A Framework for the Multi-modal Description of Learning Objects

Eva Heinrich, Jisong Chen
Massey University
E.Heinrich@massey.ac.nz, nzjschen@yahoo.com

Abstract

World-wide large repositories of learning material are created with the dual goals of global access for the learner and of re-use of material by the teacher. Content description standards have been defined which make it possible to locate appropriate material. This paper looks into complementing these standards by extending the possibilities of describing the actual content of the learning material and by extending the use of these descriptions beyond the normal search-and-retrieve tasks. A framework for the multi-modal description of learning material is suggested. The framework aims at providing multiple content description mechanisms and suggests the design of an information system to perform this content description. The content description mechanisms are text-based, that is keywords, freeform text, FSCL description, and audio-based. The system design addresses issues like ownership, public and private access rights for description and the subsequent retrieval.

Keywords: *Content description, multimedia, retrieval*

1. Introduction

Over the last years there has been an increasing number of initiatives to create systems for global accessibility of learning material. Examples are our own Technology Integrated Learning Environment, the TILE project [1], and the web-based training system GENTLE [2]. With ever growing repositories of learning material the need for some form of metadata or content description has become widely accepted. This finds expression in initiatives like 'Multimedia Annotation' [3], work like the Procedural Mark-up Language, PML [4] or Merrill's knowledge objects for the representation of instructional contents [5], and emerging standards like the Learning Object Metadata standard, LOM [6], the Dublin Core [7] (which recently combined efforts with LOM) or the IMS Learning Resource Meta-data specification [8]. Researchers are looking for uniform ways of describing learning material or learning objects. The goal is to make an increasingly large repository of learning material accessible to both students and teachers, with students gaining the advantage of locating appropriate material for their studies, teachers having the opportunity of constructing new lessons out of existing learning objects.

In our work described here we look at the description of learning resources through metadata from a different angle. This does not mean that we want to replace the above-mentioned approaches but that we want to complement them. More specifically we want to address the following aspects.

- We are looking for a rich and detailed description of the actual content of the learning objects. Standards like LOM focus largely on capturing information *about* a learning object (like the cognitive type or interaction style) and not on domain specific cognitive concepts delivered *in* the learning object (like an explanation of recursive programming or the political reforms in France in the 19th century). To describe such cognitively complex concepts we need a description mechanism much more powerful than single keywords. This leads us into looking at language-based description mechanisms for the learning object content.
- Following on from our approach of describing the actual content of our learning objects our metadata (or content descriptions) themselves become useful besides just serving for search-and-retrieval for appropriate learning material. We can design applications where these content descriptions form the basis for group discussions among students or allow an instructor to enrich material provided to the students.
- Looking at these kinds of applications for the use of content description the actual process of creating the metadata and issues about ownership and accessibility of the metadata become important. The earlier mentioned standards contain information, which specify, for example, the user access right for the described learning object. Yet it is assumed that the metadata themselves are freely accessible. In our work, where the metadata consist of rich content description and where these data will have been created by various users, we face the issue of protecting the access to the metadata themselves.

The framework introduced in this paper addresses these issues of rich content description. We suggest multi-modal description formats to give the user (who can be as well instructor as learner) the chance to select the format optimally suited to their specific requirements. As description formats we include keywords, freeform text, audio recordings and our Flexible Structure Coding Language, FSCL [9]. Our framework controls the actual process of generating content descriptions. We define sets for description data, track who is performing the description task and save the resulting descriptions accordingly. This enables us subsequently to decide which users are allowed to access certain descriptions.

The remainder of the paper is structured as follows. We first present some scenarios showing potential applications of our description framework to clarify both the requirements for multi-modal description and the management of access to the description data. After analysing access management and description modes in more detail we present a high level view on the design of a system implementing our framework.

2. Scenarios for the use of our description system

In a first scenario an instructor has already made learning material available online. A group of assistants is working on the definition of detailed content descriptions that will allow the students to accurately locate specific material. The instructor validates the content descriptions provided by the assistants and then makes these descriptions available to the students. Additionally, the students are able to annotate the learning material for their own uses. To facilitate this scenario our description system has to separate between the different sets of descriptions produced by the authorized group of assistants and the individual students. The instructor has to be able to control when the descriptions created by the assistants become publicly available.

In a second scenario the description framework is used to support the online discussion of learning material. In a history class, for example, an online version of a historic document is provided. The students are divided into groups in which they discuss this document. They use the description framework to formulate detailed comments or interpretations which they attach to specific segments of the online document. There are two requirements to facilitate this scenario: a rich, language-based description mechanism and the ability to regulate access to the descriptions based on membership to a group.

In a third scenario we use audio-based description in analogy to expert commentaries in research in social sciences or psychology. An instructor is discussing novels written by different authors from a specific literature period. The instructor has made these novels available online as learning objects. In his lecture the instructor discusses general characteristics of novels of the literature period and introduces the various authors. To illustrate and enforce his points the instructor now records an audio commentary relating to the example novels for his students. The instructor displays the novel learning objects on his computer screen. He selects specific segments in one or several of the novels and speaks his commentary into a microphone. The description information system records the commentary, keeps track of the segment selections in the learning objects and stores the resulting descriptions in a database. In this case the instructor prefers the recording of an audio commentary to the typing of text. Speaking the commentary means a faster input for the instructor than typing. Additionally, the instructor can easily add additional information that could only be expressed with additional words or special notations in written language. The instructor emphasises specific sections in the novels by adding special emphasis or expressiveness to his voice. As there is no text input area required the instructor can use the whole screen area for the display of the novels. The instructor saves his audio-based descriptions together with his standard text-based descriptions in a description set.

The students will use these audio-based descriptions slightly different than the text-based descriptions. With text-based descriptions the emphasis will be on searching through these descriptions to locate specific sections in the material provided by the instructor. With the audio-based descriptions the emphasis will be more on listening to these descriptions to gain value from the descriptions themselves. The students get access to the audio descriptions either by selecting appropriate descriptions from a list of entries on description set level or by browsing through the learning objects, the novels in this scenario, which indicate via high-lighted areas where audio descriptions are attached.

3. Requirements for the description system

An information system for the annotation of learning objects with metadata needs to have two main components, a description and a retrieval component. Our framework supports the description of leaning objects in multiple ways, using textual and audio input as outlined in a later section of this paper. Descriptions are stored in a database to build a repository for subsequent retrieval with the syntax of retrieval dependent on the modality of description. Our descriptions are performed on learning objects. The learning objects are defined and maintained in our Technology Integrated Learning Environment, TILE [1] and we access the learning objects from

there. TILE as well defines the structure of lessons and therefore which learning objects contribute to which lessons. We can use this information, together with our set concept to direct users to search the appropriate descriptions. The set concept means that descriptions are grouped depending on their purpose and relationship to learning objects. This grouping is determined by the person setting up the description. The effect of the resulting description sets is that we can point users to descriptions relevant to their courses of study. A student logging into our system studying a computer science course, e.g., can be directed to retrieve information based on descriptions relating to learning objects on computer science topics in general or more specific on descriptions relating to database topics only entered by a specific instructor.

Besides defining our search space the set concept allows us to target the issues of ownership and access rights. We envisage the following typical uses of a description and retrieval system:

- An instructor prepares descriptions for the use by students for retrieval of appropriate course sections: An instructor together with a group of assistants describes learning objects belonging to their courses. The resulting descriptions initially are regarded as private to the instructor and assistants, that is they can only be seen and modified by this group of people. After a period of review the instructor decides to change the status of descriptions from private to public to allow general access for retrieval. The modification rights will always stay with the authors of the descriptions.
- A student performs description for their own personal use: A student is working through material belonging to their course of study. The student annotates material to assist their study or possibly examination preparation with comments like 'need to study this topic further' or 'revise this for exam'. These student descriptions have the status of private and the student, not being 'in charge' of the course, cannot make these description available for public use. Yet the student can share these private descriptions with other users, e.g. a study group, by identifying these users in the description set.
- An instructor looks for existing learning objects while constructing a new course: An instructor can use the information in public description sets (as outlined above) to search for suitable learning objects. In this case the instructor is using descriptions which have been provided by colleagues to identify teaching material available online.
- A student searches for material available for a specific topic: Again, the student uses the public description sets to locate appropriate material for study. Based on their course enrolment information the student can be directed to the relevant description and retrieval sets.
- An instructor performs a retrieval and makes the retrieval results available to their students: As described before, the instructor retrieves information based on public description sets. Yet in this case the instructor does not use the retrieval results for their own purposes but makes them available for use by the students (we have an equivalent set concept on retrieval level). Doing that the instructor can point students to specific material in a course, like material on a specific topic presented in different forms across various lectures or material of importance for an upcoming examination.

In the framework, we distinguish between the different user types of instructor and student. On description set level (and equivalent on retrieval set level) we specify the owner of the set and optionally a group of trusted people (who can see and edit the descriptions in the set). Sets have the initial status of private and can only be transferred to the status of public (meaning general read access to the descriptions in the set) by the owner of the set if this owner is an instructor.

These concepts give us a number of advantages. We can target the retrieval of information to a subset of all available descriptions. This makes our retrieval more efficient as we restrict the search space. Further, we limit the problems of context for our natural language type descriptions. According to for example Sowa [10], the uncertainty of the context is one of the main problems in natural language analysis. With our set approach we can reduce or hopefully avoid these context problems. Using the course and learning objects from our TILE framework we can guide the retrieval of descriptions to sets linked to specific courses and therefore subject domains. The concept of ownership of descriptions together with the public and private settings allows us to maintain the quality of our descriptions. All publicly accessible descriptions are controlled by instructors which should ensure the quality of these descriptions. With private sets owned by students we allow these students to access our description and retrieval mechanisms for their own use in support of their learning and examination preparation.

4. Effective, multi-modal content representation

As outlined briefly in the introduction there are a number of approaches to content representation. We want to build on these approaches and extend them in two ways. Firstly, we do not want to restrict a learning object to just a very short or small sequence of material. We see the need for learning objects that have to be large enough to teach a cognitively

complex concept as a whole. The presentation of such a concept cannot necessarily be composed from very small units (learning objects). Instructional theories cannot yet give sufficiently clear instructions of how to construct a lesson for a complex topic from single components [11]. We want to relay more on the teacher to present a coherent discussion of a topic in one unit. With this approach we increase the size of a learning object. To still provide a focused description for large learning objects we attach descriptions to segments of learning objects. This means that a description for us consists of two parts: the actual description called the description sentence and the segment (or segments) the description sentence refers to.

Secondly, we want to provide a more effective way of describing the actual contents of learning objects. Current approaches largely use either keyword description or freeform natural language description. Keyword description is appropriate to describe properties of learning objects like the author, some subject categorisation, and media or cognitive type. Keywords are obviously restricted in their expressiveness but can be easily searched. To describe complex contents natural language is used (see e.g. the description tag (1.5) in LOM). Natural language is obviously very expressive and rich but it is still not possible to extract information from natural language with one hundred percent completeness and correctness. We address this issue of powerful description mechanism versus correct and complete retrieval by using the Flexible Structured Coding Language, FSCL. These ideas on content description and segmentation are outlined in more detail in [12].

Our framework contains all these options for description (keyword, freeform text and FSCL) applied to segments of learning objects. The framework could be easily extended to include other forms of description. Which form of description is to be used will be decided by the user performing the description depending on the intended use and characteristics of the learning objects. Regarding our set concept, descriptions of the various types can be combined in one set.

The use of the different types of description forms one aspect of our multi-modal description framework. The other aspect is the use of different media for description. Most if not all description approaches used to date are text-based. Textual description has obvious advantages like the ease of storage of descriptions and of searching through these descriptions. Yet in our framework we go a step further and suggest the use of audio-based description to complement text-based approaches. An audio description follows the same principles as a textual description. A segment of learning material is selected and described, in this case by recording an audio clip. The audio description has a number of advantages: the input will be faster as no typing is required (and provides an alternative for users who are not be able to type); there is no screen area required to display textual input and therefore the screen can be utilised in full for display of the learning objects; some information like the correct pronunciation of words, the intonation reading a poem or the sound of a musical instrument cannot be delivered in textual form. The third scenario presented earlier illustrates the value of audio-based description.

Within our framework we can combine the various description styles presented so the users can choose the style most suitable for their purposes and mix the different styles within or across learning material. For example, a user can define colour-coded areas in a image learning object and annotate these with audio recordings (intonating key areas of a musical notation with a searching performed in keyword-style based on the colour codes).

5. Design of the description system

According to the requirements of our framework we have completed a design for a description (and retrieval) information system. A high-level class diagram for this system is given in Figure 1. The diagram shows the following key elements:

- A description set is defined by either a staff member (or group of staff members) or a student (or group of students). A description set can include descriptions for several learning objects. A description set contains multiple descriptions, which can be of different description modes. A description set can be used in several retrieval sets.
- Three modes of description are included in the model: AudioInput which uses for voice input for description; FFTextInput which allows free-form text, natural language description; FSCLInput which handles natural language-like, structured description and keyword description. Each description contains a description sentence (of one of the three description modes) and is linked to one or several segments.
- Each segment has a reference to a learning object. The segment identifies one specific section of this learning object. The format of a segment depends on the media type of the learning object. There are time-based segments (for media types video and audio), position-based segments (for text documents) and area-based segments (for images and video pictures).
- A retrieval set is defined by either a staff member (or group of staff members) or a student (or group of students). A retrieval set is linked to multiple description sets. A link between a retrieval and description sets means that a retrieval question will be applied to the descriptions belonging to

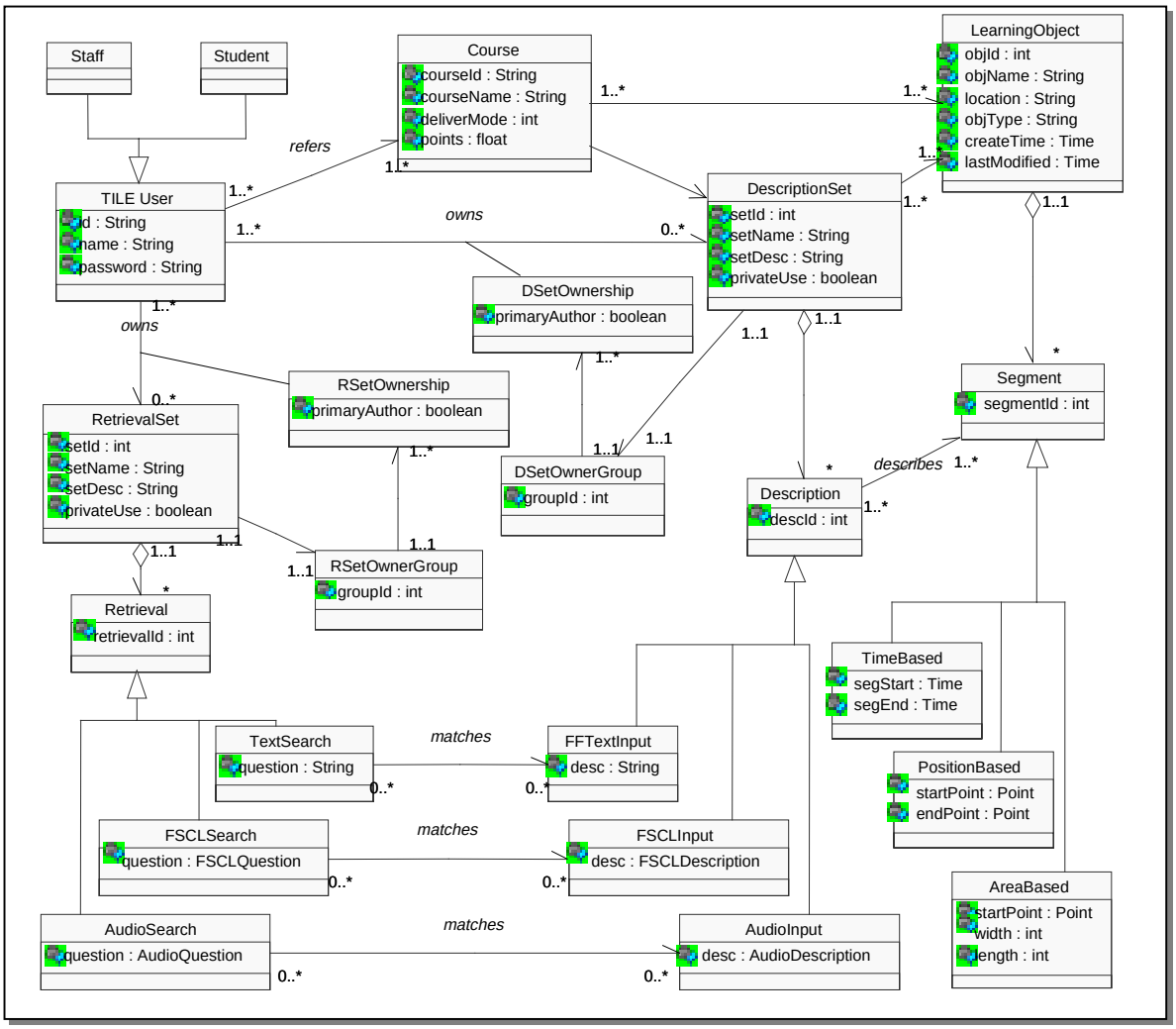

Figure 1. High-level Class Diagram for the Multi-modal Description System

these description sets. This mechanism sets the search space for the retrievals. A retrieval set can contain several retrievals. This means, that several retrieval questions can be asked for the same search space.

- The format of a retrieval depends on the format of the descriptions which are searched. An AudioSearch applies to AudioInput, a TextSearch applies to FFTextInput, a FSCLSearch applies to FSCLInput. The rational behind this design is to exploit the specific characteristics of the various description modes. A FSCL search addresses the sentence structures and vocabulary definitions incorporated into FSCL descriptions, an audio search will map sound patterns while a text search will simply look for word patterns. We can still interpret one retrieval request in various ways by, e.g., regarding a textual input as both a FSCL sentence and as a freeform textual input and therefore performing both FSCL and freeform textual search.
- A staff member/student can define description and retrieval sets. The information about which courses a staff member/student teaches/learns (which is maintained in the TILE framework) is used to direct the staff member/students to relevant description sets (and from there the link to retrieval sets in possible). The system checks in which courses the staff member/student is involved. Once these courses are identified the system can check which learning objects are involved in these courses (again, this information is maintained in the TILE project) and from there which description sets are relevant.
- One description set or retrieval set may have one group of owners, and normally have one primary owner (primary author). Only the primary owner(s) can delete the set, and maintain (add or remove) group of owners information. All owners in the group can maintain (add or delete) descriptions in the set.
- A set (description set or retrieval set) which is set

to private use means that it can only be accessed by the owner(s) of the set. A set that is set to public use means that it can be accessed by all TILE users who are involved in the same course.

We have completed an UML design of the description framework and have implemented a first version of the framework in an application called PAC. This implementation addresses the access right mechanisms described and allows for FSCL and keyword description of video documents (we have implemented the audio description in a related application that now has to be integrated with PAC). PAC is currently used for the detailed description of classroom videos and we expect valuable feedback from this work.

6. Conclusion

In this paper we have introduced a framework for the multi-modal description of learning objects. This framework allows the description of learning objects or multimedia objects in more general terms and the subsequent retrieval of information. We include various types of textual description, that is keyword, freeform text and FSCL, in the framework and add audio-based description. Via the set concept we deal with issues of ownership, public and private access to descriptions to support high quality descriptions and at the same time the flexible use of the description system. Through its design we leave our framework open for the inclusion of further description approaches. We have developed (and continue to improve) the application PAC that implements our framework and therefore allows users to create and access multi-modal descriptions. Our framework is closely integrated with the Technology Integrated Learning Environment, TILE, for the delivery and structuring of learning material.

With our work we want to complement existing approaches to content representation that focus on the description on a learning object as one entity. We want to continue the use of existing approaches to ensure exchangeability and provide additional value by accessing the cognitive concepts within a learning object.

References

[1] Gehne, R., Jesshope, C.R. and Zhang, J. Technology Integrated Learning Environment - A Web-based Distance Learning System. Accepted by IMSA 2001, Hawaii, USA.

[2] GENTLE. http://wbt-3.iicm.edu/. Accessed February 2001.

[3] Multimedia Annotation. http://www.sigmatics.co.jp/. Accessed January 2001.

[4] Ram A., Catrambone R., Guzdial M.J., Kehoe C.M., McCrickard S., Stasko J.T. (1999). PML: Adding Flexibility to Multimedia Presentations. IEEE Multimedia, April – June 1999.

[5] Merrill, M. D. (1998) Knowledge Objects. CBT Solutions Mar/Apr 1-11.

[6] LOM Standard (2000). Draft Standard for Learning Object Metadata. IEEE P1484.12/D4.0. Available from http://ltsc.ieee.org/doc/wg12/LOM_WD4.doc.

[7] Dublin Core Metadata Initiative (2001). http://dublincore.org/. Accessed February 2001.

[8] IMS Learning Resource Meta-data Specification Version 1.2 (2001). Available from http://www.imsproject.org/metadata/index.html.

[9] Heinrich E., Kemp E. and Patrick J.D. (1999). A Natural Language Like Description Language. *10th Australasian Conference on Information Systems (ACIS*). B. Hope and P. Yoong (Eds.). School of Communications and Information Management, Victoria University of Wellington, Wellington, New Zealand. P. 375 – 386.

[10] Sowa, J.F. (2000) Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks/Cole; Pacific Grove, USA.

[11] McArthur D., Lewis M. and Bishay M. (2000). The Roles of Artificial Intelligence in Education: Current Progress and Future Prospects. http://www.rand.org/hot/mcarthur/Papers/role.html, accessed September 2000.

[12] Heinrich E., Jesshope C., Walker N. Teaching Cognitively Complex Concepts: Content Representation for AudioGraph Lectures. Proceedings of EdMedia 2001. C.Montgomerie, J. Viteli (Eds.), Association for the Advancement of Computing in Education.

Acknowledgements

We would like to acknowledge the support for this project from the New Zealand government's New Economy Research Fund (NERF) under contract MAUX9911. Without this support, this project would not have been possible.