Ontology Technology for the Development and Deployment of Learning Technology Systems - a Survey

Claus Pahl
Dublin City University
School of Computing
Dublin 9, Ireland
cpahl@computing.dcu.ie

Edmond Holohan
Griffith College Dublin
Faculty of Computing Science
Dublin 8, Ireland
edmondholohan@eircom.net

Abstract: The World-Wide Web is undergoing dramatic changes at the moment. The Semantic Web is an initiative to bring meaning to the Web. The Semantic Web is based on ontology technology – a knowledge representation framework – at its core. We illustrate the importance of this evolutionary development. We survey five scenarios demonstrating different forms of applications of ontology technologies in the development and deployment of learning technology systems. Ontology technologies are highly useful to organise, personalise, and publish learning content and to discover, generate, and compose learning objects.

Motivation

The World-Wide Web is the predominant learning technology platform today. Its accessibility has made it a successful environment particular for the publication of learning material. Learning resources can be provided in a standardised format that can be accessed at any time from any location.

The Web, however, is still evolving. The current evolution of the Web will have a major impact on educational technology. This will affect instructors and learners alike. The Semantic Web initiative aims to bring semantics to the Web [W3C 2003a]. Currently, search and retrieval functionality relies on human interaction and often ad-hoc approaches to selection of documents for a given set of search criteria. Semantic annotations, which can be processed by software applications, will improve the precision of searches. This will enable accurate searches for learning resources. The changes that will occur as a result of the Semantic Web initiative, however, go beyond search and retrieval. The overall development and deployment process of educational technology can be affected. Ontology technology – the knowledge representation core of the Semantic Web – enables this wide applicability.

An area such as education, where access to information is central, depends on the representation and organisation of knowledge both for the content but also the metadata level. The Semantic Web is based on ontology technology – a knowledge representation and inference framework [Berners-Lee et al. 2001]. We will show how ontology technology can be used to support various aspects of learning technology, including the creation of content, the publication and personalisation of content, the discovery of learning objects, the generation of complex learning objects, and the composition of learning technology components in learning technology architectures.

Ontology technology has already been used [IEEE 2001; Fischer 2001; Leidig 2001, Pahl & Holohan 2003; Diaz et al. 2004] – with different purposes ranging from the definition of a terminology to the use of conceptual models and inference in the generation and composition of learning technology content and systems. We will give a comprehensive overview here, surveying a number of different scenarios. We will illustrate the scenarios based on our own experience with this technology.

Ontology Technology

We introduce a knowledge space for learning technology systems and how it can be structured through ontologies. We introduce background technologies – ontologies and the Semantics Web – and provide a classification and comparison framework for the subsequent discussion of five scenarios.

Knowledge Spaces and Ontologies

The Web creates a space in which developers, instructors, and learners contribute to and participate in learning processes. Knowledge [Sowa 2003] is a central component in this space. This *knowledge space for learning technology systems* comprises several *knowledge types* of knowledge relevant to the educational context: firstly, *subject-specific knowledge* describing the subject-related aspects of the content, secondly, *pedagogic knowledge* describing the educational aspects of the content, and, thirdly, *technology-related knowledge* describing the implementation-oriented aspects of the content. Another aspect of the knowledge space is its *purpose,* i.e. which functions are supported [Daconta et al. 2003]:

- *taxonomy (TX)* – terminology definition and classification are the central issues – it supports browsing and retrieval of educational resources,
- *thesaurus (TH)* – relationships between terms are the central issues – it constrains the use of a vocabulary,
- *conceptual model (CM)* – a formal model of some domain – it supports modelling of the subject area and technical aspects which often uses more than classification-oriented relationship types,
- *logical theory (LT)* – reasoning and inference are the central issues – it combines knowledge representation with a logic and, thus, supports reasoning within a knowledge domain.

These functions are ordered based on the degree of their semantics support. Ontologies are knowledge representations, but ontologies are often seen as intertwined with logics [Ontology.Org 2003; Daconta et al. 2003]. Ontologies provide the terminological aspects needed in logical reasoning. Pure logics are symbolic frameworks. We can divide the four ontology functions into two different purposes:

- *vocabulary* and *terminology* – taxonomies and thesauruses mainly address annotation and retrieval needs,
- *modelling* and *reasoning* – conceptual models and logical theories address requirements arising in the composition of educational resources.

Ontology technology is central as it allows us to reconcile problems arising from the fact that we have to deal with different actors, different organisations, different systems, and different locations. These different aspects create a space in which common understanding and agreement in relation to knowledge in this space is central.

Web Ontologies and Ontologies for Educational Technology

The *Semantic Web* [Berners-Lee et al. 2001; W3C 2003a] provides an ontology framework, based on an ontology language called OWL (the Web Ontology Language). OWL in turn is based on RDF (the Resource Description Framework) and XML (the eXtensible Markup Language).

- *XML*: XML and the XML Schema language provide the basic syntactic interoperability for an ontology definition through a approach to define markup languages.
- *RDF*: RDF and RDF Schema introduce semantics. This allows the description of concepts in terms of triples – subject, property, and object. A new concept is defined in terms of its properties in relation to others.
- *OWL*: OWL is a full ontology language providing logical facilities for reasoning and inference. The central reasoning concept is subsumption – the subclass relationship between concepts or properties of concepts.

Schema languages are the tools to introduce a vocabulary into an ontological framework. A logic underlying the ontology provides the reasoning facilities.

Ontological frameworks for the educational domain exist that are not directly based on RDF. A learning object annotation scheme, the Learning Object Metadata standard [IEEE 2002], is an example that introduces a vocabulary.

As already mentioned, several ontologies inhabit and organise the knowledge space in the educational context: subject, pedagogy, and technology ontologies. We can distinguish *two ontology layers* divided by their relevance to particular learning technology development and deployment activities:

- The *upper ontology layer* contains the pedagogy and subject ontologies – this layer supports the discovery, structurisation and sequentialisation of learning objects, it addresses the *why* and *what* of development. Often, educational metadata is expressed in terms of the upper ontology layer.
- The *lower ontology layer* contains the subject and technology ontologies – this layer supports the composition of learning objects, it addresses the *how* of development. Often, XML-based educational markup and other content organisation techniques are based on the lower ontology layer.

We used the term *learning object* so far to represent any digitally represented object or component part of a learning technology system.

Applications of Ontology Technology

We will illustrate possible applications of ontology technology for learning technology systems development and deployment – drawing on our experience in the development and deployment of these systems. *Five application scenarios* shall be introduced:

- the creation and organisation of content,
- the publication and personalisation of content,
- the annotation and discovery of learning objects,
- the generation and sequentialisation of complex learning objects, and
- the interaction and composition of learning technology components in learning technology architectures.

The functions of the knowledge space and the ontology layers form a *classification and characterisation scheme* for these application types – see (Tab. 1).

	Organisation and Structure	Publication and Personalisation	Annotation and Discovery	Generation and Sequentialisation	Interaction and Composition
ontology layer	upper	upper	upper	upper	Lower
ontology function	TX, TH	TX	TX, TH	CM, LT	CM, LT

Table 1: Scenarios of different forms of application and their classification.

The main case study for the illustration of the scenarios is an undergraduate computing course support system called *IDLE – Interactive Database Learning Environment* [Murray et al. 2003]. IDLE supports classical knowledge learning, but also training of activities. It is a system that has been developed and extended for almost ten years.

Content Organisation and Structure

Application Type

Learning content is usually made up of textual documents. These documents are often based on an inherent (but implicit) structure. In content documents we find definitions of new concepts, their illustration, examples, exercises, etc. XML is an ideal technology to make this structure explicit. This has a number of advantages:

- Firstly, an explicit structure supports the instructional designer in the design process. It gives guidance allowing the designer to construct content from small individual building blocks.
- Secondly, an explicit structure makes the document accessible to others. The document can be searched for particular content items; for instance, a learner can search for exercises on a particular concept.
- Thirdly, the fine-granular organisation of documents into small, classified units allows the flexible storage and assembly of these units. Adaptive delivery is an example for this approach where personalised content can be assembled from these small units. This aspect shall be discussed in the next section.

The upper ontology layer provides the knowledge support here. Taxonomy and thesaurus functionalities are used. Two ontologies describing two types of knowledge are important.

- *Educational knowledge* is needed to give structure to educational documents [Koper 2001]. It provides the primary structure. An ontology acts as a taxonomy, introducing the vocabulary of tags. This knowledge forms part of a development methodology, which is essential for the instructor as an instructional design tool.
- *Subject-specific knowledge* can be used to support the educational structuring. It is not a necessary aspect, but it adds another dimension of access to the document. It is in particular suitable for learners to search for topic-specific units. Ideally, the subject-specific knowledge is based on a common, accepted ontology for the topic domain. The development of these ontologies has begun for various domains, for example software engineering or genetics. In addition to an introduction of a vocabulary with concept classifications (a taxonomy), subject ontologies often comprise thesaurus functionality as well, supporting for instance synonyms in searches.

There is a limitation connected to markup as a structuring tool. It is essentially limited to textual resource. Even though text is central in education, learning technology systems are often multimedia environments. Markup to structure content is, due to its nature, not suitable for non-textual media.

Case Study Scenario

(Fig. 1) contains an example of the use of an XML-based structured vocabulary for a learning content unit. The vocabulary definition is combined with a hierarchical, grammatical structure definition for content documents.

Educational markup definition (excerpt from an XML Document Type Definition) of a structured vocabulary:

```
<!ELEMENT EducUnit        (UnitName?,UnitContent,Keywords?)          >
<!ELEMENT UnitContent     (Unclassif|Knowledge|Example|Exercise)*    >
<!ELEMENT Knowledge       (Unclassif|ConceptDef|ConceptDescr)*       >
```

Educational markup (excerpt from XML file) used to structure a content document:

```
<EducUnit>
  <UnitName>       SQL Background & Standards   </UnitName>
  <UnitContent>
     <Knowledge>
       <Unclassif> SQL is the Structured Query Language. ... </Unclassif>
     </Knowledge>
  </UnitContent>
</EducUnit>
```

Figure 1: Definition of educational markup and its use in IDLE.

Subject-specific markup can be used in a similar style. For instance, some terms like ‘database’ or ‘query language’ are central concepts in the database domain that would appear in any domain ontology. These could be marked up in XML.

Publication and Personalisation

Application Type

The variety of forms and the degree of personalisation of delivery can be enhanced through the use of ontology-based content organisation. The upper ontology layer can provide knowledge support and taxonomy functionality. Content organisation in terms of educational markup in XML is a prerequisite for this application.

- In contrast to HTML, XML is not presentation-oriented. This is seen as one of its major advantages. Separation of concerns is an approach to master complex design and implementations. XML separates structure from presentation [Apache 2003]. An advantage of this approach is the possibility to transform XML-based content into several publication formats. Alternatives include Web publication, i.e. the conversion into HTML, or publication for print media, e.g. the conversion into PDF or similar formats.
- XML is an abstract data and document structuring format that allows machine processing. This enables for instance the flexible storage and retrieval of XML-documents. An advantage is the possibility to adapt to specific users or user groups and to personalise the delivery [de Bra et al. 1999]. It allows content to be assembled from small units into learning objects that suit the needs of individuals or groups.

Case Study Scenario

In IDLE we have represented text-based content in XML as described above. XSL is the XML Stylesheet Language that allows us to describe transformation of XML documents [W3C 2003b]. We have used these transformations to create HTML representations for the abstract XML representation. We can use this technique to create a richly laid out version for local use or a more reduced version for distance or mobile learning.

Before addressing different output formats and properties in the translation, we could let the target learner decide on the assembly of XML units into larger learning objects. We could, for instance, generate a summary version which includes concept definitions, but no exercises.

Annotation and Discovery

Application Type

The wide accessibility of the Web makes it an ideal environment to share resources. Educational resources, ranging from simple text-based material to highly interactive systems, can be provided and accessed using Web technologies. In order to support the discovery of sharable resources by potential users, the resources need to be annotated by suitable descriptions. A prerequisite for this to work is a standardised and agreed vocabulary for these annotations. The upper ontology layer provides the knowledge support for taxonomy and thesaurus functionalities.

The Learning Object Metadata standard (LOM) [IEEE 2002] provides a metadata framework for the annotation of learning objects. The learning object notion comprises a variety of educational technology applications. LOM defines the attributes required to fully describe a learning object. It classifies attributes into nine categories addressing for example general, technical, educational, and lifecycle aspects. The provider of the learning object describes the object in terms of content and infrastructure properties. A potential user – learner or instructor – then uses a related query language (or search engine) to formulate requirements in terms of the properties described.

In contrast to the content structuring scenario this form is a black-box approach not considering the content itself. It relies on the provider to describe a learning resource adequately [SCORM 2003].

Case Study Scenario

We illustrate the LOM standard using our own database course system IDLE as an example. We focus on three of the nine categories of attributes – see (Fig. 2).

General attributes:
- `title       = "Introduction to Databases"`
- `description = "A third-level course for computing students addressing principles, models and languages for database development"`

Technical attributes:
- `format      = "text/html"` or `"audio/mp3"`
- Other attributes: `size`, `location`

Educational attributes:
- `interactivity type = "active-simulation"` or `"expositive–audio"`
- Other attributes: `learning resource type`, `interactivity level`, `semantic density`

Figure 2: Annotation of the IDLE learning object.

Generation and Sequentialisation

Application Type

Knowledge about a collection of learning objects can be used to organise individual units into a larger learning object by sequencing the units based on inherent dependencies that are derived from the knowledge. This knowledge goes beyond the annotations we discussed in the previous section. The upper ontology layer provides the knowledge support here. In addition to basic taxonomy and thesaurus functionalities, conceptual modelling and logical theory functions play an important role.

In ontologies, the standard organisational form is a hierarchy, categorising concepts into classes and subclasses. In more elaborate ontologies a variety of relationships between concepts might be represented. The knowledge represented can comprise subject and education-related aspects [Leidig 2001]:

- Subject-related knowledge is often based on a semantic concept network. A subject-related ontology can be richer than a vocabulary or concept hierarchy. Often it forms a conceptual model describing a full domain.
- Educational knowledge often involves relationships that express dependencies, e.g. *isBasedOn*. This knowledge can also comprise a vocabulary to classify educational units, such as definition, example, or exercise.

In this *composition-oriented scenario*, knowledge is separated from content. This scenario makes use of advanced ontology techniques. Different types of relationships have to be dealt with in the process of arranging the content unit in a suitable sequence [Fischer 2001]. A possibility is to prioritise relationships in this process. This generation algorithm has to use the reasoning facilities of an underlying logic to determine the ordering dependencies.

We shall briefly address another, *creation-oriented scenario*, where stand-alone knowledge such as a subject ontology (a conceptual model) is used to generate a courseware outline [Holohan 2003]. Here content does not exist prior to the generation process. Input is solely provided by the ontology. Different types of content can be generated:

- Content outlines. The knowledge represented in the ontology can be converted into course material. A concept hierarchy usually forms the backbone of such an ontology; this hierarchy also guides the sequentialisation of the concepts and their descriptions.
- Assessments. For instance, multiple choice questions (and answers) can be generated that can be used as input for an MCQ-based self-assessment tool based on related and unrelated concept terms.

Similar to the publication and personalisation this scenario is based on a transformation step, but the purposes are different (content delivery vs. content creation and composition).

Case Study Scenario

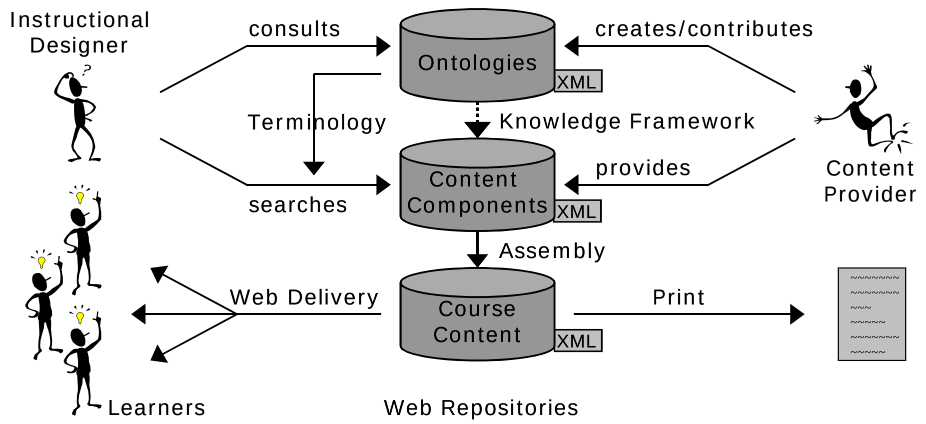

Figure 3: Generation of learning content based on ontologies.

An enhanced database ontology can support learning object generation or composition. Central concepts are defined: database object, relation, relation schema, record, table, table definition, etc. Concepts are related through a central subclass or is_a relationship. For instance, table is_a database object. Concepts are in addition related through isPartOf and isBasisFor relationships. For instance relation isBasisFor table or record isPartOf table

The subclass hierarchy is the backbone of the sequentialisation, but can not, as we can see here, resolve all dependency problems. Using a prioritisation approach – here isBasisFor as the secondary and isPartOf as the tertiary relationship, we could obtain the following order of concepts: database object, relation schema, relation, record, table definition, table. The constraints expressed in the ontology, however, might not lead to a unique solution.

(Fig. 3) illustrates the scenario. Ontologies define a conceptual model for content components. The latter are assembled to larger content objects. These can be translated into Web representations or other, print-oriented media.

Interaction and Composition

Application Type

In the previous section on generation and sequentialisation, we looked at a form of composition of learning units. Sequencing was the composition approach. The components (often passive learning objects or units) are only connected implicitly through dependency constraints. Now, we address the composition of learning objects that are connected more explicitly, through interactions of active learning objects that a learner can interact with.

In this scenario, more than abstract metadata annotations are needed. In order to compose these objects, information about their behaviour and interaction patterns is required. The lower ontology layer provides the knowledge support here. The taxonomy functionality (for syntactical aspects) and the logical theory function (for semantical aspects) are used. The composition is based on the functionality of the learning objects or components.

A reference architecture and terminology for composing learning technology systems is provided by the Learning Technology Standard Architecture LTSA [IEEE 2001]. The standard introduces an architecture for learning technology systems consisting of active components such as learner entity, coach, delivery component, evaluation component, and storage components for learning resources and the learner model. The standard distinguishes components based on their functionality within a system – in contrast to LOM where we assume a differentiation by subject topic. LTSA describes the interactions between learning technology system components. These components are independent services provided to offer functionality to the educational environment.

The LTSA can be applied to the Web platform. The Web Services Framework [Daconta et al. 2003] defines a services-oriented architecture. Its objectives are similar to those of the LOM framework – its goals are the reuse and sharing of resources, but in contrast to LOM, these resources are expected to be integrated into a services-based software system. The research in this area has recently moved into the direction of semantic Web services, i.e. services that are described semantically. Ontologies provide the knowledge representation framework for these service descriptions. For instance, we could enhance the LTSA definition by ontology-based semantic descriptions.

Case Study Scenario

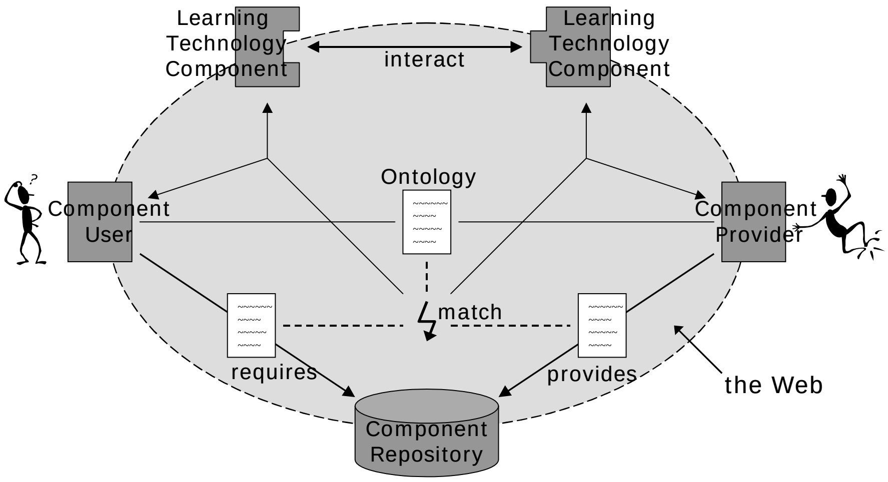

Figure 4: Scenario for the development and deployment of semantic learning service components.

Work in this area has only just begun. We are currently working on a Web services architecture for our IDLE system. We also work on a framework for semantic service description and composition. We expect this area to become of major importance in the future. Therefore, we have added this scenario to our discussion.

(Fig. 4) illustrates an environment for this type of development. A developer specifies requirements for a learning technology component, which might match descriptions of a provided component [DAML-S 2002]. Semantic descriptions within a shared and agreed ontology can be used to decide on suitability. A provided service component can then be integrated into a learning technology system under development.

Conclusions

Knowledge is of major importance for learning technology development and deployment. Content, learning objects, and learning technology system components are different notions of parts of a learning technology system. Knowledge is central for their structure, metadata, presentation, creation, and composition. Different types of knowledge – content and meta-level knowledge about content and technical aspects – can be captured in form of ontologies.

Tool support and automation are important. Various XML tools including editors and transformation tools support content structuring and publication tasks. Database support can support adaptivity and retrieval. Ontology tools can support matching activities for retrieval and composition tasks. The discussion of tools, however, is beyond the scope of this paper.

References

Apache Software Foundation. (2003). *The Apache Cocoon Project*. http://cocoon.apache.org.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*.

DAML-S Coalition. (2002). DAML-S: Web Services Description for the SemanticWeb. In I. Horrocks and J. Hendler, editors, *Proc. First International Semantic Web Conference ISWC 2002*, LNCS 2342, pages 279–291. Springer-Verlag.

Daconta, M.C., Obrst, L.J., & Smith, K.T. (2003). *The Semantic Web – a Guide to the Future of XML, Web Services, and Knowledge Management*. Indianapolis, USA: Wiley & Sons.

De Bra, P., Brusilovsky, P. & Houben, G.-J. (1999). Adaptive hypermedia: from systems to frameworks. *ACM Computing Surveys* 31(4).

Fischer, S. (2001). Course and exercise sequencing using metadata in adaptive hypermedia learning systems. *Journal on Educational Resources in Computing* (JERIC) 1(1).

Holohan, E. (2003). *Automating the Generation of Courseware*. M.Sc. Dissertation. School of Computing. Dublin City University.

Holohan, E. & Pahl, C. (2003). Semantic Web Technology for Web-based Teaching and Learning: a Roadmap. *4th Annual LTSN-ICS Learning and Teaching Support Network - Information and Computer Sciences*. Galway, Ireland. LTSN.

IEEE Learning Technology Standards Committee LTSC (2001). *IEEE P1484.1/D8. Draft Standard for Learning Technology – Learning Technology Systems Architecture (LTSA), 04/06/2001.* IEEE Computer Society.

IEEE Learning Technology Standards Committee LTSC (2002). *IEEE P1484.12/D4.0 Draft Standard for Learning Object Metadata (LOM).* IEEE Computer Society.

Koper, R. (2001). *Modeling units of study from a pedagogical perspective - the pedagogical meta-model behind EML*. Open University of the Netherlands. http://eml.ou.nl/ introduction/docs/ped-metamodel.pdf.

Leidig, T. (2001). L3—towards an open learning environment. *Journal on Educational Resources in Computing* 1(1).

Murray, S., Ryan, J. & Pahl, C. (2003). A Tool-mediated Cognitive Apprenticeship Approach for a Computer Engineering Course. *Proc. International Conference on Advance Learning Technologies ICALT'03*. IEEE.

Nejdl, W., Wolf, B., Qu, C., Decker, S., Sintek, M., Naeve, A., Nilsson, M., Palmer, M. & Risch, T. (2002). EDUTELLA: A P2PNetworking Infrastructure Based on RDF. *Proc. World-Wide Web Conference WWW2002*. ACM.

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

Pahl, C. (2001b). XML-technologies for the Support of Active Learning in Web-based Virtual Learning Environments. *Proceedings 6th WebNet Conference*, Orlando, Florida, US. AACE.

Diaz, P., Lytra, M.D., & Sampson, D.G. (Eds) (2004). Special issue on Ontologies and the Semantic Web for E-Learning. *Educational Technology & Society*. (forthcoming).

SCORM (2003). *Sharable Content Object Reference Model*. http://www.adlnet.org.

Sowa, J.F. (2000). *Knowledge Representation – Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA, USA: Brooks/Cole.

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

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