

# USING SEMANTIC WEB MINING TECHNOLOGIES FOR PERSONALIZED E-LEARNING EXPERIENCES

Penelope Markellou<sup>1,2</sup>, Ioanna Mousourouli<sup>2</sup>, Sirmakessis Spiros<sup>1,2,3</sup>, Athanasios Tsakalidis<sup>1,2</sup>

<sup>1</sup> Research Academic Computer Technology Institute, 61 Riga Feraiou str., 26221 Patras, Greece

<sup>2</sup> University of Patras, Computer Engineering & Informatics Department, 26504 Patras, Greece

<sup>3</sup> Technological Education Institute of Messolongi, Department of Applied Informatics in Administration and Economics, 30200, Messolongi, Greece

[markel@cti.gr](mailto:markel@cti.gr), [mousouro@ceid.upatras.gr](mailto:mousouro@ceid.upatras.gr), [syrma@cti.gr](mailto:syrma@cti.gr), [tsak@cti.gr](mailto:tsak@cti.gr)

## ABSTRACT

The challenge of the Semantic Web Mining technologies in the e-Learning domain can relate to the provision of personalized experiences for the users. Particularly, these applications can take into consideration the individual needs and requirements of learners. In this paper, we propose a framework for personalised e-Learning based on aggregate usage profiles and a domain ontology. We have distinguished two stages in the whole process, one of offline tasks that includes data preparation, ontology creation and usage mining and one of online tasks that concerns the production of recommendations.

## KEY WORDS

e-Learning, Semantic Web, Web Usage Mining, Ontologies, Adaptive Hypermedia, Personalization, Association Rules, Recommendations.

## 1. Introduction

Nowadays, the Web is rapidly growing and becoming a huge repository of information, with several billion pages and more than 300 million of users globally. Indeed, it is considered as one of the most significant means for gathering, sharing, and distributing information and services. At the same time this information volume causes many problems that relate to the increasingly difficulty of finding, organising, accessing, and maintaining the required information by users. All these have affected greatly the way web-based applications are designed and implemented and e-Learning systems could not comprise an exception. Besides, among all other “e” movements, e-Learning is one of the fastest growing and universally accepted.

*E-Learning* (stands for all forms of web-based learning) uses computer and computer networks to create, deliver, manage and support online learning courses. In particular, thanks to the aforementioned Web explosion, the research on e-Learning has gained more and more attention. Educational and commercial organizations demonstrate a continued interest in the area, which has been a strong

driving force behind numerous research and commercial efforts in the recent years. The variety of available e-Learning systems and applications is a solid indication of the maturity in the area [1], [2], [3].

However, in the majority of past e-Learning systems the courses and the educational materials were not dynamic enough or presented complicated structuring and consequently could not respond effectively to the needs and competencies of the learners, resulting in poor experiences. Generally, hyperlinked course material allows learners to follow any navigational path they choose and not necessarily use the structure determined by web site designers or content creators (who have a certain navigational pattern in mind). This freedom may prove a hindering factor since in many cases learners do not have the necessary maturity and skill to follow an effective path and it is often the case that they wonder around topics that are either too difficult, too easy, or just irrelevant to individual learning needs [4].

An answer to this problem that comprises also the current challenge for web-based learning systems, is their enhancement by the integration of *adaptive* features that allow for the delivery of *personalized* learning. Such systems feature as a remedy for the problems that stem from the traditional “one-fits-to-all” approach that delivers the same static learning material to everyone, despite of individual domain expertise, information needs and preferences, which may vary dramatically [5]. These advanced e-Learning applications provide high quality content, efficient structuring, as well as full support for the varied tasks of all the user profiles participating in a typical distance learning scenario [6].

To achieve this, methods and techniques from various scientific domains and application areas are used. The most well-known are Data Mining, Web Mining, Knowledge Discovery, User Modelling, User Profiling, Artificial Intelligence and Agent Technologies, etc.

Especially, *Web Mining* is defined as the use of Data Mining techniques for discovering and extracting

information from web documents and services and is distinguished as *Web Content, Structure or Usage Mining* depending on which part of the Web is mined [7]. In the majority of cases, e-Learning applications base personalization on Web Usage Mining, which undertakes the task of gathering and extracting all data required for constructing and maintaining learners' profiles based on the behavior of each user as recorded in server logs [8].

Recently, the area of the *Semantic Web* is coming to add a layer of intelligence in these applications. According to [9] "*the Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation*". While a more formal definition by the W3C [10] refers that "*the Semantic Web is the representation of data on the World Wide Web. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URLs for naming*".

The combination of Web Mining and Semantic Web has created a new and fast-emerging research area that of *Semantic Web Mining*. The idea behind using the Semantic Web for generating personalized Web experiences is to improve Web Mining by exploiting the new semantic structures [11]. With the integration of Semantic Web Mining technologies, the provided web applications and especially e-Learning will become smarter and more comprehensive.

In this paper we will investigate how Semantic Web Mining technologies and in particular *ontologies* can be incorporated in the e-Learning domain. Especially, in personalized web-based teaching and learning systems where the individual needs and requirements of the learners play significant role. Specifically, the paper structure follows: in section 2 we present basic issues from the Semantic Web Mining and e-Learning area. In section 3 we describe our approach (personalization scenario) to support personalization in a given e-Learning system, while section 4 concludes the paper.

## 2. Semantic Web Mining and e-Learning

Traditional approaches to personalization have included both content-based and user-based techniques [12]. Recommendations produced with the first technique based on content similarity to the personal profile of the users, while the second one focus on similarities to other users [13]. Their drawback concerns the difficulty to capture semantic knowledge of the application domain i.e. concepts, relationships among different concepts, inherent properties associated with the concepts, axioms or other rules, etc.

As the Semantic Web comes with new emerging standards based on evolving Web technologies, it allows the reuse of material in different contexts, flexible solutions, as well as robust and scalable handling. For achieving this, the web documents are now annotated with *meta-information* or *metadata*. This metadata defines what the documents are about in a machine processable way. *Ontologies* offer a way to cope with these heterogeneous representations of Web resources. They comprise the backbone of the Semantic Web and appear as a promising technology for implementing in particular e-Learning applications. The reason ontologies are becoming so popular is due to what they promise: "*a share and common understanding of a domain that can be communicated between people and application systems*" [14].

An ontology can formulate a representation of the learning domain by specifying all of its concepts, the possible relations between them and other properties, conditions or regulations of the domain. The development of the ontology is akin to the definition of a set of data and their structure. In this way, the ontology can be considered as a knowledge base that is used further for extracting useful knowledge and producing personalized views of the e-Learning system.

Current research on ontologies has shown the important role that they can play in the e-Learning domain. In [15] the authors outline how the Semantic Web technologies based on ontologies can be used for realizing sophisticated e-Learning scenarios and improve the management of their resources. In this case the ontologies are used for describing the semantics and defining the learning context of the material, as well as for structuring the courses.

A framework for personalized e-Learning in the Semantic Web and the way the resource description formats can be utilized for automatic generation of hypertext structures from distributed metadata is proposed in [16]. In particular, several ontologies are used for describing the features of domains, users, and observations.

An ontology-based tool suite, the Courseware Watchdog, which allows making the most of the e-Learning resources available on the Web is presented in [17]. The tool addresses the different needs of tutors and learners and organizes their learning material according to them.

An overview over the use of ontologies and metadata for e-Learning, as well as about innovative approaches and techniques is described in [18]. The authors give emphasis on relevant metadata standards, bindings, schemas and annotations, classifications for describing content/topic of a resource, etc. Then they introduce different ontologies and present a RDF-based peer-to-peer network for digital resources and for the exchange of learning objects and services.

### 3. Proposed Personalization Scenario

In our scenario for supporting personalized e-Learning, the structure of knowledge and information play a crucial role. The proposed ontology-based organization helps the structure and the managing of content related to a given course or lesson. In particular, the framework for personalization based on aggregate usage profiles and the domain's ontology and it is depicted in Figure 1. This framework distinguishes between the offline tasks of data preparation, ontology creation and usage mining, and the online personalization components.

Starting with the offline part, the preprocessing tasks result in aggregate structures such as a user transaction file computing meaningful semantic units of user activity to be used in the mining stage. Given the preprocessed data a variety of data mining tasks can be performed. In our approach, we focus on the discovery of association rules, using Apriori Algorithm.

The system uses server's log files, which describe users' navigational activity. Basically, these files encapsulate all the relative information with the usage of the e-Learning domain by the users. In this stage, server's logs should be cleaned according to site files. This task involves the removal of redundant references. It requires detailed site structure information in order to determine which page file accesses contribute to a single browser display, and

more specifically which content corresponds to each user's request.

The preprocessing tasks described above result in a set of:

$$n \text{ pageviews, } P = \{p_1, p_2, \dots, p_n\}$$

with each pageview uniquely represented by its associated URL, and a set of:

$$m \text{ user transactions, } T = \{t_1, t_2, \dots, t_m\}$$

where each  $t_i \in T$  is a subset of  $P$ .

Having the set of transactions  $T$ , the problem of mining association rules is to generate all association rules that have *support*  $\sigma$  and *confidence*  $\alpha$  greater than a specified minimum support (called *minsup*) and minimum confidence (called *minconf*) respectively. An algorithm for finding all association rules is the Apriori algorithm [19].

Apriori is going to be applied to transactions which arose above in order to discover the set of *association rules* that correspond to the specific transaction set. The algorithm initially finds groups of items (in this case are the URLs appearing in the preprocessed log) occurring frequently together in many transactions. Such groups of items are referred to as *frequent item sets*.

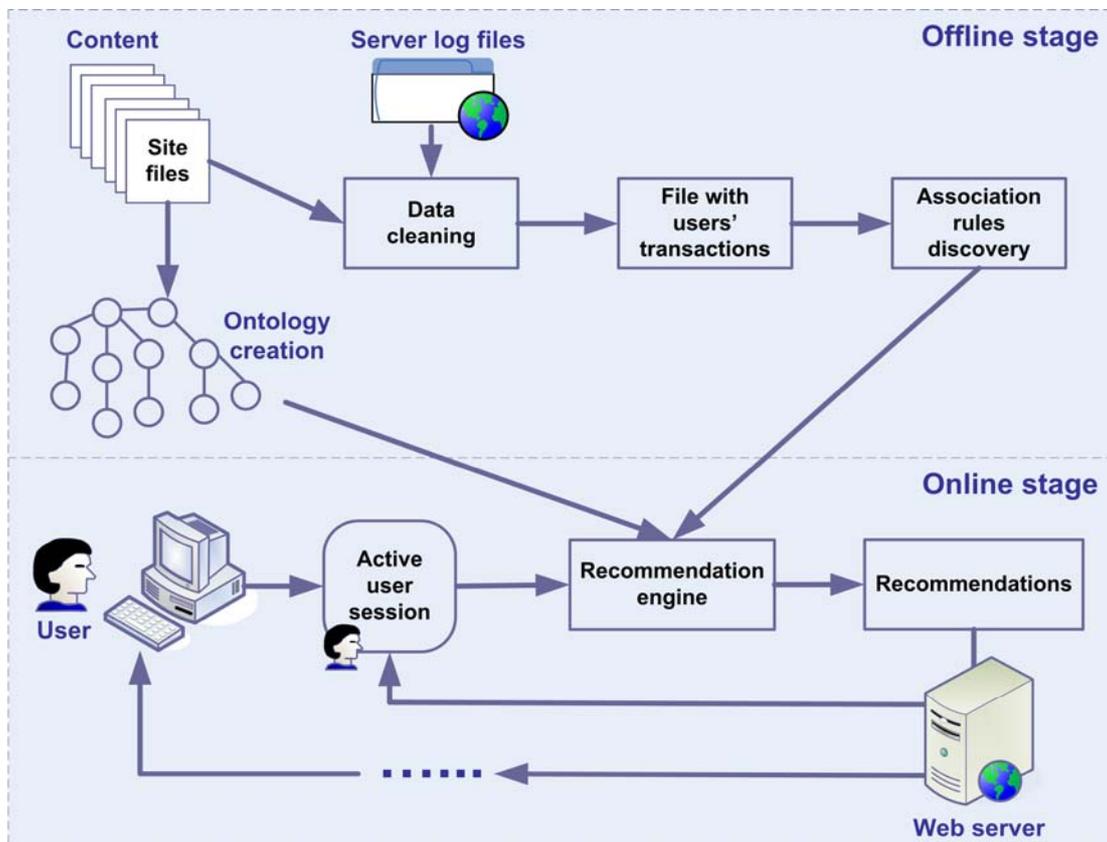


Figure 1. Proposed scenario for producing recommendations in an e-Learning system.

Given a set  $I = \{I_1, I_2, \dots, I_k\}$  of frequent itemsets, the support of  $I_i$  is defined as:

$$\sigma(I_i) = \frac{|\{t \in T : I_i \subseteq t\}|}{|T|}$$

and it represents the ratio of transactions in transaction set  $T$ , which have the frequent itemset  $I_i$ .

The support threshold (minsup) is used by the algorithm for pruning the search space and is generally specified before the mining step. Association rules capture the relationships among items based on their patterns of co-occurrence across transactions. In the case of Web transactions, association rules capture relationships among URL references based on the navigational patterns of users.

An association rule  $r$  is an expression of the form:

$$X \Rightarrow Y (\sigma_r, \alpha_r)$$

where  $\sigma_r$  is the support of  $X \cup Y$ , and  $\alpha_r$  is the confidence for the rule  $r$  given by  $\frac{\sigma(X \cup Y)}{\sigma(X)}$ . The

confidence of the rule  $r$  shows the ration of transactions, in transaction set  $T$ , that contain  $X$  will also contain  $Y$ . We are going to use frequent itemsets and association rules to provide recommendations to the learners.

For ensuring effective personalization, we combine the existence of an ontology of the content with the knowledge that comes out of the users' navigation paths. We are going to use the latter in order to infer the way that students learn the concept. Recommendations are going to be made to users according to the ontology relations and the inferences mentioned above, with respect to user's current position. The role of the ontology is to determine which learning materials are more suitable to be recommended to the user, and according to frequent itemsets (user's navigation paths) which of these choices have the maximum support.

The ontology of the e-Learning domain, is going to describe the content and the relations between the various notions. It will formulate a thorough representation of the domain by specifying all of its concepts and the existing relations. Through the ontology the system will express hierarchical links between entities.

We decide to use one common ontology and express the knowledge described in each of the corpora as subgraphs of the ontology by labelling the nodes accordingly. This approach allows us to easily compare the knowledge of a user in relation to each of the corpora. Another approach would be to separate ontology for each corpus and construct the overall ontology by ontology mapping. But this approximation doesn't give us the capability to

correspond to user's knowledge with each corpora and it is less flexible.

As it concerns to online part, the system keeps track of the active user session, which depicts the recent past user's choices. According to his current state, a recommendation engine recommends him the next more appropriate link. This engine accepts active user session and also takes into consideration the ontology of the domain and the set of association rules, which came from users' transactions during the offline part.

In particular, we base on the following for discovering the most appropriate recommendations to make:

1. *The document ontology.* We assume that documents are annotated according to standard metadata schemas for documents like e.g. Dublin Core (DC) [20], or in the area of education, according to the Learning Objects Metadata Standard (LOM) [21]. In our approach the metadata descriptions of documents are in accordance with LOM.
2. *The file with the extracted association rules.* These rules resumed from the users' transaction, during the preprocessing step.

Engine's role is to compute a recommendation set, which consists of links to pages that the user may want to visit. It essentially represents a short view of potentially useful links based on the user's navigational activity through the site. These recommended links are then added to the last page in the session accessed by the user before that page is send to the user browser.

By using a fixed-size sliding window over the current active session, we can capture the current user's history depth. For example if the current session (with a window size of 3) is  $\langle A, B, C \rangle$  and the user references the URL  $D$ , then the new active session becomes  $\langle B, C, D \rangle$ .

The factors that we should take into consideration in the recommendation process are:

- the domain ontology.
- the matching criteria with the frequent itemset.
- whether the candidate URLs for recommendation have been visited by the user in the current active session.
- the graph of the site.

Next, we are going to compute the potential recommendation set using the ontology of the domain. Our goal is to find recommendation set according to the ontology domain. This set is then filtered through frequent itemsets, which was discovered during the preprocessing stage. Frequent itemsets essentially depicts the knowledge that comes out from navigational activity of other users who act commonly with the current user.

As we have already mentioned, in the factors above we have included the site graph. The latter is going to be

used for computing the distance of a candidate URL from user's current position. As distance we consider the numbers of clicks (click stream) that the user should made in order to go from his current position to the recommended URL. The algorithm used for producing recommendations to the users is presented in Table 1.

<p><b>Input:</b>  -Active user session <math>z</math> (session window max size <math>n</math>).  -Domain ontology.  -Frequent itemsets and association rules.  -Minsup threshold <math>\sigma</math>.  -Minconf threshold <math>\alpha</math>.  -Site graph.</p> <p>Recommendation set <math>r = \emptyset</math>.  Potential recommendation set <math>w = \emptyset</math>.</p> <p>Potential recommendation set <math>w</math> according to the ontology <math>w = \{w_1, w_2, \dots, w_k\}</math>.</p> <p>For each set <math>z+w_i, w_i \in W</math>  for each frequent itemset <math>I_i = z+w_i</math>  if <math>\text{sup}(I_i) \geq \sigma</math> then      <math>c = \text{conf}(z \Rightarrow w_i)</math>      if <math>c \geq \alpha</math>          <math>w_i\_score = c * \text{click\_num}</math> (*)          <math>r = r \cup w_i</math>      end if  end if  end for  end for</p>
--

Table 1. Proposed algorithm for producing recommendations to the users.

Specifically, the proposed algorithm orders the recommendations that came out from the ontology according to the computation step (\*), and excludes these recommendations with low support and confidence. The initial recommendation set is been filtered and enhanced through frequent itemsets and association rules.

#### 4. Conclusion

In this paper, we presented firstly basic Semantic Web and Web Usage Mining notions. Then, we discussed about the application of techniques coming from the new emerging area of Semantic Web Mining in the domain of e-Learning systems and analyzed the significant role of ontologies. We expounded and argued about our proposed approach for producing recommendations to users in a given e-Learning corpus. Finally, we concluded with the description of the recommendation engine's operation and

presented an algorithm for making effective recommendations.

As shown in the paper, the proposed personalization scenario tries to integrate the Semantic Web vision by using ontologies with Using Mining techniques in order to better service the needs and the requirements of learners. We strongly believe that the combination of domain's ontology and frequent itemsets, which include all the information about users' navigational attitude, enhances the whole process and produces better recommendations.

The system first finds an initial recommendation set and then uses the frequent itemsets to enrich it, taking into consideration other users' navigational activity. In this way, we reduce the time we spend on parsing all frequent itemsets and association rules. We focus only on those sets that come out from the combination of the active user session and the ontology's recommendations. The time reduction arises because of the fact that frequent itemsets are filtered through the ontology's recommendation set resulting in a smaller searching space.

A limitation of this approach relates to that the engine doesn't always give the best results because of its straight dependence from the specific domain. Besides, the created ontology depicts the way that the e-Learning domain should be taught to the learners and based on the view of the designer. If the ontology isn't made correctly, then the initial set of recommendations would be much far away from the way that users learn the domain, and our method can not change that. Our approach doesn't add new recommendations in the initial recommendation set. It only reorders and exclude items according to the thresholds of *minsup* and *minconf*.

Future work will focus on further experiments with different combinations of the system's functionalities, further contextualization possibilities from the Semantic Web Mining area, and an evaluation of the proposed approach with respect to learning support and to open-corpus learning.

#### References:

- [1] T. Urdan, & C. Weggen, Corporate e-learning: exploring a new frontier, WR Hambrecht and Co, 2000.
- [2] T. Wentling, C. Waight, J. Gallaher, J. La Fleur, C. Wang, & A. Kanfer, E-learning: a review literature, Knowledge and Learning Systems Group, National Center for Supercomputing Applications, University of Illinois, 2000.
- [3] K. Fry, E-learning markets and providers: some issues and prospects, *Education and Training, Emerland*, 43(4), 2001, 233-239.
- [4] K. Markellos, P. Markellou, M. Rigou, S. Sirmakessis, & A. Tsakalidis, Web personalization

- for enhancing e-learning experience, *Proc. 5<sup>th</sup> International Conf. on Information Communication Technologies in Education*, ICICTE, Samos Island, Greece, 2004.
- [5] P. De Bra, P. Brusilovsky, & G.J. Houben, Adaptive hypermedia: from systems to framework, *ACM Computing Surveys*, 31(4es), 1999.
- [6] F. Mödritscher, V.M.G. Barrios, & C. Gütl, The past, the present and the future of adaptive e-Learning, *Proc. ICL2004, International Conf. Interactive Computer Aided Learning*, Villach, Austria, 2004.
- [7] R. Kosala, & H. Blockeel, Web mining research: a survey, *SIGKDD Explorations*, 2(1), 2000, 1-15.
- [8] P. Markellou, M. Rigou, & S. Sirmakessis, Mining for web personalization, *Web Mining: Applications and Techniques*, A. Scime (Ed.), Hershey: Idea Group Publishing, 2004, 27-48.
- [9] T. Berners-Lee, J. Hendler, & O. Lassila, The semantic web, *Scientific American*, 284(5), 2001, 34-43.
- [10] W3C. <http://www.w3.org/2001/sw>.
- [11] P. Markellou, M. Rigou, S. Sirmakessis, & A. Tsakalidis, Personalization in the semantic web era: a glance ahead, *Proc. 5<sup>th</sup> International Conf. on Data Mining, Text Mining and their Business Applications, Data Mining 2004*, A. Zanasi, N.F. Ebecken & C.A. Brebbia (Eds.), Wessex Institute of Technology (UK), Malaga, Spain, Southampton, Boston: PWIT Press 2004, 3-11.
- [12] H. Dai, & B. Mobasher, Integrating semantic knowledge with web usage mining for personalization, *Web Mining: Applications and Techniques*, A. Scime (Ed.), Hershey: Idea Group Publishing, 2004, 276-306.
- [13] B. Mobasher, X. Jin, & Y. Zhou, Semantically enhanced collaborative filtering on the web, *EWMF*, 2003, 57-76.
- [14] J. Davies, D. Fensel, & F. van Harmelen, Introduction, *Towards the Semantic Web, Ontology-Driven Knowledge Management*, John Wiley & Sons, 2003, 1-9.
- [15] L. Stojanovic, S. Staab, & R. Studer, E-learning based on the semantic web, *Proc. WebNet2001, World Conf. on the WWW and Internet*, Orlando, Florida, USA, 2001.
- [16] N. Henze, P. Dolog, & W. Nejdl, Reasoning and ontologies for personalized e-learning in the semantic web, *Educational Technology and Society*, 2004, 82-97.
- [17] J. Tane, C. Schmitz, & G. Stumme, Semantic resource management for the web: an e-learning application, *Proc. 13<sup>th</sup> International WWW2004 Conf. on Alternate Track Papers and Posters*, New York, USA, 2004, 1-10.
- [18] J. Brase, & W. Nejdl, Ontologies and metadata for eLearning, *Hanbook on Ontologies*, S. Staab, & R. Studer (Eds), Springer-Verlag, 2004, 555-573.
- [19] R. Agrawal, & R. Srikant, Fast algorithms for mining association rules, *Proc. 20<sup>th</sup> VLDB Conf.*, Santiago, Chile, 1994, 487-499.
- [20] Dublin Core. <http://dublincore.org/>.
- [21] LOM: Draft standard for learning object metadata. <http://ltsc.ieee.org/wg12/index.html>.