling of the particular documents. The documents are represented on three different layers as physical documents, an Object-Oriented Content Representation and a Semantic Content Representation. Analysis and modifications are performed via a generic interface that abstracts from particular document formats.

The feasibility of the re-purposing framework has been proven by the implementation of some basic adaptation processes. More sophisticated re-purposing applications, which use more of the framework's potential, are currently developed. One of the remaining tasks is to implement complex analysis methods in order to facilitate the development of those re-purposing applications. Also, additional format plug-ins for further document formats should be implemented to broaden the use of the framework.

8 Acknowledgments

This work is supported by the German Federal Ministry of Economics and Technology in the context of the project Content Sharing. We would like to thank Sonja Bergsträßer for developing the Content Ontology, Birgit Zimmermann for her implementation of an adaptation tool that runs as a first re-purposing application on top the framework and Heiko Schmitt, who implemented parts of a SCORM Format Plug-in.

References

1. Rensing, C., Bergsträßer, S., Hildebrandt, T., Meyer, M., Zimmermann, B., Faatz, A., Lehmann, L., Steinmetz, R.: Re-Use and Re-Authoring of Learning Resources - Definitions and Examples. Technical Report KOM-TR-2005-02, TU Darmstadt - Multimedia Communications Lab (2005)
2. SYSTRANS: SYSTRANS WebServer, (http://www.systransoft.com/products/client-server/webserver.html)
3. OpenOffice.org: OpenOffice filters using the XML based file format, (http://xml.openoffice.org/filters.html)
4. Rotard, M., Finsterle, L.: Creating and Publishing Courseware Independent of the Authoring System. In: Proceedings of the Technology Enhanced Learning Conference. (2003) 251–254
5. Verbert, K., Gasevic, D., Jovanovic, J., Duval, E.: Ontology-based learning content repurposing. In: WWW '05: Special interest tracks and posters of the 14th international conference on World Wide Web, New York, NY, USA, ACM Press (2005) 1140–1141
6. Bayerl, P.S., Lüngen, H., Goecke, D., Witt, A., Naber, D.: Methods for the semantic analysis of document markup. In Vanoirbeek, C., Roisin, C., Munson, E., eds.: Proceedings of the 2003 ACM Symposium on Document Engineering (DocEng-03), New York, ACM Press (2003) 161–170
7. Advanced Distributed Learning: Sharable content object reference model (SCORM) 2004, (http://www.adlnet.org)
8. Ayars, J., Bulterman, D., Cohen, A., Day, K., Hodge, E., Hoschka, P., Hyche, E., Jourdan, M., Kubota, K., Lanphier, R., Layaida, N., Hegaret, P.L., Michel, T., Newman, D., van Ossenbruggen, J., Rutledge, L., Saccocio, B., Schmitz, P., ten Kate, W.: Synchronized multimedia integration language (SMIL 2.0) specification — second edition. World Wide Web Consortium, Recommendation REC-SMIL2-20050107 (2005)