



RESEARCH ARTICLE

IMPACT OF WEB MINING IN DIGITAL LIBRARY

***Dr. Rakesh Kumar Mishra**

Central Library, D.D.U.Gorakhpur University, Gorakhpur (UP)-273009

ARTICLE INFO

Article History:

Received 21st May, 2016
Received in revised form
27th June, 2016
Accepted 09th July, 2016
Published online 20th August, 2016

Key words:

KDD, Data Mining, Web Content Mining,
Web Usage Mining, Web Structured Mining,
Information Discover, Information Supply.

Copyright©2016, Dr. Rakesh Kumar Mishra. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Dr. Rakesh Kumar Mishra, 2016. "Impact of web mining in digital library", *International Journal of Current Research*, 8, (08), 36420-36422.

ABSTRACT

In this paper we have discuss about the meaning of web mining, Digital Library, Various types of web mining, and advantages of web mining in digital library system to extract information. Web mining is part of data mining, helps to find new patterns that enable a merchant to take new decisions for business opportunity.

INTRODUCTION

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, etc. From Data-Centric View Web mining is defined as the application of data mining techniques to extract knowledge from web data, i.e. web content, web structure, & web usage data. Knowledge discovery in databases refers to the overall process of turning low-level data into high-level knowledge. Data mining is the process of finding trends and patterns in data from large volumes of raw data. Data mining is a process of extracting connotative and unknown but useful information and knowledge from the vast, incomplete, noisy, fuzzy and stochastic data. DM brings the database technology into a more advanced phase whereas Web data mining is the application of data mining technology to net information handing. It is the process that extracts the useful information or knowledge from web resource and distills the unknown mode, which contained in Web resource and has potential application value. The goal of web mining is to improve the quality of the interaction between the digital library & users.

Digital library

A Digital library can provide a single point of access to a huge quantity of structures and accessible information that is

available to a variety of users with different information needs. Digital libraries area set of electronic resources and associated technical capabilities for creating, searching and using information. The contents of digital libraries include data, metadata that describes various aspects of the data, and metadata that consists of links or relationships to other data or metadata. The definition of digital library can be defined as a set of characteristics i.e. *collection of services, collection of information objects, supporting users with information objects, organization/presentation of those objects, availability of directly or indirectly and electronic/digital availability*. Digital library is not a single entity and it requires technology to link the resources of many.

Classification of Web Mining

Based on the different mining objects, Web data mining can be classified into following categories: Web Content Mining, Web Structure Mining and Web Usage Mining.

Web content mining is a process that obtains potential and valuable knowledge or mode from the file content and information of Web. **Web structure mining** is to infer the knowledge out of Web content by using organization structure and hyperlink relationship of Web page and institution information of Web document, it comprises hyperlink mining, inter structure mining and HTML mining. To mine the log left by the users when they accessed Web server is Web usage mining, and it is also called Web log mining. **Web log mining**

discovers the browse mode of access site of users by analyzing log file of Web serve to provide multifarious information which profits the improvement of Web sites to the site administrators, such as the navigation function of Web sites and system design of Web application. **Web Usage Mining** focuses on techniques to study the user behavior when navigating the web

Reasons for web mining

As like a coin, WWW has two sides, the user and the information provider. Both the sides face problems while dealing with the web data which are described are as followed-

A. The User Problems

(1). *finding relevant information*: People, either browse or use the search service to find specific information on web. Today's search tools have two problems. (a) Low precision due to the irrelevance of various search results. (b) Low recall due to the inability to index all the information available on the web as some of the relevant pages is not properly indexed. This is a Query – triggered process.

(2) *Extracting new knowledge form the web*: This is a data triggered process. As it is hard to get relevant information, it's very hard to make sense out of it.

B. The Information Providers' Problems

Deficient in gathering information about

- What do the customers do?
- What do the customers want?
- How to personalize the individual users?
- How effectively use the web data to market products and to service the customer?

Applications of web data mining to digital library

1. Applications of Web data mining to information discovery

The applications of Web data mining to information discovery comprise Web data exploitation techniques, multilingual information discovery and interdisciplinary coordinated searches. Web data exploitation technology is developed due to the difficulty of gaining the information from Internet, the Realization process consists of information collection, document identification, document classification and others. Interdisciplinary coordinated search can send concurrent request to several even tens of databanks, and links with OPAC, interlibrary loan, document submission and electronic source. It provides uniform searches interface to users and returns to uniform results after interdisciplinary querying.

2. Applications of Web data mining to service improvement

User information and user access log are the two types of user access data that can be recorded when the user accesses digital library with browser. The user information includes user name, user access IP address, age and profession so on. User name is

inputted by user when the user enters into the library. The profession, age and other information of user are filled in when the user register the library. The user access IP address can attain via program. User access record embodies the click record of classified browse item and keywords and mode of search. The user access record also can obtain from program. To analysis the user access data, it can help the digital library understand well the user requirements in following aspects:

- (1) To guide the collection of digital library information source. The default of information source can be discovered by analyzing the data of fail demand of user in server and integrating the assemble arithmetic to direct the collection of information source, and then it makes the information source systems of digital library rational.
- (2) To improve the design of library site structure. The information demand and action rule of user can be found by using Web mining to provide reference to network site structure optimization. The measures that improve the design of library site structure include following aspects:

Investigating the user address in high access rate resource; Making mirror image in concentrated region of user; Mining, catching and confirming the frequent browser and access path by using path analysis mode, adjusting the site structure; Appending advertisement or reading guide slip in suited place.

3. Applications of Web data mining to information supply

One of the applications of Web data mining to information supply is to individual service, supplying actively the user with integrated and relative intact information or knowledge aggregate to meet the special demand of user. There are three types of individual service.

- a. According to the specific user demand, providing customizing Web page and information channel or information item to user, putting the query agent service in practice;
- b. According to specific subject, to direct literature source or supply literature full text, and to bring the individual literature information service into effect;
- c. According to specific subject, offering relative intact project knowledge and carrying individual decision-making support service.

Some Other Applications of Web Data Mining instead of digital library are-

- E-Commerce
- Search Engines and Web search
- Website Design
- Recommendation engines
- Web communities and web market places
- Web data sets can be very large
- Cannot mine on a single server
- How to organize hardware and software to mine multi-tera byte data sets?

Conclusion

Web mining enables us to screen specific data through content mining, to discover the structural summary of web sites through structure mining and to predict the behavior and interaction of the surfers' with the web through usage mining and encompass a broad range of issues. Towards this goal, in this paper, we proposed a definition of web mining, reasons and applications of web mining, and classifications of web mining.

REFERENCES

- Arms W Y. 2000. Digital Libraries. Massachusetts, London: The MIT Press Cambridge.
- Bao Junjie. 2004. Research on Web Data Mining, *Journal of Chongqing College of Education*, May.
- Cooley. R. 2000. *Web Usage Mining: Discovery and Application of Interesting Patterns from Web Data*. PhD thesis, University of Minnesota.
- He Shaozhuo. 2004. A few words about Data Mining and its Application to Library, *Library World*, Sep.,
- Liu Wenqing, Bao Juejie. 2004. Research on Web-based Data Mining Technology and Its Application, *Journal of Chongqing Three Gorges University*,
- Wang Baoji. Study and Practice on Agricultural Engineering Digital Library Based on TPI, *New Technology of Library and Information Service*, 2003
- Wang Yan. 2003. The Application of Data Mining in Digital Library. *Information Science*,
- Zhou Tao, Li Jun. 2004. The Research of Web Data Mining Technology, *Journal of Han Zhoug Teachers College (Natural Science)*, June.
