Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 2, February : 2024

ASEMANTIC BASED INFORMATION RETRIEVAL IN E-LEARNING DOCUMENTS WITH QIRSYSTEM

Avinash Seekoli* *Assistant Professor, Department of Computer Science and Engineering *St. Martin's Engineering College, Secunderabad, Telangana, India. *Corresponding Author E-mail: savinashcse@smec.ac.in

Abstract.
The advent of the internet, the evolution of the World Wide Web (WWW), coupled with the e-learning paradigm has resulted in the availability of a plethora of learning resources on the Web. However, these resources are not being fully utilized to their greatest potential. Learners, educators and researchers seeking educational content usually spend a great deal of time sorting through resources on the web without satisfactory results. Most times, this is not because the information is not available, but because the techniques being applied by major search engines do not handle the semantics and personalization required in this context. In a bid to proffer a solution to the problem of discovering relevant resources online by different categories of users, this work presents an integrated framework for personalized information retrieval of educational content. The framework exploits semantic web technologies. Further work will include the implementation and testing of the framework.
Keywords: Personalized information retrieval · E-learning · Semantic web Ontology

1 Introduction

The web has transitioned gradually from what is called Web 1.0 to Web 2.0 and now Web 3.0. The World Wide Web (WWW) at inception also known as Web 1.0 was a static web where only few publishers could upload content and majority of users were mere readers. Then came Web 2.0, which brought a new dimension to this, rather than 'read-only' which characterized Web 1.0, it became 'read-write'. Majority of users could upload their own content on the web, and this led to so much content being generated on the WWW. Web 2.0 also witnessed the advent of social networks and other forms of collaborative activities on the web. However, the large volume of useful information available on the WWW is not being utilized to its fullest potential, the information is so much that users do not go past the first page of search results returned by search engines. The most efficient utilization of this massive volume of data would be largely dependent on the ability of application programs to autonomously extract and make meaning out of the data despite the dynamic nature of information systems today. This is the problem that the semantic web also known as Web 3.0 has been brought on stage to solve [1, 2].

The Semantic Web envisions an era when the content available on the WWW will be readable, understandable and used appropriately by machines autonomously. As such, agents would be able to carry out knowledgeable tasks for users unaided, for example, schedule an appointment with a physician based on a doctor's diagnosis, one's location, personal schedule and cost preferences [3]. The semantic web however will only enable machines to interpret the meaning of semantic documents and data, and not human speech and writings.
With the arrival of e-Learning which is defined as "is the use of electronic media for a variety of learning purposes that range from add-on functions in conventional class- rooms to full substitution for the face-to-face meetings by online encounters" [4] there are a plethora of resources available for learners on the web. However, learners, educa- tors and researchers seeking educational content usually spend a great deal of time sorting through resources on the different e-learning platforms on the web without satis- factory results most times, this is not always because the information is not available, but because the techniques being applied by major search engines do not properly handle semantics and personalization. Different users seek information on the web for different purposes; however, at the present they all get the same result for their search. It is this problem that we seek to

proffer a suitable solution to via the framework discussed. Therefore, the remainder of this work is organized as follows: Sect. 2 gives the back- ground study and literature review, Sect. 3 discusses the proposed framework, it compo- nents and provides a dry run simulation of how it would work, while Sect. 4 presents further work and conclusion.

2 Literature Review

This section gives background information on the domain of the application of semantic technologies and e-learning, as well as related works.

2.1 Semantic Web Technologies

In this section, several technologies that enable and encompass the semantic web are discussed. They include: The Semantic Web Stack, Semantic Annotation, Vocabularies (Ontologies), Linked Open Data and Knowledge Graphs, and Semantic Information Retrieval.
The Semantic Web Stack. Several technologies underlie the semantic web, and most of them are properly captured in the layered architecture developed by Tim Berners Lee called the semantic web stack shown in Fig. 1 [1, 5].
They include: Unicode and Uniform Resource Identifier (URI) - The Unicode is a character encoding standard that provides a unique number for every character regard- less of the platform, program or language. This allows for consistent representation and interpretation of text. The URI on the other hand is a standardized identifier that allows for the unique identification of resources. This layer thus provides a universal way to

Fig. 1. The semantic web stack [5]

represent and identify resources on the web. eXtensible Mark-up Language (XML) - This layer defines syntax. XML which is a general purpose mark-up language together with the namespace and schema allows for a common syntax for representing structured documents and data objects on the web. Resource Description Framework (RDF) - RDF is a data description and representation language. It is used to represent metadata, which is a major requirement for the semantic web vision. It is a data model used to describe resources on the web [6]. The Resource Description Framework Schema (RDFS) - It is used to write taxonomies (hierarchy) of web objects i.e. their classes and properties, as well as lightweight ontologies. It is quite primitive for writing ontologies therefore, more powerful languages for writing ontologies, which allow for representa- tions of complex properties and relationships of web objects are presented in the subse- quent layer. The next layer comprises of the Web Ontology Language (OWL) for writing more robust ontologies, as well as Rule Interchange Format (RIF) and Semantic Web Rule Language (SWRL), which are rule languages for writing web rules that can be executed. Rule sets as well as ontologies are used to make new discov- eries on the semantic web. They differ from ontologies in that they focus on general mechanisms for the discovery and generation of new relations from existing ones while ontologies focus on classification methods. They are collections of IF-THEN statements otherwise known as rules. Another rule language is RuleML - Rule Markup Language. The RIF allows for rule exchange i.e. sharing, exchange and reusability of rules between rule systems, applications and rule engines with the semantics still preserved. The Simple Protocol and RDF Query Language (SPARQL) an SQL-like language is used to query the RDF, RDFS and OWL ontologies. It is also a protocol for accessing data in RDF format. The Proof layer involves deductive processes, representations and vali- dation of proofs while the Trust layer applies cryptographic techniques, such as digital signatures to verify sources of input. The Trust layer is very important because it is when the operations of the semantic web are trusted for quality that its full potential would be realized. Building on all these layers, user applications are then developed.

2.2 Related Works

According to [7], from the year 2000 there has been a continuous increase in researchers' interest towards the development and usage of ontologies generally as shown in the increased volume of publications having the word 'ontology' in its title or list of keywords. More specifically as seen in Fig. 2, there has also been accelerated interest for the use of ontologies in e-learning systems as indicated in searches conducted on four (4) digital libraries: IEEE, ACM, ScienceDirect, and Google Scholar.

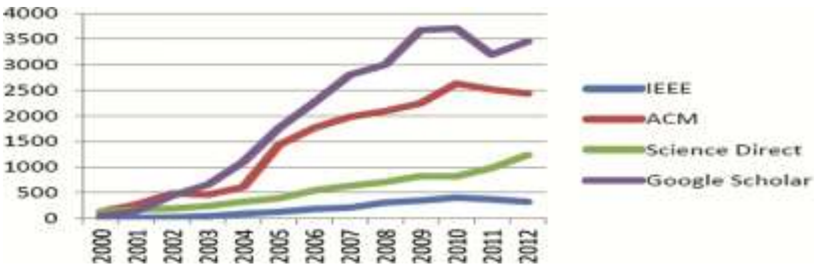

Fig. 2. Population growth for the words 'ontology' and 'e-learning' [7]

Dicheva [8] stated that learning systems using semantic technologies such as ontol- ogies are the third generation of e-learning systems. Ontologies add meaning to the information on the web, they represent knowledge in a way that insights can be inferred and information processed automatically.

Ontologies have been applied to different e-learning tasks. They have been used to augment learning management systems by providing an easier way to manage, distribute and retrieve learning materials. They create a more dynamic learning environment while enhancing personalized learning via the creation of individual learning paths. They have also been proposed to be capable of expediting social learning [9].

According to [10], they provide a common understanding of the structure of infor- mation represented both for humans and agents; when built in a robust manner, they are highly reusable while providing thorough analysis of the terms and specifications contained, they also help to distinguish between operational and domain knowledge. Their use also enhances the semantic organization of content online, and can aid in the development of personalized learning [11].

Dicheva et al. [12] in their work, classified the applications of semantic web tech- nologies in e-Learning into five (5) parts: ontologies as enabling technologies, ontologies for authoring instructional systems, instructional support and adaptation, sematic web- based intelligent learning environments and social semantic web applications.

Ontologies as Enabling Technologies. The study in [13] proposed a system to create learning resources by using an agreed vocabulary to collate and annotate learning content. The authors also worked on an algorithm to help learners determine starting points for repository exploration for efficient and faster search.

Ontologies for Authoring Instructional Systems. The study in [14] used an ontology called OMNIBUS to support a theory-aware authoring system called SMARTIES a kind of an expert system that simulates the activities on a human expert during instructional design. The study in [15] focused on

the application of ontologies to build a type of Intelligent Tutoring System called constraint-based tutors (where knowledge is repre- sented by constraints).
Instructional Support and Adaptation. Jovanovic et al. [16] applied both ontologies and semantic annotations to build a system that gives contextualized feedback on students' online activities to teachers.
Sematic Web-Based Intelligent Learning Environments and Social Semantic Web Applications. Melis et al. [17] developed the ACTIVE MATH intelligent e-learning platform with different semantic web features.
E-Learning and the Social Semantic Web. Social semantic web is the integration of the semantic web with social web approaches. It combines technologies from the social web popularly known as Web 2.0 and the Semantic Web [18].
On the other hand Al-Yahya opined that ontologies are applicable to four (4) cate- gories of e-learning tasks namely: curriculum modelling and management, description of learning domains, description of learner data and description of e-learning serv- ices [7].
Ontologies for Curriculum Management. Modelling of curriculum elements simpli- fies access to and retrieval of needed information. It also makes it possible to link learning units to their respective outcomes and objectives or to other learning units [7].
CURONTO is an example of an ontology that was developed to model the needed semantics of a curriculum in order to facilitate its review and assessment [19].
Ontologies for Describing the Learning Domain. Ontologies representing the learning domain can either be subject-domain ontologies or task ontologies. Subject-domain ontologies focus on the subject matter and knowledge elements of a domain e.g. a course of study while task ontologies represent structural elements of a learning task such as an activity, assessment, feedback, simulation or search and retrieval [7].
Oele, an ontology based e-assessment system capable of both summative and forma- tive assessment functions was developed [20]. Sameh researched on the application of ontology for effective semantic feedback and support in a e-learning system [21], while
[22] built an ontology to represent pedagogical patterns.
Ontology for Describing Learner Data. Ontologies are also used to describe charac- teristics of the learner such as learner profile, progress and performance.
The study in [23] developed an ontology to represent the student model for an Intel- ligent Tutoring System, the ontology modelled the student's profile comprising of the student's academic and personal information.

3 Methodology

This section discusses the proposed framework and how provides a solution to the identified problem.

3.1 Components of the Proposed Framework

The diagram shown in Fig. 3 represents the framework of the proposed system. Its components include a learner's profile ontology, learning objects ontology, a semantic web search engine (which will comprise of both a generic web crawler and a RDF crawler), a reasoning engine and inferencing system, a ranking mechanism and a suitable user interface.

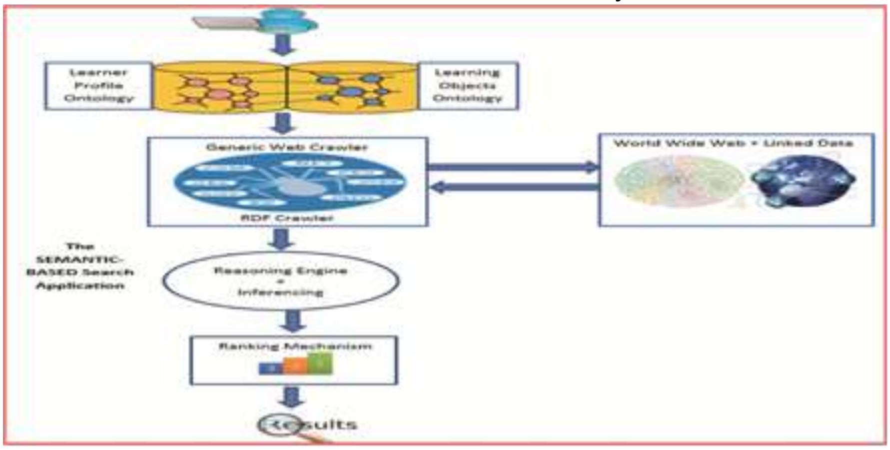

Fig. 3. The Proposed Framework

The Learner's Profile Ontology. This would be a vocabulary to define terms to describe the learner's profile. This would be used to represent the learner's preferences and be able to find a match with available resources.
The Learning Objects/Resources Ontology. This would be a vocabulary that describes various learning resources available on the web. A suitable one to be used is the Learning Resource Metadata Initiative (LRMI). LRMI created a common metadata framework to describe, tag or annotate learning resources published online. The metadata schema that was developed has also being adopted by Schema.org (a schema developed and being used by major search engines), therefore, tagging educational content with the LRMI mark-up ensures their proper recognition by these search engines. It was setup with the aim of making it easier and faster to discover and publish educational content and prod- ucts online, to facilitate personalized learning, decrease costs through standardization, address demands for standardized description of learning resources and provide param- eters that can be used for searching and filtering learning resources online. This makes it very suitable to be adapted to the framework.
A Semantic Web Search Engine. This would be responsible for searching the web of data for available resources. The results being returned to the system are then filtered based on some semantics and the user's preferences. Such search engines already exist such as Swoogle and Watson, they can be adapted and their results further filtered to fit the intended use.
The Reasoning/Inference Engine. This would be responsible for carrying out the matching of the returned resources and the learner's profile based on specified instruc- tions. It would be programmed using the Java language.
A Ranking Algorithm. This component of the system would also be implemented using a programming language; it would help to prioritize the filtered results based on pre- specified preferences and their weights or level of importance.
A User Interface. The user interface would allow for the user or learner to properly and conveniently interact with the system in submitting his queries and giving feedback on the suitability of results returned.

3.2 Operation of the Proposed Framework

The user accesses the system via the user interface on his computer, his preferences and search terms are represented using the standardized terms as represented in the vocabu- lary. These representations are then used to search, and matches meeting the pre-speci- fied criteria are filtered. They are then ranked and the best-fit results are presented back to the user, together with extra information to guide his decision. Figures 4, 5, 6, 7 and 8 explain the operation in pictorial representations.

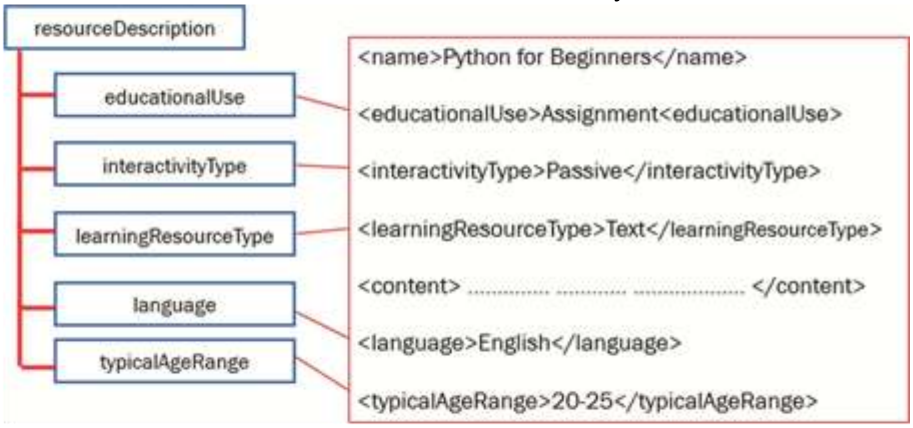

Fig. 4. Some LRMI terms representing the learning object

Figure 4 shows possible terms that will be in the vocabulary to represent the learning objects while Fig. 5 shows that of the learner's profile.
Figure 5 further shows that the learner is searching for Python tutorial materials; he prefers video resources, learns in an active mode, is 12 years old, has no previous knowledge of Python and learns in English.

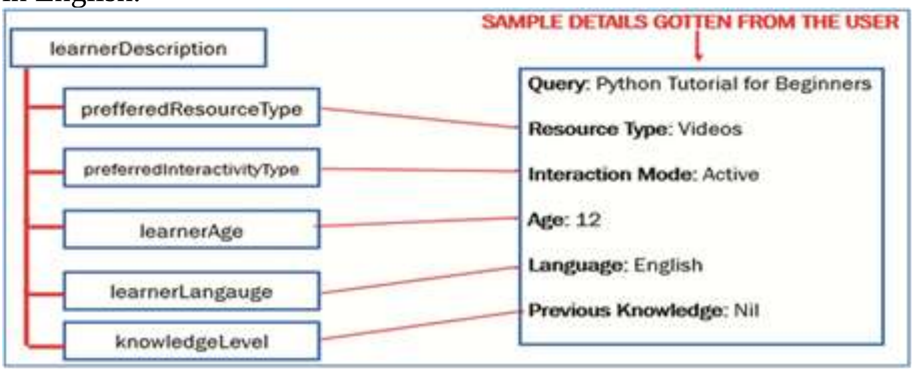

Fig. 5. A representation of the learner's profile

Figure 6 shows the semantic representation of sample web resources, which have been annotated using terms from the vocabulary. The first resource identified by number 1 is an assignment resource (as seen from the 'educationalUse' tag), it is for learners between 20-25 years (as seen from the 'typicalAgeRange' tag), it is a text material (as seen from the 'learningResourceType' tag) and preferable for passive learners (as seen from the 'interactivityType' tag). The second resource identified by number 2 is a tutorial material for children between 10 and 15, and it is a video. The third resource is also a tutorial for 22-25 years old learners, and it is a video. The fourth resource is a lesson plan for an educator.


Fig. 6. Sample web resources before filtering

Figure 7 shows the filtering; resources 1 and 4 were dropped because they do not closely match the need of the learner. Resources 2 and 3 were picked since they are video materials, and they are in line with what the learner wants. In Fig. 8, resource 3 ranks lower (75%) because it does not match the learner's age, which is 12 (see Fig. 5) while resource 2 is the best match with 95%.

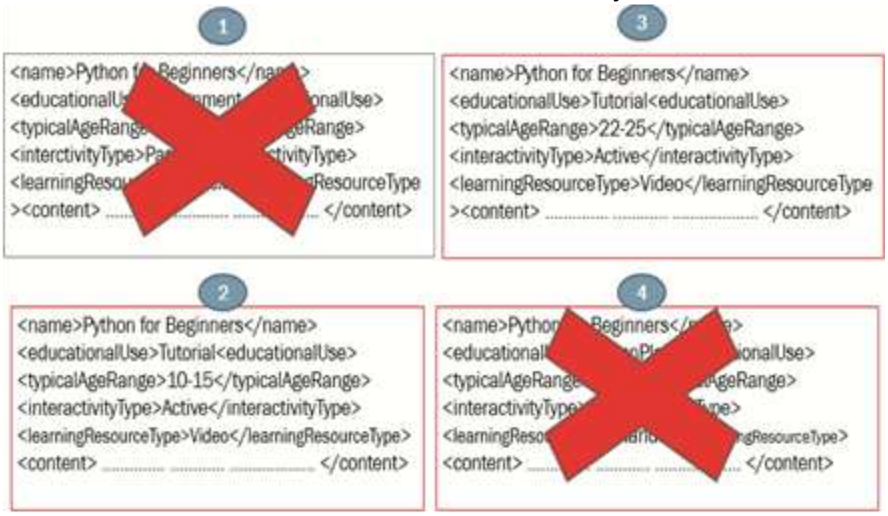

Fig. 7. Filtered web resources

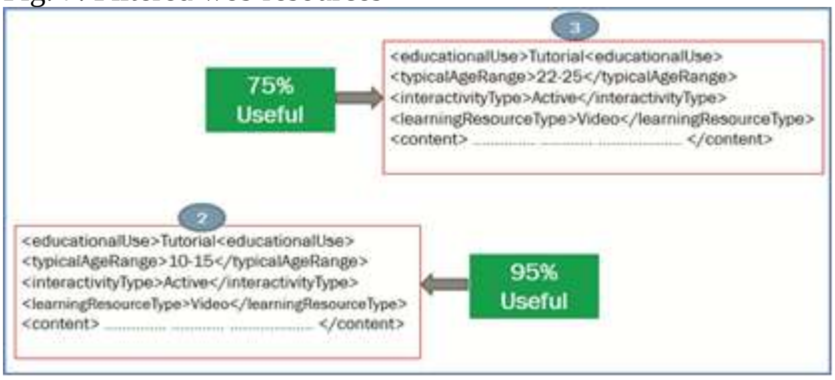

Fig. 8. Ranked result

This scenario shows how the proposed framework can enhance semantic and person- alized retrieval, thus improving learner's satisfaction on e-learning systems.

4 Future Works and Conclusion

This proposed integrated framework would be implemented using the various technol- ogies as described in Sect. 3. The framework would then be validated and tested by active users.

This framework when implemented would enhance the personalized information retrieval of educational content and resources. It exploits semantic web technologies including ontologies, standardized metadata annotation schemas, and semantic web search engines to provide users with resources that match their preferences. It also incorporates a ranking algorithm to help prioritize returned relevant results, a social web component to get user ratings of retrieved resources as well as a feedback mechanism to fine tune subsequent results. Instead of having a situation where a learner needs to choose from 571,000 documents returned in 0.35 s or 778,000 videos in 0.37 s from a Google search, using the user's preferences, instead of just retrieval, filtering and personalization will take place as well.

The proposed system is going to help to provide a better approach to use open web content in e-Learning systems. It will also help in the further synergy of the social and semantic web, which would really help to improve learner engagement and collaboration during learning.

Acknowledgements. We acknowledge the support and sponsorship provided by Covenant University through the Centre for Research, Innovation and Discovery (CUCRID).

References

1. Antoniou, G., Groth, P., Van Harmelen, F., Hoekstra, R.: A Semantic Web Primer. The MIT Press, Cambridge (2012)
2. Rubens, N., Kaplan, D., Okamoto, T.: E-learning 3.0: anyone, anywhere, anytime, and AI. In: Chiu, D.K.W., Wang, M., Popescu, E., Li, Q., Lau, R. (eds.) ICWL 2012. LNCS, vol. 7697, pp. 171-180. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-43454-3_18
3. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Sci. Am. 284, 28-37 (2001)
4. Guri-Rosenblit, S.: "Distance education" and "e-learning": not the same thing. High. Educ. 49, 467-493 (2005)
5. Obitko, M.: Semantic web architecture. http://www.obitko.com/tutorials/ontologies-semantic-web/semantic-web-architecture.html#semantic-web-layers
6. Sikos, L.: Mastering Structured Data on the Semantic Web: From HTML5 Microdata to Linked Open Data. Apress, New York (2015)
7. Al-Yahya, M., George, R., Alfaries, A.: Ontologies in E-learning: review of the literature. Int. J. Softw. Eng. Appl. 9, 67-84 (2015)
8. Dicheva, D.: Ontologies and semantic web for E-learning. In: Adelsberger, H.H., Kinshuk, Pawlowski, J.M., Sampson, D.G. (eds.) Handbook on Information Technologies for Education and Training, pp. 47-65. Springer, Heidelberg (2008). https://doi.org/ 10.1007/978-3-540-74155-8_3
9. Monachesi, P., Simov, K., Mossel, E., Osenova, P., Lemnitzer, L.: What ontologies can do for eLearning. In: Proceedings of IMCL International Conference on Mobile and Computer Aided Learning, Amman, Jordan (2008)
10. Askar, P., Kalinyazgann, K., Altun, A., Pekince, S.S.: An ontology driven model for E-learning in K-12 education. In: Kidd, T.T., Song, H. (eds.) Handbook of Research on Instructional Systems and Technology, pp. 107-116. IGI Global (2008)
11. Kizilkaya, G., Torun, E.D., Askar, P.: Restructuring e-learning with ontologies. In: Fifth International Conference on Computational Science and Applications, pp. 161-164. IEEE (2007)
12. Dicheva, D., Mizoguchi, R., Greer, J.: Semantic Web Technologies for e-Learning: The Future of Learning, vol. 4. IOS Press, Amsterdam (2009)
13. Dicheva, D., Dichev, C.: Authoring and exploring learning content: share content by sharing concepts. In: Dicheva, D., Mizoguchi, R., Greer, J. (eds.) Semantic Web Technologies for e- Learning: The Future of Learning, vol. 4, pp. 24-43. IOS Press, Amsterdam, The Netherlands (2009)
14. Mizoguchi, R., Hayashi, Y., Bourdeau, J.: Inside a theory-aware authoring system. In: Dicheva, D., Mizoguchi, R., Greer, J. (eds.) Semantic Web Technologies for e-Learning: The Future of Learning, vol. 4, pp. 59-76. IOS Press, Amsterdam (2009)
15. Suraweera, P., Mitrovic, A., Martin, B., Holland, J., Milik, N., Zakharov, K.: Using ontologies to author constraint- based intelligent tutoring systems. In: Dicheva, D., Mizoguchi, R., Greer, J. (eds.) Semantic Web Technologies for e-Learning: The Future of Learning, vol. 4, pp. 77-95. IOS Press, Amsterdam (2009)
16. Jovanovic, J., Gasevic, D., Torniai, C., Devedzic, V.: Using semantic web technologies to provide contextualized feedback to instructors. In: Dicheva, D., Mizoguchi, R., Greer, J. (eds.) Semantic Web Technologies for e-Learning: The Future of Learning, vol. 4, pp. 117-135. IOS Press, Amsterdam (2009)
17. Melis, E., Goguadze, G., Libbrecht, P., Ullrich, C.: ActiveMath - a learning platform with semantic web features. In: Dicheva, D., Mizoguchi, R., Greer, J. (eds.) Semantic Web Technologies for e-Learning: The Future of Learning, vol. 4, pp. 159-177. IOS Press, Amsterdam (2009)
18. Jovanovic, J., Gasevic, D., Devedzic, V.: E-learning and the social semantic web. In: Dicheva, D., Mizoguchi, R., Greer, J. (eds.) Semantic Web Technologies for e-Learning: The Future of Learning, vol. 4, pp. 245-259. IOS Press, Amsterdam (2009)

19. Al-Yahya, M., Al-Faries, A., George, R.: CURONTO: an ontological model for curriculum representation. In: Proceedings of the 18th ACM Conference on Innovation and Technology in Computer Science Education - ITiCSE 2013, p. 358. ACM Press, New York (2013)
20. Litherland, K., Carmichael, P., Martínez-García, A.: Ontology-based e-assessment for accounting: outcomes of a pilot study and future prospects. J. Account. Educ. 31, 162-176 (2013)
21. Sameh, A.: Ontology-based feedback E-learning system for mobile computing. In: Mastorakis, N., Mladenov, V., Kontargyri, V. (eds.) Proceedings of the European Computing Conference. LNEE, vol 27. Springer, Boston (2009). https://doi.org/10.1007/978-0-387- 84814-3_48
22. Cobos, C., Rodriguez, O., Rivera, J., Betancourt, J., Mendoza, M., León, E., Herrera-Viedma, E.: A hybrid system of pedagogical pattern recommendations based on singular value decomposition and variable data attributes. Inf. Process. Manag. 49, 607-625 (2013)
23. Panagiotopoulos, I., Kalou, A., Pierrakeas, C., Kameas, A.: An ontology-based model for student representation in intelligent tutoring systems for distance learning. In: Iliadis, L., Maglogiannis, I., Papadopoulos, H. (eds.) AIAI 2012. IAICT, vol. 381, pp. 296-305. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33409-2_31