Semantic Technologies for Learning and Teaching in the Web 2.0 Era

Thanassis Tiropanis, Hugh Davis, Dave Millard, and Mark Weal,
University of Southampton

Deploying semantic tools and services over a field of linked data could be a way to address many current challenges of higher education.

The strengths of semantic technologies for learning and teaching, and their benefits for digital libraries, virtual communities, and e-learning, have been a major topic of discussion during recent years. Experts argue that semantic technologies can enhance the advanced learning experience by using the expressive power of metadata to describe learning content, people, and services, and then matching these intelligently. However, concerns over the feasibility of ontology consensus and over annotating the enormous amount of content available on the Web have made globally available, interoperable, semantic-rich metadata for learning resources a long-term vision.

The recent emergence of Web 2.0 systems has enabled a lightweight knowledge-modeling approach (sometimes called *folksonomies*) based on techniques such as community tagging, clustering, and community authoring. Such Web 2.0 systems are already used in education.[1] In this respect, semantic technologies are already affecting the way we learn and teach, and this could have an increasing impact as the tools become more sophisticated (see http://ochre.wordpress.com/2007/11/21/semantic-web-session-at-cetis).

In the UK, the Joint Information Systems Committee (JISC) commissioned the Semantic Technologies for Learning and Teaching project (SemTech, www.semtech.ecs.soton.ac.uk) to conduct a survey on semantic technologies' use in higher education today, taking into account these recent developments. One of the first tasks in SemTech was to establish the relevance of semantic technologies to higher-education learning, teaching, and support challenges. In this article, we discuss our findings on the current use of semantic technologies in UK educational institutions. We also address issues that require further attention for these technologies to see widespread adoption in higher education.

Hard and Soft Semantic Technologies

For the SemTech survey, we distinguished between hard and soft semantic technologies. *Hard semantic technologies* provide ways to express meanings of resources

1541-1672/09/$26.00 © 2009 IEEE
Published by the IEEE Computer Society

and their relationships in machine-processable formats, and ways to draw conclusions—to reason—based on these meanings. Examples include the Resource Description Framework (RDF), the Friend of a Friend ontology (FOAF), the Simple Knowledge Organization System (SKOS), and triple stores. *Soft semantic technologies* provide ways to express the meanings of resources in formats that humans can interpret, or in formats that employ domain-specific information structures. Examples include traditional tagging tools, topic maps, and domain-specific XML schemas.

Many Web 2.0 technologies for education seem to employ soft semantic technologies such as wikis, tagging, and topic maps. At the same time, requirements for interoperability across data sources and for more advanced, efficient resource discovery seem to encourage a transition from soft to hard semantic technology use. An example of this trend is the transition from the soft semantic knowledge structure in Wikipedia to the hard semantic knowledge structure (in RDF) in DBpedia and Freebase.

Relevance of Semantic Technologies to Higher Education

The SemTech project engaged with the JISC Centre for Educational Technology and Interoperability Standards (CETIS) Semantic Technology working group and with several UK universities to organize a workshop in London in January 2009. The purpose was to identify current challenges from the perspectives of learning, teaching, and institutions of higher education.

Semantic tools and services were found to be relevant to the following learning and teaching challenges:

- course creation, delivery, and revision—assisting the workflow by recommending relevant content and people in the context of the course and the institution;
- student assistance—recommending resources that match the topics of students' assignments and people that could support their activities;
- access to teaching and learning material across institutions—supporting contextualized searches on the basis of field of study, types of teaching and learning activities, or pedagogical framework;
- group formation for collaborative work—selecting groups on the basis of students' background, personal preferences, and successful prior collaboration;
- critical thinking and argumentation—providing argument visualizations and linking relevant discussions;
- cross-curricular activities in emerging areas—matching people and resources across schools or departments within a single institution or multiple institutions;
- personalized knowledge construction; and
- group knowledge construction.

From the viewpoint of a higher-education institution, the workshop identified several additional challenges that semantic technologies could address:

- curriculum development or alignment—assisting the workflow by identifying niches for new courses and comparing courses offered by different institutions;
- student retention—efficiently supporting students and monitoring their progress; linking data currently scattered across different databases, Web pages, and spreadsheets in each institution to support advanced data analysis;
- data transparency—selectively and securely exposing institutions' information to relevant parties in semantic, interoperable formats such as RDF;
- degree program accreditation—making relevant data more readily available to professional bodies;
- interaction with funding bodies—enabling information integration, searching, and matching to make universities' knowledge capital and information on organizational resources more accessible;
- cross-institutional collaboration—providing large repositories, such as triple stores, in which multiple universities could efficiently store, search, and manage information; currently universities' relevant information systems aren't interoperable, and deploying linked-data repositories in each institution is costly; and
- intellectual property statements—providing a framework to enable each institution to differentiate its intellectual property for the resources it is eager to expose; declarative semantic formats could unambiguously state this information and encourage knowledge dissemination.

Categorizing Semantic Technologies for Higher Education

The SemTech survey on the availability of tools and services that relate to learning, teaching, and institutional challenges (http://semtech-survey.ecs.soton.ac.uk) resulted in the identification of more than 30 relevant tools and services. We coarsely classify them into four main categories on the basis of their main functionality:

- collaborative authoring and annotation tools, including semantic wikis and argumentation tools;

Semantic Technologies for Higher Education

Through the SemTech survey, we identified more than 30 semantic tools and services that could help higher education institutions meet challenges in learning, teaching, and administration. The following list includes some notable examples. Further information is available online at http://semtech-survey.ecs.soton.ac.uk.

Collaborative Authoring and Annotation Tools

AceWiki	http://attempto.ifi.uzh.ch/acewiki
Cicero	http://cicero.uni-koblenz.de/wiki/index.php/Main_Page
Mymory	www.dfki.uni-kl.de/mymory
Kiwi	www.kiwi-project.eu
Compendium	www.aktors.org/technologies/compendium
Debategraph	http://debategraph.org
PROWE	www.prowe.ac.uk

Searching and Matching Tools

ArnetMiner	www.arnetminer.org
Twine	www.twine.com
Watson	http://watson.kmi.open.ac.uk/WatsonWUI

Repositories and Virtual Learning Environments

Freebase	www.freebase.com
DBpedia	http://dbpedia.org
CIP	www.ilrt.bris.ac.uk/whatwedo/projectsaz/project?search=CIP
Project Gutenberg	www.gutenberg.org
MyExperiment	www.myexperiment.org
DSpace	www.dspace.org
EPrints	www.eprints.org

Infrastructural Tools and Services

D2R Server	www4.wiwiss.fu-berlin.de/bizer/d2r-server
TALIS	www.talis.com/platform
Virtuoso	http://virtuoso.openlinksw.com
RKBexplorer	www.rkbexplorer.com
Yahoo SearchMonkey	http://developer.yahoo.com/searchmonkey

- searching and matching tools using semantic technologies;
- repositories and virtual learning environments (VLEs) that import and export their data using semantic technologies; repositories for scientific resources that can expose metadata in RDF; and
- infrastructural tools and services that enable exposing databases or integrating data sources within or across organizations in interoperable semantic formats.

See the sidebar, "Semantic Technologies for Higher Education," for examples of tools in the four categories and relevant Internet addresses.

Regarding the value of semantic technologies in the tools and services reported in the survey, in more than four in five cases the tools' value lies in providing well-formed metadata. In almost half the cases, semantic technologies also provide data integration and interoperability. In just under two out of five cases, the tools' value lies in data analysis and reasoning.

Collaborative Authoring and Annotation Tools

Collaborative content authoring and annotation tools and services respond in several ways to the challenges we identified earlier. First, they allow precise representation of shared knowledge and can recommend related content and people for collaborative activities. Semantic wikis such as AceWiki and Kiwi and tools such as ArnetMiner are relevant examples.

Tools in this category also provide documentation and support collaborative workflows on a large or small scale for teaching and learning activities and for collaboration across departments and institutions. MyExperiment, for example, allows documentation of experiment workflows.

Another function of collaborative authoring and annotation tools is to support argumentation and visualization of arguments and relevant resources to enable critical thinking. Debategraph and Cicero, for example, support argumentation.

Finally, such tools can represent the shared knowledge capital of higher-education institutions in ways that can be accessed by different faculties, schools, institutions, and the public. Semantic wikis could provide one way to achieve this. Collaborative knowledge modeling as done in Freebase could also be relevant.

Searching and Matching Tools

Searching and matching tools can provide contextualized queries and searches across repositories of teaching material or repositories in different departments or institutions. Additionally, these tools can support people matching for collaborative activities. ArnetMiner is a relevant example from the surveyed technologies.

Semantic search frameworks such as Yahoo SearchMonkey could enable

the development of more efficient question-and-answer systems and knowledge bases for learning and teaching support.

Searching and matching tools are also appropriate for exposing the institution's expertise to the outside world to attract funding and student enrollment. ArnetMiner is currently the most representative example of such tools.

This category would seem to be where we might find tools to combine information from different sources within institutions to enable better monitoring of student progress and to provide recommendations based on declarative statements that can be validated. None of the surveyed tools, however, seem to address this challenge at the moment.

Repositories, VLEs, and Authoring Tools

Repositories, VLEs, and authoring tools can provide semantic annotation of content to support more precise knowledge construction, interoperability, and integration of repositories across institutions. Representative examples are EPrints, DSpace, DBPedia, Freebase, and Project Gutenberg.

Semantic enrichment of repository classifications can enable more efficient resource discovery and interoperability. To a certain extent, Freebase and DBpedia can be used for this purpose.

Infrastructural Technologies

Infrastructural technologies such as RKBexplorer can offer large repositories for efficient storage and search of data from different sources in different institutions and repositories.

Such technologies can also expose some organizational data to partners or the Web, support interoperable, machine-processable data formats, and integrate data from different sources. Several infrastructural tools—such as TALIS, Virtuoso, and D2R Server—enable semantic enrichment and exposure in semantic formats.

Semantic Technology Adoption in UK Higher Education

Between 10 and 20 UK universities appear to have begun using wikis on an institutional level to support learning and teaching. Current activities to develop semantic wikis indicate an awareness of the advantages of adding meaning to the relationships among wiki resources. In addition, reasoning tools to support argumentation, where the relevance between arguments can be precisely identified, can be used to navigate and visualize discussions. The survey revealed an awareness, in certain cases, of the potential of technologies for visualizing discussions, reflected in the development of tools such as Compendium. This indicates a trend for the institutional adoption of semantic wikis and argumentation tools in the near future.

Repositories that already employ semantic metadata feature searching and matching functionalities. The University of Tsinghua in China has deployed expert matching through ArnetMiner.[2] The JISC-funded project Awesome (http://awesome.leeds.ac.uk) provides software that combines semantic wiki and pedagogy-aware inline recommendations to empower academic writing. Several schools at the University of Leeds, University Coventry, and University of Bangor are using this software.

Repositories are the most widely adopted type of infrastructure featuring semantic extensions in UK educational institutions. More than 40 universities in the UK appear to employ repositories to publish their research results, conference and journal articles, presentations, or course material. They most often use the DSpace and EPrints platforms, and the fact that both of these platforms are adding support for RDF shows the potential for the adoption of semantic technologies for educational repositories.[3,4]

A handful of universities expose SPARQL endpoints—for example, the University of Southampton (http://imageweb.zoo.ox.ac.uk/wiki/index.php/DefiningImageAccess/Project/SCULPTEUR) and the University of Oxford (http://zoo-garos.zoo.ox.ac.uk/ibrg/index.php/Image_Bioinformatics_Research_Group_home_page). For the moment, however, the SPARQL endpoints are in the context of research projects. Requirements for repository integration and queries on larger data sets might encourage the use and further development of infrastructural semantic tools and services such as large linked-data repositories and semantic enrichment and data integration platforms. The School of Electronics and Computer Science at the University of Southampton provides information on entities such as its people, roles, interests, courses, seminars, and presentations in RDF format (http://id.ecs.soton.ac.uk/docs/). Visitors can obtain this information using HTTP or RKBExplorer.

The vision of the Semantic Web, or Web 3.0, has inspired significant research output. Many researchers agree that some form of the Semantic Web will inevitably result from the development of existing technologies.[5] Others place the adoption of semantic-aware applications for education about four to five years away,[6] but we believe that this activity will just be commencing during this period.

A key conclusion from our survey is the promising potential of a higher-education linked-data[7] field, populated with interoperable semantic data on a large scale; we expect this to provide significant value with regard to learning, teaching, and support challenges. Before a linked-data field across higher education reaches critical mass, however, we expect the emergence of advanced reasoning applications that will rely on a limited number of repositories and on ad hoc mappings of unstructured data to specific ontologies.

There are certain barriers to the exposure of linked data from higher education institutions: cost, confidentiality, and the availability of applications that rely on linked data. However, the availability of a plethora of tools for exposing data sources as linked data could reduce costs. In addition, most of the higher-education challenges we identified in this article could be addressed by using linked-data formats for information that is already available on institutions' Web pages and is therefore not confidential. The emergence of applications that will demonstrate the value of a linked-data field for higher education might generate a network effect and drive further developments.

It is critical that further work and research identify and address the issues related to fostering the development of a global linked-data field based on optimized metadata repositories across educational institutions. The performance of linked-data queries on a large scale will require further attention from the research community. The development and availability of tools that will assist collaborative ontology building and efficient mapping of linked data to those ontologies could be a decisive factor in the development of pedagogically meaningful semantic tools and services.

THE AUTHORS

Thanassis Tiropanis is a lecturer in the Learning Societies Lab, School of Electronics and Computer Science, at the University of Southampton. His work focuses on linked data and semantic technologies for higher education and virtual communities. His interests also include technologies and infrastructures for communication, collaboration, and learning. Tiropanis has a PhD in computer science from the University of London. He is a senior member of the IEEE, a member of the British Computer Society, and a fellow of the Higher Education Academy. Contact him at tt2@ecs.soton.ac.uk.

Hugh C. Davis is a professor and head of the Learning Societies Lab, School of Electronics and Computer Science, at the University of Southampton. He is also the University Director of Education with responsibility for e-learning. His research interests include the applications of hypertext for learning, open hypertext systems, and architectures for adaptation and personalization; more recently he has focused on Web and grid service frameworks for e-learning, with a particular concentration on the assessment domain. Davis has a PhD in computer science from the University of Southampton. He is a member of the British Computer Society and a fellow of the Higher Education Academy. Contact him at hcd@ecs.soton.ac.uk.

David Millard is a senior lecturer of computer science in the Learning Societies Lab, School of Electronics and Computer Science, at the University of Southampton. His research interests have long involved hypertext and Web research—first the area of open, adaptive, and contextual hypermedia, and more recently Web 2.0, the Semantic Web, knowledge and narrative interfaces, and the impact of Web literacy on e-learning and mobile learning. He is interested in the ways that people use information systems in the wild, and how we can use emergent social, organizational, and semantic structures to help them make sense of their world. Millard has a PhD in computer science from the University of Southampton. Contact him at dem@ecs.soton.ac.uk.

Mark Weal is a lecturer in the Intelligence, Agents, Multimedia Group and the Learning Societies Lab, School of Electronics and Computer Science, at the University of Southampton. His research interests include Web science and the application of Semantic Web technologies in e-learning, healthcare, and pervasive systems. Weal has a PhD from the University of Southampton. Contact him at mjw@ecs.soton.ac.uk.

References

1. P. Anderson, "What Is Web 2.0? Ideas, Technologies and Implications for Education," *JISC Tech. & Standards Watch,* Feb. 2007, www.jisc.ac.uk/media/documents/techwatch/tsw0701b.pdf.
2. J. Tang et al., "ArnetMiner: Extraction and Mining of Academic Social Networks," *Proc. 14th Int'l Conf. Knowledge Discovery and Data Mining* (KDD 08), ACM Press, 2008, pp. 990–998.
3. D. Huynh, S. Mazzocchi, and D. Karger, "Piggy Bank: Experience the Semantic Web Inside Your Web Browser," *Web Semantics: Science, Services, and Agents on the World Wide Web,* vol. 5, no. 1, 2007, pp. 16–27.
4. S. Harris and N. Gibbins, "3store: Efficient Bulk RDF Storage," *Proc. 1st Int'l Workshop Practical and Scalable Semantic Web Systems,* 2003, http://km.aifb.uni-karlsruhe.de/ws/psss03/proceedings/harris-et-al.pdf.
5. J. Ohler, "The Semantic Web in Education," *Educause Q.,* vol. 31, no. 4, 2008, http://net.educause.edu/ir/library/pdf/EQM0840.pdf.
6. L. Johnson, A. Levine, and R. Smith, *The 2009 Horizon Report,* The New Media Consortium, 2009.
7. K. O'Hara and W. Hall, "The Semantic Web," *Encyclopedia of Library and Information Science*, 2nd ed., M.J. Bates, M.N. Maack and M. Drake, eds., Taylor & Francis, in press, 2009; http://eprints.ecs.soton.ac.uk/17126/1/hall-ohara-elis-semantic-web.pdf.