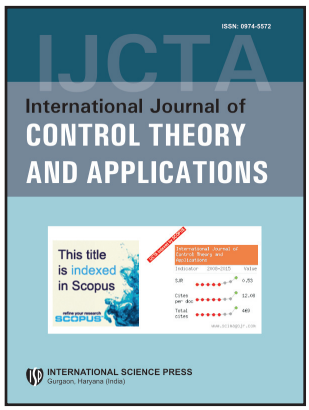

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 24 • 2017

An Enhanced Searching Technique to Improve user Navigation

Aiswarya B.[a], Geethu Prasad[b] and Remya Nair T.[c]

[a-c]Department of Computer Science & IT, Amrita School of Arts & Sciences, Amrita Vishwa Vidyapeetham. Email: [a]aiswaryababu92@gmail.com, [b]geethuprasad1992@gmail.com; [c]remybhi@gmail.com

Abstract: Web mining is one of the promising areas in the field of Data Mining due to the immense amount of data available on web services in recent years. Nowadays web structure reorganization is becoming a challenging factor in web mining. The predominant reason behind web structure reformation is that developers' perception of viewing a website is often eloquently different from that of user. The intend of our paper is to implement a data storage mechanism to enhance the user navigation effectively. For that we are providing a mathematical programming method to reorganize the website structure with minimum changes.

Keywords: Web Mining, Boolean Matrix, Page Rank.

1. INTRODUCTION

With the growth of significant information available on Internet, Web is a dominant and interactive medium to passing information all over the world. Web Mining is to discover and extort meaningful information from web. Depending on various research studies, the web mining are examined in the categories of :Web Content Mining, Web Structure Mining and Web Usage Mining. Web content mining can be defined as the process of extracting and integrating the useful information from web documents. Interpreting the information from structure of data is known as web structure mining. Web usage mining is the mechanism of discovering behavioral patterns on web data. Since Web site contains bulk amount of information, our initiative on preferring web structure mining is that, users are mostly browsing irrelevant pages. As a result user lost their interest on searching data over the website. So our aim is to develop a Website that contributes end users to receive the information they want with minimum clicks.

In spite of targeting investments in websites, identifying the essential information from a website is awkward. According to paper [1], website failure can occur due to inadequate website design. If the user is not capable in tracking their required information, then they can leave from this website and moves to a web site that offers high quality information as proposed by McKinney et. al., [2]. Web mining is an efficient method to comprehend and assess user behavior.

Web mining process consists of four important steps:

- Resource finding: Extracting data from text resources.
- Data selection and Preprocessing: It is the process of automatically selecting and pre-processing required information from relevant web data.
- Generalization: Investigating the general patterns at the individual web sites and over various sites.
- Analysis: The interpretation and validation of mined patterns.

The primary purpose of poor website design is that, developer's view of understanding a website structure will be quite different from the users view. Therefore, it is important that the website has to be organized according to the user's choice. Study of paper [3] reveals the improvement of website navigability using a Mathematical Programming Model.

This paper addresses a data storage mechanism for improving user navigation by suggesting a new field in a dataset for counting the number of occurrences of a keyword in a most optimized way.

2. BACKGROUND STUDY

Due to the massive amount of information presented on internet makes the reorganization of web structure a challenging task in the era of web mining. To find the required information for users various methods have been used in web mining applications. As in [4] web mining taxonomy deals with three aspects in which there exist plenty of techniques and issues for each phases. In web mining various algorithms and tools are available for getting better results. Also it discusses about major issues faced in web mining. Finally it examined the fields where web mining applications are used. The paper named "Web Usage Mining For extracting Users' Navigational Behavior" which is done by Divya Racha introduced a more appropriate idea about the web usage mining and extracting the browsing behavior of the users from the web logs[5].Pooja Mehtaa, Brinda Parekh, Kirit Modi, and Paresh Solanki in their work titled "Web Personalization Using Web Mining: Concept and Research Issue" first of all, a study of various web mining categories has done. After the analysis gives a brief idea about web personalization, and also it gives a vague idea about different phases involved in it[6]. Using web personalization methods[7] they form cluster of user details from web logs based on their access pattern in generating dynamic links. A research paper presented by Prasad K. Bhamare, Ravikumar R. Metkar and Amol A. Sonawane which titled "Effective User Navigation for Improving Web Structure " propose an approach called web transformation to modify the structure of the website for a set of users, but it is not for individual users[8].The paper named "Reorganization of links to Improve User Navigation" which is done by Deepashree A.Vadeyar1 and Yogish.K elaborates how the web structure reorganization can be done using data mining techniques in improving user navigation[9].The paper named "Web site Usability, Design and Performance Metrics Information Systems presented by J. Palmer indicates how a well structured website allows users to retrieve relevant data without getting misguided or need to perform backtracking[10]. Papers[11][12] proposes a mathematical approach to raise user navigation with minor variations that provide an efficiency in navigation. The paper [13][14] demonstrates a brief survey of different page ranking algorithms and their essential role in making the user navigation easier. Contents of [15] developed an experimental model for search engine and produce a concurrent data flow of relevant information from the database.

2.1. Page Rank Algorithm

Page Rank was invented by S.Brin and L.Page at Stanford University. Page Rank is an algorithm used by the popular search engine called GOOGLE for their searching purposes. The algorithm emphasis on the web pages based on their link structure.

The basic idea of this algorithm is that, if many pages are linked to a single web page then, we can provide higher rank value to that particular page. If one page is linked to other page, it is actually going to vote for the other pages. Based on the number of votes we can say whether the page is important or not. The Page Rank algorithm is described as follows:

$$PRA(P) = (1 - d_f) + d_f(PRA(T1)/C(T1) + ... + PRA(Tn)/C(Tn))$$

where,

- PRA(P) – Page Rank of page P
- PRA(Ti) – Page Rank of pages Ti which link to page P
- C(Ti) - number of outbound links on page Ti
- d_f – d_f, is dampling factor, $0 < d_f < 1$

The default value of $d_f = 0.85$

Page Rank Algorithm (PRA) = 0.15 + 0.85 × (a share of the Page Rank of every page that links to it.

2.2. Boolean Matrix

A matrix ($p \times q$ where is p is number of rows and q is number of columns) whose factors are either zero or one, such matrices are also called (0, 1)-matrix or Boolean matrix. The term 'Boolean 'makes it clear that the intramural activities on elements adhere to the Boolean law: 1 + 1 = 1.

3. PROPOSED SYSTEM

In previous system, keywords are search against the words in each document. This keyword search will return a large amount of inappropriate data. In our paper we proposed an enrichment to the existing keyword search by allowing deepest search mechanism, while examining the number of occurrences of the keyword in a most optimized way using Mathematical Programming Model. This proposed model helps the end user to view the most relevant data, with the help of page rank algorithm and the number of occurrence of the query term.

The following steps show the working of the proposed system.

Step 1: End user will gives the keywords in Search engine.

Step 2: The Proposed model will split that keyword into word phrases.

Step 3: Then it goes to the web, next we check whether that keyword match up with web page content.

Step 4: A Boolean matrix is generated based on availability of the keyword on various pages, Actually here we convert text to Boolean.

Step 5: Then doing the Boolean AND operation on the matrix values, adding that to the same matrix as a new column, these values are changed automatically after performing AND operation.

Step 6: The resultant TRUE values are plotted and count the number of words occurrences in that particular Ip address.

Step 7: Then we retrieve all the related links and stored into the large data warehouse.

Step 8: From there the crucial data will be elicited with the help of Page rank algorithm to find the highest page rank value and also we get the number of occurrences of the query term.

Step 9: Based on the highest page rank value and the number of occurrences we are able to get the most relevant link and the user is capable to recapture the most convenient data as output.

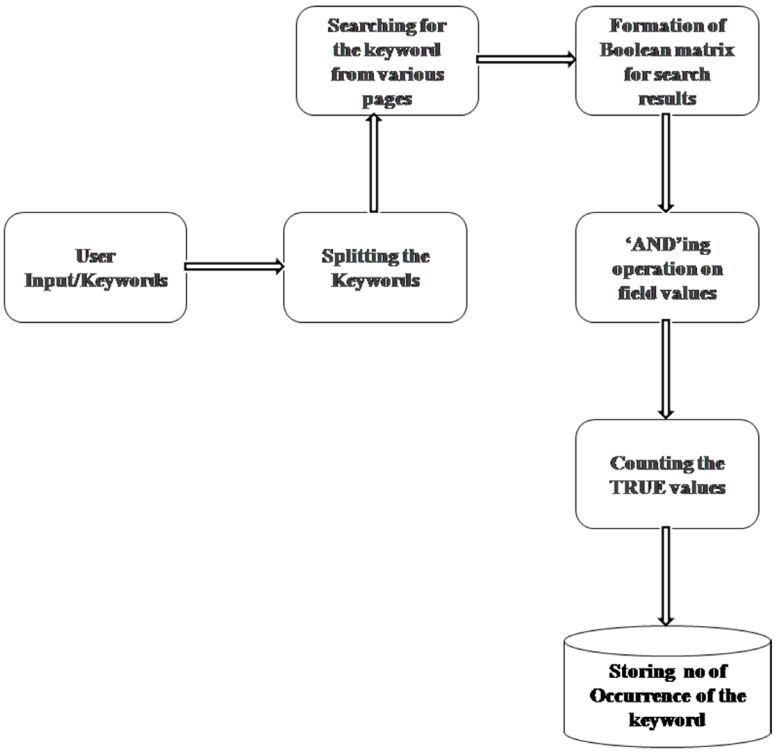

Figure 1: Architecture of the Proposed System

4. EXPERIMENTAL RESULTS

The above snapshot describes the search page where end user can search the Keywords; it displays the most appropriate link and the number of occurrences of the keyword along with Page ranking.

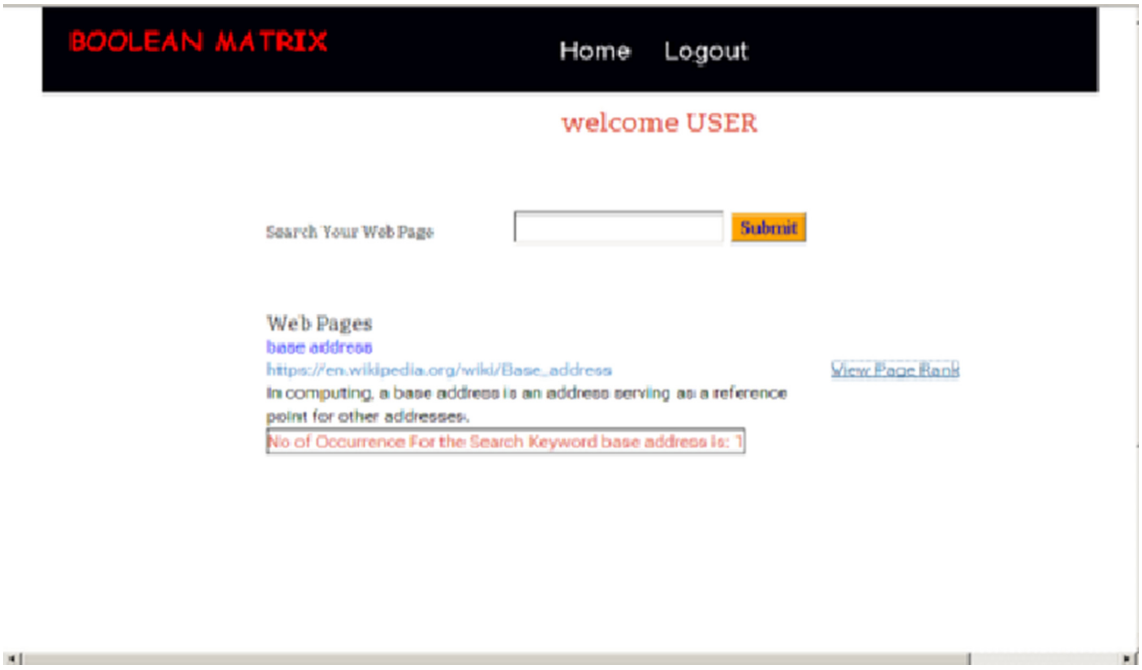

Figure 2: Screenshot representing search page

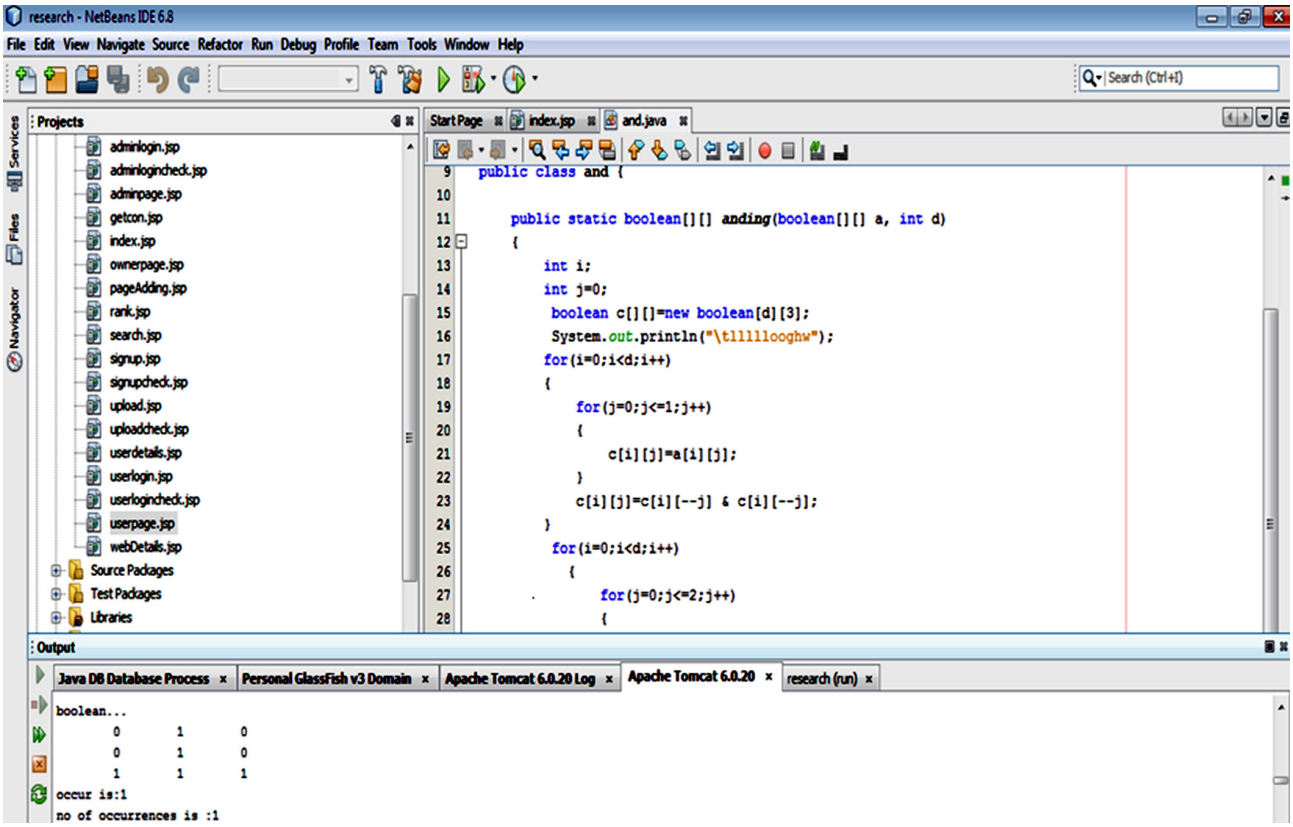

Figure 3: Boolean Representation of the above searched result

5. CONCLUSIONS

In existing system keyword search will return a large number of inappropriate data. To overcome this drawback we proposed an enhanced searching technique to improve user navigation and also this model provides a mathematical programming approach with the help of Boolean matrix. Based on the highest page rank value and no of occurrences we are able to get the most relevant link and user is able to retrieve the most useful data as output.

REFERENCES

[1] J. Palmer, "*Web Site Usability, Design, and Performance Metrics*,"Systems Research, Vol. 13, No. 2, pp. 151-167 2002.

[2] V. McKinney, K. Yoon, and F. Zahedi, "The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach," Information Systems Research, Vol. 13, No. 3, pp. 296-315,2002.

[3] Deepshree A.vadeyar et. al.,/(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1), 2014, 834-837.

[4] Dr.S. Vijiyarani1 and Ms. E. Suganya "Research Issues in Web Mining " International Journal of Computer-Aided Technologies (IJCAx) Vol.2, No.3, July 2015.

[5] Divya Racha "Web Usage Mining For extracting Users' Navigational Behavior "International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 3 Issue 5 may,2014 PageNo.5989-599.

[6] Pooja Mehtaa, Brinda Parekh, Kirit Modi, and Paresh Solanki " Web Personalization Using Web Mining: Concept and Research Issue "International Journal of Information and Education Technology, Vol. 2, No. 5, October 2012.

[7] Magdalini Eirinaki and M. Vazirgiannis "Web Mining for Web Personalization"

[8] Prasad K. Bhamare, Ravikumar R. Metkar, Amol A. Sonawane," Effective User Navigation for Improving Web Structure" International Journal of Computer Applications (0975 –8887)Volume 128–No.17, October 2015.

[9] Deepashree A.Vadeyar1, Yogish.K " Reorganization of links to Improve User Navigation.

[10] J.Palmer "Web site Usability,Design and Performance Metrics Information Systems Research; Jun 2002; 13, 2; ABI/ INFORM Globalpg. 151.

[11] Chhaya N. Shejul1, Prof. B. Padmavathi "Website Structure Improvement through Web Mining International JournalofEngineering Research and General Science Volume 3, Issue 2, Part 2, March-April, 2015 ISSN 2091-2730.

[12] Min Chen and young U. Ryu "Facilitating Effective User Navigation through Website Structure Improvement" IEEE KDD Vol. No. 25. 2013.

[13] Yogita Garg, Mr. Vinod Jain "A Brief Survey of Various Ranking Algorithms for Web Page Retrieval in Web Structure Mining", International Journal of Engineering Trends and Technology (IJETT), V21(3),168-172 March 2015. ISSN:2231-5381.

[14] Parveen Rani "A Review of Web Page Ranking Algorithm "International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 3, March 2013.

[15] Shaheen Parveen, Ajay Kushwaha " An Approach of Deep Web Mining For Data Extraction" IJESAT] INTERNATIONAL JOURNAL OF ENGINEERING SCIENCE & ADVANCED TECHNOLOGY Volume-2, Issue-6, 1653 – 1656.