UNIVERSITY
OF TRENTO

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

38050 Povo – Trento (Italy), Via Sommarive 14
http://www.dit.unitn.it

DE-FRAGMENTING KNOWLEDGE: USING METADATA FOR INTERCONNECTING COURSES

Marco Ronchetti, Alberto Giuliani
and Paramjeet Saini

May 2003

Technical Report # DIT-03-019

DE-FRAGMENTING KNOWLEDGE: USING METADATA FOR INTERCONNECTING COURSES

Marco Ronchetti, Alberto Giuliani and Paramjeet Saini
Dipartimento di Informatica e Telecomunicazioni
Università degli Studi di Trento, 38050 Povo (Trento)
Italy

ABSTRACT

E-learning systems are often based on the notion of "course": an interconnected set of resources aiming at presenting material related to a particular topic. Course authors do provide external links to related material. Such external links are however "frozen" at the time of publication of the course.
Metadata are useful for classifying and finding e-learning artifacts. In many cases, metadata are used by Learning Management Systems to import, export, sequence and present learning objects. The use of metadata by humans is in general limited to a search functionality, e.g. by authors who search for material that can be reused.
We argue that metadata can be used to enrich the interconnection among courses, and to present to the student a richer variety of interconnected resources. We implemented a system that presents an instance of this idea.

KEY WORDS
Metadata, Knowledge Management

1. INTRODUCTION

E-learning systems are often based on the notion of "course": an interconnected set of resources aiming at presenting material related to a particular topic. Frequently courses become "islands", and the e-learning environment becomes an archipelago of disconnected entities. Course authors do provide some form of "external interconnections", e.g. links to related material. Both internal and external links are "frozen" at the time of publication of the course: however the external world changes continuously and there might be new related resources that remain unknown to the static world of the course. Of course they can be updated every now and then, but this requires that the author is aware of what related material is available. Therefore an e-learning platform ends up being a container of these knowledge fragments, which are weakly interlinked. The main structural link often remains the fact that courses belong to a predefined path that is the student's curriculum. Links between similar courses belonging to different curricula in the same university are however often completely missing, even when the covered topics are similar. This is probably more likely to happen in university systems strongly based on Schools (like in Italy) rather than primarily the on Departments (like in USA).

A tool for allowing authors to find related material exists: metadata. In recent years several standards have come out to define a common convention for metadata definition and representation (IEEE LTSC P1484.12, AICC AGR, ADL SCORM, Dublin Core Metadata Initiative etc.). Some metadata are "container oriented", i.e. they are meant to inform the learning management system about format, sequencing etc. of the material. These metadata are objective, and therefore rather easy to define and use, but do not solve our problem. Other metadata are meant to facilitate the use of the material by teachers, by curriculum designers and by producers. Among these, there are data description about the type of granularity of objects, pedagogical purpose, assessments and learning objectives, etc.: all of these metadata represent subjective interpretations of resources, leading to serious problem in the definition of an agreed upon value. An interesting point of view is expressed by Nilsson et al. [1,2]: they argue that the image of meta-data as being objective information about data is wrong, or at least incomplete. They support the existence of multiple, even conflicting descriptions, and maintain that the RDF technology allow to implement and use such subjective metadata.

We are interested here in metadata that describe the content i.e. that are able to convey the notion of "what the learning object is about". Such metadata present both the problem of objectivity, and the problem of describing them using a shared language (a common ontology). We discuss this last point in another paper [3], and assume here that this problem is solved: i.e. that there is a common language for describing the argument that a learning object is about. We will also assume in the following a given granularity for a learning object: for us it will be a self-contained portion that explains one or more concepts. With such choice of granularity, an image by itself will not be considered a significant learning object.

In such scenario, it would be possible to solve the knowledge fragmentation problem: authors would be able

to search for material that is related to the topic they discuss, and they could put references that interlink their courses with related ones.
The problem is that this author-driven process is costly, and that the result is static. In the present paper we present an approach that allows automating the process, and that produces a result that remains up to date in time.

2. USING METADATA FOR AUTOMATIC GENERATION OF KNOWLEDGE INTERLINKING

We outline here our idea. The main assumption is that we have an ontology (that at minimum should be a taxonomy), and that it is possible to use it to index lectures or lecture fragments. Whenever a student requests a page containing a lecture fragment, the page is automatically decorated by adding (e.g. at the end) an active point (e button, or a link) that connects to another page that is generated on the fly. The process that generates the new page goes through the following steps:

1. look up the indexes, and retrieve the content description metadata M, i.e. the particular metadata that describe the content of the original page in terms of the ontology;
2. look for all the other Learning Objects that have the same value of the metadata M, and build a list of them
3. explore ontological space by traversing the relations that link the content description metadata M with a set of other content description metadata $\{M_1 \ldots M_k\}$
4. for each of the values $\{M_1 \ldots M_k\}$ look for all the other Learning Objects that have the same value for the content description metadata, and build a list of them.
5. if requested, iterate though points 3 and 4.
6. summarize the information that has been found and present it to the requestor.

In other terms, what happens is the following:
Step 1 finds out what is the topic T of the page.
Step 2 finds all the pages that are known to the system, that are about the same topic T.
Step 3 finds the set of topics $\{T_1 \ldots T_k\}$ that are related to the original topic.
Step 4 finds all the pages that are about each of the related topic $\{T_1 \ldots T_k\}$
Step 6 presents the result. The resulting page contain a list of URLs (links) ordered by relation.

Since the resulting page is generated on the fly, it is always u to date: as soon as a new page is added to the system and it is labeled with suitable metadata, it becomes visible as "related topic". The process in not obtrusive from the user point of view, because it is only triggered on demand (i.e. when the user presses the button that might say something like "show me related material".

The nature of the relations that are traversed depends on the ontology. At the minimum the ontology should classify material as belonging to a certain topic, and present relations among topics (e.g. taxonomical relations), but certainly it would be preferable to have a richer ontology that contains other relations (like for instance "is used by", "requires knowledge of" etc.). In the case of such a richer ontology, the system could offer to the user the option of selecting one particular kind of relation.

Moreover, the production of the final page could be filtered according to some user's choices. For instance, the user could ask to see only the pages that belong to the same university, or that are produced by a certain author, or that were added to the system after a certain date.

Finally, the system could include the possibility of explicitly navigating the ontology by using a "concept map navigator"-like tool. We would expect that such kind of navigation would not be useful for students, being probably too abstract, but that it might be useful for teachers and authors for exploring the whole learning management system space.

3. THE ARCHITECTURE

We implemented a prototype [4] that allows exploring most of the ideas outlined in the previous section.
A keystone of the system is the chosen ontology. We focused on a particular domain: the subset of Computer Science that is thought during undergraduate studies. We used the work that some of us had previously done, i.e. the definition of ontology for such domain [3]. Our ontology was extracted from the huge work performed by a task force created by the Association for Computing Machinery (ACM) and described in the "ACM Computing Curricula 2001" [5]. Details about the reason why we chose the definition of suggested Computer Science curricula as our starting point are given elsewhere [3]. Here we only say that the resulting ontology is at present a taxonomy based on Areas, Units and Topics. The only relation is therefore at present a "part-of", but we plan to work to enrich the taxonomy by adding other types of relations among topics and among units. For our goal, that is for showing a prototype that implements our ideas, such ontology is sufficient. Also, it turns out to be very useful in practice (we are currently using it for indexing all the on-line material in the area of Computer Science at our university).
The key architectural ingredients for our prototype are:

?? The LMS: a standard learning management system that provides access to learning resources (in the simplest case a simple WebServer). It acts as Learning Object Storage
?? The MD-Storage: a database for collecting metadata;

?? The Engine: a collection of JSP pages that interrogate the MD-Storage to dynamically creates the pages for presenting the related topics

?? The Broker: a dynamic web server that interacts with the MD-Storage

?? The Navigator: an applet that allows to directly navigate the ontology.

The user has two possible starting points. The simplest is the Navigator. The user can navigate the applet passing through areas, units and topics. For each topic (the lowest hierarchical level in the taxonomy) the Navigator can ask the Engine to retrieve from the MD-Storage a list of the Learning Objects that are related to it. This list is then formatted by the Engine, that produces the final view: a (possibly filtered) collection of Learning objects related to the chosen topic.

We implemented two versions of the Navigator: one presents a taxonomical view (well suited to the present stat of the ontology) and the second allows navigating in the style of concept maps. This second option is better suited to a richer ontology, which exhibits the structure of a graph structure rather than that of a tree. Figure 1 shows a snapshot of this second option.

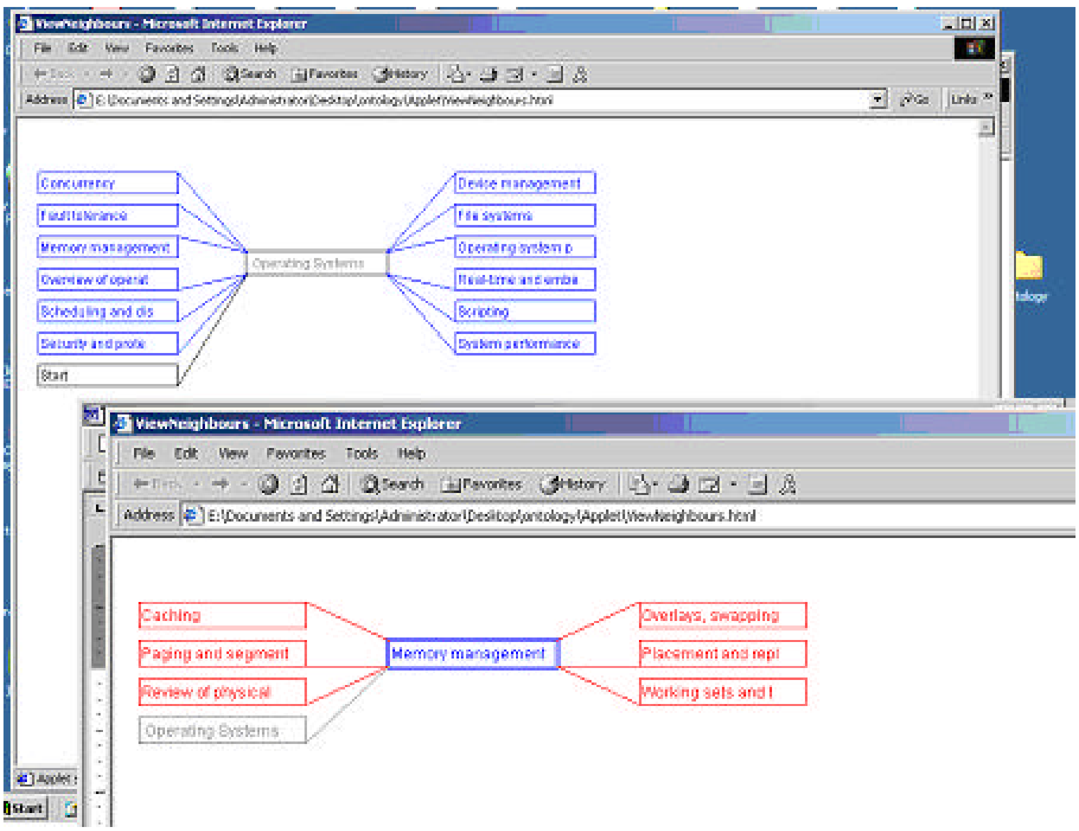

Figure 1: Navigating metadata in the ontological space with our tool.

The other possible starting point for the user is a common page served by the LMS. In our architecture, the page is never directly requested to the LMS, but is requested to the Broker. The Broker retrieves the page from the LMS, and decorates the page with an active element: a button or a link that the user can exploit to talk back to the Broker for requesting additional information. When the user decides to exploit this possibility, the Broker activates the Engine, which in turn starts a process that interrogates the MD-Storage following the sequence of steps 1 to 6 described in section 2. At the end of the process the Engine is able to present (in an additional Browser window) the list of URLs representing the related learning objects (e.g. lectures or lecture fragments) that the user expects. Each of these URLs interacts with the Broker, starting a new iteration of the process that we described here.

Such system allows using the LMS in a direct and traditional way, without the benefit of the knowledge extracted from the metadata, or through the Broker that enables the whole process we described.

We should also say that the page with the referenced material also presents a link to the Applet, so that the user can at any time switch between the explicit metadata navigation allowed by the Applet and the implicit one provided by the Broker.

We must also add that having a single LMS is not a requirement: our architecture can rely on several heterogeneous LMS's, and it is able to interconnect learning material regardless its location. All that is requested is that the resources metadata are defined, i.e. that the MD-Storage has a notion of the available learning objects and knows their location.

A possible extension that we are presently considering is to have a distributed MD-Storage, possibly based on peer-to-peer technology. The only requirement is that the ontology be shared. Even such requirement could be released by introducing the notion of a Mapper, which interrelates different ontologies.

At present we implemented the architecture we described and used the system for indexing approximately 1000 learning objects. Considering that our ontology consists of approximately 1000 topics, this means that at present we have some sufficiently well covered areas, while other areas are not covered by learning objects (or they are not yet described by the content description metadata. Our plan is to progress in parallel to the project of putting on line the entire curriculum of CS at our university. In the meantime, we intend to use our prototype to study its acceptance and practical usefulness for our students.

The present work is somehow similar, in aim, to initiatives connected with applications of the Semantic Web to e-learning (see e.g. [6]). Our approach however does not rely on the typical Semantic Web technology, and its implementation is available today. Moreover, often the use of metadata is aimed at helping author discovering resources for reusing them as material for producing new lectures (see eg. [7]). Our approach is rather to aid students to find material that presents different points of view or different ways to explain thing.

3. CONCLUSION

We described a way to automatically generate glue that interconnects "knowledge fragments". We started from the consideration that frequently in on-line learning initiatives the focus is on providing self-contained courses, which often have little or no relation with other courses. Even when the relation is present, it is typically static; that means than new material that is added to the repository can relate to existing material while the reverse is not true. Also, finding and implementing such relations requires human interventions and is time expensive.

Therefore we designed a system that can discover relations and exposes them, provided that metadata describing the content of the learning objects exist and that they are defined according to an agreed-upon ontology.

We described the logical architecture of the system, and presented an implementation that indexes lectures on Computer Science at our university, according to an ontology that we describe in detail elsewhere.

Our system can be interfaced with any learning object repository, even distributed multiple ones. At present our system relies on a centralized metadata repository, but we suggested some possible extensions in the direction of peer-to-peer distributed architecture.

REFERENCES

[1] Nilsson, Mikael, Palmér Matthias and Naeve Ambjörn, Semantic Web Metadata for e-Learning - Some Architectural Guidelines, *WWW2002 International Conference*

[2] Naeve, Ambjörn, Nilsson, Mikael, Palmér Matthias, E-learning in the Semantic Age, Proceedings of the 2nd European Web-based Learning Environments Conference (WBLE 2001), Lund, Sweden 201

[3] Ronchetti, M., Saini, P., Ontology-based metadata for e-learning in the Computer Science domain, to be published.

[5] Giuliani, A., Tesi di Laurea in Informatica, Universita' di Trento, 2003

[5] Computing Curricula 2001 Computer Science, *ACM Journal of Educational Resources in Computing, 1*(3):1-240, Fall 2001.

[6] Stojanovic, L. Staab S., Studer R. E-learning based on the Semantic Web In WebNet2001 - World Conference on the WWW and Internet, Orlando, Florida, USA, 2001

[7] Lytras M.D., Semantics for E-learning: An advanced Knowledge Management oriented metadata schema for leaning purposes, Procedings of the E-Learn 2002, Montreal, Canada, 2002