International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 6– No. 9, March 2014 – www.ijais.org

An Ontology Framework based on Web Usage Mining

Ahmed Sultan Al-Hegami
Sana'a University
Yemen – Sana'a

Mohammed Salem Kaity
Al-andalus University
Yemen – Sana'a

ABSTRACT

Finding relevant information on the Internet became a real challenge. This is to some extent due to the volume of data available and the lack of structure in many Web sites. Web usage mining is an important area and fast developing mining on the Internet. The purpose of Web mining is the development of techniques and systems to detect patterns of things and processes on the World Wide Web and the Internet for performance systems that appear to adapt.

Ontology is some knowledge that can be used to describe the information on the Web.

In this work we propose a framework for generating ontology based on web usage mining . We have implemented all stages of the system which are data acquisition, web mining and ontology creation. Our ontology learning framework proceeds through ontology import, extraction, pruning, refinement, and final review of the ontology .

Keywords

Data Mining, Web mining, Web usage mining, Web content mining, Web structure mining, Ontology, Clustering, Sequential Pattern.

1. INTRODUCTION

Fast growth of e-commerce has made both business community and customers face a new situation [10]. Number of Internet users has grown dramatically over the past decade and continues to increase. [2] With the continued growth and the explosion of e-commerce, Web services, and other information systems that rely on the Internet, the volume of data that has been collected user clicked and organizations on the Internet in their daily operations and reached a large volume [3][10].

Analysis of these data can help these organizations to make a decision for the value of life-time customers, and design marketing strategies across products and services, and evaluate the efficiency of marketing campaigns, and improve the functions of Web-based applications, providing content more personalized for visitors, and create a logical structure the most effective space their own web . This type of analysis involves the automatic discovery of patterns with meaning and relationships of a large group of data semi-structured in the first place, and store frequently in the Web site and server applications access logs, as well as the sources of operational data relevant [3] [2].

Is also used widely for e-business, and technology is adapted to web data mining of e-business to provide personalized e-business and better meet the requirements of users. Starting from the concept of personalized information services.

Also with the huge growth of information available on the World Wide Web, and it has become more difficult to obtain related information from the Web. One possible way to resolve this problem is to use a Web mining. It aims to discover patterns and interesting user access of data usage on the Internet based on ontology [4] [5] [6]. Then these discovered can be used as the model to analyze and predict future user access behavior.

Several techniques such as association rules and hierarchical clustering have been proposed to deal with ontology. However, majority of these approaches have focused on generating concept hierarchy for building ontology from text documents [7][2][3].

Although ontology generation is important, we need to deal with the technical issues on how to define activities access to the Internet, the discovery of hierarchical relationships of activities access to the Internet, and automatically converted into ontology. Furthermore, we can deduce from the use of ontology knowledge [8][4]. This research proposes a web usage mining approach that generating ontology.

2. MOTIVATION

Understanding customer needs and customers behavior are necessary to keep customers in the store on the Internet for e-commerce applications over the Internet as competitors are just one click away. Web usage mining can help an e-commerce solution to improve the selling . by analyzing the clickstream and customer purchase data through data mining techniques [9]. Recently, web usage mining has attracted the researchers' attention for e-business professionals and it offers many of the benefits to an e-commerce web sites.

With the huge amount of data available on the World Wide Web, it has become a difficult task to access relevant information from the Web. So our observation that trends to make use of Web usage Mining techniques for generating ontology. This paper aims to solve these problems by using web usage mining technique to generate some domain knowledge that could be used to describe information on web.

3. WEB USAGE MINING

Data Mining: refers to extracting or "mining" knowledge from large amounts of data [10]. It aims to use the data and information to extract interesting knowledge [2] .

Others view Data Mining as a stage in a process called KDD *(knowledge discovery database.* The process of Knowledge discovery as shown in Figure 1, includes the following steps [10] :

1. Data cleaning .
2. Data integration .
3. Data selection .
4. Data transformation .
5. Data mining .
6. Pattern evaluation .
7. Knowledge presentation.

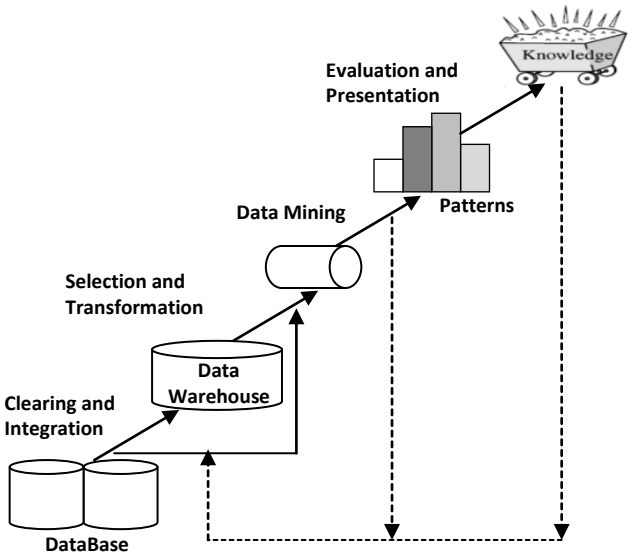


Figure 1. Data mining as a core step in the process of knowledge discovery [3].

Web mining: consists of a set operations defined on data residing on WWW data servers. Web mining aims to discover interesting patterns or knowledge from the Web hyperlink structure, page content, and usage data. Although Web mining uses many data mining techniques, web mining tasks can be categorized into three types: Web structure mining, Web content mining and Web usage mining [3] [4][5].

- ***Web content mining*** (WCM) is a form of text mining applied to Web pages. This process allows discovering relationships related to a particular domain via co-occurrences of terms in a text for example.
- ***Web structure mining*** (WSM) is used to examine data related to the structure of a Web site. This process operates on Web pages' hyperlinks. Structure mining can be considered as a specialization of graph mining.
- ***Web usage mining*** (WUM) is applied to usage data such as those contained in logs files. A log file contains information related to the queries executed by users on a particular Web site. Web usage mining can be used to modify the Web site structure or give some recommendations to the visitor. Personalization can also be enhanced by usage analysis [4].

Web usage mining refers to the automatic discovery and analysis of patterns in **clickstream** and associated data collected or generated as a result of user interactions with Web resources on one or more Web sites[3]. The goal is to capture, model, and analyze the behavioral patterns and profiles of users interacting with a Web site. The discovered patterns are usually represented as collections of pages, objects, or resources that are frequently accessed by groups of users with common needs or interests.

4. THE ONTOLOGY

Since ontology deals with existence of entities, their hierarchical properties and relations it has been a topic of interest for all science branches and information science is no different in that aspect. According to [2] "*A specification of a representational vocabulary for a shared domain of discourse — definitions of classes, relations, functions, and other objects — is called an ontology"*. From the point of view of computer scientists ontologies are very good candidates for sharing information about a specific domain in a formal way [11].

Ontology however should not be confused with taxonomy where entities are arranged in a way that only takes the generalization and specialization properties into account. The well known online encyclopedia Wikipedia for example categorizes and sub categorizes the topics in their database in a manner where finding information is easier. The result of this categorization is taxonomy. Ontology also defines many relations between entities, restrictions and also the way these relations are to be used [2].

5. RELATED WORKS

There are many previous studies of this subject, most of these studies and research focused on the extraction of patterns and decision-making. And there are a number of theses has developed frameworks for the web usage mining [10][2][3].

One of these methodologies are based on the knowledge discovery techniques mainly from HTTP Web logs and want at confronting the discovered knowledge in terms of usage with the existing ontology in order to propose new relations between concepts [4][5]. These frameworks are divided into two main phases, the batch stage and online stage [6]. Figure 2 shows the view of the previous steps that depend on the previous two phases.

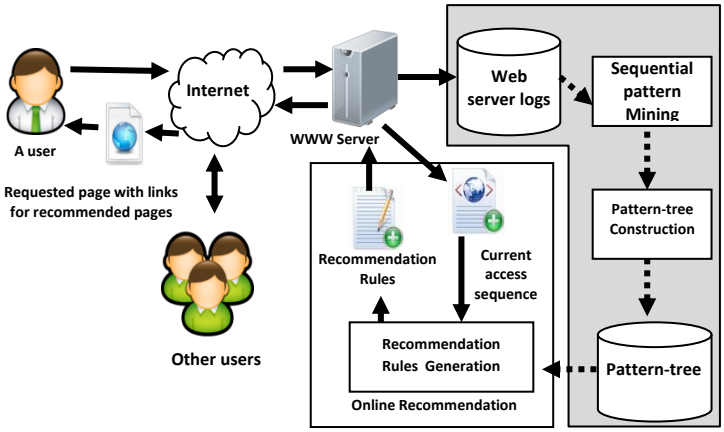


Figure 2. A web usage mining architecture [27]

There is another look for the extraction of the ontology of web pages, where the ontology of the finest knowledge structure the complex definitions for the web and reach a common understanding [2][4][13]. According to this view, we need to generate ontology from the pages of the Internet. Also, there are efforts to generate ontology by analyzing the text or re-write web pages and the addition of ontology through the RDF or XML, but it is a waste of time. Some approaches transform html pages into a semantic structured hierarchy encoded in XML, taking into account html regularities [14].

The data sources can be texts, semi-structured data, relational data, etc. In the following, we explain some approaches fanatical to knowledge extraction from web pages.

Ontology construction can be performed manually or semi-automatically. In the first case, this task is hard and time-consuming. This is the reason why many methods and methodologies have been design to semi-automate this

process. A survey on ontology learning methods and tools can be found in the Ontoweb web site Many methods or methodologies have been proposed to enrich an existing ontology using web documents [15][16] However, these approaches are not specifically dedicated to web knowledge extraction.

The approach proposed by [17] attempts to reduce the conceptual confusion by learned the concepts and relationships from a set of web sites using the Ontolearn tool. The main steps are: terminology extraction from web sites and web documents data warehouse, semantic interpretation of terms and identification of taxonomic relationships.

Finally, we can also point out some approaches only dedicated to ontology construction from web pages without using any a priori knowledge.

The approach described in [34] is based on extract some keywords representative of the domain, after that find a collection of web sites related to the previous keywords , then exhaustive analysis of each web site, after that the analyzer searches the initial keywords in a web site and finds the preceding and following words; these words are candidates to be concepts, for each selected concept, a statistical analysis is performed based on the number of occurrences of this word in the web sites and finally, for each concept extracted using a window around the initial keyword, a new keyword is defined and the algorithm recursively iterates.

In [18], a method is proposed to extract domain ontology from web sites without using a priori knowledge. This approach takes the web pages structure into account and defines a contextual hierarchy. The data preprocessing is an important step to define the more relevant terms to be classified. Weights are associated to the terms according to their position in this conceptual hierarchy. Then, these terms are automatically classified and concepts are extracted.

In [19] the authors define an ontological architecture based on a semantic triplet, namely: semantics of the contents, structure and services of a domain. This paper focuses on the domain ontology construction and is based on a meta-ontology that represents the linguistic structure and helps to extract lexico-syntactic patterns. This approach is a hybrid one, based on statistical and linguistic techniques. A set of candidate concepts, relationships and lexico-syntactic patterns is extracted from a domain corpus and iteratively validated using other web corpus. Experiments have been realized in the tourism domain.

6. AN ONTOLOGY FRAMEWORK BASED ON WEB USAGE MINING

Our approach relies on the usage analysis of the chosen Web site, in complement of the existing approaches based on content analysis of Web pages. Our methodology is based on the knowledge discovery techniques mainly from Web logs and aims at confronting the discovered knowledge in terms of usage with the existing ontology in order to propose new relations between concepts. One major contribution of this thesis is thus the application of usage analysis to support ontology evolution and/or web site reorganization.

Our work aims at providing a solution to this problem by using an existing or semi-automatically built ontology intended to enhance information retrieval. After that using Web Usage Mining methods like classification and sequential pattern mining techniques to analysis the log files for extract pattern. Finally, generating Web usage ontology and / or Updating the exist ontology with extracted knowledge.

Our proposed framework is divided into two stages, Batch stage and Ontology stage. These stages shown in Figure 3 and are described below.

BATCH STAGE

The first stage consists in building the domain of ontology of the considered Web sites or in enriching the existing ontology if one is already available. The construction of this basic ontology can be achieved through knowledge extraction that divided into three steps :

Step 1: Ontology Construction Methods :

Ontology construction can be performed manually or semi-automatically. In the first case, this task is hard and time-consuming. This is the reason why many methods and methodologies have been design to semi-automate this process. The data sources can be texts, semi-structured data, relational data, etc. In the following, we describe some methods dedicated to knowledge extraction from web pages. A survey on ontology learning methods and tools can be found in the Ontoweb web site . Ontology Construction Methods based on content and structure mining (not our focus) [4].

Step 2: Web Logs Pre-processing Methods:

This stage consists of pre-processing the logs (raw data) in a rich usage data warehouse based on various notions such user session, user visit etc. pre-processing consists of :

- Acquisition of data
- Cleaning data
- Identifying page views
- Identifying Users
- Identifying sessions
- Path Competition

Step 3: Web Usage Mining Methods:

The last step in this stage aims at applying data mining techniques on users' visits on these Web sites through the analysis of log files in order to generating or / and update the current *ontology*. Web Usage Mining techniques allow us to define visit or navigation profiles and then to facilitate the emergence of new concepts, for example by merging several existing ones. Two different methods will be emphasized and used to analysis usage: *Clustering and Sequential Pattern Mining*. Finally, we using the information obtained from these two methods to ontology construction .

In the following subsections, we describe the data mining methods we applied on Web usage data in order to illustrate our Framework :

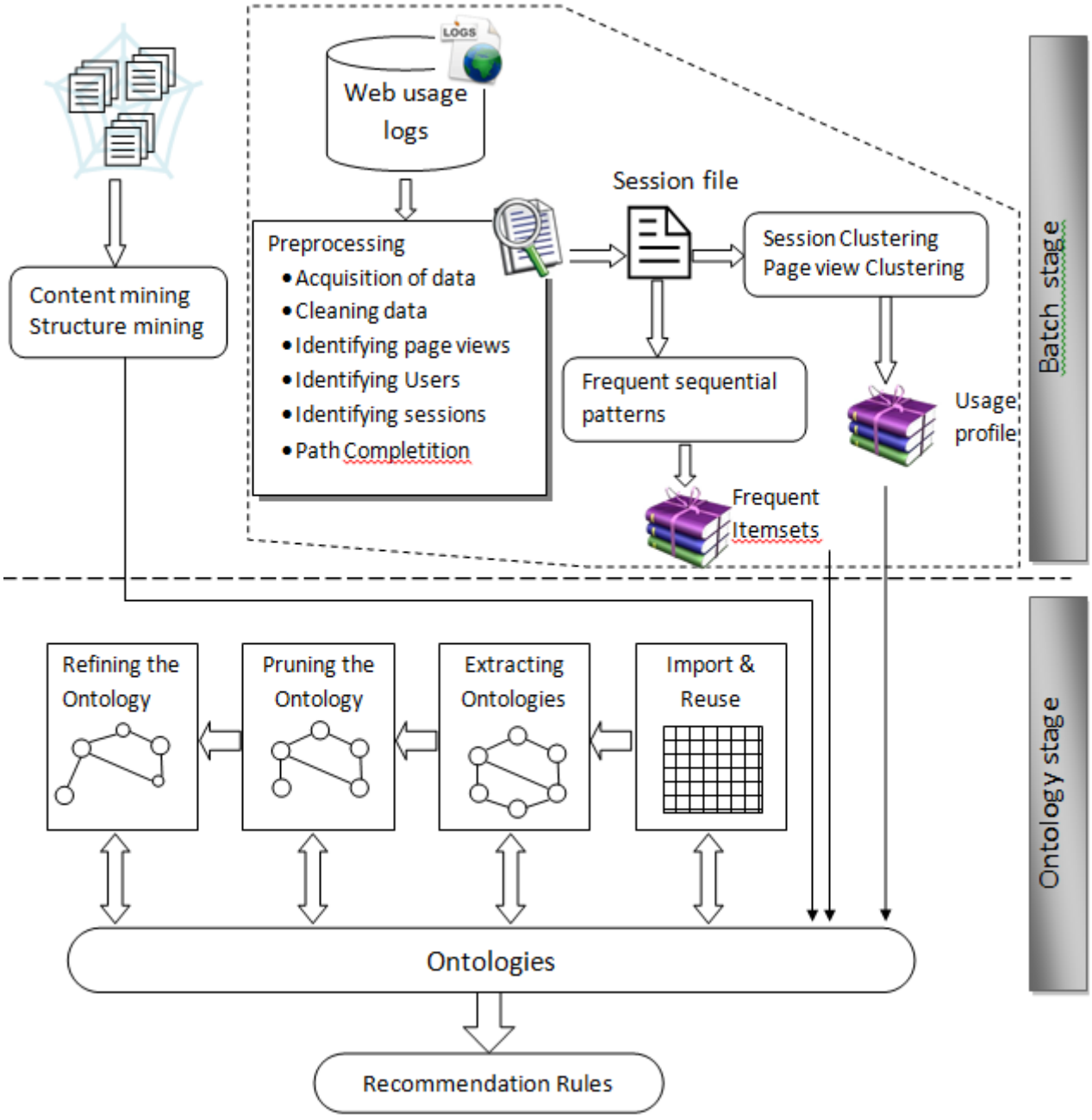


Figure 3. The Ontology Framework based on Web Usage Mining

Clustering Methods- Appropriate use of a clustering algorithm is often a useful first step in extracting knowledge from a database. Clustering, in fact, leads to a classification, i.e. the identification of homogeneous and distinct subgroups in data , where the definition of homogeneous and distinct depends on the particular algorithm used: this is indeed a simple structure, which, in the absence of a priori knowledge about the multidimensional shape of the data, may be a reasonable starting point towards the discovery of richer and more complex structures.

Sequential Pattern Extraction Methods- Sequential pattern mining deals with data represented as sequences (a sequence contains sorted sets of items). Compared to the association rule extraction. Due to the notion of time embedded in the data, applications for sequential pattern extraction are numerous and the problem definition has been slightly modified in different ways. Associated to elegant solutions, these problems can match with real-life time stamped data (when association rules fail) and provide useful results.

ONTOLOGY STAGE

The result of usage analysis provides classes of pages and visits. The clustering relies on the site's structure, visit classes

(also called navigation profiles) are based on clusters of visited pages and pages labeling is achieved syntactically with regard to the directories corresponding to the Web sites' structure. In order to see the impact on the ontology, a matching must be done between these syntactic categories and the ontology's semantic concepts.

Our idea is generating and updating concern ontology extension which does not completely modify the initial ontology:

- Addition of a leaf concept in a hierarchy,
- Addition of a sub-tree of concepts in the hierarchy,
- Addition of a relation between two concepts.

A new concept may appear as a consequence of the emergence of a class of user navigations (from the classification process).

In the same way, a new relation between two concepts may be identified through the extraction of sequential patterns.

The proposed approach for Web Usage Ontology Generation, which consists of the following components:

- Import & Reuse
- Extracting Ontologies
- Pruning the Ontology
- Refining the Ontology

Import & Reuse- In the first part of the import & reuse step, the schema structures are identified and their general content needs to be discussed with domain experts. Each of these knowledge sources must be imported separately. The ontology structure and exist ontology are import from last steps by discovered knowledge from frequent sequential patterns and session and page view clustering .

In the second part of the import & reuse step, imported conceptual structures need to be merged or aligned in order to constitute a single common ground from which to take-off into the subsequent ontology learning phases of extracting, pruning and refining.

Extracting Ontologies- In the ontology extraction phase of the ontology learning process, major parts, i.e. the complete ontology or large chunks reflecting a new subdomain of the ontology, are modeled with learning support exploiting various types of (Web) sources. Thereby, ontology learning techniques partially rely on given ontology parts. Thus, we here encounter an iterative model where previous revisions through the ontology learning cycle may propel subsequent ones and more sophisticated algorithms may work on structures proposed by more straightforward ones before.

Describing this phase, we sketch some of the techniques and algorithms that have been embedded in our framework and implemented in our ontology learning environment .

Pruning the Ontology- we aim at a model that captures a rich conceptualization of the target domain, but that excludes parts that are out of its focus. The import & reuse of ontologies as well as the extraction of ontologies considerably pull the lever of the scale into the imbalance where out-of-focus concepts reign. Therefore, we pursue the appropriate diminishing of the ontology in the pruning phase.

Refining the Ontology- refining plays a similar role as extracting. Their difference exists rather on a sliding scale than by a clear-cut distinction. In principle, the same algorithms may be used for extraction as for refinement. However, during refinement one must consider in detail the existing ontology and the existing connections into the ontology, while extraction works more often than not practically from scratch.

IMPLEMENTATION AND EXPERIMENTAL

In this section we apply the proposed framework. We have chosen Al-andalus University site (http://andalusuniv.net/) to apply the framework .The site is applicable to the e-Learning ontology.

This site is still under development, Also got this site named best educational site in Yemen.

For this experiment, we used an existing ontology and modified it according to the structure of the studied Web site. This ontology, edited with Protégé 5.2 [12] .

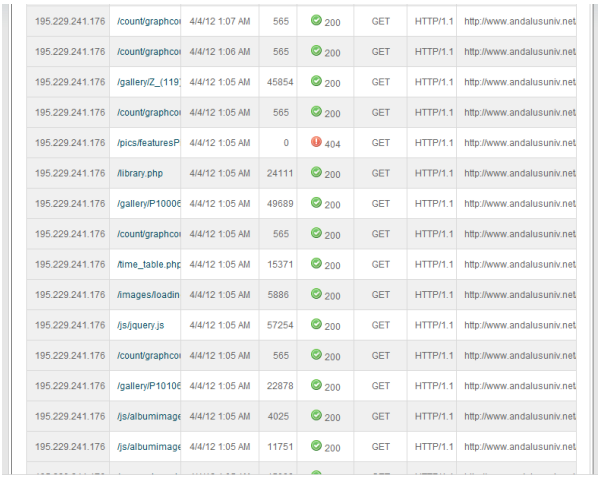
The general steps of the framework have the basics of any other web usage mining system . There is the data acquisition part where data is obtained. There is a data cleaning part where log files are cleaned of irrelevant information. There is an offline mining part .

STEPS AND METHODS

Pre-processing Web Logs

In this step we prepare the data for mining. The used log data consists in 2 HTTP log files with a total of 5243 requests for page views. Each log file contains all the requests cover a continuous 5 days period starting from 25th -3- 2012 until the 30th -3- 2012.

The log files contain the seven following fields , show Figure 4 :



195.229.241.176	/count/graphcou	4/4/12 1:07 AM	565	200	GET	HTTP/1.1	http://www.andalusuniv.net
195.229.241.176	/count/graphcou	4/4/12 1:06 AM	565	200	GET	HTTP/1.1	http://www.andalusuniv.net
195.229.241.176	/gallery/Z_(119	4/4/12 1:05 AM	45854	200	GET	HTTP/1.1	http://www.andalusuniv.net
195.229.241.176	/count/graphcou	4/4/12 1:05 AM	565	200	GET	HTTP/1.1	http://www.andalusuniv.net
195.229.241.176	/pics/featuresP	4/4/12 1:05 AM	0	404	GET	HTTP/1.1	http://www.andalusuniv.net
195.229.241.176	/library.php	4/4/12 1:05 AM	24111	200	GET	HTTP/1.1	http://www.andalusuniv.net
195.229.241.176	/gallery/P10006	4/4/12 1:05 AM	49689	200	GET	HTTP/1.1	http://www.andalusuniv.net
195.229.241.176	/count/graphcou	4/4/12 1:05 AM	565	200	GET	HTTP/1.1	http://www.andalusuniv.net
195.229.241.176	/time_table.php	4/4/12 1:05 AM	15371	200	GET	HTTP/1.1	http://www.andalusuniv.net
195.229.241.176	/images/loadin	4/4/12 1:05 AM	5886	200	GET	HTTP/1.1	http://www.andalusuniv.net
195.229.241.176	/js/jquery.js	4/4/12 1:05 AM	57254	200	GET	HTTP/1.1	http://www.andalusuniv.net
195.229.241.176	/count/graphcou	4/4/12 1:05 AM	565	200	GET	HTTP/1.1	http://www.andalusuniv.net
195.229.241.176	/gallery/P1010(	4/4/12 1:05 AM	22878	200	GET	HTTP/1.1	http://www.andalusuniv.net
195.229.241.176	/js/albumimage	4/4/12 1:05 AM	4025	200	GET	HTTP/1.1	http://www.andalusuniv.net
195.229.241.176	/js/albumimage	4/4/12 1:05 AM	11751	200	GET	HTTP/1.1	http://www.andalusuniv.net

Figure 4. The log files fields

- **IP address:** the computer's IP address of the client making the request;
- **Date:** the time of the request;
- **Request:** the requested resource on the server;
- **Status:** The HTTP status code returned to the client such as success, Failure, redirection, forbidden access…;
- **Size:** the content-length of the document transferred;

- **User Agent:** the user agent.

The rest of the pages are pictures, style sheet documents or robot access document. We clean the log file of all these non content details.

Also discarding the fields that will not be used in mining, we are left with a simpler format. We have

- *The timestamp value:* This value is converted to seconds passed since 00.00.0000 for easier processing.
- *IP address:* Will be stored as a string.
- *Page address:* Will be stored as a string.
- *User Agent:* Will be stored as a string.

IP address	*Page address*	*Timestamp*	*User agent.*
195.229.241.176	/images/favion.ico	4/4/12 1:04 AM	Mozilla/5.0
195.229.241.17	/count/graphcount.php	4/4/12 1:07 AM	Mozilla/5.0
195.229.241.22	/count/graphcount.php	4/4/12 1:07 AM	Mozilla/5.0

Table 1: Format of the Requests

Step through the pre-processing of data we will exclude a number of records. And therefore we will collect orders from the same IP address. Where he became the remaining 1839 session, each session is divided into several visits. And end of each visit when we set in advance for 30 minutes. The number of visits is equal to 2633 visit. We have excluded the failed orders. So we accepted only successful orders with (code 200).

After that, we have compiled the visits and the number of pages per visit as is the position in the table:

Visits	Pages (number of requests)
V_1	http://andalusuniv.net/dlc/moodle/, http://andalusuniv.net/news, http://andalusuniv.net/events.php,...
.....	.....
V_256	http://andalusuniv.net/engit.php, http://andalusuniv.net/events.php, http://andalusuniv.net/objectives.php,...
.....	.....

Table 2: Quantity of pages requested during each visit

Web Usage Mining Methods

- **Clustering Methods**

It is known that e-learning sites frequented by many users. We have used the tools of clustering for the classification of users with pages. We got the result shown in Table 3 and Figure 5 which illustrates these intersections.

Users were divided according to the site considered to student, teacher, manager and examiner. Page was divided by the visit to a group of classifications.

Partition	Ps	Pt	Pm	Pe	Total
Vs	1442	4	0	0	1446
Vt	64	724	4	0	792
Vm	0	0	231	0	231
Ve	0	0	0	164	164
Total	1506	728	235	164	2633

Table 3 : crossed clustering between visited pages and users pages

The pages of the studied web site are divided into :

- (Ps) contains all the pages of Students ,
- (Pt) contains all the pages of Teachers ,
- (Pm) contains all the pages of Manager ,
- (Pe) contains all the pages of Examiner .

And the visits of the studied web site are divided into :

- (Vs) contains all the visits of Students ,
- (Vt) contains all the visits of Teachers ,
- (Vm) contains all the visits of Manager ,
- (Ve) contains all the visits of Examiner .

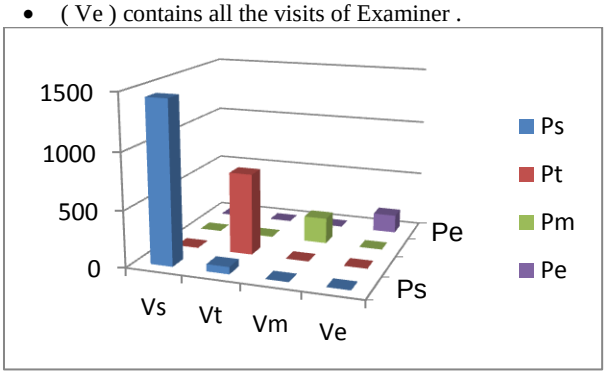


Figure 5 : Page Groups and Visit Classes

The clustering analysis in Figure 5 is shows that the visits associated with the pages related to the user type. So, these relationships between users are shown in Figure 6. We can note that by clustering the pages are divided by users .

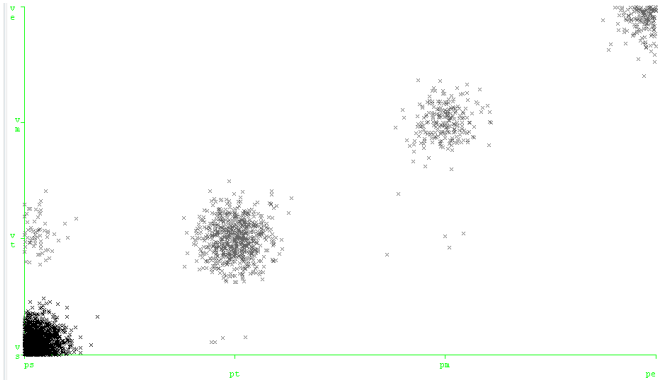


Figure 6. The clustering analysis of visits and pages

- **Sequential Pattern Extraction Methods**

Sequential pattern mining deals with data represented as sequences. So, after the preprocessing step, the log file of studied Web site contains 167 URLs and 1678 visit sequences. These sequences have an average length of 10 itemsets (requested URLs). The extracted sequences reflect the frequent behaviors of users connected to the site. They have been obtained on the whole log file with minimum supports. We details in this section a few sequential patterns extracted from the log file:

Pattern 1:

http://andalusuniv.net/engit.php

→ http://andalusuniv.net/events.php

→http://andalusuniv.net/objectives.php

Minimum support: 0.1 (263 instances) and

Minimum metric <confidence>: 0.9

Pattern2 :

http://andalusuniv.net/dlc/moodle/

→http://andalusuniv.net/dlc/moodle/meo

→http://andalusuniv.net/dlc/moodle/diode

Minimum support: 0.1 (261 instances) and

Minimum metric <confidence>: 0.9

The ontology stage

The explanation of usage analysis results derived from clustering and sequential pattern mining allow us to make suggestions in order to update the ontology.

Our vision is generating and updating concern ontology extension which does not fully modify the initial ontology:

- Addition of a leaf concept in a hierarchy,
- Addition of a sub-tree of concepts in the hierarchy,
- Addition of a relation between two concepts.

We used an existing ontology and modified it according to the structure of the studied Web site. . This ontology, edited with Protégé 2000 [38], is represented in Figure 7 .

- **Addition of a leaf concept in a hierarchy**

We observed that relationships between pages and users .So, we suggested to add a concept derived from usage analysis by gathering several existing concepts. The concept of specialties does not exist in our domain ontology and could be added.

- **Addition of a relation between two concepts.**

The sequential pattern mining suggests that a relation exists between users and resources due to the extraction of various patterns linking a page of the users and resources. This relationship extracted from usage analysis does not exist as a hyperlink on the Web site.

THE ONTOLOGY DESCRIPTION

The e-learning web site in this study is fairly simple in terms of data diversity. It only learning in this site. So, we did not need a complex ontology structure. The "e-learning ontology" is used in this study. We used the Protégé that utilized as the ontology creating tool. It uses OWL as the ontology description language.

Figure 7. The Domain Ontology

7. CONCLUSION

We have proposed a web usage mining approach for generating the ontology . This approach is consisting of three stages beginning from using an existing or semi-automatically built ontology intended to enhance information retrieval. After that, using Web Usage Mining methods like classification and sequential pattern mining techniques to analysis the log files to extract pattern. Finally generating Web usage ontology and / or Updating the existing ontology using the extracted knowledge.

8. FUTURE WORK

The objective of our research was to accelerate and to improve the Ontology development process by semi-automatically generating a hierarchal ontology. This work can be expanded to include many topics such as :

- Extracting recommendation rules of our framework .
- Partner to build the semantic web from the generated ontology
- Convert this framework to be incremental framework that is useful to benefit of previous ontology .

REFERENCES

[1] Abraham A. , " Business Intelligence from Web Usage Mining" . Journal of Information & Knowledge Management, c Publishing Co. Vol. 2, No. 4 (2003) 375-390.

[2] Yilmaz H.," Using Ontology Based Web Usage Mining And Object Clustering For Recommendation" .school of

natural and applied sciences of middle east technical university. (May 2010).

[3] Liu B. , "Web Data Mining Exploring Hyperlinks, Contents, and Usage Data" Springer-Verlag Berlin Heidelberg (2007)

[4] Trousse B., Aufaure M., Grand B., Lechevallier Y. , Masseglia F. , " Web Usage Mining for Ontology Management ", IGI Global , (2008).

[5] Arayaa S., Silvab M., Weber R. , " A methodology for web usage mining and its application to target group identification" Fuzzy Sets and Systems 148 (2004) 139–152.

[6] Berendt B., Mobasher B. , & Spiliopoulou M. , "Web Usage Mining for E-Business Applications", ECML/PKDD-2002 Tutorial, (2002).

[7] Eirinaki M. , Vazirgiannis M., "Web Mining for Web Personalization", ACM, http://doi.acm.org/10.1145/643477.643478 (2003) .

[8] Lim E. and Sun A. , "Web Mining - The Ontology Approach". Research Collection School of Information Systems. Paper 900 http://ink.library.smu.edu.sg/sis_research/900 . (2005)

[9] Hu Xiaohua, Cercone N., "A Data Warehouse/Online Analytic Processing Framework for Web Usage Mining and Business Intelligence Reporting " ,

[10] Han J. & Kamber M. "Data Mining: Concepts and Techniques", 2nd Edition, University of Illinois at Urbana-Champaign. Harcourt India Private Limited. (2001) .

[11] Ontology - From Wikipedia, the free encyclopedia http://en.wikipedia.org/wiki/Ontology.

[12] protege is a free, open-source ontology editor http://protege.stanford.edu/

[13] OWL Web Ontology Language , Overview ,W3C Recommendation http://www.w3.org/TR/owl-features/

[14] Davulcu, H., S. Vadrevu, S., & Nagarajan, S. ,"OntoMiner: Bootstrapping and Populating Ontologies from Domain Specific Websites" . Proceedings of the First International Workshop on Semantic Web and Databases (SWDB 2003), Berlin. (2003).

[15] Agirre, E., Ansa, O., Hovy, E., & Martinez, D. "Enriching very large ontologies using the WWW" . In: Proceedings of ECAI Workshop on Ontology Learning. (2000).

[16] Faatz, A., & Steinmetz, R. "Ontology enrichment with texts from the WWW" . Semantic Web Mining 2nd Workshop at ECML/PKDD-2002, Helsinki, Finland. (2002).

[17] Navigli, R., & Velardi, P. "Learning Domain Ontologies from Document Warehouses and Dedicated Web Sites". Computational Linguistics, 30(2), MIT press, pp. 151-179. (2004).

[18] Karoui, L., Aufaure, M.-A., & Bennacer, N., "Ontology Discovery from Web Pages: Application to Tourism" . Workshop on Knowledge Discovery and Ontologies (KDO), collocated with ECML/PKDD, Pisa, Italy, Sept., pp. 115-120. (2004).

[19] Ben Mustapha, N., Aufaure, M.-A., & Baazhaoui-Zghal, H. "Towards and Architecture of Ontological Components for the Semantic Web" . Proceedings of Wism (Web Information Systems Modeling) Workshop, CAiSE 2006, Luxembourg, pp. 22-35. (2006).