SEMANTIC WEB TECHNOLOGY FOR WEB-BASED TEACHING AND LEARNING: A ROADMAP

Edmond Holohan
Dublin City University
School of Computing
Dublin 9, Ireland
Edmond.Holohan@dcu.ie

Claus Pahl
Dublin City University
School of Computing
Dublin 9, Ireland
Claus.Pahl@dcu.ie
http://www.computing.dcu.ie/~cpahl

ABSTRACT
The World-Wide Web has become the predominant platform for computer-aided instruction. Content-orientation, access and interactive features have made the Web a successful technology. The Web, however, is still evolving. We expect in particular Semantic Web technology to substantially impact Web-based teaching and learning. In this paper, we examine the potential of this technology and how we expect it to influence content representation and the work of the instructor and the learner.

Keywords
Web-based instruction, Semantic Web, Ontology, XML, Content representation, Design and delivery.

1. INTRODUCTION
The emergence of the World-Wide Web has changed the way we represent content and the way we deliver courses. However, the development of the Web is not complete. The Web will continue to evolve dramatically in the future and we can expect more changes for teachers and learners alike.

One of the current major activities in the Web community that we consider as the most important one for the educational context is the Semantic Web initiative [2,10]. The Semantic Web initiative aims to bring semantics to the Web. Clearly defined concepts and the possibility to reason about concepts and other forms of knowledge are the objectives. Based on an underlying, already extremely successful technology – the eXtensible Markup Language XML [11] – it provides knowledge presentation and inference techniques for a shared, distributed environment such as the Web.

4th Annual LTSN-ICS Conference, NUI Galway

Our objective in this investigation is to outline a roadmap to the possible future uses of Semantic Web technology. This is based on experience we have already had using XML and Semantic Web technology for educational and other applications.

Today, the Web is the predominant infrastructure platform for computer-aided teaching and learning. It is the ideal environment for distance education through its ubiquity. Moreover, the Web is also an ideal platform to support or substitute classical forms of teaching and learning.

Distance education and other forms of teaching and learning in a distributed environment make use of the Web as a world-wide information access infrastructure for course delivery. In the future this aspect of distribution and sharing will also impact the instructional design of Web-based courseware. The Semantic Web will in particular be beneficial for the way we create, provide, and deliver content.

2. METADATA AND STRUCTURE

2.1 XML
The foundation on which the Semantic Web is based is XML, the eXtensible Markup Language. XML [11] is a framework that allows users to define application-specific markup languages. The key objective is interoperability of documents and data.

Central for XML are Document Type Definitions (DTD), which allow us to define an application-specific vocabulary (in the form of tags) and a grammar for the syntactical use of these tags (in terms of hierarchy, sequence, repetition, option, etc). Documents or data are then described based on the defined vocabulary and rules in XML files.

2.2 Metadata, Content, and Instruction
A number of different applications of XML in teaching and learning immediately come to mind:

- Metadata. Metadata is data about data, e.g. data about learning objects such as courses. These learning objects are described in terms of attributes such as author/instructional designer, institution, subject, or expected audience and required skills and knowledge [3].

Metadata information aims to support retrieval and comparison of learning objects. Course providers can describe their courses in abstract terms. Made available in accessible repositories these specifications can be searched by instructional designers or by learners in search for suitable courses on a particular subject. The IMS project [3] for example defines a learning object metadata standard based on XML

- Content. It is possible to structure and standardise content using XML. Subjects are usually based on an accepted set of concepts and principles. XML can be used to create a vocabulary of concepts and principles and also some grammatical rules that govern their use.

  XML-based subject markup allows us for example to search for specific terms in a learning object. The vocabulary can act as an index to the course material.

- Instruction. Instruction comes in different variants. Knowledge and skills are taught in the form of tutorials, lectures, exploratory exercises, projects, to name just a few.

  These instructional forms can be defined as tags in an educational markup language for learning object design. The vocabulary and the grammatical rules provided in an XML DTD provide structure for learning objects. Several design markup languages for learning objects are already suggested, e.g. [4].

The combination of content and instructional markup provides a powerful structuring tool. It gives guidance to instructional designers and allows learners to perform accurate searches on courses and topics. Ideally, standardised learning object formats can be created from these structures.

2.3 Experience

We have used the technologies described above in a Web-based course system for 3rd-level computing students, which is taught by the second author. We converted existing HTML-material into an XML-format based on an educational markup language that we defined for this purpose. This language is based on elements (tags) such as concept definition, example, exercise, etc. Essentially, this is instructional markup, but we also explored the use of subject markup in conjunction with instructional markup.

We illustrate this by a small subject-specific markup example. This is taken from our database course. We have used XML-tags, such as `DBQuery`, in this excerpt that are specific to the course subject.

```
<DBQuery>
 <Number>   1                </Number>
 <Name>     SIMPLE QUERY      </Name>
 <QueryStr> Get colour and city for
            parts from Rome    </QueryStr>
 <QuerySQL> select colour, city
            from  parts
            where city='Rome'  </QuerySQL>
</DBQuery>
```

The DTD – Document Type Definition – defines the tags to describe a database query:

```
<!ELEMENT DBQuery
          (Number,Name,QueryStr,QuerySQL)>
<!ELEMENT Number   (#PCDATA)          >
<!ELEMENT Name     (#PCDATA)          >
<!ELEMENT QueryStr (#PCDATA)          >
<!ELEMENT QuerySQL (#PCDATA)          >
```

Tree-like structures created by DTDs correspond to the hierarchical structures of course content with units, aspects, definitions and examples.

In order to support instructional design and the flexible configuration of course material, we stored the XML-based content in a hierarchical structure in a database. This architecture allows powerful database retrieval technology to be used to retrieve and configure material, which can support the educator in configuring material for particular purposes (such as creating revision material) as part of the instructional design.

The database architecture is also beneficial for course delivery. Content is extracted from the database, assembled, converted into a HTML representation, and presented to the learner dynamically – this would allow a personalised presentation based on a learner profile.

2.4 Publication and Personalisation

In contrast to HTML, XML is not presentation-oriented. Its major achievement is the separation of content and presentation. This has several advantages.

- XML-based content can be transformed into several publication formats, e.g. a HTML-based Web-presentation for online learning or a PDF-based print version for course notes, using XSL (XML Stylesheet Language) transformations [1].
- Content can be personalised. Both content and presentation can be made dependent on individual learner profiles. Content can be selected from a repository, assembled, and suitably presented for a particular learner.

In particular the personalisation aspect in adaptive systems is a promising aspect. XML-based standard formats for learner profiles already exist.

The combination of these technologies enables better learner support through automated mentoring, which can complement the existing set of educational measures.

3. SEMANTICS

XML allows us to define terms and grammatical rules. However, the terms have no meaning. Technically, they are syntactical elements. Terms representing concepts in general will only have a meaning if they are defined in terms of other well-defined concepts.

3.1 Knowledge Representation

Knowledge representation is a research area that addresses the problems arising in the semantical definition of concepts [9]. Triples are used to define concepts. Triples are similar to subject, verb and object in an elementary sentence. Two examples:

- (Database, consists_of, Tables) is a subject-specific triple that defines a database (subject) as a collection (verb/property) of tables (object).
- (Concept, has, Definition) is an instruction-oriented triple that associates a definition to a concept.

Various knowledge representation systems exist. However, these lack interoperability and shared understanding. The Web provides its own knowledge representation framework – the Resource Description Framework RDF.

3.2 RDF

RDF allows us to express the semantics of concepts using triples of subjects, properties, and objects [6]. Subjects, properties and objects are identified by URIs (Uniform Resource Identifiers), i.e. newly defined concepts are associated with a unique identifier. Firstly, anyone can define a new concept, and, secondly, concepts that are syntactically the same can be defined differently, associated with different URIs.

It is said that the Semantic Web is about adding logic to the Web, which means that rules are used to make inferences. For example, based on facts (there is a concept example, e.g. a database table) and inference rules (an example is always preceded by a definition – formulated as a triple) we can infer that there is a definition of a database table for a table example, and we can search the content using the corresponding definition and table tags.

As one of the World-Wide Web Consortiums standardised technologies, RDF has the potential of becoming a widely accepted standard. In particular the lack of shared understanding and standardisation of knowledge representation has so far hindered the success of this technology.

4. WEB ONTOLOGIES

RDF allows us to express semantics. However, in order to be applicable, a further problem needs to be addressed. There can still be a number of knowledge repositories that use different URIs to identify the same concept.

4.1 Ontologies

Ontologies solve the problem of finding out whether two concepts actually mean the same thing [7]. Ontologies formally define the relationship among concepts. Typically, an ontology consists of a taxonomy that relates terms and a set of inference rules to deduce new knowledge. Ontologies can be represented as documents that can be published, exchanged and shared on the Web.

Searching the Web can be improved through ontologies. A search can be based on a particular definition of a concept. Information on a page (e.g. representing course content) can be related to associated knowledge structures and inference rules – see Figure 1. In the educational context, content found can be embedded into subject and instruction-related knowledge structures.

Currently, some effort is being made to define an ontology language for the Web – the Ontology Web Language OWL. Ontologies allow content providers and content requestors to cooperate. Properties of provided content need to satisfy the requirements of the requestor. These properties are complex concepts. Ontologies allow different terminologies to be reconciled. Their inference systems allow us to decide whether a provided learning component matches the requirements.

4.2 Experience

Even though the technology is still very immature, the second author has had some experience using multiple ontologies to formalise different aspects of an application domain. This experience has demonstrated the potential for the discovery, matching, and assembly of material based on ontology descriptions. We have used this for matching and retrieval in a software object repository where users and object providers share a description notation though an ontology.

The first author has recently completed work on a courseware generation system that, based on an ontology for a given subject, generates (i) a slide show outline, which can form the basis of more worked course material, and (ii) a multiple choice test based on the subject. A simple concept hierarchy is a classification of animals into categories and subcategories. Simple concept hierarchies can be converted into several slides – one for each major concept – each containing hierarchically presented bullet-point lists. Generated tests that can ask questions about the association of subconcepts to more general concepts.

The quality of the results depends, of course, on the quality and detail of the ontologies available. For example for the computing domain, knowledge

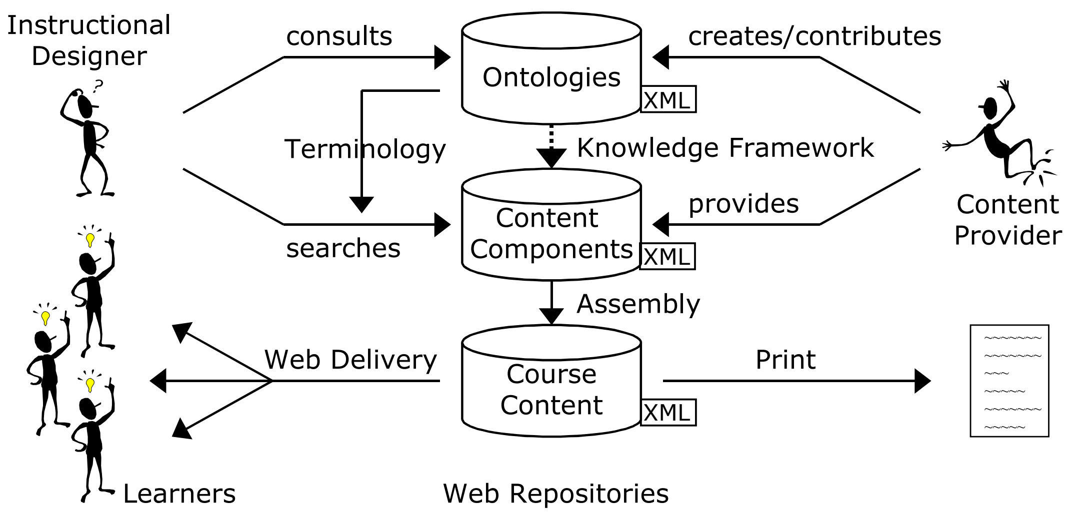

Figure 1. Semantic Web-based Instructional Design.

description frameworks and their ontological representations are under development.

We envisage the use of multiple standardised ontologies – e.g. metadata, subject, and instruction-oriented – in the design of courseware in the future.

5. DESIGN AND ARCHITECTURES

Semantic Web technologies described will impact the instructional design of Web-based, but also other forms of teaching and learning – see Figure 1. From our own experience with XML technologies we expect substantial improvements.

5.1 Components – Reuse and Sharing

The Web is an infrastructure that encourages sharing of information. This philosophy does not only apply to users of online courseware, it will also apply to designers of courseware. More than ever, we will share and collaborate in the design of educational content. This is a process that is already more apparent in another discipline using the Internet and the Web as a development environment. Software engineers have started to use software repositories to search for and assemble software from these repositories. The Web can support, however, also other forms of digital content such as educational content.

Modularity and reuse are the keywords that describe this new approach to design. The Web can act as a marketplace that brings providers, developers, and users of educational content together – Figure 1. New courseware is assembled from existing components. The components can feature in different forms:

- whole courses or learning objects,
- topic-oriented units within learning objects,
- course outlines and knowledge infrastructure, i.e. the structure and instructional aspects.

All of these artifacts are described and structured in terms of XML-based technologies, including RDF.

- Course content providers will make structured content and metadata available. Providers will either link themselves to existing ontologies or provide their own ontology support.
- Users of the provided material and services include instructional designers and also learners. Based on their requirements specification, aligned with existing ontologies that provide the terminology, the most suitable components will be extracted.

Different roles are important in these processes, including ontologists (knowledge engineers) and course designers (both provider and user of repository material).

5.2 Architecture

The architecture required to support such a development style consists of the following components [5,8] – see also Figure 1:

- Web-based ontology repositories make various ontologies available to users.
- Content repositories make course content and service components available – possibly subject to payment, contracts, etc.
- Retrieval and assembly facilities provide the functionality a user (requestor of educational content) needs to search for suitable components and to integrate these into her/his own environment.

Automatic processing is crucial for the Semantic Web. So-called software agents – independent and mobile software entities – will automatically search various repositories, gather information, and provide results to their clients. This technology will take the main burden off the users.

6. Discussion and Conclusions

Semantic Web technology will have an impact on the way we develop courses – for the Web and also for other platforms of presentation and delivery.

We can summarise the overall Semantic Web architecture in form of layers – see Figure 2. At its basic XML layer the Semantic Web offers structure for content and metadata annotations for learning objects. At the RDF layer it enables the semantical definition of concepts and reasoning about semantics. At the ontology layer it provides features to share and reconcile different knowledge representations.

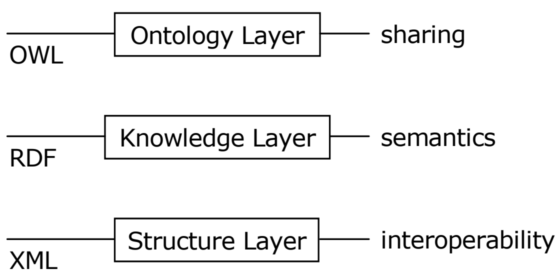

Figure 2. Semantic Web Layers.

The lowest layer provides an accepted format for education- or subject-specific markup languages. We have used this in our undergraduate course system. RDF adds the possibility to defines concepts – for example those that occur in tags in educational or subject-specific markup – in form of semantic nets such as hierarchies. Ontologies provide the infrastructure to share these definitions. Our courseware generation tool is based on publicly available subject ontologies.

The combined use of these technologies provides two central artifacts. Firstly, structured educational content with precise and shared semantics, and secondly, a knowledge repository in form of ontologies that formalises the knowledge aspects related to educational content.

With the use of standardised subject and instruction ontologies, an instructional designer or learner can search content repositories. Highly accurate search is the Semantic Web promise. This enables a component approach to instructional design based on individual, reusable modules (such as content topic or course layouts and structures).

Tool support is of course essential. XML-tools are available, and RDF tools are beginning to mature at the moment. However, more time is needed before our scenario will be fully implementable. Automation is an essential prerequisite.

We have limited this discussion here to the Semantic Web initiative. The Web is evolving in a number of directions which could also be beneficial for the educational context. Whereas the Semantic Web is about the interoperability of data and document structures and semantics, for example the Web Services framework is about interoperability on a service and computation level. We expect a potential here for teaching and learning, in particular for interactive features and their integration in computer-based instruction.

7. References

[1] The Apache Project. Cocoon. http://xml.apache.org/ cocoon. (2002).

[2] T. Berners-Lee, J. Hendler, O. Lassila. The Semantic Web. Scientific American. (May 2001).

[3] IMS Global Learning Consortium. IMS Web Page http://www.imsproject.org. (2003).

[4] R. Koper. Modeling units of study from a pedagogical perspective: the pedagogical meta-model behind EML. http://eml.ou.nl/ introduction/docs/ped-metamodel.pdf. (2001).

[5] T. Looms and C. Christensen. Emerging and Enabling Technologies for the Design of Learning Object Repositories. http://xml.coverpages.org/ADLRepositoryTIR.pdf (2002).

[6] W. Nejdl, B. Wolf, C. Qu, S. Decker, M. Sintek, A. Naeve, M. Nilsson, M. Palmer, T. Risch. EDUTELLA: A P2PNetworking Infrastructure Based on RDF. Proc. WWW2002. ACM. (2002)

[7] Ontology.Org. http://www.ontology.org. (2003).

[8] SCORM – Sharable Content Object Reference Model. http://www.adlnet.org. (2003)

[9] A. Stutt and E. Motta. Knowledge Modelling: an Organic Technology for the Knowledge Age. In M. Eisenstadt and T. Vincent. The Knowledge Web. Kogan Page, (1998).

[10] W3C – the World Wide Web Consortium. The Semantic Web Initiative. http://www.w3.org/2001/sw. (2003).

[11] W3C – the World Wide Web Consortium. Extensible Markup Language (XML). http://www.w3.org/XML . (2003).