Multibook: Metadata for Webbased Learning Systems

Achim Steinacker, Cornelia Seeberg, Stephan Fischer, and Ralf Steinmetz
Industrial Process and System Communications
Darmstadt University of Technology
Merckstr. 25, D-64283 Darmstadt, Germany

GMD IPSI
German National Research Center for Informtion Technology
Dolivostr. 15, D-64293 Darmstadt, Germany
{Achim.Steinacker,Cornelia.Seeberg,Stephan.Fischer,Ralf.Steinmetz}@kom.tu-darmstadt.de

Keywords

WEBBASED LEARNING SYSTEMS, METADATA, LEARNING OBJCETS

Abstract

The Multibook project of the Technical University of Darmstadt builds a webbased adaptive hypermedia teaching and learning environment for multimedia and communication technology. Besides information gained from the interaction with the user, the system uses standardized content relations and meta-information to adaptively compile a selection from the set of available information units. In this paper we will compare the meta descriptions we have specified for the connections between information resources with the suggestion from other organizations currently developing meta-data standards for the Web. We will show, that relations tailored to the needs of learners and instructors like the ones we are using in Multibook, can help to adequately identify the resource someone is looking for and can increase the benefit of working with resources.

1 Introduction

One of the main reason for the success of the WWW was and still is, that it is very easy to provide information. Futhermore additional information already available on the web can be integrated into the own pages, with just adding a simple link. Unfortunately the easiness and because of this the exponential growing of obtainable information is meanwhile also the reason for the so called "information overload" in the Web these days. Search engines usually offer thousands of hits to a user as a result of a query, if the keywords the user provided are very popular or generic. The problem is that it is still not possible to describe the content of HTML pages, videos or applets in an adequate way. What is needed is information about information also called metadata - labeling, cataloging and descriptive information structured in such a way that allows Web pages to be properly searched and processed [9]. With metadata users can describe much more accurately what kind of information they actually want to find. This would be especially useful for webbased learning systems. Besides using a standard web browser as a user interface for these systems, one of the main advantages of web based learning is the possibility to use the openness and link facility of the Web, to integrate and reuse already available teaching material, when building own learning lessons. Metadata for teaching material should be tailored to the needs of learners and teachers. Consequently a description facility should not only offer the ability to describe for example the difficulty or the implemented learning strategy of a resource. If we think about an example where a user is looking for a document explaining the Discrete Cosinus Transformation and the search engine provides an adequate document. If the metadata for this document would

contain links to document explaining or even visualizing Image Compression or the concrete example JPEG, the user could more easily understand the content of the document because the metadata is offering additional links, where the user can learn something about the context the DCT is used in or some concrete examples where it is used. The same is true for an instructor. Imagine someone has built a lesson which gives an overview about the DCT and the appropriate metadata. With adding links from his own description to other documents describing other forms of compression, or to a document going into the mathematical details of the DCT, available somewhere else on the web, the instructor can enrich his own lesson, without implementing all necessary documents itself. These small examples show that the relation between metadescriptions for learning resources can play a key role for learners and instructors.
The paper is structured as follows. In Section 2 we will provide an overview of the most common meta data project, the Dublin Core and a short introduction in a meta data project for learning material from the IEEE, the Learning Objects Metadata (LOM). After identifying the inadequacy of the way these metadata projects describe the relations between metadata descriptions, in Section 3 we will introduce the metadata we are using in the Multibook project. We will show that the small number of relations we have developed can help both instructors and learners to better identify the teaching resources for reuse and combining or learning. Section 4 will conclude the paper.

2. Metadata for the Web

The frustration of retrieving thousands of hits, for someone seeking for information at todays Web search engines, has increased the interest in metadata standards and practices. Metadata usually describes an Internet resource: what it is, what it is about, where it is, and so on. Metadata is not limited to describing textoriented documents. Any resource (e.g. video, image, audio, etc.) can be described with an appropriate metadata element set. The first aspect which must be addressed is what set of information is to be captured by the metadata. This will depend on the type of the resource and on the purpose of the metadata. A metadata scheme must be sufficiently flexible to capture useful information about a wide variety of resources for a range of purposes. Ideally, a single metadata scheme should be used as this minimizes the cost of using metadata. The second aspect is the production of metadata. Metadata is essentially a summary of the data produced by various levels of "intelligence". Using humans to generate these summaries is expensive and metadata systems attempt to reduce this cost by making humans more productive by automating as much of the process as possible. The final aspect of metadata concerns how the metadata is accessed and used. It must be retrieved in a form which can be processed with its semantics preserved. An important use of metadata is as a mechanism for resource location in distributed networks like the Internet. Metadata can provide information for the user to identify which resources they might be interested in. Once a resource has been identified, metadata provides the information to allow the resource to be accessed. The basic model used for metadata is known as "attribute type and value" model. Metadata is represented as a set of facts about the resource (e.g. Title, Author). Each fact is represented as an attribute (also known as an element). An attribute contains a type (which identifies what information the attribute contains) and one or more values (the metadata itself).A very basic and simple approach to describe web-resources with metadata is the DublinCore which will be briefly described in the following.
The Dublin Core elements are text-oriented and therefore also readable for humans and understandable for machines.

2.1 Dublin Core

Dublin Core is one of the best-known schemes and was invented by the library community. The name of the schema comes from the town of the first meeting of the working group, Dublin, Ohio, USA. Dublin Core is a specialized vocabulary associated with bibliographic information and consists of fifteen elements [1]:
1.Title
2.Author or Creator
3.Subject and Keywords
4.Description
5.Publisher
6.Other Contributor
7.Date
8.Resource Type
9.Format
10.Resource Identifier
11.Source
12.Language
13.Relation
14.Coverage
15.Rights Management

As noticed, this metadata-standard has its roots in describing traditional library resources because it favours document-like objects. Additionally, some attributes like author, publisher or contributor may be of crucial importance for publications lists or copyright questions, for describing and identifying a learning resource they are less useful.

2.2 Learning Objects Metadata

A more appropriate metadata-standard for describing learning resources than the Dublin Core is currently developed by the IEEE P1484.12 Learning Objects Metadata Working Group. Learning Objects are defined as any entity, digital or non-digital, which can be used, reused or referenced during technology-supported learning. Examples of technology-supported learning include computer-based training systems, interactive learning environments, intelligent computer-aided instruction systems, distance learning systems, and collaborative learning environments. The motivation for and benefits from metadata describing learning resources among others are [2],[5]

- To enable learners or instructors to search, evaluate, acquire, and utilize learning objects.
- To enable the sharing and exchange of learning objects across any technology supported learning systems.
- To enable the development of learning objects in units that can be combined and decomposed in meaningful ways.
- To enable computer agents to automatically and dynamically compose personalized lessons for an individual learner.
- To enable education, training and learning organizations, both government, public and private, to express educational content and performance standards in a standardized format that is independent of the content itself.
- To enable the development of education standards could consist simply of itemized lists of required domain modules for given grade levels.

These learning objects metadata also contain a relation-attribute to express correlation of different resources. Unfortunately the possible values for this attribute are the same as for the DublinCore The values are:

{isPartOf, HasPart, IsVersionOf,HasVersion, IsFormatOf,HasFormat, References IsReferencedBy, IsBasedOn, IsBasisFor, Requires IsRequiredBy}

The values of the relation attribute are also text-document oriented. The "partOf"-value for example represents the fact, that a text or an image is a physical part of another document. It is not possible to model the situation that a concept explained with this resource is "a part Of" a more general concept. For example one cannot map that the Discrete Cosine Transformation (DCT) is a part of Image Compression. Other values like "IsBasedOn" allow the description of a semantic relation between two resources. But again, the roots of these values limit the possibility to express relations to bibliographic tasks. The values of the relation attribute in the IEEE metadata standard are therefore insufficient for modeling relations between resources a learner or an instructor is interested in.
Probably someone has not found what she/he was actually looking for, although she/he has described his/her requirements with metadata. If the resources which were returned from a search query would also offer links to related topics in a meaningful way, this could help to better find what someone was looking for. Furthermore even if the results of a query fulfilled the users claims, it might be interesting for her/him to find links to an alternative or a more detailed explanation or a visualization of the same topic.
Additionally a requirement like the possibility to combine learning objects in a meaningful way can be easily fulfilled, if the resources itself provide links to additional but related material, that can be used to build a learning lesson.
These features are not only useful for instructors which build lessons and teaching material for the web, the same is also true for a learner. In the next section we want to introduce the metadata we use for relations between resources in the Multibook. We will show that just a small amount of possible values between relations enables the system to offer suitable informations for users and is even able to generate complete lessons itself with combining the information reassures.

3 MULTIBOOK

The Multibook project developed by the Technical University of Darmstadt and the Fern-Universität Hagen is a webbased adaptive hypermedia teaching and learning environment for multimedia and communication technology. Hereby, the demands of diverse user groups, user levels and especially of diverse learning strategies are taken into account. Besides information gained from the interaction with the user, the system uses standardized content relations and meta-information to adaptively compile a selection from the set of available information units (media bricks). To be able to accomplish this task, the content and metadata concerning the content must be represented in a formal way, that enables the system to operate on it. As a result the system can play the role of a guide for the user.

3.1 MULTIBOOK'S Knowledge Base

There are two spaces where the information and the metadata are modeled: a domain model, which we call the ConceptSpace, and the MediaBrickSpace, where knowledge is represented in

small modular pieces of text, images, audio, video, or animation. There are objects (concepts and media bricks), relations (semantic in the ConceptSpace and rhetorical in the MediaBrickSpace) and attributes of the media bricks. The concepts and the actual content are therefore separated. While a teaching method reflects the concepts and especially the order in which concepts are displayed, the actual content represented by the media bricks transforms user preferences into a suitable document. In other words, the ConceptSpace represents the domain, the MediaBrickSpace is a set of possible explanations of this domain. Formalization of the content to be learned lies at the core of the Multibook approach. However, we cannot aspire to substitute or completely control the process of learning from text or multimedia material rendered for human consumption and typically not available in a formal description. Our approach is to compose individualized lessons using the network of facts and concepts from the subject domain (ConceptSpace) as an access structure to a body of text and multimedia material rather than to generate lessons directly by turning facts from the domain model into sentences, diagrams, questions and tests. Being used as an index to the original learning material, an approach much closer to publishing practice, related to structuring and re-using assets etc., the facts in the ConceptSpace will not mirror each assertion that is made in the media bricks. It is of crucial importance for the approach to adequately choose the level of granularity, i.e. the detail and extent to which the subject field (and the learner's knowledge about it) is modeled. As a consequence of not formalizing in detail every statement of each media brick, the information we have about which bricks to combine into a lesson is a very general thematic information. If we want to take into account additional characteristics of the media bricks, e.g., their level of difficulty, or how the bricks fit into a certain argumentative strategy this information has to be given explicitly as attributes. We also established relations between the media bricks, these relations are not semantic, but rhetorical.
These considerations influenced the modeling of both the ConceptSpace and the MediaBrickSpace which will be described in the following.

3.2 ConceptSpace

As a rule of thumb, we choose the level of granularity such that objects of the model identify topics that could serve as chapter headings in a multimedia study book. The difference to traditional chapter headings is evidently that the information is held independently from the texts and other media bricks, with the effect that the topics are "centralized" instead of being scattered across the body of material. Formally, this is realized as an entity relationship model, where each object appears only once and accumulates information. Conversely, every topic that is to be related with some other piece of information has to be modeled as an object. While a few instance objects may be relevant (some concrete person or company, etc.) the main part of the network will be formed by abstract concepts.
Such a focus on topics/concepts suggests that our ConceptSpace will be a terminological ontology rather than an axiomatic ontology, i.e. an ontology where concepts and relations have associated axioms and definitions that are stated in logic or some computer oriented language that can be automatically translated to logic. [6]. We have, however, to draw a clear distinction between the generic and the partitive hierarchy relation (a distinction, the necessity of which is discussed in the area of domain modeling [3]). We are building upon this distinction when identifying, e.g., a component (AEpart) of a media server, a certain type of file server (generic) as a topic to be explained but want to start the explanation of the component, say a disk controller, with principles applying to controllers in general. Up to a certain point, deciding whether talking to the learner about controllers at a given stage of the file server lesson makes sense, assumes a conclusion

across a path composed of different relations. To this end, we found it necessary to look at the AEpart relation as a statement relating the set of instances of one concept with the set of instances of the other and to additionally quantify this statement. In general, it must be possible to query the network starting from arbitrary points. This makes it necessary to also think of the terminological relation that we use as statements/assertions that have to be correct no matter which way you read them. They may not be based on implicit assumptions in which order the material is presented or any other context information.
We employ the following set of relations for the ConceptSpace (a more detailed explanation of the relations can be found in [7]:

- Superconcept
- AEpart
- EEpartOf
- InverseProcedure
- Follows
  ProblemSolution
- Partition: Difference (Rahmstorf Relations)

There are different possibilities to choose a set of relations. The granularity of the model relates to the fact that we are interested in roughly structuring the domain rather than completely representing all facts. In the vocabulary of Knowledge Engineering, this makes our ConceptSpace a terminological rather than an axiomatic ontology [6]. We are aware of the benefits that the extended use of verb concepts, and the expression of a greater variety of facts they allow, could have for our domain. However, the current model is, for simplicity reasons, mainly built around noun concepts [3]. Semantic relations strongly guide the composition of lessons but do not predetermine a unique set of topics and the order of these, to map a single order of specific topics onto a teaching method.

3.3 MediaBrickSpace

Media bricks can either be text or other multimedia elements such as images, graphics, video and audio streams and - with the main focus - animations implemented as Java applets. Also these multimedia elements satisfy the requirements of modularity, this means that it is necessary that the format of the media bricks enables the system to describe the content, grade of detail, and the underlying pedagogical concept of the media brick. Thus it is possible to integrate a media brick in a lesson, independent of the kind of media. A specific problem is the level of difficulty which is not a matter of a single media brick. Whether a media brick is difficult to understand depends on various aspects, such as background knowledge of a specific user or the media bricks the user has seen/ learnt before. Examples for such rhetorical relations are "deepen" or "explain". It is a major task of the system, to make these relations explicit. We currently use the following relations:

- **example:** MediaBrick B contains a text describing the runlength algorithm, MediaBrick A contains a concrete example with an uncompressed data stream and the runlength encoded stream.
- **illustrates:** MediaBrick B contains a text with some numbers, MediaBrick A contains a diagram illustrating these numbers.

- **instance:** MediaBrick B contains a graphic where all video compressing processes are listed. MediaBrick A describes MPEG.
- **restricts:** MediaBrick B contains a description of a theory, MediaBrick A describes cases where the theory fails.
- **amplifies:** MediaBrick B contains a video stream describing a CSCL system. MediaBrick A contains a text describing how CSCW systems work.
- **continues:** MediaBrick B contains an uncompressed image. MediaBrick A shows the same image runlength encoded.
- **deepens:** MediaBrick B contains an animation visualizing how the DCT works. MediaBrick A contains the formula of the DCT.
- **opposite:** MediaBrick B states a theory of a specialist. MediaBrick A contains a statement of another specialist contradicting the first one.
- **alternative:** MediaBrick B contains an image making clear what a pixel image is. MediaBrick A contains a text explaining a pixel image.

In [8], it is shown that with the relations we are using in both spaces, it is possible to compile lessons for individual learners. Furthermore the system can generate individual table of contents, as a guidance for the learner through the material. This approach can help closing the gap between free navigation and static guidance in an adaptive hypermedia system. We are aware, that the ontology we propose has not the fine granularity as the one described in [4]. The main reason for our model is the openness and extensibility of our system. We believe that the majority of learning resources available on the web are rather complete units examples or tests, than just sentences or phrases, so a more detailed modeling is not necessary. Another reason for using our model is the fact that authors and teachers using a knowledge based learning system must be willing to add the necessary meta-information to both spaces to integrate a learning resource. It is therefore better to reduce the amount of needed informations for describing a resource as low as possible.

4 Conclusion

Metadata standards which are currently under development can help learners and instructors to search, evaluate, acquire and utilize learning objects, but the development and utilization of learning objects in units that can be combined and decomposed in meaningful ways, needs a richer vocabulary for the relation between the information resources than just a bibliograhic view on the material. Relations, tailored to the needs of learners and teachers like the ones we are using in Multibook, can help to identify in a reliable way the resource someone is looking for and can increase the benefit of working with the resources.

Acknowledgments

The authors would like to thank the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) which partially funded the research project and the Volkswagen Stiftung. We also owe thank to Lilo Kolb and Moni Jayme for their constant support.

References

[1] Dublin Core Metadata for Resource Discovery. Stuart Weibel, John Kunze, Carl Lagoze, and Misha Wolf, RFC2413, September 1998

[2] IEEE Learning Technology Standards Committee (LTSC) ,Learning Object Metadata (LOM),

Draft Document v2.1, http://grouper.ieee.org/groups/ltsc/ltscdocs/wgc/LOMdoc2_1.html

[3] Fischer D.H., From Thesauri towards Ontologies?, in Hadi M., Maniez J. and Pollit St. A.; Structures and Relations in: Knowledge Organization, Proceedings of the 5th ISK Conference, 1998

[4] Mizoguchi R., Task Ontology Makes It Easier To Use Authoring Tools ,Proc. of IJCAI-97, Nagoya,Japan, pp.501-506, 1997.

[5] Schoening J., A Case and Strategy for Developing Standardized Educational Domain Modules and Ontologies, http://grouper.ieee.org/groups/ltsc/ontol.htm, 1997

[6] Sowa J., Ontologies for Knowledge Sharing, Manuscript of the invited talk at TKE'96, 1996

[7] Seeberg C., et al. From the users needs to adaptive documents, to appear in the Proceedings of the IDPT'99

[8] Steinacker A., et al. Dynamically Generated Table of Contens as Guided Tours in Adaptive Hypermedia Systems, Proceedings of the EdMedia, Seattle 1999

[9] W3, Metadata Activity Statement, http://www.w3.org/Metadata/Activity.html, 1999

2nd International Conference on New Learning Technologies

Multibook: Metadata for the Web

Achim Steinacker and Cornelia Seeberg and Stephan Fischer and Ralf Steinmetz

The Multibook project of the Darmstadt University of Technology builds a webbased adaptive hypermedia teaching and learning environment for multimedia and communication technology. Besides information gained from the interaction with the user, the system uses standardized content relations and meta-information to adaptively compile a selection from the set of available information units. In this paper we will compare the meta descriptions we have specified for the connections between information resources with the suggestion from other organizations currently developing meta data standards for the Web. We will show, that relations tailored to the needs needs of learners and instructors like the ones we are using in Multibook, can help to better identify the resource someone is looking for and can increase the benefit of working with the resources.

BibTeX entry