Improving Retrieval of Reusable Learning Resources by Estimating Adaptation Effort

Birgit Zimmermann[1], Marek Meyer[1], Christoph Rensing[2], Ralf Steinmetz[2]

[1] SAP AG
SAP Research CEC Darmstadt
Bleichstr. 8, 64283 Darmstadt
{Birgit.Zimmermann, Marek.Meyer}@sap.com

[2]Fachgebiet Multimedia Kommunikation
Technische Universität Darmstadt
Merckstr. 25, 64283 Darmstadt
{Christoph.Rensing, Ralf. Steinmetz}@kom.tu-darmstadt.de

Abstract. Due to slightly changed scenarios of usage a one-to-one reuse of existing E-Learning resource often is not possible. Actual approaches for retrieving existing E-Learning resources for reuse mostly do not take this into account. But to achieve a high quality often it is necessary to adapt the existing resources to the new usage scenario. In this paper an approach for retrieving existing learning resources is presented that takes into account the difference between the desired usage scenario and the original usage scenario of existing resources.

Keywords. Re-Use, Re-Purposing, Re-Authoring, E-Learning, Metadata, Similarity Measurement, LOM, Learning Resource, Learning Object Retrival

1 Introduction

Creating appropriate E-Learning material is a costly task. Hence the reuse of existing material is an important topic in current research [2]. But to be able to reuse existing material users must find suited resources. Many approaches deal with reuse of existing E-Learning material but all these approaches do not consider that often a one-to-one reuse is not possible due to a changed usage scenario. In these cases adaptations to the changed usage scenario are needed [11]. It is a challenge to find appropriate E-Learning resources that can be reused with a minimum of adaptation effort. In addition, it would be desirable to know which adaptations are needed.

In this paper a system is presented that is based on a new recommendation mechanism, which ranks learning resources according to their adaptation effort. In addition the system names the needed adaptations.

The paper is structured as follows: The first section lists definitions and related work. An approach to estimate the costs needed to adapt an existing learning resource to a required profile representing the new usage scenario is presented in the second section. Section three focuses on the implementation of this approach. Section four provides a summary and gives an outlook on future work.

1.1 Re-Purposing of E-Learning Content

Re-Purposing means that existing E-Learning content is adjusted to new needs or situations. Rensing et al. [7] define it as follows:

"Re-Purposing is the transformation of a Learning Resource to suit a new learning or teaching context. This means especially that the Learning Resource is transformed to suit a new learning objective or a new target group, which is different from the learning objective or target group the Learning Resource was created for."

Re-Purposing consists of modularization, adaptation, and aggregation of content. Modularization means to split a learning resource in smaller units that are suited for reuse. Adaptation means to change a learning resource in order to adjust it to a new usage context. Aggregation means to combine small learning resources to a larger learning resource. In this paper we mainly focus on adaptation. But this may also lead to a need for modularization and aggregation.

1.2 Related Work

Many different approaches deal with the reuse of E-Learning content. Some focus on learning objects as reusable elements like [10]. Others focus on technical aspects like providing the needed infrastructure [1, 4]. But not one of them explains how adequate resources for reuse can be found if a one to one reuse is not possible.

An interesting approach is the one of Sicilia and Sánchez-Alonso [9]. They propose a concept for the reuse of learning objects based on Design by Contract, a concept used in Object-Oriented Software Engineering. The authors assimilated it to the context of learning objects. In their approach they emphasize that strictness of metadata is needed to enhance machine understandability. But they also do not consider that for a reasonable reuse it might be necessary to adapt learning resources.

Similarity measurement measures the similarity between images, words, documents etc. The distance between two objects is calculated to express their similarity. There are lots of methods to calculate this distance, e.g. Euclidean Distance, Mahalanobis Distance, or Chord Distance. One work in this area is WordNet::Similarity that measures the similarity of concepts. "It provides six measures of similarity, and three measures of relatedness, all of which are based on the lexical database WordNet." [6] Pedersen et al. compare two concepts, and calculate a numeric value representing the degree to which the concepts are similar.

2 Adaptation Cost Calculation

As stated before, reusing existing E-Learning content often leads to a need for adaptation. But how can users find the one existing learning resource that needs least adaptation effort to meet their requirements? To overcome this problem this paper proposes an approach that uses metadata similarity measurement to deduce the needed adaptations. Based on this, the costs for the adaptations are calculated. In the following the prerequisites and a detailed description of the approach will be given.

2.1 Prerequisites

To be able to find the resource that best fits the user's requirements a requirements profile has to be provided, which describes the desired course. It consists of a metadata description of the ideal learning resource. To determine the similarity between the perfect course and existing courses, the profile is compared with the metadata of the available existing learning resources.

A prerequisite to the solution proposed here is the use of strictly formalized metadata. This means that metadata fields have to be comparable and normative instead of being descriptive. This is necessary for a quantifiable comparison.

Thus, strictly formalized metadata should be used to characterize existing and desired courses. A standard often used to characterize learning resources is the Learning Object Metadata (LOM) standard [3]. One of its metadata fields is the target age range field that describes the age range of the intended learner. It requires character strings as input. This allows, for example, values like "over 21" as well as "suited only for adults" which, although they might mean the same, may not be comparable because of their different formats. Therefore they are not suited for a similarity measurement. Hence, existing and desired courses should be characterized with normative metadata, as described by Sánchez-Alonso and Sicilia [8].

2.2 Solution Approach

First the user has to create a requirements profile. This profile contains the metadata describing a fictive ideal learning resource. E.g. you are looking for a learning resource in English with a high semantic density to the educational objective "TCP/IP". The learning resource should be created for your company "xy AG". This would lead to the following requirements profile:

(Language=English, Semantic Density=High, Educational Objective=TCP/IP, Contribute=xy AG).

When the requirements have been entered the system calculates the resource that needs least adaptation effort and thus matches best the requirements. The calculation of the adaptation effort and the cost estimation are performed in three steps:

1. Field-wise calculation of metadata deviation
2. Calculation of involved adaptation types
3. Estimation of resulting adaptation costs

First for each element of the metadata record it is calculated how much it differs from the corresponding field in the requirements profile. Based on the metadata type the adequate function out of a set of similarity functions is chosen.

For some metadata, e.g. language, the outcome of the function is a binary value. If the languages of a desired profile and an existing resource are different, this leads to "1", which means, there is a need for adaptation (in this case translation). If the languages of the elements are identical, this leads to "0", which means that no adaptation is needed.

Then there are values that can vary more or less (e.g. semantic density). The higher the difference is the higher is the distance value. For some of these values, if the difference it too high, it is impossible to make the resource meet the requirements. Then the distance is set to infinite. This might be the case if the learning objectives are totally different (e.g. "TCP/IP" and "Modern Art").

The outcome of the similarity measurement is a distance vector (DV), describing for each metadata field the difference between a given resource and a fictive ideal resource. Let us assume that you have three learning resources that match at least some of the metadata of the requirements profile described before. Table 1 shows the metadata of the three learning resources:

Table 1. Metadata for learning resources.

	Language	Semantic Density	Educational Objective	Contribute
First resource	English	High	Modern Art	xy AG
Second resource	English	High	TCP/IP	ab GmbH
Third resource	German	Medium	TCP/IP	xy AG

Calculating the similarity between the three resources and the requirements profile leads to the following distance vectors:

$$\text{DV1:} \begin{pmatrix} 0 \\ 0 \\ \infty \\ 0 \end{pmatrix}, \quad \text{DV2:} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad \text{DV3:} \begin{pmatrix} 1 \\ 0{,}25 \\ 0 \\ 0 \end{pmatrix}$$

Fig. 1. Distance vectors for learning resources metadata.

In the second step the needed adaptation types are determined by multiplying each distance vector by a so called Adaptation Types Involvement Matrix (ATIM). The ATIM contains for each adaptation type and for each metadata field if a difference in the metadata value leads to a need of the particular adaptation. If this is true the matrix contains "1", otherwise it contains "0".

The multiplication of the ATIM with a DV results in a vector containing the estimated effort for each adaptation type. In our example we use the ATIM shown in Fig. 2. The metadata fields are (from left to right) language, semantic density, educational objective, and contribute. The adaptations are (from top to bottom) translation, adaptation of semantic density, adaptation to educational objective, adaptation of (corporate) design, and adaptation of terminology. Different metadata values for Language lead to a need for translation (marked by "1"). Different metadata values for Semantic Density lead to a need for an adaptation of semantic density. Different metadata values for Educational Objective lead to a need for an adaptation to the educational objective. Different metadata values for Contribute lead to a need for both an adaptation of (corporate) design and an adaptation of terminology.

$$ATIM = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Fig. 2. Adaptation Types Involvement Matrix.

Multiplying the distance vectors of the first step with this matrix leads to the adaptation effort vectors (AEV) shown in Fig. 3:

$$AEV\ 1: \begin{pmatrix} 0 \\ 0 \\ \infty \\ 0 \\ 0 \end{pmatrix}, \quad AEV\ 2: \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad AEV\ 3: \begin{pmatrix} 1 \\ 0{,}25 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Fig. 3. Adaptation effort vectors.

Finally in the third step, the adaptation cost function (ACF) takes the adaptation effort vectors that have been generated in the second step as an input to calculate the estimated adaptation costs.

The cost function is a multiplication of the adaptation effort vector by a cost vector, containing a cost value for each adaptation type. The cost value depends on the available tools. For example, if a tool supports easy translations, the cost function will return low cost values for needed translations; if the available tools support translations only insufficiently or not at all, the resulting costs will be higher. Or if a tool is used that executes a (corporate) design adaptation automatically the values for (corporate) design adaptations are very low. If (corporate) design adaptations have to be executed completely manually, as no tool is available, the values are set to a higher value. The outcome of this step is a cost value for each effort vector.

Notice that the cost function calculates an absolute cost value for the adaptations, but this is only an estimated value, relative to a reference adaptation. The real adaptation costs depend on the length and quality of the learning resource, as well as on further processing details that cannot be calculated exactly. However, the estimation provides a magnitude that helps the user in choosing the resource with the least adaptation effort. In addition the tool tells the user which adaptations are needed.

2.3 Expandability and Possibilities for Customization

The approach introduced above provides three possibilities for customization and expandability to future needs: The Distance Functions for metadata fields, the Adaptation Types Involvement Matrix and the Adaptation Cost Function:

There might be a need for customization of the Distance Functions if the metadata types change. E.g. if a new metadata is added to the metadata record this new metadata has to be added as well to the distance function. In addition a method has to be provided how to calculate the distance for this metadata type.

The Adaptation Types Involvement Matrix could be subject to changes if either new metadata fields are introduced (in this case the matrix has to be enlarged) or the supported adaptation types change in number or definition.

The Adaptation Cost Function probably is the component that is most likely to be customized. It depends heavily on the available tools, as different tools support the same adaptation type in very different ways. The cost function reflects how efficient the available tools support the different adaptation types. Hence, each user has to customize the cost function in a way that it reflects the tools that are available to the user to perform the adaptations.

3. Implementation

As explained in the previous chapter, three steps are needed to calculate the costs for adapting a learning resource. This leads to the following system architecture (Fig. 4):

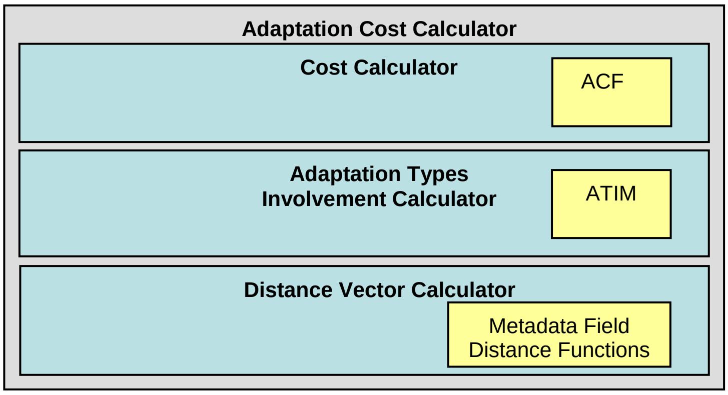

Fig. 4. System architecture.

Based on this architecture and the approach described before a prototype of the adaptation cost calculator (ACC) has been developed. The ACC is integrated into a tool called Module Editor [5]. The Module Editor is a tool that supports users in performing all reauthoring tasks (modularization, adaptation, and aggregation). Within this tool the ACC helps users to find the learning resource that needs least adaptation effort to be adapted to the user's requirements.

The ACC is invoked via the menu. First users have to enter the requirements profile (see Fig. 5). At the moment four metadata fields are taken into account: language, contribute, semantic density, and educational objective. The calculation can be started by clicking on the button "Next".

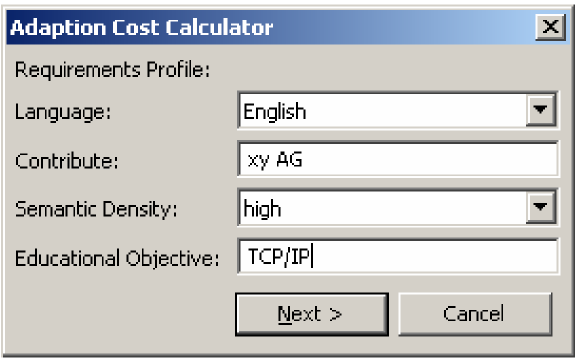

Fig. 5. Entering the requirements profile.

To calculate the adaptation costs the system compares each available learning resource to the user's requirements. The needed adaptations are calculated as described in the example in chapter 2.2. Based on this the costs are estimated.

On the basis of the considered metadata the following adaptations might be needed: translation, adaptation of (corporate) design, adaptation of terminology, adaptation of semantic density, and adaptation to educational objective.

At the moment the adaptation tool provides (partially) automated support for the (corporate) design adaptation, translation, printability and terminology. The other adaptations have to be performed manually. Therefore the costs for adapting a resource to a changed (corporate) design are the lowest, followed by changing terminology. Adapting a course to another semantic density causes higher costs. The highest costs are caused by an adaptation to a changed educational objective. Translation is positioned between adaptation to semantic density and adaptation to a changed educational objective.

The outcome is an ordered list of the three resources that match best the user's requirements (see Fig. 6). The user is also told which adaptations are needed to adapt the resource to the requirements.

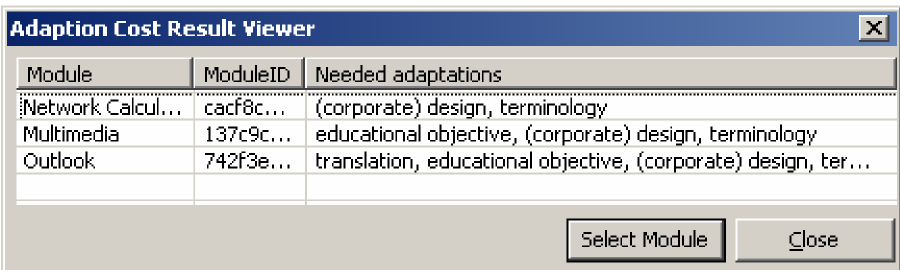

Fig. 6. Results of the calculation.

4. Conclusion and Future Work

We presented an approach that allows estimating the costs needed to adapt existing learning resources to a fictive ideal resource. Our approach has been implemented as a prototype that is integrated in the Content Sharing Module Editor. First evaluations of the prototype are promising as the tool helps users in finding the existing learning resource that fits best their needs. The tool also tells users which adaptations have to be performed to adapt the existing resource to the user's requirements.

At the moment we are working on getting reliable data on the adaptation costs to provide users accurate cost estimations. For the future we plan to take into account additional metadata and adaptation types.

Acknowledgements This work is supported by the German Federal Ministry of Economics and Technology in the context of the project Content Sharing.

References

1. ARIADNE Foundation: Ariadne Foundation for the European Knowledge Pool. http:// www.ariadne-eu.org.
2. Duval, E., Hodgins, W. "A LOM research agenda". In: Proceedings of the twelfth international conference on world wide web, 2003.
3. IEEE Learning Technology Standards Committee, Learning Object Metadata (LOM), IEEE 1484.12.1-2002.
4. Merlot: Multimedia Educational Resource for Learning and Online Teaching. http://www.merlot.org/.
5. Meyer, M. et al.: Requirements and Architecture for a Multimedia Content Re-purposing Framework. Proceedings of First European Conference Technology En-hanced Learning (EC-TEL), 2006.
6. Pedersen, T., Patwardhan, S., Michelizzi, J.: WordNet:: Similarity - Measuring the Relatedness of Concepts. In Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI-04), July 25-29, (2004).
7. Rensing, C., et al.: Re-Use, Re-Authoring, and Re-Purposing of Learning Resources - Definitions and Examples. Technical Report KOM-TR-2005-02, Darmstadt University of Technology, (2005).
8. Sánchez-Alonso, S., Sicilia, M.A.: Normative Specifications of Learning Objects and Learning Processes: Towards Higher Levels of Automation in Standardized e-Learning. In International Journal of Instructional Technology and Distance Learning, Vol 2. Issue 3, (2005).
9. Sicilia, M.A., Sánchez-Alonso, S.: On the concept of learning object "Design by Contract". In WSEAS Transactions on Systems, Vol. 2, Issue 3, 2003, 612-617.
10. Wiley, D. A.: Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In: The Instructional Use of Learning Objects. Bloomington, Indiana. Agency for Instructional Technology and Association for Educational Communication & Training, (2002), 3-23.
11. Zimmermann, B., Bergsträßer, S., Rensing, C., Steinmetz, R. A Requirements Analysis of Adaptations of Re-Usable (E-Learning) Content. In: Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications (2006).