

Reusing Learning Resources based on Semantic Web Technologies

Sylvain Dehors¹, Catherine Faron Zucker², Rose Dieng Kuntz¹

¹Acacia, INRIA Sophia Antipolis, 2004 route des Lucioles, BP93, 06902 Sophia Antipolis, France
{Sylvain.Dehors, Rose.Dieng}@sophia.inria.fr

²Mainline, I3S, 930 route des Colles, BP 145, 06930 Sophia Antipolis cedex, France
faron@essi.fr

Abstract

Reusing existing web resources for e-learning is a very promising and highly promoted idea in the research field of web-based education, especially for intelligent or adaptive systems where the cost of authoring is tremendous. However methodologies and tools are still lacking and feasibility at reasonable cost is a pending issue. In this paper we propose an original approach to reuse learning resources based on knowledge engineering and Semantic Web technologies. The QBLS platform reuses learning resources available on the web and pedagogical ontologies to create a convenient learning system for assignment sessions. QBLS is being used on actual curricula at the EPU engineering school of Sophia Antipolis.

1. Introduction

Today, the spread of “intelligent” or “adaptive” systems for learning is not as advanced as expected. On the one hand e-learning standards, like LOM and SCORM, have emerged to annotate and package learning content. But they mainly deal with technical aspects and do not express much information about pedagogy. On the other hand, most of advanced systems need a set of annotated learning resources, where annotation is based on very precise and rich models. Knowledge of the domain to learn is coupled with knowledge of the planned pedagogy and the cost of human effort necessary to create models and connect them to learning resources seems excessive for a large diffusion.

In the vision of the Semantic Web [4], humans and machines share semantic data across the world. Transferred in the context of e-learning this would mean the reuse and exchange of courses [7]. The investigation proposed in this paper shows how knowledge engineering and semantic web technologies allow some reusability. We do not propose theoretical guidelines to use semantic web technologies in e-learning [11] but rather an original and realistic solution for reusing learning resources. Our aim is to *reuse pedagogical material by annotating it* [13].

We have developed the QBLS (Question Based Learning System) platform to demonstrate this approach. Both learning resources and pedagogical ontologies available on the web are reused, relieving the teacher from the burden of authoring from scratch a complete dedicated course content. QBLS is supporting lab sessions, helping students accessing and understanding their course material to perform exercises. QBLS is being used on actual curricula at the EPU engineering school of Sophia Antipolis¹.

2. A Real World E-Learning Problem

We address the following general scenario of a teacher preparing a course. He searches the web, uses learning object repositories [10] or contact colleagues for available pedagogical material. Among the results of his search he selects a slide show that, as far as he can tell, contains all the important concepts of the course. In order to help students stuck with paper handouts and experiencing troubles finding information, the selected learning content should be made accessible through an “intelligent” web system. A presentation of the pedagogical content based on explicit semantics should foster students in reading and understanding the course. This general scenario is quite common and covers a wide range of real situations in higher education, like the ones we have met at the EPU engineering school of Sophia Antipolis.

By now there are few annotated learning resources available on the web and, above all, annotations may not perfectly suit the contexts in which they could be reused. As a result we claim the need for *teacher input* when reusing learning resources, to *adapt or extract semantic information*. The amount of information (and work) required from the teacher must be kept to a minimum. It is unrealistic to foresee that e-learning systems will dispose of large amounts of manually authored semantic metadata.

Tools exist that help teachers authoring courses and annotating them [2]. Experiments have shown the

¹ <http://www.essi.fr>

inefficiency of form-based annotation tools [3], whereas markup-based annotation appears much more acceptable. However, a generic annotation tool like Annotea for example, is too complex and not straight forward enough to be used in that scope. Based on this observation we propose a solution offering a reasonable alternative to the development of custom *markup tools*.

3. Reusing by Annotating

The proposed methodology advocates *markup-based annotation* of learning resources. For this, we rely on existing editing tools. Teachers are familiar with tools like MS Word or OpenOffice writer. Using automatic styles enables to annotate the content of learning resources through *visual markers*. This connects the semiotic level to the semantic one. As described in Figure 1, the process is threefold: (1) The learning resources to be reused are preprocessed to be displayed in a standard editing tool with an XML storage format and sufficient functionalities (for example slides have to be transformed into text documents). (2) The teacher annotates the content of the resources using the editing tool. Annotations are based on a pedagogical ontology (displayed as a paragraph style hierarchy) and a domain vocabulary (key concepts are identified through text styles put on some terms: bold, italics, etc.). This distinction between pedagogical and domain knowledge follows [9]. (3) The XML of the edited text is transformed with an XSLT stylesheet: The semantic information captured in text and paragraph styles are automatically translated into RDF annotations (see Figure 3) and the content of the resource is translated to XHTML. This translation phase is the key to make learning resources *reusable and accessible in a specific educational context*.

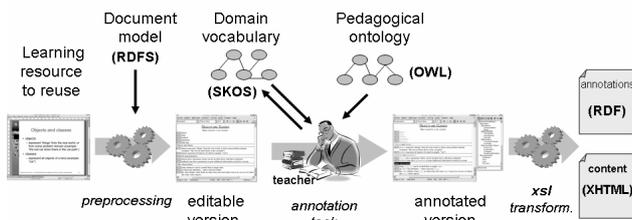


Figure 1. Overview of the annotation method

This methodology is absolutely generic and can be applied in various situations. Moreover its modularity guarantees flexibility. We have applied it successfully in two different experiments, in one case [6] by using Microsoft Word documents and in the other case, detailed in this paper, by using slides edited with OpenOffice. The QBL system that implements our methodology needs little to no adjustment at all to be deployed again. Parts of the process we propose have already been experienced in other works [2][3][13]. But our approach takes advantage of semantic web technologies to unify the process and exploits the possibilities of editing tools much further.

Figure 2 illustrates this annotation task. On the right window, paragraph styles are organised in a hierarchy reflecting the main organisation axis of the pedagogical ontology (see section 4). On the left, a short resource is being annotated by selecting the title “Method and Parameters” and setting its paragraph style to “Definition”.

In the underlying XML format, the style attributes identify the nature of the element and allow associating it with a semantic description referring to concepts from an ontology or a vocabulary. This way of annotating is a very powerful but somehow limited technique as it requires the occurrence in the textual content of a term identifying this concept. Content authored using ellipses (when the subject is under-stated) could not be annotated this way. Also, a resource must play a single role, whereas in some cases, a slide plays several roles with respect to a concept. The teacher has then to choose the most prominent aspect. Despite these limitations our experience is that a large number of courses perfectly suits this process.

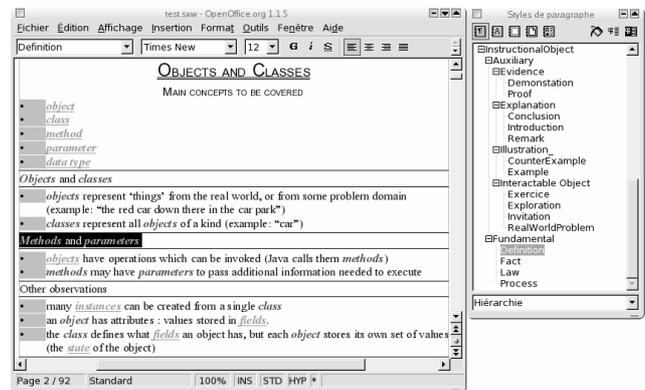


Figure 2. OpenOffice as an annotation editor

Word processors offer more graphical possibilities than standard HTML (especially drawings), thus we developed a XSLT transformation that renders some of them, like arrows and text-boxes. This is by far the most technical aspect of this extraction mechanism, but the XSLT stylesheet we have written is generic, and re-applying our method on different courses does not require any change to it. The stylesheet for the RDF transformation is also completely generic and only relies on the naming of the styles. Applying it to another course only requires to follow the same document organisation (set of paragraphs separated by titles) and style naming conventions.

The Semantic Web languages (RDF, OWL, SKOS) provide us with the means to express those annotations and their conceptual meaning. They act as pivotal formalisms to exchange models and knowledge across the web. It must be noted that other classical design models for learning systems, identified in [3], can perfectly be expressed using these formalisms. So our experiments demonstrate the feasibility of turning an existing generic editor into an annotating tool using Semantic Web languages.

4. Ontological Design

In this section we describe the ontological foundations of the models used in our experiences.

Design Model and Document Model: Following the original document structure, we distinguish three levels of granularity: the course, the chapter and the slide. This is in essence a very generic model and guarantees that it could be applied to a very wide range of courses. The indexing model follows the “fragment indexing approach” [3] and uses two properties proposed by the SKOS schema [15]: “skos:primitiveSubject” and “skos:subject”. They differentiate between resources explicitly related to a concept and the ones just referencing this concept.

Pedagogical Model: Part of the advantages of intelligent web based learning systems lies in their capacity of supporting a given pedagogical strategy. This strategy emerges through an ontology that defines a conceptual approach to learning. Such a model can be more or less generic and may rely on existing taxonomies, like Bloom’s [1]. When annotating learning resources the teacher creates an expression of this model. In the context of the semantic web, he creates instances of the concepts defined in the ontology.

In our experiment we reused the ontology proposed by [12]. It is expressed in OWL and describes the “particular instructional role a learning object can play”. Generally speaking the choice of the *knowledge model* must be motivated by both the system requirements and the limitation of the annotator’s workload. Reusability at the ontological level is one of the Semantic Web visions. Anyone can refer to a concept of an ontology by using its URI. It does not solve contextual problems with semantic annotations [2]. But at least we proved that a *pedagogical conceptual model can be reused*.

Domain Model: This model expresses the conceptualization of the domain to learn from the teacher’s point of view. The author of the original document might not completely share the same domain model. Thus finding a trade-off between these two conceptualizations illustrates the idea of ontological consensus necessary in knowledge management. The way the domain is structured supports a vision that should help understanding the learning resources and thus learning with them. This is a quite different goal from giving a well structured vision of the domain that could be “The Ontology” of the domain. On the one hand, generic ontologies might not be suitable for beginners [14]. They often contain too much detailed information or are structured along axes not yet known to learners. On the other hand, the ontologies we found in the domains of our experiments (Java [8] and Signal analysis) are not generic ones but are closely related to specific documents or applications. These examples lack the generality to be reused. Thus it is still undetermined if the domain model can be shared and to what extent. Figure 3 shows an annotation in RDF that references the

pedagogical ontology (edu-onto namespace), the domain ontology (java-skos namespace) and uses relations from the document model (edu:number and edu:belongsTo).

```
<rdfs:Resource rdf:about="&prog;chap1fic3.xhtml">
  <rdf:type rdf:resource="&edu-onto;Definition"/>
  <edu:number>3</edu:number>
  <dc:title>Methods and parameters</dc:title>
  <skos:primarySubject rdf:resource="&java-skos;method"/>
  <skos:primarySubject rdf:resource="&java-skos;parameter"/>
  <skos:subject rdf:resource="&java-skos;Object"/>
  <edu:belongsTo rdf:resource="&prog;ObjectsAndClasses"/>
</rdfs:Resource>
```

Figure 3. RDF annotation generated from styles

5. Deploying Semantic Web Technologies

Once annotated according to the specific pedagogical context in which we want to reuse them, the learning resources are exploited by Semantic Web technologies. One of the advantages of this approach is that external tools can be easily integrated. For example in QBLs the organisation, the retrieval of the resources and user adaptations are all based on the use of the Corese semantic search engine [5]. Corese is based on conceptual graphs for internal reasoning and takes as input RDFS or OWL-Lite ontologies and RDF statements. It accepts the SPARQL² query language for RDF and answer semantic queries with RDF graphs either in RDF/XML serialisation or in the XML binding defined in SPARQL.

No modification on this tool was necessary to import the ontology, neither was there any problem to load the definition of the SKOS domain vocabulary or our generated RDF annotations. The architecture of the QBLs system is shown in Figure 4. Students access the course content through an interface displayed in a web browser. The browser sends requests over the web to a distant server using JSP/servlet technology. The server builds pages by querying its embedded Corese instance. The XML output given by the engine is processed through an XSL stylesheet to generate XHTML finally piped to the result page.

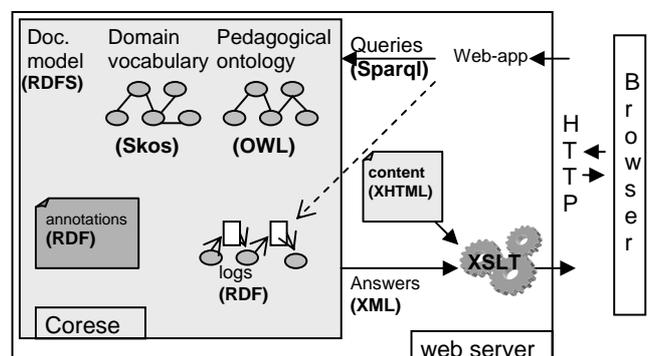


Figure 4. Architecture of QBLs

² <http://www.w3.org/TR/rdf-sparql-query/>

5.1. Dynamic display of learning resources

The SPARQL query used to dynamically construct the web page retrieves resources relative to a given domain concept and belonging to a given chapter, as shown in Figure 5. In this query we ask for (1) a given concept (<%=concept%> stands for the value of the concept parameter), (2) its label, (3) the documents whose primary subject is this given concept or a narrower concept in the SKOS vocabulary, (4) the type and title of these documents, (5) we precise that the documents must belong to the current given chapter. The interest of using a semantic search engine is to take advantage of the semantics of the model. For instance, it automatically computes the transitivity of the skos:broader property. Thus in (3) we reach all the descendants of the current concept in the vocabulary hierarchy, no matter how deep they are in this hierarchy.

- ```
(1)?concept = <%=concept%> .
(2)?concept skos:prefLabel ?conceptLabel .
(3)FILTER (?doc skos:primarySubject ?concept or
?doc skos:primarySubject ?c2 . ?c2 skos:broader ?concept))
(4)?doc rdf:type ?t .
?t rdfs:label ?docLabel .
?doc dc:title ?docTitle .
(5)?doc edu:belongsTo <%=chapter%> .
```

Figure 5. A semantic query to retrieve resources

### 5.2. Dynamic hyperlink between resources

When concepts appear in the content of the resource we want to create hyperlinks towards those concepts, provided they are relevant in the current context. So we further query Corese to establish whether appearing concepts are primary subjects of other resources belonging to the same chapter.

- ```
(6)?doc skos:subject ?ext_concept .
(7)?ext_doc edu:belongsTo <%=chapter%>
{?ext_doc skos:primarySubject ?ext_concept or{?ext_doc
skos:primarySubject ?c3 . ?c3 skos:broader ext_concept}} .
```

Figure 6. A semantic query for dynamic hyperlinks

5.3. User adaptation

User profiles are used to determine the ordering of the resources in the interface depending on the pedagogical role assigned to each of them. In a given profile a type of resource is associated with a ranking node (in RDF). The integer value hold by this node is used to sort the results of the whole query. Values are determined automatically. Each profile specifies pedagogical relations like “Fundamental resources have priority-upon Auxiliary ones”. Then a set of rules exploits the ontology to complete this information. For example Definition, a subtype of Fundamental, is inferred to be prior upon Illustration which is a subtype of Auxiliary. For any couple of resource types the Corese engine can answer which one is to be displayed first. This creates an order that is used to generate ranking numbers. Figure 7 presents the query to Corese enabling to sort the resources retrieved for one domain concept. If the type (?t) of the resource does not determine the order, the original order in which the slides were sequenced is used.

- ```
(7) OPTION{?user = <%=user%> . ?user edu:profile ?profile .
?profile edu:ranking ?r . ?r edu:concept ?t . ?r edu:order ?order}
?doc edu:number ?number
```

Figure 7. A semantic query to adapt the organisation of the learning resources

As a matter of fact the presentation layer does not perform any logical inference that would be hard-coded in XSLT. This makes QBLs a very generic system. A good proof of its generality is that we did only look and feel adaptations between two different experiments on totally different domains (Java and Signal Analysis).

To monitor the activity and propose more adaptation (displaying the history, computing “conceptual” back links) the actions of the user are recorded and added directly into an RDF graph. Corese is then able to answer if a given user has already visited a document and when.

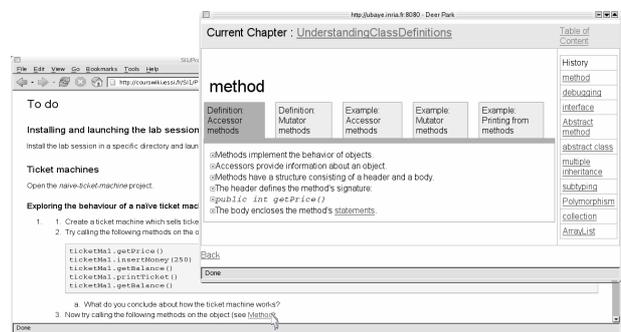


Figure 8. Accessing QBLs from a wiki page

## 6. QBLs at Work at the EPU

QBLs<sup>3</sup> is actually deployed at the EPU engineering school of Sophia Antipolis to support Java programming lab sessions. It reuses the available BlueJ pedagogical material, a set of 430 slides divided into 15 chapters. This material was annotated to provide access to the course during assignment sessions to 80 first year engineering students. The sessions were bi-weekly 2h labs in small groups, supervised by a teacher.

Labs are collaboratively written by teachers using a wiki. QBLs is then accessed through wiki pages where teachers collaboratively add links to provide entry points to the course content in QBLs. Entry points are either hypertext links to chapters or specific concepts. Clicking them opens QBLs in a new browser window (see Figure 8). Many students accessed QBLs outside lab hours, showing the interest of a web based system. The recorded daily usage during a whole semester and the number of users demonstrates the scalability of the approach.

The whole annotating process took around 20 hours to the teacher. Considering the number of resources

<sup>3</sup> This version of QBLs is freely accessible on the web at <http://ubaye.inria.fr:8080/prog101/cours>

annotated, this is a quite efficient process. Nonetheless it requires a substantial effort from the teacher to perform this task. We could not have asked her to give more human input than what is described here, confirming the idea of focusing the annotation task and reusing as much as possible. A comfortable display of the ontology helping the teacher in discovering and getting familiar with the ontology would also be appreciated. This is yet another strong argument for the integration of third party tools, thanks to the Semantic Web, like an ontology browser in this case.

## 7. CONCLUSION

In this paper we address the problem of reusing learning resources in intelligent or adaptive learning systems. To answer it we propose a methodology, applied in a real system, to reuse learning resources by extracting semantic information from them. We focus on offering a scalable, reusable and reasonable annotation method, a question that has not been fully considered yet in research on intelligent web based systems [8] [14].

Our methodology recommends markup-based annotation relying on existing editing tools. Annotations are based on ontologies represented through the style hierarchy of the editing tool. Most editing tools are now provided with an XML format for storage. We used it to transform the semantic information captured in text and paragraph styles into RDF annotations. This is the key to make learning resources reusable in a specific educational context. The reuse itself of the annotated learning resources in our QBLS platform is based on an extensive use of semantic web technologies to organize, retrieve and display the learning resources.

This is an original research by its connection with a real world application and the integration of the latest semantic web technologies [11]. This method does not only apply to this system but to most systems relying on a pool of annotated resources of that kind. The extraction mechanism is also generic and could be reused for other systems provided we can map the models used in them to the ones presented here, which should be the case for a large number of them. More generally many existing results in adaptive hypermedia could be reproduced using this single generic approach and technology.

Our ongoing work addresses the problem of adaptation with user profiles to enhance the system adaptive features. We want to base adaptation on the extensive use of the inference capabilities offered by the semantic search engine Corese. Our aim is to collect more knowledge about the users, especially activity traces. We currently express those

traces into RDF graphs. As Corese loads them it has access the whole user history. We also have in prospect the integration of heterogeneous and possibly redundant learning resources into QBLS.

## REFERENCES

- [1] Anderson, L.W., Krathwohl, D.R. (Eds.) A taxonomy of learning, teaching, and assessment: A revision of Bloom's taxonomy of educational objectives, New York, Longman, 2001.
- [2] Azouaou, F., Desmoulins, C., Semantic Annotation for the Teacher: Models for a Computerized Memory Tool. SWEL workshop, AIED2005, Amsterdam, The Netherlands, 2005.
- [3] Brusilovsky, P. Developing adaptive educational hypermedia systems: From design models to authoring tools, in T. Murray, S. Blessing and S. Ainsworth (eds.): Authoring Tools for Advanced Technology Learning Environment. pp. 377-409, 2003.
- [4] Berners-Lee, T., Hendler, J., Lassila, O. The Semantic Web. Scientific American, May, 2001.
- [5] Corby, O., Dieng-Kuntz, R., Faron-Zucker, C. Querying the Semantic Web with the CORESE search engine. Proc. of the ECAI'2004, subconference PAIS'2004, pp 705-709, Spain, 2004.
- [6] Dehors, S., Faron-Zucker, C., Giboin, A., Stromboni, J.P. Semi-automated Semantic Annotation of Learning Resources by Identifying Layout Features. In SWEL workshop, AIED2005, Amsterdam, The Netherlands, 2005.
- [7] Devedzic, V., Education and the Semantic Web, in International Journal of Artificial Intelligence in Education, Vol.14, pp. 39-65, 2004.
- [8] Henze, N. Personal Readers: Personalized Learning Object Readers for the Semantic Web, in AIED2005, Amsterdam, The Netherlands, 2005.
- [9] Mizoguchi R., Bourdeau J. Using Ontological Engineering to Overcome AI-ED Problems, in Int. J. of Artificial Intelligence in Education, Vol.11, No.2, pp. 107-121, 2000.
- [10] Quemada, J., Huecas, J., Miguel, G., Salvachua, T., Fernandez, J., Simon, B., Maillet, B., Law, K. EducaNext: A Framework for Sharing Live Educational Resources with Isabel, in Int. World Wide Web Conference, USA, 2004.
- [11] Stojanovic, L., Staab, S and Studer, R. E-Learning Based on the Semantic Web, in Proceedings of WebNet2001 - World Conference on the WWW and Internet, Florida, USA, 2001.
- [12] Ullrich, C. Description of an instructional ontology and its application in web services for education, in SWEL Workshop, ISWC'2004, pp. 17-23, Japan, 2004.
- [13] Verbert, K., Jovanovic, J., Gaevic, D., Duval, E., Repurposing Learning Object Components, in OTM'2005, Workshop on Ontologies, Semantics and E-Learning, Cyprus, 2005.
- [14] Winter, M., Brooks, C., Greer., J. Towards Best Practices for Semantic Web Student Modelling, in AIED2005, Amsterdam, The Netherlands, 2005.
- [15] SKOS Core Guide, <http://www.w3.org/TR/2005/WD-swbpskos-core-guide-200505>