The Semantic Web and Higher Education: What Should We Know About It?

Introduction

Background

Imagine the following scenario. John just graduated from the University of Central Florida with a Master of Arts in Instructional Technology. John has been out of school and working for six months, and while John takes satisfaction in being finished with school, he has already identified some particular skills he could need to improve for work. John sees that if he could be better at matching pedagogical approaches with new technology trends, not only would he be more competitive for other jobs, but he would really stand out from his workmates, and his boss would take notice. At home, John turns on his PC and checks his email. With that done, John launches his web browser and navigates to *learnbot.com*. John logs into his account and navigates to his learning profile. Within his profile, John first modifies his learning style preferences from *read* and *test facts* to *read*, *hear*, *watch*, *problem-based*, *test concepts*, and *apprenticeship*. Then John sets his initial competency level from *level 7* to *level 3* (i.e., lower number equates to lower competency), and his expediency settings from *casual* to *necessary* and *expedited*. John selects the *OK* button and then selects *My Learning Target*. A new web page displays, and in the field labeled, "*My learning goal is to…*" John enters the following:

> "Learn about new technologies that can be used as educational tools. Learn new pedagogical strategies and review existing strategies. Match new and existing pedagogical strategies with new technologies. Create realistic training scenarios for sales team."

John selects the *Analyze* button and waits while LearnBot considers the information. After a brief moment, a user data field displays below the "My learning goal is to…", and this field is labeled, "Additional clarifying notes, if you please…" There is already text in the field: it is the text John entered but each sentence John entered is on its own line with a comment in a blue colored font. Following John's first sentence is the comment, "*John, I need more details regarding technologies as educational tools. Please provide additional details.*" John places his text entry cursor after the comment and enters the following: "Communication one-to-one, one-to-many, and many-to-many. Text, pictures, audio, video compatible. Synchronous and asynchronous. Public domain, trial based, or service is less than $20 per month per user. Supports high level of interaction." John adds additional notes after the remaining comments and selects the *Analyze* button. A graphic symbolizing a process is in progress displays and moves towards completion, and John waits. A few minutes pass, and then a new web page displays. This page is rather simple, but it has in large, friendly letters, "*Learn when ready*."

The rest of the scenario would describe the various presentations of technical products or systems with appropriate facts, concepts, specifications, and the like, together with manufacturer contact details. Depending on the results, perhaps several of these might be presented, or just one, and then a test would follow to ensure that John was paying attention. If John makes a mistake, the relevant material is replayed until John's competency reaches the target level. Then because John routinely reviews pedagogy and strategies, his profile has a catalog of links to material that gets rechecked for updates or modifications and new pedagogical strategies are also searched and cataloged. These become assembled into a series of presentations similar to the previous technology sequence, and John is tested on the pedagogical approaches similarly.

In the final phase, LearnBot assembles, based on descriptions or specifications, its best guess to matches between technology and pedagogy, but it refrains from presenting it to John. Instead, LearnBot formulates a series (in increasing difficulty) of problems for John to find matches, which he must be able to argue why they would fit together or not. At each step, LearnBot, depending on John's right or wrong choice or qualifying argument, may present teaser responses that will point John in the right direction. If John's choices and qualifying arguments differ from LearnBot, and John specifies the nature of the relationship between the two and why he believes he is right, LearnBot might in the background go back out into the Web to see if there may be substantiation. Perhaps then LearnBot finds the substantiation and then informs John that he is indeed correct, and LearnBot updates its own catalog for John. Or, perhaps LearnBot does not find the substantiation, but it lets John know in any case and updates its catalog for this match as *provisional* until John can provide evidence otherwise, or if the match is substantiated by someone else, which LearnBot can then verify. Perhaps an hour passes while this learning scenario plays and completes, or perhaps it requires several days where John spends a couple of hours each day until the scenario is complete. Once finished, John returns to work and begins assembling his plans for modifying the training strategy for the sales team, with qualifying arguments for his manager, which he can verbally argue successfully and provide substantiated material for his approach.

Back to now. This scenario should be of interest to anyone in the field of education. The technology behind this scenario is being developed. Many call it the *Semantic Web*, or *Web 3.0*. Much of the driving force (at this point), however, is from the consumer and healthcare industries, where scenarios surround businesses selling travel or entertainment services and related products to consumers through the Web (Borland, 2007a), or as research tools to match drug treatment programs on a personal basis (Feigenbaum, Herman, Hongsermeier, Neumann, & Stephens, 2007). The issue focus for this essay surrounds the impact this technology might have on the field of education. The guiding question is simple: "What is the Semantic Web, and should institutions of higher education, particularly schools of education and educational research, be interested in it?"

Connections

Charles Reigeluth, David Merrill, Gloria Gery, among others, who for the last 15 years, have repeatedly and independently made calls for "software agents" capable of finding and assembling training or learning materials for presentation to individuals in need of personalized learning. Their visions and subsequent work in the field have yielded significant impact on the conceptualization of how these automated agents would function. In several cases, these individuals have designed and built technical systems that were steps in the right direction. However, none of these individuals have yet fulfilled this vision, but I believe it is safe to state that the field would be much less developed in this direction without their work and repeated calls for action. While reading Reigeluth's edited book, *Instructional-Design Theories and Models (vol. 2)* (Reigeluth, 1999), I began thinking about his calls for such a technology, and so I wondered about the state of the art. After conducting some Internet searches, I found a research and development area called the *Semantic Web*. I became fascinated. You should be also. The game is about to change.

Source Materials

A list of the articles used in this mini-study follows.

Table 1: Source Articles Used to Derive Information on the Semantic Web

Count	Art. No.	Title	Author(s)	Pub. Date	Journal	Publication Format
1	1	A Smarter Web	Borland, J.	3/4 2007	Technology Review	

2	2	Semantic Web technology gains steam	Rapoza, J.	6-4 2007	eWeek	Weekly
3	3	The Semantic Web: Meaning And SOA	Kobielus, J.	10 2007	Business Communications Review	
4	4	An intelligent semantic e-learning framework using context-aware Semantic Web technologies	Huang, W.; Webster, D.; Wood, D.; & Ishaya, T.	2006	British Journal of Educational Technology	Peer Review
5	5	Semantic e-learning: synthesizing fantasies	Lytras, M.; Naeve, A.	2006	British Journal of Educational Technology	Peer Review
6	6	A prototype Semantic Web-based digital content exchange for schools in Singapore	Shabajee, P.; McBride, B.; Steer, D.; & Reynolds, D.	2006	British Journal of Educational Technology	Peer Review
7	7	e-Learning for depth in the Semantic Web	Shafrir, U.; Etkind, M.	2006	British Journal of Educational Technology	Peer Review
8	8	The Semantic Web in Action	Feigenbaum, L.; Herman, I.; Hongsermeier, T.; Neumann, E.; & Stephens, S.	12 2007	Scientific American	Monthly Journal

What is the Semantic Web

The vision of the Semantic Web is that it will become increasingly a *Services Oriented Architecture* (SOA) (Kobielus, 2007). Where the first phase of the Web (i.e., Web 1.0, if you will) saw a technology that permits content to be available everywhere that is readable, searchable, and understandable by human consumers, the second phase (i.e., Web 2.0) improved on the first by expanding the range of media forms that can be published to the Web: we saw the introduction of blogs, wikis, podcasting, and mashups (Kobielus, 2007). The Web in this second phase also saw the development of social networking technologies, such as Facebook, MySpace, and social web page tagging Del.icio.us. The anticipated next phase, (i.e., Web 3.0, again, the version number reflects no official recognition of technical state, but rather it is used as a differentiator for the sake of convenience) will see an increase in non-human content consumers, such as "… services, applications, bots and other automated components…" (Kobielus, 2007, p. 30).

The jump from publishing for people to publishing for the non-human will present profound advantages as the vision, approach, and supporting technologies mature. The scenario described at the beginning of this paper is a small example of the sought after benefits. The initial challenge is recognizing the scope of what it means to organize content on the Web. The fruits of Melvil Dewey's labor to organize books, periodicals, and all the other published content in a library with the Dewey Decimal System will not work to organize the Web: the essence of the Dewey Decimal System takes advantage of the fact that materials *must be checked into the library* before it can be accessed. The check-in process includes applying the categorization scheme to the item before it is placed on the proper shelf and in the proper location (and then becomes ready for *a reader to find it*). By contrast, content on the Web, is not "checked-in" by any organizational authority, which would ensure the content is properly categorized. By the early 1990s, as young as the Web was, it was already growing faster than any attempts to organize it could keep pace (Borland, 2007b).

Descriptions of the technologies that are being developed around the globe that support the Semantic Web involve digesting an alphabet soup. Extensible Markup Language (XML) code can be added to web pages as metadata that can provide instructions for how to find and display content, but this coding language and approach is insufficient to provide order to the Web (Borland, 2007b). The World Wide Web Consortium (W3C) began to address this problem in early 1996 by assembling a team of researchers from MIT and a graduate researcher from Ohio State, Eric Miller. Their initial task was to find a way to tag sexually explicit material in order to enact safeguards to prevent minors from accessing it. Miller and the team saw this as an opportunity to create a technical infrastructure to classify a lot of different information, such as pricing, subject, or title of a book for sale online (Borland, 2007b). The infrastructure that resulted is called the *Resource Description Framework* (RDF), which was published through the W3C in 1999. Following this publication, the question of how to define relationships between categories came up, and led by Tim Berners-Lee, the lead developer of the Web, the group discussed the idea of a Semantic Web that could provide categorization, as well as translation capabilities of the relationships between categories; these translation dictionaries and thesauri are called *ontologies*. An ontology will permit a machine to know that *post code* and *zip code* mean the same thing. Ontologies can be created to label a *cat* as an *animal* and a *veterinarian* as an *animal doctor*, which a machine could read and make use of (Borland, 2007b). The team's work led to the development of a *Web Ontology Language* (OWL) that supports rich descriptions of resource properties, classes, relationships, equality, and typing. "Within an RDF ontology, statements consist of well-defined 'subjects,' 'predicates,' and 'objects'" (Kobielus, 2007, p. 31).

There is significant growth in the development of the range of Semantic Web standards and specifications under the W3C. What follows is a list without specification definition or explanation, so that the reader might simply appreciate the direction to where this work is heading:

- OWL
- Gleaning Resource Descriptions from Dialects of Languages (GRDDL)
- Semantic Mediation
- Semantic Tools
- Semantic Repositories
- SPARQL – Query Language for RDF
- Semantic Modeling
- Semantic Mining
- Semantic Engines
- Semantic Controls

Who (or what businesses) are involved with developing the Semantic Web

The range of Semantic Web vendors is growing. The following is a sampling of businesses and what they are doing (as of October 2007) in the field of the Semantic Web (Kobielus, 2007). (Note: this information is completely attributed to the research by James Kobielus, 2007.)

Cycorp, with its headquarters in Austin, TX, develops turnkey solutions in artificial intelligence, knowledge representation, machine reasoning, semantic data integration, and information management and search. Their technology solutions include an ontology (now in the public domain) with a knowledge base, inference engine, natural language interfaces, and a semantic integration bus. Cycorp offers a no-cost license to its semantic technologies development toolkit to the research community. Cycorp is also conducting R&D into scenarios where users create lightweight ontologies, which go through a process of elaboration and enrichment and then have their semantic inferencing engines map the results to more formal global ontologies.

Oracle announced this past July that they are incorporating Semantic Web support into the latest "11g" generation of their database management systems. Oracle is incorporating a data management platform for RDF-based applications which will support new object types to manage RDF data in Oracle. The Oracle approach will support scalable semantic-based enterprise applications.

Other enterprise developers who are investing significantly in these technologies include IBM, Sandpiper Software, TopQuadrant, Modus Operandi, and Revelytix (Kobielus, 2007). These organizations can be looked at as infrastructure leaders. The problems they are facing and resolving can translate into users of their systems being able to work with their information in similar ways to the scenario described in the beginning of this paper.

The Scientific American article (Feigenbaum et al., 2007) describes a number of organizations that are involved in leveraging the technologies of the Semantic Web. **Cincinnati Children's Hospital Medical Center** uses Semantic Web technologies to extrapolate underlying genetic causes of cardiovascular diseases. **Eli Lilly** scientists use Semantic Web technologies to build a complete picture of the most likely drug targets for a given disease. **Pfizer** is using Semantic Web technologies to merge data sets regarding protein-protein interactions in order to reveal difficult to find correlations that might aid in the development of medications. The British firm **Garlik** uses Semantic Web technologies to compare databases that were previously incompatible so they can alert their subscribers to identity theft risks. At the **University of Texas Health Science Center in Houston**, a system named SAPPHIRE (Situational Awareness and Preparedness for Public Health Incidences using Reasoning Engines) uses Semantic Web technologies to integrate a broad range of data from local healthcare sources, as well as from hospitals, the EPA, and scientific literature to allow health officials to assess health risks, such as the spread of influenza or the treatment of HIV cases. SAPPHIRE has improved the efficiency, accuracy, and speed by which healthcare information is processed and then placed into the hands of decision makers.

In the field of education, who is doing what with research into the Semantic Web

The Infocomm Development Authority (IDA) of Singapore commissioned the development of two prototype Semantic Web portals to facilitate educators with finding useful educational material amongst the tens of thousands of valuable resources held in the many different web-based multimedia collections in the country. Some of the sites where the material is stored are not indexed by Google or Yahoo! Material forms include still images, audio/video clips, and interactive animations. Some material belongs to commercial vendors and is only available through paid subscriptions, and to know if the material exists requires subscription and login. The usual Web search engines are not able to work with the educator's context, and therefore cannot prioritize the results (Shabajee, McBride, Steer, & Reynolds, 2006). The developed prototype permits the user to "…locate resources by using a range of educationally relevant facets (e.g., I'm looking for 'images' that relate to the 'Japanese Occupation')," as well as to "…relate external information to a current search (e.g., topical news items and local educational events that are related to this curriculum topic" (Shabajee et al., 2006, p. 468).

Associated with the University of Toronto and Ryerson University in Canada, Uri Shafir and Masha Etkind developed a new semantic analysis methodology with two new e-learning tools that for the first, allows learners to explore semantic content through interactive concept discovery learning and, for the second, the Meaning Equivalence Reusable Learning Objects (MERLO) that focuses learners' attention on meaning (Shafrir & Etkind, 2006). In this article, Shafir and Etkind (2006) describe what is meant by a *semantic search*, and they introduce an operational definition of the word *concept* as it would be used within context of semantic analysis; this operational definition leads to the authors' *concept parsing algorithms* (CPA), which is "…a generic semantic procedure that identifies the lexical labels and building blocks of concepts" (Shafrir & Etkind, 2006, p. 426; Shafrir & Etkind, 2005). The authors describe evaluative implementations of the approach and the application of the technologies in multiple disciplines within high school and college level courses in Russia (the Russian Academy of Sciences and the Lycee Technical-Physical High School of Ioffe Institute in St. Petersburg) and Canada (Ryerson University, Ontario Institute for Studies in Education, George Brown College in Toronto, and the Material and Manufacturing Ontario Center of Excellence): the course areas include Math, Physics, Biology, History of Western Architecture, Foundation of Learning and Development, English as a Second [or Foreign] Language, and in Risk Management in the Supply Chain. Among the eight positive results, four of the most interesting (this author's opinion) include the following:

> "Iterative Concept Discover Learning is intuitive, learner centered and empowers active learning; it exposes learners to different points of views and multiple representations of conceptual content, accommodates instruction geared to different learning styles and augments language proficiency of new immigrants (ESL) and students with reading difficulties;"
>
> "MERLO pedagogy is effective across different populations of researchers, instructors and learners and across disciplines;"
>
> "The new semantic e-learning tools provide detailed formative and summative assessments and enhance learners' conceptual thinking;"
>
> "MERLO is technologically scalable and can be implemented in low-tech classrooms as well as online" (Shafrir & Etkind, 2006, pp. 442-443).

Miltiadis Lytras and Ambjön Naeve (Lytras & Naeve, 2006) in their article titled, "Semantic e-Learning: synthesising fantasies" identify that semantic e-learning represents an important convergence of the worlds of learners, pedagogues, and technologists. Lytras and Naeve (2006) present the technological framework of the field as a staircase graphic, which includes (from the bottom up) Unicode and URI (Uniform Resource Identifier); XML + NS (Namespace) + XMl Schema; RDF + RDF Schema; Ontology Vocabulary; Logic; Proof; and Trust (whereby RDF+RDF Schema through Proof are bound by a digital signature for content validation and authenticity). The authors present an e-learning roadmap (to see the integration of the Semantic Web) from 2004 through 2010. Then Lytras and Naeve (Lytras & Naeve, 2006) present an overview process map that illustrates the knowledge management life cycle, how that life cycle engages with pedagogy, and from there how context is assembled, whereby learners find the access point. Lytras and Naeve also bring forward provocative questions to bind the future of research into the Semantic Web and e-learning. The two key questions are: "Does e-learning promote effective learning?" and "Can you prove it?" (Lytras & Naeve, 2006, p. 484). Lytras and Naeve offer the following response:

> "In trying to answer this question, a lot of effort has been invested in the modeling of the learning process with a parallel intensive analysis of the content requirements. However, this 'context versus content rivalry' seems to miss a critical factor for the effectiveness of technology-enhanced learning. Learning is a unique experience for learners. So, if we make general hypotheses about the learners' readiness to learn, then we in fact decontextualise the problem from its key ingredient: the individual learning needs. Personalised and adaptive learning that takes into account such individual learning needs is a very difficult but important issue that is being addressed, e.g., within the PROLEARN network of excellence (www.prolearn-project.org)" (Lytras & Naeve, 2006, p. 484).

There are a few additional articles worthy of presentation within this category, but in the interest of brevity and time, this author elects to not present them.

Discussion

This paper begins with a vision, presents a working reference for the elements involved with the Semantic Web, and then presents examples of some of the efforts within the field, both in commercial development of the infrastructure, as well as examples of the technology as being practiced in healthcare, and a lone educational example. The paper then presents two critical research efforts by different teams as the field might impact the field of education. An unverified review of the literature would seem to indicate that nearly all of the effort, as pertains specifically to the field of education, is taking place outside of the United States. Commercial investment within the United States would seem formidable, given the short list derived through the short research of this paper with names such as Oracle and IBM.

Given what the technology is purported to be capable of achieving, and given the fact that most efforts are not focused on educational applications, the question of whether institutions of higher education, and specifically schools of education and research, should be interested becomes obvious: schools of education and educational research should most emphatically take interest and become involved. At a time when the primary focus is not on education, the technology development process will lack the depth of research experience and perspective of researchers and practitioners within the fields of educational psychology, educational research, and instructional design and technology. Learning theories with pedagogical frameworks need to be part of the developmental dialog into the Semantic Web.

An important shortcoming of this paper is the lack of additional reviews: the initial literature search yielded 117 articles. There most likely exist additional articles that may improve the current perspective of the overall level of activity by educational researchers on the development of the Semantic Web. In addition, further research should include a count of those educational research programs and note the number associated with institutions of higher education within the United States. And a follow-up should be conducted on the tools and approaches of CPA by Shafrir and Etkind (2006) to determine if they might be useful or applicable to the college of education in the instructional technology track.

Schools of education and research within higher education must become interested and be involved, such that we can contribute to the quality of the developing Semantic Web. Lytras and Naeve (2006) present the proper perspective: "In our opinion, creating such an ecosystem of quality is the most valuable contribution that the Semantic Web can provide for learning because it can provide enriched approaches to organizational learning and can effectively enhance the quality of the global, lifelong learning process" (Lytras & Naeve, 2006, p. 489).

References

Borland, J. (2007a). A smarter web. *Technology Review, 110(2)*(March/April), 64-71.

Borland, J. (2007b). A smarter web: New technologies will make online search more intelligent-and may even lead to a" web 3.0". *TECHNOLOGY REVIEW-MANCHESTER NH-, 110*(2), 64-71.

Feigenbaum, L., Herman, I., Hongsermeier, T., Neumann, E., & Stephens, S. (2007). The semantic web in action. *Scientific American, 297(6)* (December), 90-97.

Kobielus, J. (2007). The semantic web: Meaning and SOA. *Business Communications Review, 37(10)* (October), 30-36.

Lytras, M., & Naeve, A. (2006). Semantic e-learning: Synthesising fantasies. *British Journal of Educational Technology, 37(3)* (May), 479-491.

Reigeluth, C. M. (1999). What is instructional-design theory and how is it changing? In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory* (II ed., pp. 5-29). Mahwah, NJ: Lawrence Erlbaum Associates.

Shabajee, P., McBride, B., Steer, D., & Reynolds, D. (2006). A prototype semantic web-based digital content exchange for schools in singapore. *British Journal of Educational Technology, 37(3)* (May), 461-477.

Shafrir, U., & Etkind, M. (2005). Concept parsing algorithms: Mapping the conceptual content of disciplines. *Toronto: PARCEP,* (January, 2005)

Shafrir, U., & Etkind, M. (2006). e-learning for depth in the semantic web. *British Journal of Educational Technology, 37(3)* (May), 425-444.