A Study of Web Personalization Using Semantic Web Mining

[1]Mr. Kanani Hiren M, [2]Prof.Virani Sagar

[1]ME Scholar, [2]Assistant Professor
Computer Engineering Department, VVP Engineering College, Rajkot.
[1]Hiren.kanani010@gmail.com , [2]sagarvirani@gmail.com

Abstract - **Semantic Web Mining is combination of two areas. One is Semantic web and second is Web Mining. These two areas improve the result of www (World Wide Web). The Semantic Web can make Mining much easier and Web Mining can build new structure of Web.WebMining applies Data Mining technique on web content, Structure and Usage. Method of Usage Mining can profit from enriched description of web pages visited. This will provide better utilization of web pages, better utilization of web pages and recommendation and personalization of website. Web Personalization may include the provision of recommendation to the users, the creation of new index pages or generation of target advertisements using semantic web mining. This Paper presents overview of web personalization using semantic web mining.**

Keywords - **Semantic Web, Web Mining, Semantic Web Mining, Ontology**

I. INTRODUCTION

A. Semantic Web

The World Wide Web [1], has grown in the past few years from a small research community to the biggest and most popular way of communication and information dissemination. World Wide Web has large amount of Data which have unstructured. From Unstructured human readable web data semantic web is effective and efficiently creating a machine understandable data.WWW is useful for exchanging information, multimedia or educational content. Semantic web work only with data and information. Semantic web can build more efficient and effectively data which has been understandable by machine. Today it is almost impossible to retrieve information with a keyword search when the information is spread over several pages. The following steps show the direction where the Semantic Web is heading:

1. Providing a common syntax for machine understandable statements.
2. Establishing common vocabularies.
3. Agreeing on a logical language.
4. Using the language for exchanging proofs.

B. Web Mining

Web is a collection of hyperlinked documents on one or more Web servers[2]. Web mining is data mining techniques used to extract knowledge from Web mining is a helpful tool in the process of transforming human understandable content in to machine understandable semantics.
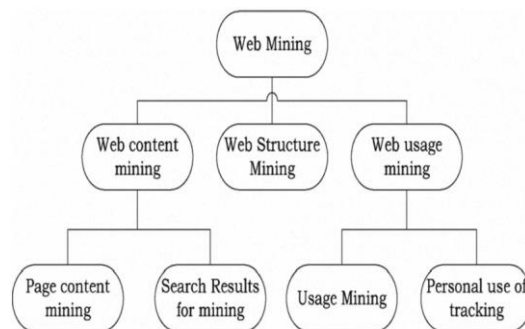

Fig 1. Classification of Web Mining Techniques

i. Web Content mining: Content of web such as audio, video that are extracting information from web. That type of information on web we used data mining technique [1].Content of pages and search keyword that type of information mining process can be done.
ii.Structure Mining: is the process of mining structure information from the Web. Web structure mining is used to improve the structure of the web pages. For example, links pointing to a document indicate the popularity of the document, while links coming out of a document indicate the richness or perhaps the variety of topics covered in the document.

iii. Web usage mining: Web usage mining is the process of extracting information from server logs i.e. user's history and web user behavior. The logs can be examined by client perspective or server perspective. This information takes as input the usage data, i.e. the data exist in in the Web server logs showing the visits of the users to the Web site[2]. Web usage mining is the process of identifying browsing patterns by analyzing the user's navigational behavior. To attain the concept, Web data (usage, content, structure) are represented by using developing model of representation, ontology's. This representation had the gap between Semantic Web and Web Mining areas, to create a research area, which of Semantic based Web Mining.

Data/information extraction: Our main Goal is extracting Useful information from web and then filtering this information.

C. Semantic Web Mining

Semantic based web mining is combination of two fast developing research areas. Semantic web and web mining. Semantic web is make data machine and user understandable data. While web mining extracting information, hidden data and making available data. The Semantic Web will provide an infrastructure that enables not just web pages, but databases, services, programs, sensors, personal devices, and even household appliances to both consume and produce data on the web. Semantic web mining is essentially mining the information pertaining to the semantic Web.

D. *An Ontology*

Ontology is collection of Unicode Resource Identification. Ontology is vocabulary of Semantic web. For small Web sites with only static Web pages, it is feasible to construct a domain knowledge base manually or semi-manually[4]. When there is no direct source for acquiring domain ontology's, machine learning and text mining techniques must be employed to extract domain knowledge from the content or hyperlink structure of the Web pages. In Semantic based web mining also mine the ontologies on the web, using ontologies on the web provide that it more intelligent.

II. WEB PERSONALIZATION

Web Personalization is any action that adapts information or services provided by a Web site to the needs of a user/set of users, taking advantage of the knowledge gained from the user's navigational behaviour.
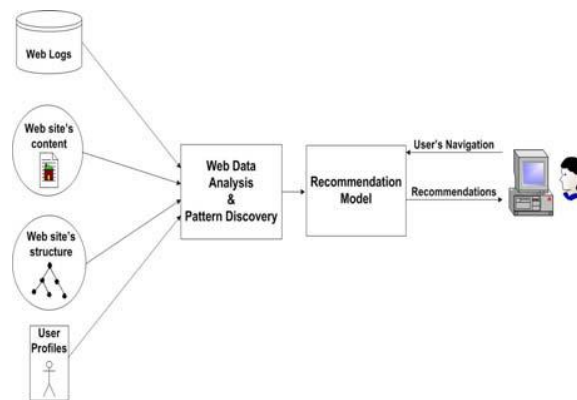

Fig 2.1 Web Personalization Process

The web is a directed labeled graph, including a plethora of hyperlinks that interconnect its web pages. Both the structural characteristics of the web graph, as well as the web pages' and hyperlinks' underlying semantics are important and determinative factors in the users' navigational process. The main contribution of this paper is a set of novel techniques and algorithms aimed at improving the overall effectiveness of the web personalization process through the integration of thecontent and the structure of the web site with the users' navigational patterns. In the first part of our work we present the semantic web personalization system standing for Semantic Web Personalization that integrates usage data with content semantics in order to compute semantically enhanced navigational patterns and effectively generate useful recommendations.

A. Data Preprocessing

The preprocessing phase is the first component in the architecture. Web server log file, which is the main source of input, generally contains noisy and irrelevant data. Preprocessing phase consists of data cleaning, user's identification and session identification tasks. Preprocessing is responsible for processing the original web usage logs in order to identify all user access sessions for each individual user. The first issue in the preprocessing phase is data preparation. Depending on the application, Web log data may need to be cleaned from entries involving pages that returned an error or graphics file accesses. Furthermore, crawler activity can be filtered out, because such entries do not provide useful information about the site's usability. Another problem to be met has to do with caching.

Usage data: The log data collected automatically by the Web and application servers represents the fine-grained navigational behavior of visitors. Depending on the goals of the analysis, this data needs to be transformed and aggregated at different levels of abstraction.

Content data: The content data in a site is the collection of objects and relationships that are conveyed to the user. For the most part, this data is comprised of combinations of textual material and images. The data sources used to deliver or generate this data include static HTML/XML pages, image, video, and sound files, dynamically generated page segments from scripts or other applications, and collections of records from the operational database(s).

User data: The operational database(s) for the site may include additional user profile information. Such data may include demographic or other identifying information on registered users, user ratings on various objects such as pages, products, or movies, past purchase or visit histories of users, as well as other explicit or implicit representation of users' interests.

B. Pattern Discovery

Log analysis tools (also called traffic analysis tools) take as input raw Web data and process them in order to extract statistical information. Such information includes statistics for the site activity (such as total number of visits, average number of hits, successful/failed/redirected/cached hits, average view time, and average length of a path through a site), diagnostic statistics (such as server errors, and page not found errors), server statistics (such as top pages visited, entry/exit pages, and single access pages), referrers statistics (such as top referring sites, search engines, and keywords), user demographics (such as top geographical location, and most active countries/cities/organizations), client statistics (visitor's Web browser, operating system, and cookies), and so on. Some tools also perform clickstream analysis, which refers to identifying paths through the site followed by individual visitors by grouping together consecutive hits from the same IP, or include limited low-level error analysis, such as detecting unauthorized entry points or finding the most common invalid URL. These statistics are usually output to reports and can also be displayed as diagrams.

C. Pattern analysis

After discovering patterns from usage data, a further analysis has to be conducted. The exact methodology that should be followed depends on the technique previously used. The most common ways of analyzing such patterns are either by using a query mechanism on a database where the results are stored, or by loading the results into a data cube and then performing OLAP operations. Additionally, visualization techniques are used for an easier interpretation of the results. Using these results in association with content and structure information concerning the Web site there can be extracted useful knowledge for modifying the site according to the correlation between user and content groups. Sensitive information or also can interfering with the operation of other nodes. performed the target nodes become occupied for time being. The malicious node masquerades as one of end nodes of the session and hijack the session.

III. CONCLUSION

In this survey paper Two Fast growing research area's information given: Web Mining and Semantic Web. Data Mining Technique work on web is called web mining. In this survey paper gives brief introduction about structure mining, content mining and usage mining. In this survey paper We use semantic web and web mining technique on web personalization process. How can work web personalization process more effectively with the help of semantic web and web mining technique. In this paper gives introduction and process of web personalization .Paper gives knowledge of ontology work on web mining process.

REFERENCES

[1] M.Venu Gopalachari,Dr. P. Sammulal, "A Survey on Semantic Web and Knowledge Processing," International Journal of Innovative Research in Computer and Communication EngineeringVolume: 1 Issue: 2.

[2] V. Sitha Ramulu, Ch. N. Santhosh Kumar, K. Sudheer Reddy, "A Study of Semantic Web Mining: Integrating Domain Knowledge into Web Mining", International Journal of Soft Computing and Engineering (IJSCE), vol. 2, Issue:3.

[3] T. KRISHNA KISHORE," Probabilistic Semantic Web Mining Using Artificial Neural Analysis" (IJCSIS) International Journal of Computer Science and Information Security,Vol. 7, No. 3.

[4] GerdStumme,Andreas Hotho, Bettina Berend "Semantic Web Mining State of the Art and Future Directions" STUMME, HOTHO, AND BERENDT: SEMANTIC WEB MINING — A SURVEY

[5] Aarti Singh, "Agent Based Framework for Semantic Web Content Mining", International Journal of Advancements in Technology.

[6] MahendraThakur,GeetikaS.Pandey." Performance Based Novel Techniques for Semantic Web Mining" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 1, January 2012