Journal of Computational Intelligence in Bioinformatics
ISSN 0973-385X Volume 10, Number 1 (2017) pp. 15-22

http://www.rfgindia.com

Adaptive and Personalized System for Semantic Web Mining

Sunny Sharma[1], Gurvinder Singh[2] and Vijay Rana[3]

[1]Research Scholar, Department of Computer Science & Engineering, Arni University Kathgarh, Indora, HP, India.

[2]Research Scholar, Punjab Technical University, Punjab, India.

[3]Department of Computer Science & Engineering, Arni University Kathgarh, Indora, HP, India.

Abstract

Semantic Web Mining is the combination of two fast-developing research areas Semantic Web and Web Mining. It has been extensively researched by the researcher that the results of Web Mining can be improved by exploiting semantic structures in the Web, and they make use of Web Mining techniques for building the Semantic Web. Furthermore, these results can be reranked for much accuracy by analyzing the user's navigational behavior. Web personalization is the process of customizing a Web site to the needs of specific users, taking advantage of the knowledge acquired from the analysis of the user's navigational behavior. In this paper, we examine user profiles to identify the requirement of web user in order to improve the semantic web search.

1. INTRODUCTION

Research on intelligent system can be traced back to the late 1950s, and a significant number of retrieval techniques have been invented during the period. Specially, intense evolution in this research field has occurred along with the birth and

proliferation of the World Wide Web. Today, search engines built on various Information Retrieval techniques have thoroughly changed the ways that people search and acquire information. Nevertheless, in many situations search engines have difficulties in retrieving relevant and quality information despite the fact that they make strenuous efforts to expand indices and rectify ranking functions. The challenge is amplified by the large scale and the continuous growth of the Web.

In the last few years, a lot of attention has been devoted to improving web search and recommender systems through web mining. It allows sifting through large quantity of data for useful information. Web mining [2] gives an innovative direction for scientific research and pushing web technology to making the meaningful information and exploits some data mining techniques[3] to automatically mine valuable information from the World Wide Web. It makes an environment where the information available on the web can be semantically interpreted. It assembles more feature to built web personalize interaction and customizing a web site according to the requirements of users, obtaining advantage of the knowledge attained from the study of the user's browsing behavior.

The Semantic Web mining [6] adds structure to the Web, while Web Mining can learn implicit structures. This is an interesting way for Semantic Web Mining to create itself as the dependence between the Semantic Web and Web Mining increases. The outcome to this combination benefits many areas of industry such as "e-activities", health care, privacy and security, and knowledge management and information retrieval. Our research work is related to the field of Semantic Web Mining. In this work, we analyze data stored in web search engines' logs to discover usage patterns, and the aim is to enhance performance of search tools as well as to help users to find information on the web.

2. THE STATE OF ART: WEB MINING& SEMANTIC WEB

Web mining exploits the data mining [3] techniques to automatically discover and retrieve desire information from the web and gave agreeable outcome to users. It uses three basic techniques content, structure and usage mining to extract meaningful information. Content mining is a superlative tool for retrieving meaningful information from the content of web resources. Structure mining is the process of extracting knowledge from the interrelated hypertext document on the web and usage mining an imperative technique of automatically searching and analysis the user interaction patterns with web servers.

Web Content Mining

Web content mining is the mining, removal and combination of useful data,

information and knowledge from Web page content. It describes the detection of constructive information from the web credentials. In web content mining the content may be manuscript, picture, aural, video, metadata and hyperlinks etc. Web content mining also distinguishes individual home pages with other web pages. Research in web page mining encompasses resource discovery from the net, document categorization and agglomeration, and knowledge extraction from websites.

Web Structure Mining

Web structure mining [5] is the process of discovering structure information from the web. Structured information is the graphical representation of the pages. A classic web graph consists of web pages as nodes, and hyperlinks as edges relating related pages. This can be further divided into two kinds: Hyperlinks and Document Structure. A hyperlink is a structural unit that connects a site in a web page to a different site, either within the same web page or on a different web page. In document structure technique the content within a Web page [14] can also be organized in a tree structured format, based on the various HTML and XML tags within the page. Mining efforts here have focused on automatically extracting document object model (DOM) structures out of documents.

Web Usage Mining

Web usage mining [4] is the involuntary discovery of user access outlines from network servers. Organizations assemble large volume of data in their daily operations, generated automatically by network servers and collected in server access logs. Further sources of user data contains referrer logs that contain data concerning the referring pages for every page reference, and user registration or examine knowledge gathered via CGI (common entranceway interface) scripts. Analyzing such data can help organizations to determine the life time value of customers, cross marketing strategies across products, and effectiveness of promotional campaigns, among other things. It can also deliver information on how to reorganize a Web site to generate a more actual organizational presence, and shed light on more effective management of workgroup communication and organizational infrastructure. Promoting advertisements on the World Wide Web, analyzing user access patterns helps in targeting ads to specific groups of users.

Semantic Web

The Semantic Web [1] (SW) intentions towards revolution of information oriented web into knowledge concerned with web. The semantic web is an apparition which mines information from the web and crafts it possible to ease machines to identify

intricate human queries. It brings the suggestion of structuring information accessible through the web in a substantial way enhancing search technique and thus resulting user satisfaction.

The idea of SW was coined by Tim-Berner Lee in 'The Scientific Americans' in 2001. The article described the development of a web that comprised largely of documents containing data and information for computers to operate. Berners-Lee proposed four versions of Semantic Web architecture.

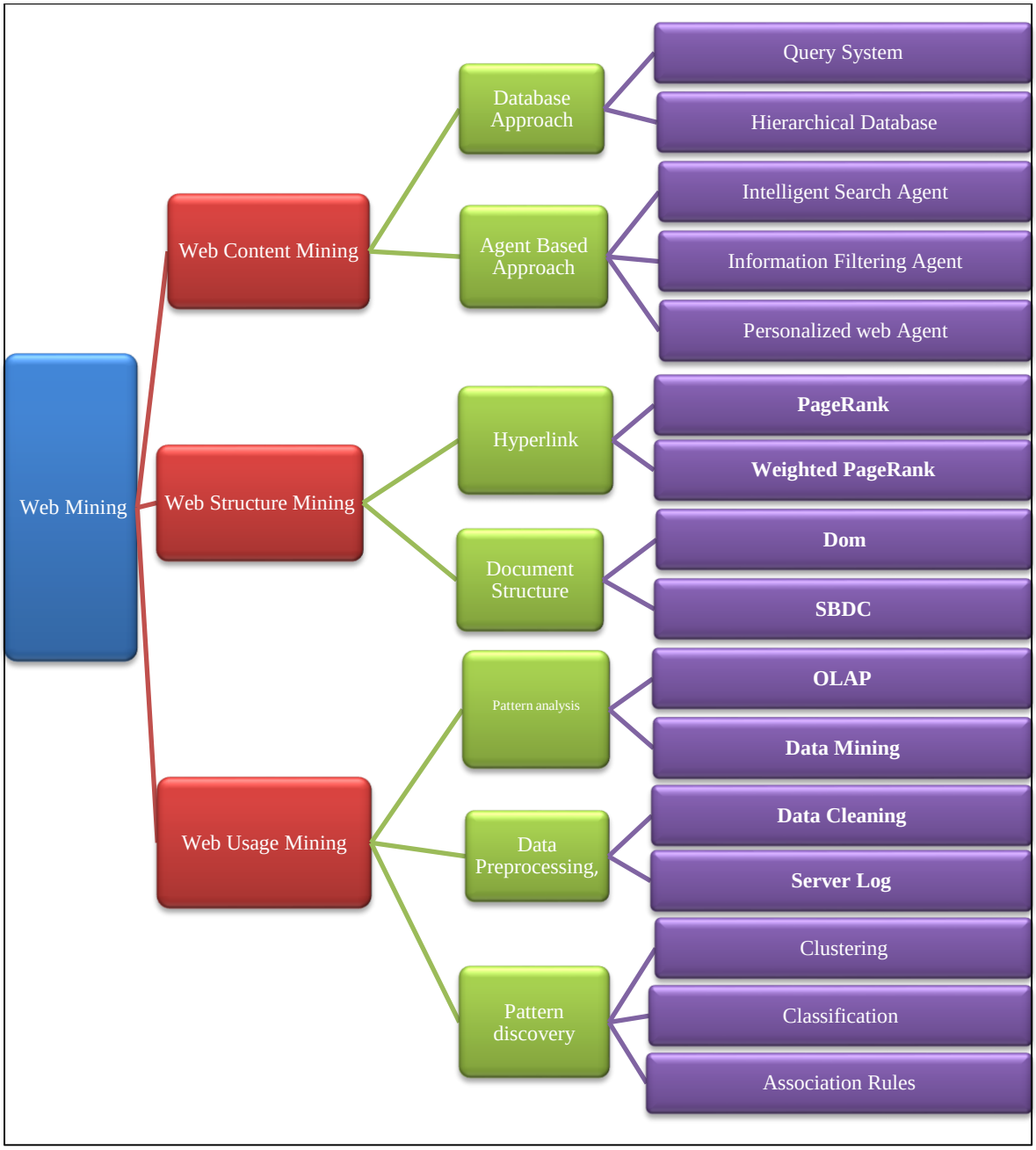

Figure 1: Classification of Web Mining

3. PROBLEM DEFINITION

A major problem on the Semantic Web Mining is inability to predict the user web surfing behavior appropriately on web and gain new knowledge through this interaction. Recently the web mining communities have focused on classifying standards that evaluating user browsing behaviors on e-commerce websites but failed to enhance user's satisfaction towards ambiguous queries from different perspectives. Ambiguous problem arise whenever two queries share the similar kind of information and numerous intended meanings are related with the same word, which leads by the semantic similarity problem of ontology context. The semantic similarity problem arises when those contexts are concepts are parallel in meaning but have dissimilar names (Truck and Lorry). If we can automatically identify the semantic similarity between ontologies contexts, it is probable to apply appropriate tools to handle particular kind of queries, instead of for all.

4. RESEARCH METHODOLOGIES:

The basic approach of this research study is to develop a standard consistent Semantic Web Mining Interface for predicting the user browsing behaviors on web and to reduce the redundancy of accessed data. Our objective is to incorporate Web mining and semantic web in order to enhance the effectiveness of web personalization by giving a solution that efficiently combines techniques used in user profiling. The view of the proposed System for the computation of user's behavior and for the reduction of the redundancy of accessed data is recommended below in figure 2.

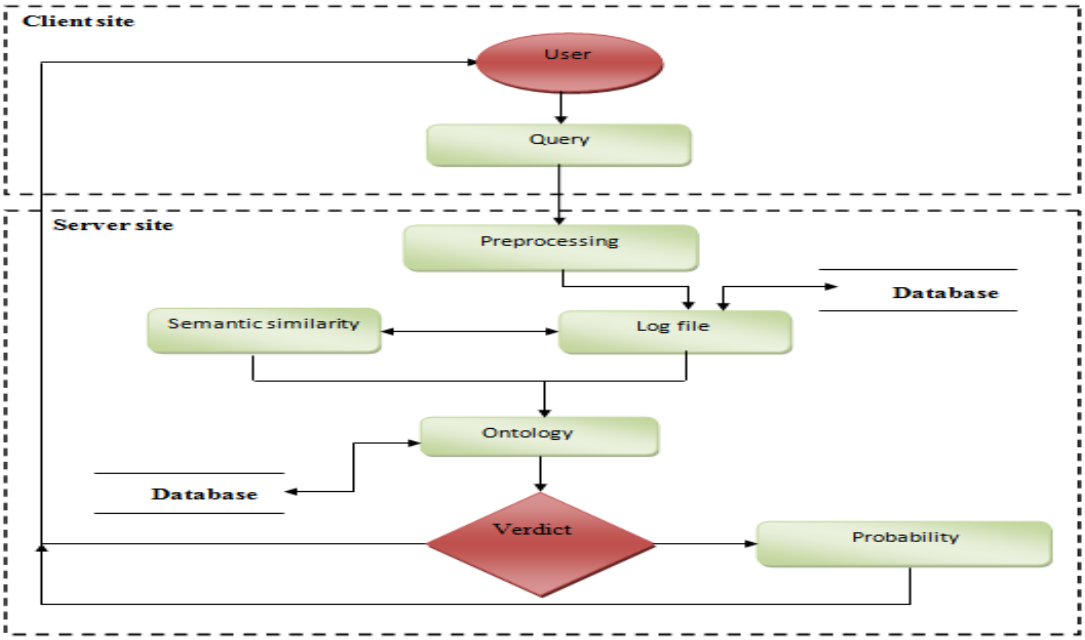

Figure 2: View of Research Methodology

Behavior Analysis through Semantic Annotation (BATSA)

BATSA framework comprises of four modules namely query processing module (QPM), data analysis module (DAM), context similarity and domain analysis module (CSADAM) and Query match module (QMM).

Query Preprocessing Module (QPM): In this set, a single or set of key words is given as an input to QPM, which describes the user information needs. In contrast to existing search engines that retrieves only the results on the basis of the probable search while ignoring the semantics of the user requirements, QPM uniquely contributes a different and novel algorithm that focus on finding the relevant meaning that describes the user's desires. The algorithm tries to find the user behavior and prevents from the repetition of accessed data. Thus QPM is an intelligent module as it improves the probability of success by finding the appropriate results. It performs several tasks to achieve the preprocessing phase: Tokenization and part of speech.

Tokenization is the process of splitting up a query string into a set of tokens or words. It usually splits words by blank, punctuation and quotation marks at both sides of a sentence. The tokens not only considered as words but also numbers, punctuation marks, parentheses and quotation marks. Parsing is the process of analyzing a string of symbols, either in natural language or in computer languages, conforming to the rules of a formal grammar.

Data Analysis Module (DAM): The second module of (BATSA) is DAM, which contains the Log Files, cookies and session identification. A log file is a file that records either events that occur in an operating system or other software runs, or messages between different users of communication software. Logging is the act of keeping a log. For Web probing, a transaction log is an electronic record of interfaces that have happened during a searching between a Web search engine and users searching for information on that Web search engine. More recent entries are typically attached to the end of the file. Evidence about the request, including client IP address, request date/time, page requested, HTTP code, bytes served, user agent, and referrer are typically added. This data can be combined into a single file, or separated into distinct logs, such as an access log, error log, or referrer log.

Context Similarity and Ontology Matching Module (CSOMM): The data after preprocessing is transferred to the log files where the IP address of the user is collected and the data accessed by the user is determined from the database. The match between the users input query and the accessed query by the semantic

similarity using the Information Content(IC) factor and the W-path algorithms. If the similarity found between the input and the cookies is similar then the user receives the result directly from the previous search. There is no need of accessing again and again the same inputs. The data of the cookies is stored in the database of the log files. So fetching the data from the database if the probability of the similar data is high then the desired result is given to the user, taking the data from the log files database.

Entity Matching Module (EMM): If the probability of the similar data is low then the query is transferred to existing probability and PageRank Algorithms. The basic approach of this research study is to develop a standard consistent ontological Interface for the semantic web

CONCLUSION

Web personalization is the process of personalize a Web site to the specific and individual needs of each user, without requiring them to ask for it explicitly. In this article we present a survey of Web mining and Semantic Web. Understanding the use needs is one of the key factors of an online project. If these needs are quickly identified, the customer can be offered the best products immediately. More specifically, we introduce the modules that comprise a Semantic Web personalization system.

REFERENCES

[1] Berners T, Hendler J and Lassila O, (2001), "The Semantic Web", Scientific American: Feature Article, Vol.284, No5. pp. 1-4

[2] Vijay Rana, Singh G, (2014) "Analysis of Web Mining Technology and Their Impact on Semantic Web", DOI: 10.1109/CIPECH.2014.7019035, IEEE Xplore, pp 5-11

[3] Karakatsanis et.al [6] Data mining approach to monitoring the requirements of the job market: A case study (2017) www.elsevier.com/locate/is

[4] Wu, K., Yu, P. S., & Ballman, A. (1998). Speed- tracer: A web usage mining and analysis tool. IBM Systems Journal, Vol.37, No.1, pp

[5] Wenpu Xing and Ali Ghorbani, (2004) Weighted PageRank Algorithm: Proceedings of the Second Annual Conference on Communication Networks and Services Research (CNSR'04© 2004 IEEE

[6] Mathieu Aquin and Enrico Motta, (2011), Watson. More than a Semantic Web search engine, INRIA, 2011- ACM Digital Library, pp 55-63

[7] Sharma S, Rana V. Web Personalization through Semantic Annotation

System. Advances in Computational Sciences and Technology. 2017;10(6):1683-90.

[8] Mahajan, Sunita, Sunny Sharma, and Vijay Rana. "Design a Perception Based Semantics Model for Knowledge Extraction." International Journal of Computational Intelligence Research 13.6 (2017): 1547-1556.