Semantic Web Mining in E-Commerce Websites

Zulqurnan Aslam

Lahore, Pakistan

ABSTRACT

Semantic web is gaining interest of lot of people due to its great power. Semantic web makes the data understandable for machines. It enables machines to interpret the data published by the web in machine interpretable form. Though lots of businesses are moving towards Ecommerce websites, but there are few problems with the current ecommerce systems due to the lack of proper standards. Daily thousands of people search for the product they want to buy and due to the inefficiency of the systems it wastes lot of time and resource of the users. Semantic web has that power to overcome those problems and it can accelerate the business to further higher level where ecommerce websites are playing very important role. This paper gives the introduction of semantic web mining and its techniques and then problems of the current ecommerce systems and how those problems can be avoided using semantic web. It discusses the detailed semantic web ecommerce architecture.

Keywords

Data Mining, Web Mining, Semantic Web, Semantic Web Mining, E-Commerce

1. INTRODUCTION

E-Commerce Websites are always handy for your business. As the world is growing, more and more companies are moving towards ecommerce websites. In ecommerce website we can sell, buy different kind of things without any tightness of physical visit. The main advantage of ecommerce sites is that you can open online store with ease without any physical cost of store location. [1] It is accessible all over the world all the time. Extracting useful information from the web is called web mining. We are drowning with data, so we need some kind of mechanism to extract the useful patterns that can help us in improving our websites. Conventional search methods fail here and semantic web comes here to overcome those problems. Semantic web is the tool that automatically finds data from websites [2]. Semantic Web Mining is an integration of two important scientific areas: Semantic Web and Data Mining [3]. This paper will give you the introduction of Web mining and Semantic web mining and how it can be used in ecommerce websits by discussing the problems in the available ecommerce websites and then it will prove how we can avoid those problems using this new architecture.

2. WEB MINING OVERVIEW

Extracting useful information from the website which is previously unknown is called web mining. Web mining is great research field and lot of research has been done and still different research is going on. Web mining can be divided into three main parts.

(1) Web Content Mining
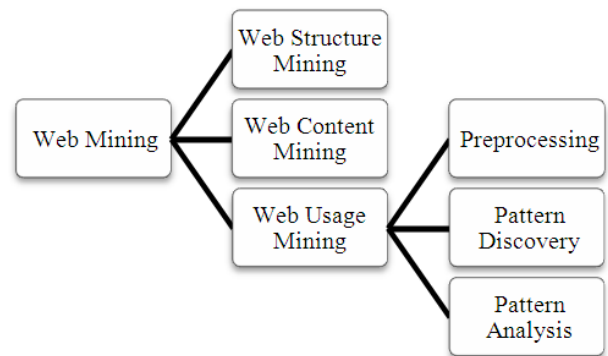(2) Web Structure Mining
(3) Web Log Mining/Web usage Mining

Fig. 1. Web Mining Overview [4]

2.1 Web Content Mining

Mining the content of the web is called web content mining. Images, texts, multimedia etc. are called the content of websites. So when we play with these things it is called content mining. HTML document can be organized and may be unorganized. So it is very easy to extract information when data is organized but it becomes very tough when data is unorganized. Due to the huge amount of data, we cant do this manually because doing manually will require huge amount of time and resources.

2.2 Web Structure Mining

Extracting information from the hyperlinks is called structure mining. Web structure mining is the process of using graph theory to analyze the node and connection structure of a website [1]. It can be divided into two main types

(1) Extracting from Hyperlinks: Extracting information from the navigational links.
(2) Extracting from document Structure: Analyzing the structure tree of the page.

The goal of Web Structure mining is to generate the structural summary of the website and webpages [1].

2.3   Web Log Mining/Web Usage Mining

Extracting useful information from the logs of the servers. This is little challenging task because server logs are not easily available. We try to extract useful patterns from the server log like we can track the user traffic and then we can also track the destination of that traffic. What is the main peak load time of server? People are coming from which part and what kind of information they are searching. So by analyzing these patterns we can suggest improvements in the websites.

3.   SEMANTIC WEB MINING

Semantic web is the extension of the standard web where semantic web makes the content understandable by the machines. Semantic web gives the power to the system to search the product which is most relevant to the users requirement. It is the future of web in few years because handling lot of data by human is not possible. Using semantic web we can overcome issues like search and other issues which we face in traditional dynamic websites [2][5][6] .

3.1   Semantic Web Stack

This Semantic stack contains the following elements as shown in figure 2, and this paper will discuss few of them which are important here.
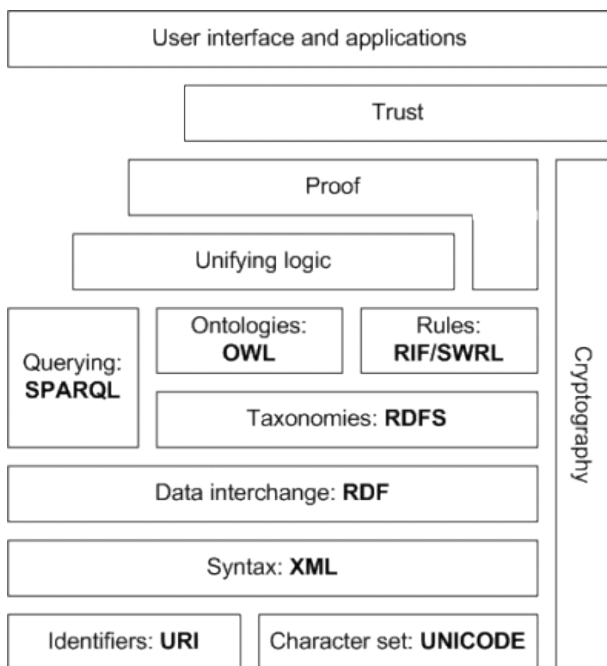

Fig. 2.   Semantic Web Stack

1- Bottom layer is **Unique Resource Identifiers (URI)** These are used for the identification of the Resource.

2- **Extensible Markup Language (XML):** It is used for the description of different type of data. Human and machine can understand this and it is the basis for other language like RDF, OWL etc.

3- **Resource Description Framework (RDF):** Foundation for processing data. It has three statements, Subject, Predicate and Object. Resource L has property M with value of O. The Color of the Car is Gray Subject is Car, Predicate is Color and Object is Gray. Similarly River has length of 600KM and it is situated near china sea.
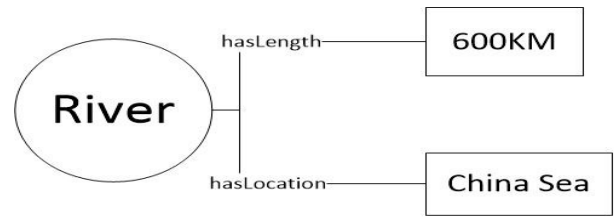

Fig. 3.

Fig. 4.   RDF Graphs

4- **Resource Description Framework Schema (RDFS):** Used for defining the vocabularies, classes, properties and restrictions.

5- **Web Ontology Languages (OWL):** Used to make the ontologies. It is W3C recommendation. It consists of XML and RDF. RDF has some limitations but OWL covers those limitations.

6- **SPARQL:** It is used to extract the information from ontologies. Like we use SQL to extract data from Database.
Semantics add the meaning to the data and web mining use to filter the data. Combination of both these techniques are called semantic web mining. The explained techniques regarding semantic web mining above have some problems. Like searching the data, when we search some data we do not get exact results. We get lot of irrelevant results as well. So semantic web enables us to remove that gap [1][2][5][6].

3.2   Pattern discovery methods for E-Commerce websites

1- **Clustering:** In this we make the groups of same and similar objects. Objects those are same, put in one cluster and different, in other cluster. So we make clusters of data.

2- **Filtering:** Filtering is the technique to filter the relevant data. This is relatively tough task because there is lot of noise in the data and we have to remove that noise to get relevant results.

3- **Path Analysis:** We analyze the path of the user while visiting the websites. Like what path user took to reach particular page.

4- **Association Rules:** We try to find the association patterns, like if A->B and B->C then A->C.

4. PROBLEMS WITH CURRENT E-COMMERCE SYSTEMS

4.1 Information Symmetry Problem:

This is the situation where different websites show different price for the same brand. Like this research took two very famous ecommerce website of Pakistan, one was Daraz.pk and other was kaymu.pk. It searched for the same perfume on both website and got different prices. **Hugo Boss - Boss Bottled** was searched on both sites and there were price differences. Now assume that if a person is new, he/she come to the website and search this perfume and order it and get it after paying high price. Now if an experienced educated user comes and search then he will search intelligently and then he will find the product with lesser cost. So now whats wrong with user 1 that he paid higher cost than the second one? Because these websites are not following same standards. So there should be some standards where everyone get same piece of information like price [5] [11].

4.2 Search Problem:

This is the situation when we search for a product and did not get proper relevant results because product generic attributes like price, color, size etc. are not considered and hence user did not get proper results. Now take an example, a consumer is searching on Google for: " 'mobile phone' with camera and email option and then it should have the download ringtones feature". Now when user will search it he will get different results. Some sites will contain mobile phones with cameras and then there will be some sites who will list the phone that can send email etc. But no one will give mobile phone which contains all these features, which is the need of the user [5].

5. ARCHITECTURE OF SEMANTIC WEB E-COMMERCE

Below is the detailed architecture of the semantic e-commerce system in fig 5. Architecture for only two users has been showed here and tried to map the example discussed above. This architecture is extendable for N users. Process will remain same. User1 and User2 are the consumer who are using the website and doing shopping and searching for the product. Producer is the manufacture of the products; here in our case Hugo Boss is the company. User will login the system and then he/she will search the Hugo boss perfume. That query will be transfer to the Agents. Agents are there to help the system to reduce the workload [12]. In the architecture there are two types of agents. Search agents and ontology agents. Query transfer to the search agents and then it will be transferred to the JENA OWL API which is written in java. Then from that API SPARQL query will be generated to get the price from Product Owl Ontology and finally results will return to the users. Sample SPARQL query is written below:

```
Select ?price ?name where {
?x Product:Brand Hugo .
?x Product:Price ?price .
?x Product:Name ?name.
}
```

Producer companies will publish their product and service descriptions in standard way of ontologies. Like they will publish the price
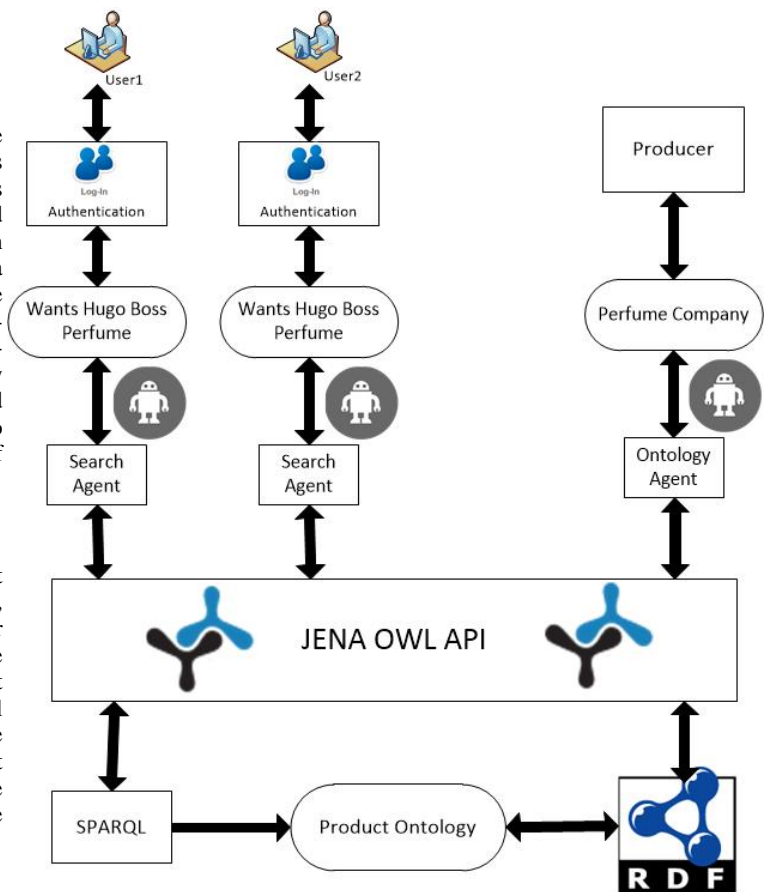

Fig. 5. Semantic Web E-Commerce Architecture

of the Hugo Perfume, so when different users will search for this perfume, the system will query from the standard ontology and then every user will get the same problem. Similarly while making ontologies of products its every aspect will be covered, like its features, color, size etc. so when user will search the product it will get more relevant results.

6. CONCLUSION AND FUTURE DIRECTIONS

World Wide Web (WWW) is growing rapidly that now every business is moving towards the e-commerce websites. So we are getting bulk of data that we are almost dying with this data. We need some kind of mechanism to make the best use from that data. So this paper discussed the Web Mining techniques that how it can be used to get useful information. Then it showed the problem in the current ecommerce system and then how semantic web can overcome those problems. This paper overviewed the detailed architecture of the semantic web that can help to reduce the problems. Now there is one main hurdle in achieving this goal, how we can achieve these standardized ontologies. We need some kind of mechanism where companies can make standardize ontologies then we can achieve this goal. Lot of search has been done and different researchers are doing research in this field of semantic. Semantic has that power that it will be the future of the next web generation. So in future this field can give twist towards the web personalization. Like we can extract useful patterns, behaviors of the users using web min-

ing techniques and then we can give the more personalized user interfaces to the users. This is the step toward the adaptive interfaces, which is very interesting field. Researchers have already been working on this area of adaptive interfaces.

7. ACKNOWLEDGEMENTS

8. REFERENCES

1- Zulqurnan Aslam Rohail Shezad SURVEY ON WEB MINING AND ITS USAGE IN E COMMERCE SITES - PATTERN DISCOVERY, ISSUES AND APPLICATIONS. International Journal of Engineering Applied Sciences and Technology, 2016 Vol. 1, Issue 2

2- Babu, K. Ganapathi, et al. "Web Mining using Semantic Data Mining Techniques." International Journal of Soft Computing and Engineering (IJSCE) ISSN (2012): 2231-2307.

3- O. Mustapaa, A. Karahoca, D. Karahoca and H. Uzun- boylu, Hello World, Web Mining for E-Learning, Pro- cedia Computer Science, Vol. 3, No. 2, 2011, pp. 1381- 1387.

4- Jafari, Maryam, Farzad SoleymaniSabzchi, and Shahram Jamali. "Extracting UsersNavigational Behavior from Web Log Data: a Survey." Journal of Computer Sciences and Applications 1.3 (2013): 39-45.

5- Heidari, Karim, and A. Dehghan. "The impact of semantic web on e-commerce." World Academy of Science, Engineering and Technology 51 (2009).

6- VijayaLakshmi, B., et al. "Perspectives of Semantic Web in e-commerce." International Journal of Computer Applications 25 (2011).

7- Extensible Markup Language (XML) https://www.w3.org/XML/

8- Rasheed M. Al-Zahrani, The Ontology problem in ecommerce ap p lications, Information Sy stemsDept., KSU.

9- SPARQL Query language for RDF http://www.w3.org/TR/rdfsparql- query/

10- Jena RDF API `http://jena.sourceforge.net/tutorial/RDF_API/index.html`

11- Tanya Gupta, Abir Qasem,Reduction of price dispersion through Semantic E-commerce: A Position Paper. SemanticWeb Workshop 2002 Hawaii, USA.

12- Oscar Martinez, Federico Botella, Building E-Commerce Web Applications: Agent- and Ontology-based Interface Adaptivity, Operations Research Center, University Miguel Hernndez of Elche, Avda.Universidad, s/n, 03202 Elche, Spain.

9. ABOUT AUTHOR

Zulqurnan Aslam is an IT professional and holds MS degree in Computer Science. He is Senior Software Engineer and have more than 5 years of development experience in ERP systems and Web Developments using different languages like JAVA,.NET,PHP. He is also EVS MS SQL SERVER 2008 Certified specialist. He is researcher by passion and his research interests are in HCI, Data Mining, Web Mining, Semantic Web, and Adaptive Interfaces.