International Journal of Research in Engineering, Science and Management
Volume-2, Issue-8, August-2019
www.ijresm.com | ISSN (Online): 2581-5792

Web Mining Techniques

Padmini Priyadharsini[1], S. Amaresan[2]
[1]*Student, Department of Computer Science, Prist University, Thanjavur, India*
[2]*Associate Professor, Department of Computer Science, Prist University, Thanjavur, India*

***Abstract*: Web Mining is a part of Data Mining. Web Mining can also be referred as web data mining. In today's web world, millions of information are added to the web every day. Information's already added are either deleted or modified due to current needs. This information is abundant and are in many forms either structured or unstructured. To obtain relevant information needed by us/any organization we use web mining techniques.**

***Keywords*: Web mining, web content mining, web structure mining, web usage mining.**

1. Introduction

Now-a-days many organizations store information in electronic format which leads to the development of information systems. This information establishes effective linkage to their customers, suppliers and other partners who are helpful for them in various activities. This in turn leads to the development of data warehousing. To handle these large amount of information we use data mining techniques.

In simple words, we can define data mining as a collection of techniques used to pull out relevant information needed by the user/organization from data warehouse or large databases. The patterns used for data mining should be actionable to be used in an organization's decision making.

Now we can come for web mining. For past two to three decades we all witness a dramatic increase in information storing in electronic formats shortly called as e-format. Day to day more number of organizations are growing and more amount of information are stored in web. These information amount doubles every 20 months and the number and size of the databases are still increasing at a faster pace. In e-commerce environment pulling out volumes of data and doing accurate analysis are beyond the reach of best human domain expert. To extract hidden knowledge, interesting patterns and new hidden business rules from huge electronic databases are not possible with this data mining technique. So we use some higher level techniques for web data or e-data called web mining.

2. Web Mining Techniques

Web revolution has a major impact on the way we find and search information at home and work. Web is an open medium and also became an important tool for communicating ideas, conducting business and entertainment. No one monitors the content published in web and there is no mechanism for quality control. There is no catalogue to search in the web but we have search engines, directories, portals and indexes to help us browse more variety of information.

We divide web mining into three categories as follows.

A. Web Content Mining
B. Web Structure Mining
C. Web Usage Mining

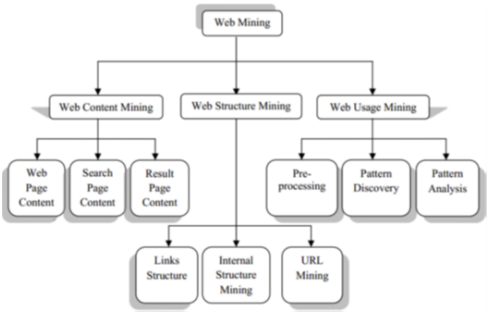

Fig. 1. Web Mining Taxonomy

A. Web Content Mining

This search is mostly done on web pages. The web pages may be structured, semi-structured or unstructured. Due to openness and no regulations on the web there is no regulations on the web data. Only pages which are journals or conference papers have been referred and published. So they are mostly trustable other pages are searched mostly with words. Web content mining can be done using agent-based approach and database approach.

Fig. 2. Web content mining classification

Using artificial intelligent community searching tools are developed called intelligent web agents (agent based approach) to meet the user's requirements. In database approach query approach is used. Search engines like Google, Yahoo etc., are also used. The techniques used here are mostly web document clustering, finding similar web pages, finger printing etc.,

B. Web Structure Mining

Web structure mining deals with discovering and modeling

the link or structured information from the web. Here, the particular web page chosen is called a node and the hyperlink goes from this page is called an edge. This method helps in discovering similarity between the sites and also finds important sites for a particular topic or authority sites for that topic. Sometimes this authority sites are omitted by search engines. The two methods used to find structured information are:

- Hyperlinks
- Document Structure

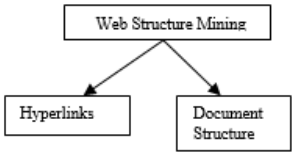

Fig. 3. Web Structure Mining Classification

C. Web Usage Mining

The goal of web usage mining is to understand and predict the user behavior in interacting with the web or website in order to improve the quality of service. Web usage mining other aims are to obtain information and discover usage patterns that may assist a fresh website design or redesigning a website for assisting user's navigation through the site. Using web server logs routine information are collected about the access, referrer and agent.

Web usage mining automatically discovers patterns in click streams and associates data collected or generated as a result of user interactions within/outside the web sites. Analyzes the behavioral patterns and profiles of users interacting the web sites.

There are three phases used in web usage mining. They are:

- Data pre-processing
- Pattern Discovery
- Pattern Analysis

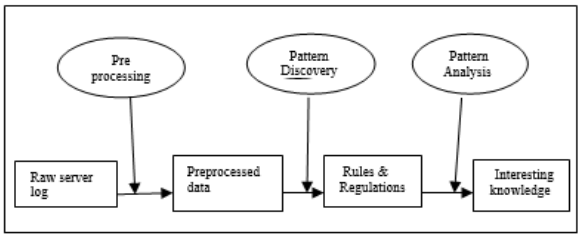

Fig. 4. Phases in web usage mining

3. Conclusion

There is a rapid growth in opportunity to analyze the web data and extract all manner of useful knowledge from it as the web and its usage continues to grow. The emergence of web mining techniques has grown rapidly in the past five years due to the effort of research community as well as various organizations practicing it. In this paper we have briefly described the key computer science contributions made by the field, a number of prominent applications, and outlined some promising areas of future research.

References

[1] Introduction to Data Mining with case studies by G.K. Gupta, third edition, PHI Learning Private Learning, Delhi.
[2] Jain Pei, Jiawei Han, Behzad Mortazavi_asl and Hua Zhu, Mining Access Patterns Efficiently from Web Logs, Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'00), Kyoto, Japan, 2000, 396-407.
[3] R. B. Geeta, Shashikumar G. Totad & Prasad Reddy PVGD, Topological Frequency Utility Mining Model Springer International Conference, SocPros 12, 2011, 505-508.
[4] Jia-ching Ying, Vincent S. Tseng, Philip S. Yu IEEE International Conference on Data Mining workshops IEEE Computer Society, 2009
[5] Jaroslav Pokorny, Jozef Smizansky, Page Content Rank: An Approach to the Web Content Mining.
[6] Bing Liu, Web Data Mining (New York, Springer International Edition, 2008).