A Survey on "Web Link Suggestion's based on Log Mining"

Swati B. Ghatage[1], Tejashri A. Jadhav[1], Priyanka S. Koli[1], Nikita S. Mane[1], Nikita E. Powar[1],

Prof. J. N.Jadhav[2]

B. E Student, Dept. of Computer Engineering, Dr. D.Y. Patil COET., Kasaba Bawada, Kolhapur, Maharashtra, India[1]

Assistant Professor, Dept. of Computer Engineering, Dr. D.Y. Patil COET., Kasaba Bawada, Kolhapur, Maharashtra,

India[2]

ABSTRACT: Primary purpose of the data mining is summarizing the data in the useful form. User has to search websites which contain required information. A Web Usage Mining is designed to operate on web server logs which contain user's navigation. Hence, recommendation system using Web Usage Mining can be used to forecast the navigation pattern of user and recommend those to user in a form of recommendation list. In this paper, we propose for capturing user's intuition in the form of recommendation list containing pages visited by user and pages visited by other user's having similar usage profile.

KEYWORDS: Data mining, Frequent Pattern Mining, web usage mining.

I. INTRODUCTION

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. This information can be used to increase revenue, cut costs or both. Web usage mining system should be able to gather useful usage data thoroughly, filter out irrelevant usage data, establish the actual usage data, discover interesting navigation patterns clearly, analyze and interpret the navigation patterns correctly, and apply the mining results effectively. Information gathering is time consuming process for a new user. He/she may get experience of other web users in format of suggestions. Data Mining techniques will help users to get information of next web link based on his/her current web pages & recorded web logs of previous user logs.

World Wide Web is a huge storehouse of web pages and links. It offers large quantity of data for the Internet users. User's accesses are recorded in web logs. Web usage mining is a kind of mining techniques in logs.
This application is develop for overcome some problems of pattern structure by using unordered structure is use for creating the tree structure
Media Queries-w3c
https://www.w3.org/TR/css3-mediaqueries/.
For example-A page about"Avoiding Melanoma" might use technical jargon to describe ways to prevent skin cancer. But a search engine might skip or not rank that page highly if people are instead searching for "skin cancer prevention tips". Your content needs to be written in the right ''language'-the language your user is using when searching.

II. LITERATURE SURVEY

Recently, several Web Usage Mining systems have been proposed to predicting user navigation behavior and their preferences given to new user based on logs. In the following we review some of the most significant WLSLM systems and architecture that can be compared with our system is structured according to an offline and an online component. The off-line component build session clusters by analyzing past user activity recorded in database log files. Here we maintain the logs of websites which are visited by the user. According to those logs we provide suggestions to new user. On those logs we apply preprocessing for maintaining the logs of those links

Which are mostly visited and the links which are not mostly visited those are removed from the database. Then the online component builds active user sessions which are then classified according to generated model. The classification allows to identify pages related to the ones in the active session and to return the requested page with a list of suggestions. This approach has several limitations, related to scalability. The analysis is based on anonymous usage data combined with the structure formed by hyperlinks of the site. Data mining techniques (i.e. clustering, sequence pattern discovery and association rules) are used in preprocessing phase in order to obtain aggregate usage profiles. In this phase Web server logs are converted into clusters of visited pages, and cluster made up of set of pages with common usage characteristics. Matching entries are used to compute a set of recommendations which will be inserted into last requested page as list of hypertext links.

A. Query Recommendation using Query Logs in search Engines

For search engine in this paper author proposed a scheme which given a query submitted and a list of queries which are related are suggests. On previously issued queries the related queries are depends, and for the search engine to tune or redirecting the search process which can be issued by the user. Based on a process of query clustering the method proposed in which there are identified groups of semantically similar queries. The content of historical preferences of user's are uses by the clustering process which are registered in the search engine's query log. The method not only ranks the related queries, but also discovers them according to a relevance criterion. Finally, we show with links which effectiveness of the method over the query log of a search engine.

- Some disadvantages of query recommendation
  1. Difficulty in interfacing:
      Interfacing on SQL databases is more complex than adding a few lines of code.
  2. More features implemented in propriety way:
      Although SQL database confirm go for propriety extensions to standard SQL to ensure vendor lock-in.

III. PROPOSED SYSTEM

Web module is composed of apache tomcat Web server which provides service to user requests. This is a multithreaded server provide service to number of users concurrently. It receives requested link from user (client), then forward to actual internet web server receive response from that server and forward this response to requested client. It also keep session track of each user and keep track of requested links into logs database called as web logs.

Frequency Pattern mining module performs data mining activity on web logs of all users. After processing it, outputs and association rules in form of patterns. This module composed of main three sub-modules.

Pre-processing is act as sub-module in frequent pattern mining which processes web log database before actual pattern mining. This module performs cleaning process to remove noise and reduce errors from input data to frequent pattern mining module.

Pattern mining sub-module works on data mining algorithm that is Apriori algorithm which takes error tree and noise free data as input and produce patterns which will further used for rule creation. Rule Creation This sub-module performs conversion of associative to the rules for decision tree. These rules will be further used for suggestion mechanism.

Suggestion module take current user request and processed patterns as input and provide most appropriate suggestion to used based on support of each pattern item and requested link. These suggestions will be in form of web links to the user.

Tree Representation is supportive module for system administrator to explore web logs and processed patterns in form of tree. It display final pattern in tree format to analysis purpose.
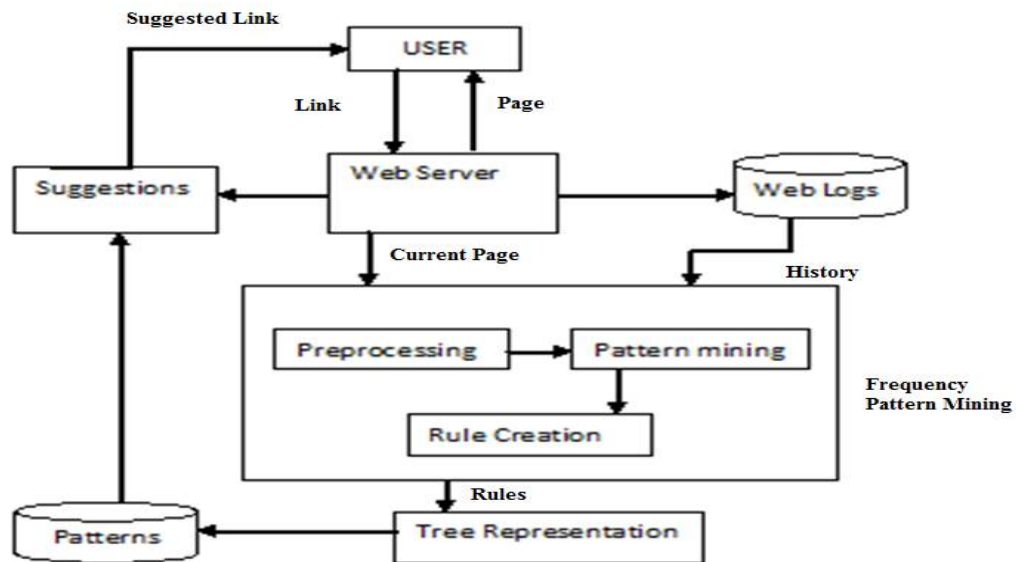
- **Modules:**
  This System architecture contains 6 modules.
  1. Frequency Pattern Mining
  2. Pre-processing
  3. Pattern Mining
  4. Rule Creation
  5. Suggestion Module
  6. Tree Representation Module

  1. **Frequency Pattern Mining**:-
     This module performs data mining activity on web logs of all users. After processing it, outputs and association rules in form of patterns. This module composed of main three sub-modules.
  2. **Pre-processing**:-
     This sub-module processes web log database before actual pattern mining. This module performs cleaning process to remove noise and reduce errors from input data to frequent pattern mining module.
  3. **Pattern Mining**:-
     This sub-module works on data mining algorithm that is Apriori algorithm which takes error tree and noise free data as input and produce patterns which will further used for rule creation.
  4. **Rule Creation**:-
     This sub-module performs conversion of associative to the rules for decision tree. These rules will be further used for suggestion mechanism.
  5. **Suggestion Module**:
     This module take current user request and processed patterns as input and provide most   appropriate suggestion to used based on support of each pattern item and requested link. These suggestions will be in form of web links to the user.
  6. **Tree Representation Module**:
     This is supportive module for system administrator to explore web logs and processed patterns in form of tree. It display final pattern in tree format to analysis purpose.

IV. CONCLUSION

The goal of this project is to make use of Web site access logs to make intelligent link recommendations for an organization's Web site. Several methods are used to make recommendations with varied parameters. Each method is evaluated based on whether automatically recommended pages are accessed later in the same   browsing session. This paper uses the web links that the user may visit in the near future will be predicted based the suggestions.

REFERENCES

1.   IEEE TRANSACTIONS ON CYBERNETICS, VOL. 45, NO. 6 JUNE 2015,
2.   Hussain Tasawar, Asghar Sohail and Fong Simon, "A hierarchical cluster based preprocessing methodology for Web Usage Mining", 6th International Conference on Advanced Information Management and Service (IMS), Pp. 472-477, 2010.
3.   T. Asai et al., "Efficiently mining frequent substructures from semistructured data," in Proc. Int. Workshop Inf. Elect. Eng., Suwon, Korea, May 2002, pp. 59–64.
4.   V. Losarwar, Dr. Madhuri Joshi, "Data Pre-processing in Web Usage Mining", International Conference on Artificial Intelligence and Embedded Systems (ICAIES'2012) July 15-16, 2012 Singapore
5.   M. Yadav, Mr. P. Mittal, "Web Mining: An Introduction", International Journal of Advanced Research in   Computer Science and Software Engineering Volume 3, Issue 3, March 2013.Sannella, M. J. 1994 Constraint Satisfaction and Debugging for Interactive
6.   Brown, L. D., Hua, H., and Gao, C. 2003. A widget framework for augmented interaction in SCAPE.
7.   Cyrus Shahabi • Farnoush Banaei-Kashani, ―Efficient and Anonymous Web-Usage Mining for Web Personalization