© 2015 IJSRSET | Volume 1 | Issue 6 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Secure and Intelligent Decision Making in Semantic Web Mining using XML, XSLT and Xquery

Roshni Dubey, Nitesh Namdeo
Shri Ram Institute of Technology, Jabalpur, Madhya Pradesh, India

ABSTRACT

Web mining is the application of data mining technologies to automatically interact and discover information from web documents, which can be in structured, unstructured or semi- structured form. 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. We present an enterprise framework regarding semantic web mining in distance learning, which can be used to not only improve the quality of web mining results but also enhances the functions and services and the interoperability of long distance educational information systems and standards in the educational field. For on line distance education system we propose an Ontology-based approach to share online data and retrieve all relevant data about students and their courses. Thus semantic web ontology help build better web mining analysis in educational institute and web mining in-turns helps contract basis more powerful ontology in distance learning. Since the majority of the online data considered as private data we need various mechanism for privacy preservation and control over the online presence data. We propose privacy protection in semantic web mining using role back access control.

Keywords: Ontology, Data Mining, Web Mining, HTML, XML, World Wide Web

I. INTRODUCTION

Briefly stated, data mining refers to extracting or "mining" knowledge from large amounts of data. Data mining can be performed on a variety of data stores, including relational databases, transactional databases and data warehouses. A comprehensive data mining system usually provides multiple mining functions. Association is one of the key features that can be found in such systems.

Nowadays, the World Wide Web has becoming one of the most comprehensive information resources. Due to the increasing amount of data available online, the World Wide Web has becoming one of the most valuable resources for information retrievals and knowledge discoveries. Web mining technologies are the right solutions for knowledge discovery on the Web. The knowledge extracted from the Web can be used to raise the performances for Web information retrievals, question answering, and Web based data warehousing. In the short span of its existence, the World Wide Web has resulted in a revolution in the way information is transferred between computer applications. It is no longer necessary for humans to set up channels for inter-application information transfer; this is handled by TCP/IP and related protocols. To almost any topic one can think of, one can find pieces of information that are made available by other internet citizens, ranging from individual users that post an inventory of their record collection, to major companies that do business over the Web. It is also no longer necessary for humans to define the syntax and build parsers used for each kind of information transfer; this is handled by HTML, XML and related standards. However, it is still not possible for applications to interoperate with other applications without some pre-existing, human created and outside-of-the-web agreements as to the meaning of the information being transferred.

The goal of data mining is to extract or mine knowledge from large amounts of data. Data mining involves the

use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. These tools can include statistical models, mathematical algorithms, and machine learning methods (algorithms that improve their performance automatically through experience, such as neural networks or decision trees). Consequently, data mining consists of more than collecting and managing data, it also includes analysis and prediction. Data mining can be performed on data represented in quantitative, textual, or multimedia forms.

II. METHODS AND MATERIAL

RELATED WORK

Semantic Web Mining aims at combining the two fast-developing research areas semantic Web and Web Mining. Though the Web is rich with information, gathering and making sense of this data is difficult because the documents of the Web are largely unorganized. In addition the Web is growing at an exponential rate as Web information is published on a daily base throughout the world, and the Web is regarded as the most popular information source and distribution medium available at present. It is plausible to state that, since its inception, the Internet and the associated Web technologies revolutionized the way in which our society generates, disseminates, accesses and uses information The Internet was originally designed as an information space usable by Both humans and machines [8]. However, the current Web is a publishing mechanism for information targeted mainly at human users. In 2001, Berners-Lee, Hendler and Lassila [9] presented a vision of a Web that is an information space usable by, in particular, machines, thus coining the term Semantic Web for this envisioned Web . Instead of attempting to process and manipulate Web information, a user would have a personal software agent on his/her computer that would solve problems related to information overload, acquisition and discrepancy resolution [10].

The agent would execute the first level of information management and the user would only access or manipulate results. The Semantic Web is at the time of writing mainly an international research effort with the goal to make Web content available for intelligent knowledge processing. The concept of the Semantic Web captured the imagination of Web users, as well as the interest of academia and industry even though several technological issues still have to be resolved.

The Semantic Web is the name used to encapsulate the 2001 vision presented by Berners-Lee et al. [11] of a new Web as an information space usable by machines rather than humans. Web researchers realized that the rapid adoption of the Web and the associated information overload would necessitate alternative solutions and technologies where autonomous programs or machines assist humans to manage the information available on the Internet. Several definitions and descriptions of the Semantic Web were published since its inception as practitioners and researchers adopted the notion of a Web semantically enriched [10]. Some of these definitions are provided below in order to highlight the diversity thereof. The list is by no means exhaustive. A Web enriched with semantic meta-data that enables agents to exe- cute complex information management tasks on behalf of its users a mechanism that contributes towards data, information and knowledge exchange and integration across communities and applications; a comprehensive architecture of meta-data language functionality that can be instantiated with different technology standards and specifications. The biggest challenge in the next several decades is how to effectively and efficiently dig out a machine-understandable and queriable information and knowledge layer, called Semantic Web, from unorganized, human-readable Web data.

The Semantic Web [1] is designed to let users make explicit statements about any resource, and maintain that data themselves in an open and distributed manner. Several standards such as the Resource Description Framework (RDF) [3] and Web Ontology Language (OWL) [4] have been developed to realize the layer cake of the Semantic Web. World Wide Web has resulted in a revolution in the way information is transferred between computer applications. A common underpinning is especially important for the Semantic Web as it is envisioned to contain several languages, as in Tim Berners-Lee's "layer cake" diagram first presented at XML 2000. The diagram depicts a Semantic Web Architecture in which languages of increasing power are layered one on top of the other. Unfortunately, the relationships between adjacent layers are not specified, either with respect to syntax or semantics.

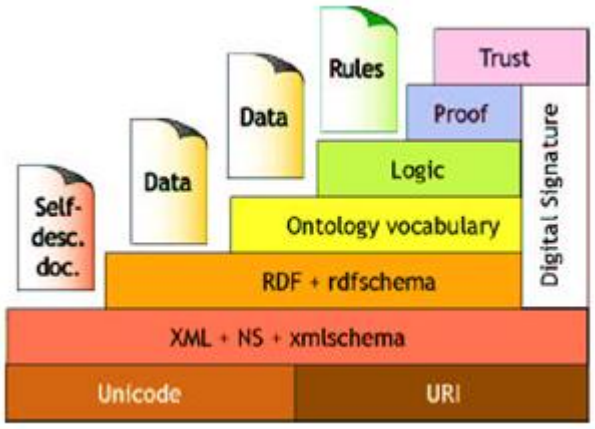


Figure 1: Semantic Web Mining Architecture

III. RESULTS AND DISCUSSION

PROPOSED WORK AND RESULTS

Proposed Algorithm

Input :- Data set in HTML or XML form and Attributes set
Output:-Data set in XML form
Step:-1 Start
Step:-2 Take input Data set in HTML form
Step:-3 While Data set is available doCreate a node root
If related root node data is found then create a new node in root path
Else
New node=root
End if
Replace the node with new node
End loop
Step:-4 Repeat for root to deepest node
If node data is not matched then
Drop the node
End if
End loop
Step:-5 Union all the nodes from root to deepest node
Step:-6 Display the Data set in XML form
Step:-7 Stop
And flow chart as shown below

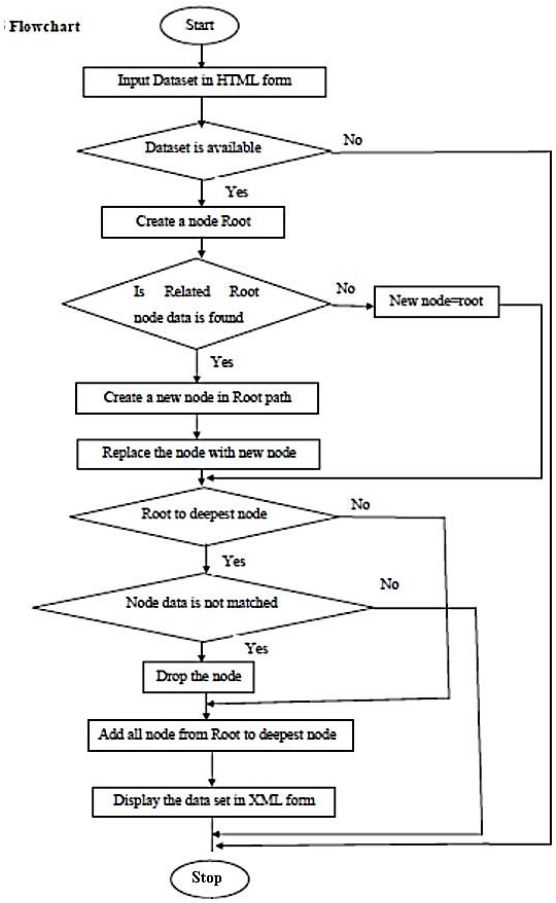
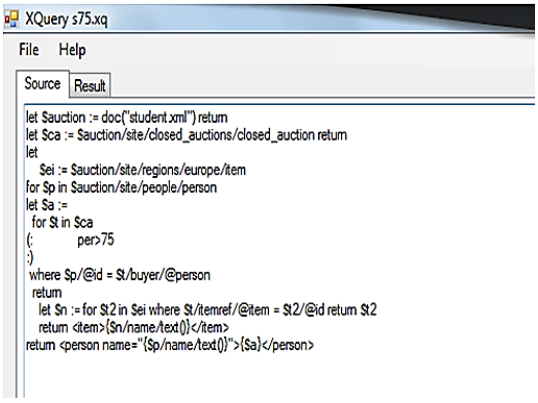


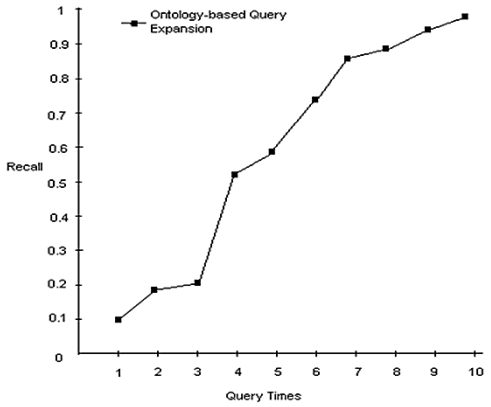
Figure 2 : System Flow Chart

A prototype system for simulation is constructed based on the framework of semantic web mining and dot net, protégé technologies. Data from HTML pages are gathered and input to the program. The output is given in table below.



User input the query words in ontology as expansion words and its performance can be showed through precision and recall ratios that are calculated from experimental results. Through 10 times different

information requests, we compute recall and precision ratios. The experimental results are as follows.



Graph 1 : Recall ratio of query methods

IV. CONCLUSION

In this paper focus on Data mining technology has been playing an important role in education related field for many years. Web mining is the use of data mining techniques to automatically discover and extract information from the web documents which can be structured, unstructured or semi structured from. XML has become very popular for representing semi structured data and a standard for data exchange over the web. The data based on XML is self-described; it can be exchanged and handled without internal description. The core technique of semantic web mining is ontology. Ontology represents a set of precisely defined terms about a specific domain and accepted by this domain's community, ontology is an explicitly specification of a conceptualization. The RDF is a simple Meta model for defining and exchanging information on the semantic web. We have developed semantic web mining for an educational domain. For this we have developed an enterprise web framework that uses semantic web mining, Resource Description Framework, Ontology and XML technology. The system helps to find suitable semantic data related to students, faculties and courses for the clients.

V. REFERENCES

[1] Ian Horrocks, Peter F. Patel Schneider, Three Theses of Representation in the Semantic Web, ACM, 2003

[2] A new document clustering algorithm based on association rule, Jiang-chun song, Jun-Yi shen, Qing-Bao song, Proceedmgs of the Thud Internauonal Conference on Macbme Learmng and Cybemeucs, Shangha, 26-29 August 2004

[3] Cheng Wang, Ying Liu, Lihang Jain, Peng Zhang"A Utility-based Web Content Sensitivity Mining Approach", IEEE, International Conference on Web Intelligent Agent Technology, 2008.

[4] Andre Vizine, Landro N. De Castro, Richord Gudwin, An Evolutionary Algorithm to Optimize Web Document Retrieval", IEEE, KISMAS, 2005

[5] Jung-Won Lee, Kiho Lee, Won kim, Preparations for semantics-Based Mining, IEEE 2001

[6] Pance Panov, Larisa N. Soldatova, OntoDM: An Ontology of Data Mining,IEEE International Conference on Data Mining Workshops, 2008

[7] Wang Jain, Li Zhuo, Research and Realization of Long Distance Education platform based on Web Mining, IEEE 2009

[8] Association Rules Mining: A Recent Overview, Sotiris Kotsiantis, Dimitris Kanellopoulos GESTS International Transactions on Computer Science and Engineering, Vol.32 (1), 2006, pp. 71-82

[9] Eficient Algorithm for Hierarchical Online Mining of Association Rules Kishore B Kumar1 and Naresh Jotwani2 International Conference on Management of Data COMAD 2006, New Delhi, India, December 14-16, 2006 °c Computer Society of India, 2006

[10] Kiavash Bahereini and Atilla Elci ,A New Software Architecture for J2EE enterprise environments via semantic access to web mining by distributed intelligent software agents, IEEE international software and Applications conference, 2008

[11] Zhang Yofeng, Wu Jinhong, Wang Cuibo, Automatic Competitive Intelligence Collection Based On Semantic web mining IEEE 2007

[12] Jiang-Chun Song, Jun-Yi-Shen, Qing-Bao Song, " A New Document Clustering Algorithm Base On Association Rule ", IEEE, Third International Conference on Machine Learning and Cybernetics, 2004

[13] An Immune-based Model for Web Data Mining Wang Feng Xuwei Li Zhu Hong (© 2005 IEEE Xplore

[14] A Classifier-CMAC Neural Network Model for Web Mining Somaiyeh Dehghan Amir Masoud Rahmani 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology

[15] Kiavash Bahreini and Atilla Elici,SDISSASA: A multiagent-Based web mining via semantic access to Web resources in Enterprise Architecture,IEEE 2008.