Implementation of Semantic Web Mining on E-Learning

Oğuz Mustapaşa*, Dilek Karahoca, Adem Karahoca, Ahmet Yücel, Huseyin Uzunboylu

*Department of Software Engineering, University of Bahçeşehir, Istanbul-Turkey

Received November 15, 2009; revised December 3, 2009; accepted January 25, 2010

Abstract

Semantic Web is a product of Web 2.0 (second generation of web) that is supported with automated semantic agents for processing user data to help the user on ease of use and personalization of services. Web Mining is an application of data mining which focuses on discovering patterns from Web logs and data. The semantic structure can be built with the pattern or relation results discovered via web mining. By combining those two applications of both disciplines, it’s possible to achieve Semantic Web Mining which is a recent hot topic in educational research. This paper gives an overview of current applications of Semantic Web Mining on e-learning which already became a base component of education.

Keywords : Semantic web; web mining; e-learning; distance learning; personalization.

1. Introduction

As the World Wide Web spreads around the world and number of internet users are increasing every day, web pages and web services are getting more popular. In recent ten years online web portals for e-learning, e-government and e-commerce became a very popular part of Web. For a better quality of service and personalization of services, web portals were in need of a semantic structure and an intelligent logic. This goal is achieved by combining two disciplines: Semantic Web and Web Mining. Nowadays Web Mining and Semantic Web are popular topics in Web. These two disciplines can be used together since they both are addressing each other to fulfill the aim of understanding or examining large amount of data in an automated way and also discovering and obtaining meaningful results. By applying Semantic Web Mining on educational purposes, especially on distance learning and course management systems where both can be used as a support to traditional education and distance learning intentions. The aim of discovering students’ learning model and personalization of services over current e-learning portals and course management systems are achievable via semantic tools such as Web Services or Semantic Web Agents. In early sections of the paper, a brief explanation of Web Mining, Semantic Web, Semantic Web Mining and their applications on educational systems are given. In this study, previous works are addressed and advantages
and disadvantages of e-learning implementations are also considered. Nonetheless, a discussion on Semantic Web Mining on E-learning is given in the section of discussions and conclusions.

2. Background

The World Wide Web (WWW or W3) was developed to be a pool of human knowledge, which would allow collaborators in remote sites to share their ideas and all aspects of data of a common project (Wardrip-Fruin & Monthfort, 2003). The idea of an automatic notification is when new material of interest is available came up, Web had its chance to meet RSS (Really Simple Syndication)-Feeds and Web2.0. This automation achieved via web services having a semantic structure behind. In the time, personalization and visualization of data depending on user interest became very popular due to fashion of web technology. This popularity also brought functionality and usability with itself so that W3 became widespread on e-commerce, e-government, and e-learning areas. Considering e-commerce and e-government and such applications are not hard to implement, application on e-learning was very interesting since that would be something new for education field to teach without any face-to-face interaction in real life.

If we think of a traditional education scenario, an educator should track each student’s situation for any lecture and also should try to improve the learning method for the student just to let him/her more successful. When we take this into account for the need of personalization, there had to be some technology which would allow e-learning users to have personalized material in their interest or profile. In the time being, many works and implementations took place for a semantic, intelligent web agent or service for e-learning. As the research topic continued to get into area of interest for many researchers, worldwide organizations standardized the techniques that can be used to build a semantic learning mechanism for educational fields.

In connection with personalization on e-learning, it has been proposed to organize content into independent units, named as Learning Objects (LO), which can be combined dynamically for building personalized learning portals (Stumme, Hotho, & Berendt, 2005; Junuz, 2009). “LO is defined as any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning” (WG12 : Learning Object Metadata). Also a standard called Learning Object Metadata (LOM) is found by IEEE that is defined as “a metadata instance for a learning object describes relevant characteristics of the learning object.” (IEEE LTSC 1484.12.1, Draft Standard for Learning Object Metadata, 2002). In IEEE Learning Technology Standards, IEEE 1484.12.3, LOM is pointed as: “Where applicable, Learning Object Metadata may also include pedagogical attributes such as; teaching or interaction style, grade level, mastery level, and prerequisites. It is possible for any given Learning Object to have more than one set of Learning Object Metadata” (WG12 : Learning Object Metadata ; IEEE LTSC 1484.12.3 Draft Standard for eXtensible Markup Language(XML) Schema Definition Language Binding for Learning Object Metadata). An international consortium, named IMS Global Learning Consortium, had been conducive to improvement of early versions of IEEE LOM, supported early drafts of the data model as a part of IMS Learning Resource Metadata specification (IMS Learning Resource Metadata Specification, 2004). In the third version of IMS LRM, IMS data model and IEEE LOM were adapted and it proposed to use IEEE XML as base structure. Also IMS LRM provided an implementation guide and an XSL that might enables migration from old versions of IMS LRM XML into IEEE LOM XML. As a result, we can describe LOs by the means of XML (Stumme, Hotho, & Berendt, 2005; World Wide Web Consorcium-W3C). For the standardization of LO content models, many researchers proposed to use ontology.

Ontology builds a backbone for artificial intelligence and semantic web purposes and they are defined by Resource Description Framework (RDF) and XML (World Wide Web Consorcium (W3C)). Basically, ontology represents the relations between concepts which underlies in a domain. With the use of ontology, combining information retrieval techniques and web agents, semantic structure in data can be found by applications of data mining techniques. This goal is achieved by Web Mining (WM), Semantic Web (SW) and a mixture of both as Semantic Web Mining (SWM) so that we can lead the way trough personalization of content and personalization of service. The following sections of this paper briefly summarize concepts of Web Mining (WM), Semantic Web (SW), Semantic Web Mining (SWM) and E-Learning.
3. Methodology

In this section, we briefly describe the concepts Web Mining, Semantic Web and Semantic Web Mining. In section 3.1, a description of Web Mining is given. In 3.2, Semantic Web and Semantic Web Mining are explained. Nonetheless, in section 3.3, previous applications of Semantic Web Mining on E-Learning systems are explained with their advantages and disadvantages.

3.1. Web Mining

Web Mining (WM) is the applied data mining on web logs, web contents and web structures. Thus it is “the nontrivial process of identifying valid, previously unknown, and potentially useful patterns (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). As given in the definition, WM has three different types of analysis specs; Web Usage Mining (WUM), Web Content Mining (WCM) and Web Structure Mining (WSM). The specific analysis types of WM for e-learning are WUM and WCM.

Web Usage Mining tries to find out what users are looking for while they are using Web, and WUM also helps to find the patterns for a particular group of people belonging to a region or depending on their interest. Web Content Mining is a kind of text mining application on Web content. This method could be used for creating metadata for Learning Objects (LO) for building ontology and semantic structure.

3.2. Semantic Web

Semantic Web (SW) derives from W3C director Tim Berners-Lee’s vision of Web as a universal medium for data, information and knowledge exchange (Junuz, 2009). The word semantic web is a product of Web2.0 (second generation web) which makes the web itself to understand and satisfy the user requests and web agents or machines to use the content of web (Berners-Lee, Hendler, & Lassila, 2001; W3C Semantic Web, 2008).

3.3. Semantic Web Mining and E-Learning

The term Semantic Web Mining is described well by Stumme and etc. all. as “Semantic Web Mining aims at combining the two areas Semantic Web and Web Mining. This vision follows our observation that trends converge in both areas: Increasing number of researchers work on improving the results of Web Mining by exploiting semantic structures in the Web, and make use of Web Mining techniques can be used for mining Semantic Web itself. The wording Semantic Web Mining emphasizes this spectrum of possible interaction between both research areas: It can be read both as Semantic (Web Mining) and as (Semantic Web) Mining.” (Stumme, Hotho, & Berendt, 2005).

Pointing at the definitions given, it’s possible to use web logs for any course available on any course management system or e-learning portal for investigation of semantic information. In a case study on Moodle, case studies for applications of data mining techniques are given by (Romero, Ventura, & Garcia); (Moodle). In these studies, the possible techniques for data retrieval and management, educator has to run third party programs manually for information retrieval, for educator are explained briefly. For a semantic and real time system, web services and web agents were announced to be useful. Also, trustworthiness of the data is very important since it can lead the algorithms or mining techniques in wrong or inadequate results. At this point, we can assume that the data we get from student’s answers or information left on CMS is reliable or we can also run data mining algorithms to fetch conflicts on answers to filter them somehow.

As we described in earlier sections of this paper, usage of LO, LOM and ontology are the requirements for mining user data on a portal. We also described the ways of implementing each technique and it can be seen that XML is widely used for implementing LO, LOM and even ontology. It’s mentioned to use RDF or XML for ontology design but in many applications of RDF, it’s written in XML syntax. While speaking of ontology, ontological needs have to be defined well. We saw that in the reference papers, domain ontology and student ontology are commonly in use. When e-learning via mobile devices included, we can use another ontology preferably named as device ontology and when the pedagogical information is more important another ontology named as pedagogical ontology might be applied.
This paper presents an overview of current techniques and applications of Semantic Web, Web Mining on E-Learning which already became a base component of education. We believe that student centralized education, in other words personalization depending students’ learning style, is the key on educational field. In situations where distance learning or e-learning is applied, personalization appears as a problem since it’s not a desirable way to teach and also not a desirable way of learning in a restricted way of education in any topic. Also a dynamic and semantic program has to be setup for minimizing the gap between traditional learning and e-learning. De Meo and etc. all. announced advantages and disadvantages at e-learning phenomenon and pointed that personalized and flexible learning programs can be constructed dynamically to fulfill these gaps (De Meo and etc. all., 2005).

They also mentioned, e-learning portals are risk-free pointing at students can try new things, make mistakes without disclosing themselves, e-learning portals provides a better personalized learning mechanism for all, information is distributed at the same level of quality, students can use the service for getting knowledge out of available information on their own schedule at any time they want and it’s much more easier to update content for any topic.

In connection with education field, pedagogical findings have to be include and examined well by data mining techniques and applications for a more realistic educational simulation that is desired to be closer to traditional learning mechanism such as educator’s tracking mental conditions of learners where the learner’s disorders are decreasing the capability of learning and teaching. Also trying to fix issues via with a powerful support on LOM and an ontological support for pedagogy, may put the current systems in a better situation in pedagogical manner. Developing and implementing these pedagogical information and ontology will not be easy but they have to be taken in consideration more than they had been.

Considering the fact that learner might have different learning styles and capabilities from the others, displaying and forcing students or learners to get the useful information out of same content, mostly a text or some images supported with text based information with them, will not be a good solution for all in e-learning. At this point, there comes the idea of building a semantic structure (Harper, Chen, & Yen, 2004).

References


