Linked Data in E-Learning: A Survey

Katrin Krieger, Dietmar Rössner

Knowledge Based Systems and Document Processing research group, Department of Computer Science, Otto-von-Guericke-University, P.O. Box 4120, 39106 Magdeburg, Germany
E-mail: \{kkrieger, roesner\}@ovgu.de

Abstract.

The evolution of the Web to a place for semantic data integration and the development of Linked Data infrastructures offer new possibilities and promise better and more effective ways of interaction, adaption, data sharing and re-use of data between applications.

E-Learning and Web-based education have become essential building blocks of education in general, delivering course material over the Internet anytime and anywhere into a matter of course. With the Semantic Web having a special impact on how we deal with seeking information online, it will also have a particular impact on the field of e-Learning. The outlook to being able to integrate shared learning resources augmented with meta-data, intelligent solutions to information searching and knowledge management shape the vision of an Educational Semantic Web.

In this paper, we will give an overview of the aspects of the Semantic Web as well as possible effects and recent applications of Semantic Web technologies in Web-based educational environments.

Keywords: Linked Data, Semantic Web, web of data, e-learning, metadata, education, ontologies

1. Introduction

The World Wide Web (WWW) is a major global space for publishing and sharing Web documents. Its open, generic and extensible nature is a key feature for its unconfined growth ([19]).

On the Web, documents written or generated in HTML are interconnected via hyperlinks that enable the user to navigate and browse through the global document space. Since they are untyped, those connections have no explicit semantics. The HTML data format does not provide any means to connect specific entities in a document by typed links to other related entities in another document. A hyperlink expresses a relation between two documents, but it does not say anything about the contained entities or the kind of their relations.

1.1. The Semantic Web

Humans are able to extract desired information from heterogeneously structured Web documents, but such information is undecipherable for machines with respect to knowledge discovery and logical inference. Berners-Lee stated this problem as follows: “The Web was designed as an information space, with the goal that it should be useful not only for human-human communication, but also that machines would be able to participate and help. One of the major obstacles to this has been the fact that most information on the Web is designed for human consumption, and even if it was derived from a database with well defined meanings (in at least some terms) for its columns, that the structure of the data is not evident to a robot browsing the Web.” ([4]).

To overcome these obstacles, the concept of the Semantic Web has been introduced as an extension to the existing WWW, allowing “computers to intelligently search, combine, and process Web content based on the meaning that this content has to humans.” ([13], also cf. [4], [5]). Ohler states “the Semantic Web con-
verts ‘display only’ information to meaningful information [...] and allows users to find relationships between tagged information [...] using inference rules and data organizational tools [...]” (225).

The purpose of the Semantic Web is to enable machines to access more information automatically. This is only possible, if the semantics of Web resources are explicitly specified in a format that is processable by computers. Strictly speaking, this requires a “markup” for the recognition of semantics as well as the existence of formal semantics allowing computers to draw conclusions.

Hitzler et al. (18) identified three main topics as core concepts of the Semantic Web:

1. **Building models** in order to describe knowledge about the world in abstract terms,
2. **Calculating with knowledge** in order to use these knowledge models and their encoded information with reasoning machines that can draw meaningful conclusions,
3. **exchanging complex information** in order to distribute, interlink, and reconcile knowledge on a global scale.

The exchange of complex information has already been addressed by classical Web technologies (i.e., transport protocols like TCP/IP and file formats like RSS). The tasks of modeling knowledge and calculating with such encoded knowledge have a long history, however, in the context of the vision of the Semantic Web, they still form difficult challenges that have to be addressed in particular.

Devedžić expands this view by pointing out four categories of issues related to the Semantic Web: languages, ontologies, semantic markup of pages on the Semantic Web, and services the Semantic Web is supposed to provide (13).

We will go into detail of these aspects when examining applications of the Semantic Web in the field of e-learning (see Section 2 and 3).

As a Web of Data, the Semantic Web adds machine-understandable data about documents and their entities and how they are related to each other. The means for the construction of such an infrastructure are formats and technologies like the Resource Description Framework (RDF), RDF Schema (RDFS), Web Ontology Language (OWL), and SPARQL, often referred to as the Semantic Web layer cake or Semantic Web technology stack (see Fig. 1).

![Fig. 1. The Semantic Web technology stack (source: http://www.w3.org/DesignIssues/diagrams/sweb-stack/2006a.png)](http://www.w3.org/DesignIssues/diagrams/sweb-stack/2006a.png)

### 1.2. Linked Data

Berners-Lee coined the term **Linked Data** describing a set of best practices for publishing and connecting structured data on the Web (3, 6). These practices are the foundation of the evolution of the Web of Documents to the Web of Data, a global data space connecting data from a multitude of different domains.

In order to become part of a single global data space, Berners-Lee proposed a number of (technical) rules for publishing data on the Web that have become known as **Linked Data Principles** (3):

1. Use URIs as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
4. Include links to other URIs, so that they can discover more things.

The opinions on the relationship between Linked Data and the Semantic Web differ. Sometimes both terms are used synonymously, others, however, refer to the Semantic Web as being made up of Linked Data. Berners-Lee, having coined both terms, has described Linked Data as “the Semantic Web done right” or “while the Semantic Web [...] is the goal or the end of this process, Linked Data provides the means to reach

\[\text{http://www.w3.org/2008/Talks/0617-lod-tbl}\]
that goal.” (6). The distinction between the Semantic Web as a concept and Linked Data as its realization as described above goes with this description.

In summary, the Semantic Web is about data integration (cf. 23). Before data can be integrated, it has to be shared. There has to be a shift from data silos containing unstructured or semi-structured data to open data repositories with semantically enriched data, that can be accessed and interlinked. In the area of e-learning, such data can be shared as learning objects (LO) or learning entities of any kind, respectively.

1.3. Implications for e-learning – visions of the Educational Semantic Web

Dicheva identifies three generations of Web-based educational systems (14). The systems of the first generation provide a central entry-point for accessing learning materials and online course, e.g., LMS and educational portals. The systems of the second generation employ Web and AI technologies to intelligently support personalization and adaption. Such systems are called educational adaptive hypermedia systems. The third generation of Web-based educational systems is a class of ontology-aware software, using and enabling Semantic Web standards and technologies in order to grant scalability, reusability and interoperability of educational material that is distributed over the Web.

As the Semantic Web affects the nature and behavior of the Web as it is known today, it will have special impact on particular fields, that use Web technologies, e.g., e-learning. The vision of an Educational Semantic Web arises.

Wheeler states that the Semantic Web will “not only promote learning that is more richly collaborate, it will also enable learners to come closer to ‘anytime anywhere’ learning and will provide intelligent solutions to web searching, document management and organisation of content” (30).

According to Anderson and Whitelock, the Educational Semantic Web “is based on three fundamental affordances. The first is the capacity for effective information storage and retrieval. The second is the capacity for nonhuman autonomous agents to augment the learning and information retrieval and processing power of human beings. The third affordance is the capacity of the Internet to support, extend and expand communications capabilities of humans in multiple formats across the bounds of time and space” (2).

The ontological graph showing relations between concepts.

Fig. 2. An ontological graph showing relations between concepts.

1.4. Structure of this article

The Semantic Web and Linked Data technologies raise high expectations with respect to providing solutions to open problems, especially in a field like e-learning.

The Semantic Web and its technologies are part of an evolutionary process, ment to develop and grow. Hence, the fullscale impact it has on the field of e-learning can only be estimated at this point of time.

This paper gives an overview on the application of Semantic Web technologies and Linked data to e-learning as they exist today.

The remainder of this article is structured as follows: Section 2 investigates the special application of models in the field of e-learning. Section 3 discusses the application of Linked Data in two particular areas of e-learning, namely adaption and personalization, and assessment. Finally, we will give a conclusion and future outlook in Section 4.

2. Building models with educational ontologies

In order to structure data in a comprehensible and machine-accessible way, the Semantic Web relies heavily on conceptual structures, i.e., formal models. Those models have to provide a controlled vocabulary of explicitly defined concepts describing individuals (instances), classes (concepts), attributes, and relations (see Fig. 2). Ontologies “help people and machines to communicate concisely — supporting semantics exchange, not only syntax” (22).

Cubric and Toric state that “the work on ontologies in e-learning domains has been primarily concentrated
on ontological formalization of learning objects, instructional processes and learning designs [...]” (11).

In order to create Web-based learning material that is ontologically annotated, there is a need for standard ontologies that cover different aspects of teaching and learning. Besides standard domain ontologies dealing with knowledge about particular domains, ontologies about pedagogical knowledge, e.g., curriculum sequencing, student modelling, grading and other pedagogical issues are required (13).

Devedžić identifies a lack of a standard vocabulary for education and instructional design as a particular problem. In fact, there are some efforts on developing such official standard vocabularies, i.e., extending existing vocabularies like Learning Objects Metadata (LOM) or Sharable Content Object Reference Model (SCORM) for usage in Semantic Web applications. Different working groups, including the IEEE Learning Technology Standards Comittee, the IMS Global Learning Consortium, Inc. or the ISO/IEC JTC1/SC36 Standard Comittee are working on standards for educational and instructional vocabularies to ensure reusability and interoperability of educational applications based on Linked Data. Section 2.2 gives an overview of work that has been done in that area.

2.1. Educational domain ontologies

Hatala and others identified in 2009 a lack of special domain ontologies formalizing the content of learning courses: “However, the main problem with all approaches that have shown the benefits of ontology adoption in e-learning systems is that they are assuming that required ontologies are available. However, this is not a realistic assumption, at least not for domain ontologies, i.e., ontologies that formalize the subject matter of learning courses. The lack of these ontologies is to be attributed to their creation process which is overly difficult and time consuming for educational practitioners.” (17). Nevertheless, they also point out, that generic domain ontologies, that can be found in ontology libraries like Swoogle or OWLSeek could be searched for ontologies for a particular domain to be used in educational contexts.

There are some projects concerned with domain ontologies to be used particularly in learning contexts. One of these projects aims at publishing math lecture notes as Linked Data (David et al. [12]). Therefore they annotated their material written in sTeX (semantically enhanced TeX), so that content entities can be interlinked with arbitrary metadata vocabularies. This annotated material could then be used as input for a process, involving various tools for the creation and publication of RDF data as well as human-readable output. As default vocabularies a custom ontology stored as OMDoc (Open Mathematical Document) is used, which is inspired by the Hypertextual Electronic Library of Mathematics (HELM).

ChemCloud uses a different approach ([29]). In the chemistry and life sciences domain, various ontologies as well as knowledge bases are available. Those resources cause a flood of data from heterogeneous sources that require semantic integration. In order to make such data available as Linked Data, proper RDF-ization of knowledge base content and semantic integration with other ontologies for compatibility reasons pose a significant challenge. Therefore, extracted entities are inserted into the ChemCloud ontology and adjusted to established ontologies from the chemistry domain, e.g., CheBI (Chemical Entities of Biological Interest), ChemAxion or Bio2RDF.

2.2. Paedagogical ontologies

For the purpose of sharing, locating and re-using educational resources on the Educational Semantic Web, it is not only important to know about the content of these resources with respect to their domain, but also to their paedagogical structure, e.g., which prior knowledge is necessary to understand this piece of learning content or what type of interaction with the learner is associated with this resource.

There exist different educational standards for describing learning resources with respect to their content and educational structure, among them LOM,
SCORM and IMS Question and Test Interoperability specification¹⁴ (IMS QTI). These standards aim at reuseability and interoperability of learning content.

LOM is an IEEE standard for describing learning objects, a piece of learning content. The data model provides a vocabulary for characterizing such learning objects with respect to attributes like title, keywords, difficulty, or interactivity type. The standard also describes bindings of the LOM data model, i.e., representations of LOM records in XML or RDF [24].

SCORM is a collection of standards for Web-based learning defining communication between learning content providers, content packaging and sequencing of learning objects. SCORM data is expressed in XML, bindings to RDF, however, do not exist so far.

IMS QTI is a standard for describing assessment content and results for the exchange of such content between different Web-based learning systems. The standard comprises a data model describing the structure of assessment content and an XML binding. Dolog proposed also a binding of the main QTI elements to RDF [15].

These three standards don’t stand for their own, moreover they build on each other or contribute to the other standards, respectively.

For example, the IMS Learning Resource Metadata Information Model is based on the IEEE 1484.12.1 LOM standard, while the SCORM Metadata Information Model is a reference to the IMS Learning Resource Metadata Information Model.

Hitherto, there is no ‘native’ RDF binding for SCORM. Qu and Nejdl discussed an approach to bridge the complex SCORM data model and the RDF data model by generating “XQuery-enabled triple-like SCORM meta data views” [27]. In this work from 2003, they also report that “the triple-like SCORM metadata data view is not 100% compatible with the IMS Learning Resource Metadata RDF binding, which is expected to become the potential SCORM Metadata RDF binding.” However, the recent status of such an RDF binding for SCORM remains unclear.

Cheniti-Belcadhi et al. introduced a framework for e-assessment that uses resources annotated with Semantic Web metadata (i.e., RDF) and LOM or QTI metadata, encoded in RDF as well (see Listing 1) [9]. In this framework, the combination of Semantic Web technologies like RDF or RDFS and learning standards like LOM (for learning resources) and QTI (for assessment resources) for the annotation of educational resources yields a metadata description that has been extended with vocabulary from educational standards. This description is then deployed for a dynamic and adaptive assessment process. The authors formalize this process with First-Order-Logic, a reference implementation is done with the rule language TRIPLE [15].

2.3. Semantic user models

For individual and personalized learning environments, it is not only important to model domain knowledge, but also to formalize learners and their attributes.

There exist several approaches to map persons with their attributes, activities, and relations to other people (i.e., personal/social networks) to formal models.

The Friend of a Friend (FOAF) ontology [16] is a project aimed at describing social networks in a decentralized manner. FOAF uses RDF and OWL to express the provided vocabulary. Data (i.e., instances of foaf:person) is stored in FOAF profiles (see Listing 2), that can be used to retrieve information about people, e.g., all people I worked with on a project or all 3rd-degree friends of a particular person.

The Semantically-Interlinked Online Communities (SIOC) project [17] provides a vocabulary for describing

---

¹¹ http://www.imsglobal.org/question
¹² http://www.foaf-project.org/
¹³ http://sioc-project.org
data from the Social Web. This ontology is often used in conjunction with FOAF.

Wilson investigated the use of FOAF for building e-portfolios \[31\]). An e-portfolio – being a composite aggregated by multiple feeds (interest feeds, goal feeds, achievement feeds, competency feeds, folio feeds, reflective feeds) – can be constructed using the FOAF vocabulary: “One solution for constructing an e-portfolio is to aggregate a range of feeds to create a composite record; the FOAF (Friend of a Friend) model is useful here, as it contains properties that can hold most of the information you might want, such as contact information, goals, interests, and publications.” \(\) \[31\]).

Abel and others use FOAF – together with other vocabularies like Grapple Core \[18\] – for their framework U-Sem, “a people modeling service [...] for the semantic enrichment and mining of people’s profiles from usage data on the Social Web” \(\) \[11\]). The application of U-Sem in an e-learning context aims at a better adaption of the behavior of an e-learning system to the individual learner and his prior knowledge and learning performance. For building learner profiles U-Sem exploits data that is publicly available from social network applications like Twitter \[19\] or Delicious \[20\].

Softic and others investigated how social semantic e-learning platforms like ELGG \[21\] can be interlinked and enhanced with semantic tagging and the deployment of Linked Data principles (see \[12\]). They employ modelling vocabularies like FOAF, SIOC or MOAT \[22\] (Meaning of a tag) in conjunction with DBPedia as a reference dataset for interlinking \(\) \[28\].

User modeling, especially learner modeling has a long history (cf. \[21\]). With sophisticated ontologies describing user modeling aspects like described above, such models can be enhanced with semantic meta data and thus can exploit personal data available on the Web for adaption and personalization purposes in e-learning environments.

3. Educational Semantic Web services

With resources enriched with meta data and knowledge expressed in (educational) ontologies, Educational Semantic Web services are able to "calculate with knowledge" and thus generate, organize and personalize e-learning content.

In the following section we will present work related to two particularly important issues in the field of e-learning: adaption/personalization and assessment.

3.1. Adaption and Personalization

Adaption and personalization of learning content presentation and navigation has become a crucial issue in the domain of e-learning. Each individual learner can be characterized by a particular set of properties with respect to personal preferences, learning styles, or prior knowledge.

Including these aspects into Web-based learning environments pays off with a better learning outcome than a "one-size-fits-all” approach (cf. \[8\]).

The project OPAL (Ontology-based framework for personalized adaptive learning) provides adaptive features using ontologies and learner models for a Web-based Java programming course \(\) \[10\]). The learner model is assembled using preferences set by the learner himself and through (pre-)tests provided by

\[18\] http://wis.ewi.tudelft.nl/rdf/grapple-core.owl
\[19\] http://twitter.com
\[20\] http://www.delicious.com
\[21\] http://www.elgg.org
\[22\] http://moat-project.org/
the system to save learning progress and performance. The ontologies represent learning object content retrieved from distributed learning objects repositories as well as course relationships between the LOs, e.g., part-subpart structures or prerequisites LOs for particular LOs. Furthermore, also assessments are modeled as ontologies. The Java rule engine Jess is deployed for the generation of personalized tests and course sequencing, based on available ontologies and models.

Kim describes a different approach to personalization in e-learning systems deploying ontologies ([20]). Their system TagSES (Tagging-based social e-learning system) utilizes a tag ontology created with the help of user-provided tags (folksonomies) for personal recommendation of users with similar interests and relevant materials for particular lecture topics. Here, ontologies are used to model relationships between students, lecture materials and tags. A prototype of the TagSES system has been implemented for a course in music history. The system is based on three different models: a domain ontology to model concepts from the domain of music history, a tagging ontology and a rule base for reasoning rules. The tagging ontology represents the tagging activities of a user by connecting tags with these users and lecture materials. The tags from the folksonomy are therefore mapped to concepts or instances in the domain ontology. Students as well as instructors tag lecture material and user-generated content. The system can then recommend (a) users with similar interests, e.g., for forming project groups, and (b) relevant material about particular topics. Therefore, reasoning rules are defined in three steps: (1) clustering of tags associated with concepts from the domain ontology, (2) clustering of users based on the clusters in the domain ontology as a basis for development of the MCQ format specification. The ontology is based on the standard IMS (Instructional Management Systems) Test and Question information model (IMS, 2002) and it gets populated with question instances during the process of question generation.” ([11]). These question templates form a semantic mapping between the domain ontology and the MC test ontology. The mappings allow to generate questions of different types such that they correspond to levels from Bloom’s taxonomy of educational objectives (knowledge, comprehension, application, etc.) ([7]).

In earlier works Fernández-Breis and others discuss an approach to apply ontologies to support the evaluation of open questions-based tests ([16]), i.e., questions, that cannot be automatically evaluated in an automatic way. Such questions are considered a better means of evaluation than MC tests, since “they are the most natural and they produce a better degree of thought” ([15]), but they ‘cost’ more time for marking and grading. In the approach introduced by Fernández-Breis and others the course knowledge, expected responses to particular test questions and the actual responses by students to these questions are semantically annotated. The semantic annotations of the responses include a set of ontological entities like concepts, properties and relations. Similarity measures are applied to compare both response types with respect to ontological entities of the same category. The actual evaluation of the correctness of a student’s answer is calculated

---

1 [http://www.jessrules.com](http://www.jessrules.com)
by computing the similarity between entities from the course ontology, the expected answer and the actual answer. The similarity values must not exceed a particular severity threshold value.

4. Conclusion and Outlook

The evolution of the Web to a Web of Data that can be "understood" by machines has raised high expectations. The vision of an Educational Semantic Web outlines these expectations for the application of Semantic Web technologies and their benefit in the field of e-learning. Being able to access, browse, integrate and re-use learning material from distributed sources creates new ideas on how to shape and use Web-based learning systems. A lot of work has been done in this field in the last decade. This article gave a short overview on the possibilities and applications on how the Educational Semantic Web could look like.

However, there is still a lot of work to be done, not only in the e-learning sector, but in the Semantic Web community in general. For example, it has proven to be a different task to chose a vocabulary of concepts to agree upon. Ontology engineering has become a special discipline, providing tools and techniques to create and merge ontologies.

Deploying the Semantic Web means data integration. This raises other issues, e.g., trust and credibility in data that comes from a remote repository, or even (digital) rights management, which seems to be especially crucial when it comes to learning and teaching. If I let a system integrate data from a remote learning objects repository, how can I be sure, that the information the student learns is correct? How can I share, but at the same time ensure persistence of my copyright of my shared learning objects?

The Educational Semantic Web – the application of Linked Data to the domain of e-learning – is still in its infancy and there is space for a lot of new ideas, techniques and tools.

References


