

INTRODUCTION TO WEB MINING & ITS APPLICATIONS

Yogesh Bhatia, Sanjeev Verma

Department of Information Technology,

Dronacharya College of Engineering, Gurgaon, Hr, India

Abstract- In this paper the concept or usage of web mining is summarized & its importance and different applications are illustrated. As the Web is the largest storehouse of knowledge of our time it has become essential to learn how to exploit its potential. Web mining is the application of data mining techniques to extract knowledge from web data, including web documents, hyperlinks between documents, usage logs of web sites. It is concerned mainly with its web contents, web usage & web structure mining. This paper deals with the discussion about web mining & its applications which are helpful in future for different various purposes.

I. INTRODUCTION

Web mining is the applying knowledge of data or information mining techniques to extract knowledge from net data, as well as net documents, hyperlinks between documents, usage different kinds of websites, etc. Internet has become an essential part of our lives currently a days therefore the techniques that are useful in extracting data present on the web is a remarkable space of analysis. These mining techniques helps us to extract data from net knowledge, during which at least one in every of structure or usage (Web log) knowledge is employed within the mining method. According to analysis targets, net mining may be divided into 3 differing types, that are net usage mining, Web content mining and net structure mining. Web mining is that the term of applying data processing techniques to mechanically discover and extract useful information from the globe Wide Web documents and services. Although Web mining puts down the roots deeply in data processing, it's not resembling data processing. The unstructured feature of net information triggers a lot of complexness of net mining. Web mining research is really a convergency area from many analysis communities, like

information, Information Retrieval, computing, and additionally scientific discipline and statistics also.

II. WEB MINING CATEGORIES

WEB USAGE MINING: It is discovery of meaningful pattern from data generated by client server transaction on one or more web localities. A web is a collection of inter related files on one or more web servers. It is automatically generated the data stored in server access logs, refers logs, agent logs, client sides cookies, user profile, meta data, page attribute, page content & site structure. Web mining usage aims at utilize data mining techniques to discover the usage patterns from web based application. Its technique to predict user behavior when it is interact with the web. It is categorized into three phases which are:

Preprocessing- According to client, server and proxy server it is first approach to retrieves the raw data from web resources and processed the data. It is automatically transformed the original raw data.

Pattern Discovery- According the data preprocessing discovered the knowledge and implements the techniques to discover the knowledge like as machine learning and data mining procedures are carried out at this stage.

Pattern Analysis- pattern analysis is the process after pattern discovery. Its check the pattern is correct on the web and how to implement on web to extract the information on your web search / extract knowledge from the web.

III. WEB CONTENT MINING

Web Mining is basically extract the information on the web. Which process is happen to access the information on the web. It is web content mining. Many pages are open to access the information on the web. These pages are content of web. Searching the information and open search pages is also content of web. Last accurate result is defined the result pages content mining. Web content mining is the mining, extraction and integration of useful data, information and

knowledge from Web page content. Web content mining is differentiated from two different points of view: Information Retrieval View and Database View. R. Kosala summarized the research works done for unstructured data and semi-structured data from information retrieval view. It shows that most of the researches use bag of words, which is based on the statistics about single words in isolation, to represent unstructured text and take single word found in the training corpus as features. For the semi-structured data, all the works utilize the HTML structures inside the documents and some utilized the hyperlink structure between the documents for document representation. As for the database view, in order to have the better information management and querying on the web, the mining always tries to infer the structure of the web site to transform a web site to become a database. It is used in data confirmation and validity verification, data integrity and building taxonomies, content management, content generation and opinion mining.

IV. WEB STRUCTURE MINING

We can define web structure mining in terms of graph. The web pages are representing as nodes and Hyperlinks represent as edges. Basically it's shown the relationship between user & web. The main aim of web structure mining is generating structured summaries about information on web pages/webs. It is the process of using graph theory to analyze the node and connection structure of a web site. Web structure mining has a nature relation with the Web content mining, since it is very likely that the Web documents contain links, and they both use the real or primary data on the Web. Its usage minimize two main problems of the World Wide Web. The first one is irrelevant search results. Relevance of search information become misconstrued due to the problem that search engines often only allow for low precision criteria. The second problems is the inability to index the vast amount of information provided on the Web. This causes a low amount of recall with content mining.

V. BENEFICIAL AREAS OF WEB MINING

- E-Learning: Web mining can be used for improving & enhancing process of learning in e-learning environments. Applications of web mining to e-learning is usually web usage based-online and not offline. Machine learning

techniques and web usage mining enhance web based learning environments.

- Digital libraries: Digital libraries services provide precious information distributed all around the world, eliminating the necessity to be physically present at different libraries in different parts of the world.
- E-Government: Organizations that interact with the citizens of the country lead to better social services. The main characteristic of e-government system is related to the use of technology to deliver services electronically, focusing on the citizens needs by providing better information and enhanced services in support of government.
- Security & Crime Investigation: Web mining techniques are also useful in protection of user system or logging information against cyber crimes. Clustering & classification techniques of web mining can reveal identities of cyber criminals whereas neural networks, decision trees can be used to trace crime patterns and network visualization on websites

VI. CONCLUSION

In this paper we have seen that web mining has wide range of applications which are used all over the world. We have also studied types of web mining which are web content mining, web structure mining & web usage mining. Organizations find a new and better way to do business: E-commerce through the Internet, but these organizations have to implement Web mining systems to understand their customers' profiles, and to identify their own strength and weakness of their E-marketing efforts on the web through continuous improvements.

REFERENCES

- [1] http://en.wikipedia.org/wiki/Web_mining
- [2] http://www.academia.edu/484699/Web_Mining_Today_and_Tomorrow
- [3] file:///C:/Users/HCL/Downloads/Documents/P2004_4.pdf
- [4] file:///C:/Users/HCL/Downloads/Documents/wang.pdf
- [5] Web Engineering by Pankaj Sharma
- [6] file:///C:/Users/HCL/Downloads/Documents/V3I3-0335.pdf