Creating and Using an Education Standards Ontology to Improve Education

Sabbir M. Rashid and Deborah L. McGuinness

Rensselaer Polytechnic Institute, Troy, NY, 12180, USA

Abstract. We present an education standards ontology that has the potential for simplifying lesson planning for teachers, providing support for students by linking relevant resources, and providing a potential terminology for use in a lingua-franca for communicating with multiple communities about education components. We address the United Nations sustainable development goals concerned with education, with the support of semantic technologies. The goals of ensuring that all children graduate from primary and secondary schools and ensuring literacy and numeracy are our targets. We have created an education standards ontology along with a methodology for automatically generating this ontology. In this submission, we review literature related to educational applications that use semantic technologies and describe a web scraping method that we designed and used to automatically generate an education standards ontology. We also describe some potential uses of this ontology to assist in the UN education goals and reflect on next steps on uses of these existing educational resources to aide in improving literacy and numeracy in students and enabling future potential services and impact for an Educational Semantic Web.

Keywords: Educational Semantic Web · education · semantic web · ontology generation · education standards ontology · Common Core

1 Introduction

Shortly after the 2001 article presenting the vision of the Semantic Web [5], a discussion on the future of education on the *Educational Semantic Web* arose with the 2004 special issue of the Journal of Interactive Media in Education [3]. The authors presented visions of the effect of Semantic Web technologies on the coming decade of education, discussing possible advancements in information storage and retrieval, agents, and communication. Educational practitioners, in turn, provided reactions and rebuttals, voicing concerns with practicality and implementation.

Since the Semantic Web 3.0 has more features than its hypertext-based Web 2.0 predecessor, educators may not be versed in the added functionality, or they may not have enough motivation and time to perform manual annotations of resources that may be needed to effectively enable use of some of the advanced Web 3.0 resources. For humans, education is traditionally an interactive, personal experience, or as one educator describes, "an artistic, social interchange rather than one waiting for enhancement and possible substitution by a human-machine interaction [3]." Nevertheless, advances toward an Educational Semantic Web have the potential to reduce workload on instructors, create useful learning opportunities, improve content accessibility for society as a whole, and provide a well defined vocabulary for discussions about learning objectives and skills, thereby enhancing intercommunication across communities.

The Common Core standards for mathematics [14] and English Language Arts (which includes reading, writing, history, and the sciences) [13] are a set of guidelines designed to outline learning objectives for students at each given grade, from primary to secondary education (Kindergarten through 12th grade). They define the essential knowledge foundation and skills each student must acquire by the end of each school year in "core" subjects. While the benefits of standards in education have been the topic of ongoing debate [9], forty-one US states, the District of Columbia, and four territories have adopted the Common Core[1].

In order to assess one's knowledge, students often take examinations throughout the academic year to determine if educational standards are being met. Problems frequently arise when students prepare for these tests, as students often find themselves frustrated as to which methods to use to approach certain problems [7]. As education systems evolve, parents may have difficulties helping when they have been taught differently how to approach a problem, or are unfamiliar with the concepts being taught. A Common Core standards-based ontology can aid with the design of online tutoring systems and can reduce ambiguity of terminology by encoding explicit computer and human understandable definitions. Making such platforms publicly accessible can provide benefit to the "Social Good" by contributing to the UN Sustainable Development Goals related to quality education[2], helping to ensure literacy, numeracy, and that all students are able to graduate from primary and secondary schools. A standards-based ontology can support parents, educators, and students in communication using a common vocabulary. The precise definitions and encodings of the standard can also be used by those unfamiliar with specific common core methods to create their own methods of meeting the standard. We further reflect in Section 4.1, where we describe personalized learning environments tailored toward individual students.

Our contributions include a literature review on the Educational Semantic Web domain, an automatic extraction method to generate an education standards ontology from web content, and the resulting open source ontology. The Python-based extraction code is available on GitHub[3]. We also have published

[1] http://www.corestandards.org/standards-in-your-state/
[2] https://sustainabledevelopment.un.org/sdg4
[3] See https://github.com/tetherless-world/education-standards-ontology

the generated education standards ontology in order to aid with the linking of education standards in semantic education applications[4]. Finally, we describe a vision of a semantically-enabled education of the future.

2 Literature Review

Semantic Web for education may leverage work from a wide range of areas. We identify some foundations in the areas of learning objects, semantic grid frameworks, content modeling, mobile learning, data science-driven learning, and open scholarship.

2.1 Learning Object

Building on earlier notions of a Learning Object (LO), in the context of technology aided e-learning [12], refines the definition to be "any digital, reproducible and addressable resource used to perform learning or support activities, made available for others to use." While the XML based Education Modeling Language (EML), developed to describe learning objects, was not quite RDF, this definition could be used to describe many resources on the Semantic Web today. EML provides the basis for IMS Learning Design (LD) specification [4], a metadata based language that distinguishes between learning objects and activities in order to describe both content and processes from varying perspectives. The authors consider both the importance of individual, personalized learning as well as collaborative, competence-based, or problem based learning. EML includes *Learner* (Student) and *Staff* (Teacher) roles, nested activity structures, Learning Object environments and services. This allows for conditions to inform flow of activity, producing outcomes and notifications for the learner, resulting in an interactive learning environment.

2.2 Semantic Grid Frameworks

Other e-learning systems in the 2000s began to leverage both education grid systems as well as ontologies. The Semantic Grid-based E-Learning Framework (SELF) [1] uses the idea that Grid computing (distributed open service-oriented network architecture) can leverage semantics to create a Semantic Grid for e-learning. As well as Grid components, the SELF architecture supports the idea of open learning by considering collaborative tools, personalization and inference services, and software agents.

OntoEdu [11] addresses the need for new technologies to be used to make e-learning more mobile. The OntoEdu platform architecture for e-learning includes user adaptation service and content models, and an education ontology. The value provided by the ontology includes reasoning to produce activity descriptions for the user to plan a sequence of actions. For example, the inferred

[4] Available at https://archive.org/services/purl/purl/ontology/edu/

activity description can be used by a teacher to restrict a sequence of actions such that a student must select a lesson before selecting a homework set. Recent work in the e-learning direction includes an OWL ontology for metadata of Interactive Learning Objects (ILO), created using the ILO Standard to permit user interaction in eTutor applications [16]. In the context of semantic network architectures, the JuxtaLearn project uses content analysis along with domain ontologies to check the understanding of science concepts [8]. Science concepts are acquired through video creation and sharing, performing network text analysis, and constructing a multimodal network of categorized concepts. This work reflects on the role of videos as a media for learning, which may be aided by the categorization of pedagogical, domain, and general concepts, as well as tools and actors.

2.3 Ontologies for Education Content Modeling

In order to consider the impacts of the Semantic Web (Web 3.0) on education, it is fruitful to reflect on the impacts of the hypertext-based web (Web 2.0). Almost all levels of education today use at least Web 2.0 with support ranging from managing course websites to discovering research articles on the Internet, and much more. The Semantic Web implies a change in web architecture to go beyond static HTML. As people have learned to use the hypertext model of Web 2.0, they can similarly learn how to use Semantic Web RDF model as they see the value proposition, eventually resulting in a potential improvement on future education [6].

Educator Greg Kearsley argues however that busy teachers will not have time to become 'hyperkrep (hypertextual knowledge representation) hackers' and adoption of technologies will require simplicity from the end user perspective [3]. We agree both that given a clear value proposition, teachers and other relevant users will adopt more semantic web methodologies, and simultaneously agree with a version of Kearsley's position, that simplified tools and services are needed.

Even before the aforementioned journal discussion, ontologies were beginning to be used to model hypermedia content, such as with Ontoportal, a framework for building web-based educational portals [24]. Ontoportal aims to address the concern that educational materials on the web are usually weakly interlinked. Interlinking of web-based educational resources can be achieved by combining ontological and hypermedia research principles. The authors of Ontoportal describe two use cases where concepts can be designed to help educators gather resources or populate course information for students.

The teacher TPortal includes concepts aimed at experts and institutions, including concepts related to teaching, assessment, and support materials. The student XPortal additionally includes tutorials, courses, evaluation materials, and standards. Similarly, a recent educational domain ontology designed for a case study for the University of Palestine [18] includes concepts such as course, schedule, employee, student, and organization.

2.4 Mobile Learning

With the rise in popularity of mobile phones and the development of hand-held technology for education, the idea of Mobile- or M-Learning naturally emerges [21]. However, despite digital applications improving education, they are not readily available in all countries or languages.

A recent M-Learning System aims to address this concern for the Arabic language by designing a mobile story-telling application that links to related images by using Natural Language Processing (NLP) techniques for Arabic text extraction [15]. The system performs segmentation, tagging, and grammar linking, while also using Semantic Web constructs to aid in semantic query construction for question answering and semantic extraction, where ontology instances are generated. Combining Semantic Web and NLP techniques for M-learning can improve query navigation, the speed of information retrieval, and instance consistency.

2.5 Data Science-driven Education Initiatives

The Big Data to Knowledge (BD2K) initiative to facilitate both in-person and online learning has resulted in the Educational Resource Discovery Index (ERuDIte). ERuDIte has been applied in the context of general data science [2] as well as domain specific biomedical data science [22]. Also, knowledge resources are imported from many sources, including Massive Open Online Courses (MOOC), videos of tutorials or research conference talks, textbooks, blog posts, and web pages. In this effort many metadata resources are integrated, including the Dublin Core, Learning Resource Metadata Initiative (LRMI), IEEE's Learning Object Metadata (LOM), eXchanging Course Related Information (XCRI), Metadata for Learning Opportunities (MLO), and Schema.org vocabularies.

The Data Science Education Ontology (DSEO) uses data science methods to teach data science, and includes an ontology of learning domains. We plan to build a tutoring system that connects concepts from the DSEO with concepts from our ontology. By connecting mathematics standards to statistical terms, we can determine practice problems for the student to solve.

2.6 Open Scholarship

As the Semantic Web is decreasing ambiguity by supporting explicit term definitions for use by both humans and machines, and a major goal of Web 3.0 is related to data integration and interoperability, the implications on education are plentiful [19]. Possible impact areas include knowledge construction, personalized learning network maintenance, personal educational administration, and collaborative education tools. It is important for researchers and educators to collaborate in the creation of content, as there are inherent biases associated with the design of content.

An open scholarship environment, which includes sharing articles, code, data, and education resources, can improve research and education [17]. Sharing of resources can result in positive contributions to society, as it improves the dissemination of knowledge. Further interconnected education tools that "understand" each other's usage of terminology have the potential to build community learning environments.

3 Methodology

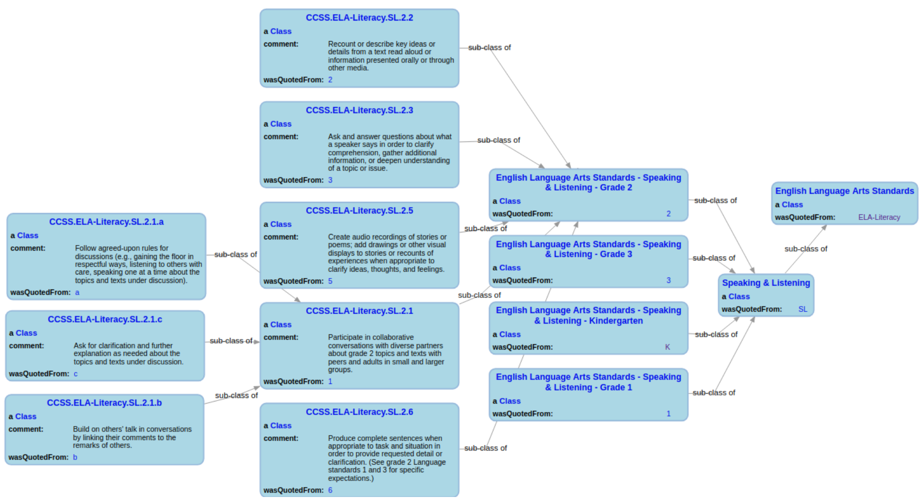

Fig. 1. Subsection of hierarchy generated from the English Language Arts Common Core speaking and listening standards

While there are numerous resources on the web related to education, many were designed under the hypertext paradigm of Web 2.0. For example, the Common Core Standards website[5] is written in Javascript and HTML and does not leverage the more expressive RDF representation, resulting in difficulties and/or limitations in machine understanding. Even though these websites are typically not encoded in a Semantic Web language (that would encourage more structure and links), we still observe some hierarchy and structure often including sets of organized links. By observing the HTML structure of a web page, it is possible to extract information in a structured way by leveraging web scraping libraries such as BeautifulSoup [20]. We take this approach to automatically construct an education standards ontology from the Common Core Standards in less than 100 lines of Python code.

[5] www.corestandards.org

The Common Core website is divided into two main portals, one for mathematics topics[6] and another for English Language Arts[7]. Each of these pages contains a well-structured HTML sidebar, which links to each set of standards for given grade levels and topics. By selecting specific HTML tags and list items from the sidebars, we are able to traverse each link and extract relevant information that is used to generate ontology classes with labels, comments, and the subsumption hierarchy. We also capture provenance by connecting the URL of each link to the corresponding concept using *prov:wasQuotedFrom*. Using this methodology, we are able to generate 1906 ontology classes. A subset of the concepts from the English Language Arts standards associated with speaking and listening is shown in Figure 1.

4 Discussion

An oft-quoted proverb states, "Give a man a fish, and you feed him for a day. Teach a man to fish, and you feed him for a lifetime." Analogously, we believe students will benefit from learning *how* to learn. We expect open access learning resources will aide in self-learning. The Education Standards Ontology can be used to help students align the knowledge they gain with standards in the Common Core. Self-guided learning will not only support learning required in standard education systems but will also provide a foundation that will support "continuous learning" beyond primary or secondary education. Further, in today's interdisciplinary world where many people often become self-taught experts or near-experts in new areas, we expect increased need for evaluating expertise and communicating that evaluation. Open ontologies and ontology-enabled environments that use well defined terms and that are aligned with published standards can provide a foundation for communication and comparison.

4.1 Personalized Learning

In today's teaching approaches, we often find that lessons are taught in a standardized way, despite a multitude of students with varying learning rates and interests. This results in difficulties for students who learn at a slower pace and disengagement (or boredom) for higher performing students. Furthermore, individuals may have different learning styles (auditory, kinesthetic, or visual), often based on the way their brain processes information. We foresee an increase in personalized learning environments, where the approaches to conveying knowledge are targeted towards individual students' needs, sometimes in response to special needs students.

In order for precision systems such as personalized learning environments to be successful, it is important to evaluate how well a student is receiving the

[6] http://www.corestandards.org/Math/
[7] http://www.corestandards.org/ELA-Literacy/

material being taught. Unfortunately, representative measurements of student progress are difficult to obtain, as much is left out in the examination process and complicating factors such as test anxiety sometimes skew results. A wide range of additional learning assessment technologies, including for example biometric measures of physiological indicators, to evaluate information reception may be included. Semantic annotations can provide support for connecting and integrating new assessment modalities.

4.2 Question Answering

In order for teachers to adopt new technologies, the value of the system has to outweigh the learning barriers or ease of use issues. One such example includes reliable automatic grading systems, as we are already beginning to see today. Such platforms have been adopted because they are not overly complicated to use, and are beneficial as they reduce the manual teacher workload. Question answering capabilities of intelligent systems are continuously improving, from search engines like Google to virtual assistants like Cortana [23] and Siri [10]. We are likely to see further improvements in machine dialog capabilities with advances in text to query syntax conversion, such as derivations of SPARQL queries from natural language sentences. Ontology-enabled environments can help improve query formation and result integration.

4.3 Learning Resources

Current day access to learning resources, including teaching materials, literature, and lesson plans, are often limited to students who are part of an academic institution. We predict expansion and improvement of remote access to learning resources. Physical class materials may be replaced by interactive materials, such as web-based learning resources.

Open access learning resources can benefit institutions, teachers, students, and parents by decreasing subscription fees, supporting customized lesson planning, and improving opportunities for learners globally. Also open-source resources help allow for collaborative learning and communication across communities.

We already see significant strides in terms of cloud-based education, where teachers (and sometimes parents) have direct access to student homework, grades, and projects. Students have direct access to course and library resources, and can remotely attend lectures or watch video recordings. We are likely to see intelligent classroom environments leverage sensor networks, as the Internet of Things (IoT) [25] holds promising opportunities. By mapping these devices to ontology concepts, semantic and social networks may be used to aid sensor networks.

4.4 Virtual Classrooms

Augmented reality techniques may be applied to learning environments, where virtual reality headsets can be used to add visual demonstrations to augment the physical classroom. Future augmented reality virtual classrooms may allow students and/or teachers to connect to a virtual reality environment where an educational lesson can take place. Of course, this does not have to be a classroom in the traditional sense, as virtual field-trips may provide methods for learners to witness world wonders, the outer reaches of space, or interactions at the molecular level. With virtual applications, learning can be fun rather than a chore; game-like simulations can provide enriching, interactive, and enjoyable environments. The effects of a virtual classroom are likely to have a huge impact in terms of social networking, in a sense the Facebook of the future. Ontologies will be used to semi-automatically create and connect virtual environments.

5 Future Work

We introduce an approach for automatically generating an education standards ontology from the Common Core website. It should be noted that the extraction code is specific to the Common Core website, as we observe the HTML structure of the relevant web pages in order to know from where to extract concepts and descriptions. Despite minimal generality in the Common Core-specific script, the described approach of observing the HTML web structure of a page to automatically generate an ontology can be applied to other web pages. Future work includes exploring methods of automatic web scraping techniques that are robust or adaptable to the structure of the web page. More importantly, we believe a useful, relatively simply-generated ontology is produced.

In order to evaluate the value of our ontology, we plan to develop a smart tutoring framework that takes advantage of earlier education ontologies and uses the education standards ontology to link course concepts and questions to Common Core standards. This will help provide useful lessons to students who are in school districts that have adopted the Common Core, thereby potentially greatly simplifying lesson planning for teachers and providing much more support for students in the form of additional relevant resources. Such work has potential to improve literacy, numeracy, and graduation rates for students.

6 Conclusion

We reviewed literature related to Semantic Web and education applications. We identified active research for at least the last 15 years, with researchers publishing methods related to the use of learning objects, semantic grid frameworks, content modeling, mobile learning, and open scholarship. We describe the automatic extraction method we used to generate an education standards ontology. We have made both the extraction script and generated ontology publicly available in order to encourage collaboration and reuse. We also discuss some ways in which the education ontology and an ontology-enabled infrastructure can advance realizing the vision of the Educational Semantic Web.

Acknowledgements

This work was partially supported by the RPI Tetherless World Constellation. We thank the members of the Constellation for their suggestions, in particular Rebecca Cowan.

References

1. Abbas, Z., Umer, M., Odeh, M., McClatchey, R., Ali, A., and Farooq, A. A semantic grid-based e-learning framework (self). In *Cluster Computing and the Grid, 2005. CCGrid 2005. IEEE International Symposium on* (2005), vol. 1, IEEE, pp. 11–18.
2. Ambite, J. L., Fierro, L., Geigl, F., Gordon, J., Burns, G. A., Lerman, K., and Van Horn, J. D. Bd2k erudite: The educational resource discovery index for data science. In *Proceedings of the 26th International Conference on World Wide Web Companion* (Republic and Canton of Geneva, Switzerland, 2017), WWW '17 Companion, International World Wide Web Conferences Steering Committee, pp. 1203–1211.
3. Anderson, T., and Whitelock, D. The educational semantic web: Visioning and practicing the future of education. *Journal of interactive Media in Education 2004*, 1 (2004).
4. Berlanga, A., and García, F. Ims ld reusable elements for adaptive learning designs. *Journal of Interactive Media in Education 2005*, 1 (2005).
5. Berners-Lee, T., and Hendler, J. Publishing on the semantic web. *Nature 410*, 6832 (2001), 1023.
6. Clark, K., Parsia, B., and Hendler, J. Will the semantic web change education? *Journal of Interactive Media in Education 2004*, 1 (2004).
7. Cooney, T. J. A beginning teacher's view of problem solving. *Journal for research in mathematics education* (1985), 324–336.
8. Daems, O., Erkens, M., Malzahn, N., and Hoppe, H. U. Using content analysis and domain ontologies to check learners' understanding of science concepts. *journal of computers in education 1*, 2-3 (2014), 113–131.
9. Delandshere, G., and Arens, S. A. Representations of teaching and standards-based reform: are we closing the debate about teacher education? *Teaching and Teacher Education 17*, 5 (2001), 547–566.
10. Gruber, T. R. Siri, a virtual personal assistant—bringing intelligence to the interface, 2009.
11. Guangzuo, C., Fei, C., Hu, C., and Shufang, L. Ontoedu: a case study of ontology-based education grid system for e-learning. In *GCCCE2004 International conference, Hong Kong* (2004), pp. 1–9.
12. Hummel, H., Manderveld, J., Tattersall, C., and Koper, R. Educational modelling language and learning design: new challenges for instructional re-usability and personalized learning.
13. Initiative, C. C. S., et al. Common core state standards for english language arts and literacy in history/social studies, science, and technical subjects. *Washington, DC: National Governors Association* (2012).
14. Initiative, C. C. S. S., et al. Common core state standards for mathematics. washington, dc: National governors association center for best practices and the council of chief state school officers, 2010.

15. Karkar, A., and Al Ja'am, J. An educational ontology-based m-learning system. *International Journal of Interactive Mobile Technologies (iJIM) 10*, 4 (2016), 48–56.
16. Luz, B. N., Santos, R., Alves, B., Areão, A. S., Yokoyama, M. H., and Guimarães, M. P. An owl ontology for metadata of interactive learning objects. *International Association for Development of the Information Society* (2015).
17. McKiernan, E. C. Imagining the "open" university: Sharing scholarship to improve research and education. *PLoS biology 15*, 10 (2017), e1002614.
18. Naser, S. S. A., Atallah, R. R., and Hamo, S. Building an ontology in educational domain case study for the university of palestine. *International Journal of Research in Engineering and Science (IJRES) 3*, 1 (2015), 15–21.
19. Ohler, J. The semantic web in education. *EDUCAUSE quarterly 31*, 4 (2008), 7–9.
20. Richardson, L. Beautiful soup. *Crummy: The Site* (2013).
21. Sharples, M., Taylor, J., and Vavoula, G. A theory of learning for the mobile age. In *Medienbildung in neuen Kulturräumen*. Springer, 2010, pp. 87–99.
22. Van Horn, J. D., Fierro, L., Kamdar, J., Gordon, J., Stewart, C., Bhattrai, A., Abe, S., Lei, X., O'Driscoll, C., Sinha, A., et al. Democratizing data science through data science training.
23. Warren, T. The story of cortana, microsoft's siri killer. *Published on: Apr 2* (2014).
24. Woukeu, A., Wills, G., Conole, G., Carr, L., Kampa, S., and Hall, W. Ontological hypermedia in education: A framework for building web-based educational portals.
25. Xia, F., Yang, L. T., Wang, L., and Vinel, A. Internet of things. *International Journal of Communication Systems 25*, 9 (2012), 1101.