



# Implementation of Web Structure Mining with Breadth First Search and Depth First Search

Renu  
Student  
CSE, ITM University

Tanu Verma  
Student  
CSE, ITM University

Deepti Gaur  
Associate Professor  
CSE, ITM University

## ABSTRACT

When a user search a query through web, then he has to pass through many web pages of the internet, then filter the data and download all information as per his requirement. The whole process of downloading and searching the data is very time consuming.[2] To save the time for searching the data by a user, a web link extraction tool has been designed and implemented in .Net. In this paper, we have implemented the web link extraction tool with Breadth first search as well as with Depth first search. This has been presented and analyzed with various URL. The user can view the total links (i.e. inlink as well as outlink) and total number of links.

## Keywords

Breadth First Search, Depth First Search, Web Structure Mining, Web Extraction, Downloading, Graphical User Interface.

## 1. INTRODUCTION

Text The World Wide Web acts as a large data repository which contains huge amount of information. While searching through the internet, the main challenge of a web user is to find the relevant and required information in an effective and efficient manner with less effort and time. So, By applying Data Mining techniques to Web data, Web Mining try to solve this problem[1].

Web Mining is an application of Data mining techniques to extract useful and knowledgeable information from web data. Web data may be web content, web structure and web usage data. Depending upon the type of data to be mined, Web Mining is divided into three types, i.e., Web Content Mining, Web Structure Mining and [4]Web Usage Mining. [10]Web Content Mining is used to extract useful information from the content of the web pages which contains different type of data, for e.g., audio, video, textual, metadata (data about data), image and hyperlinks. [7]Web data may be Structured (in the form of table, tree, list and database), Unstructured (in the form of text-document) and Semi- Structured which do not have a predefined structure (XML and HTML).Web Structure Mining is used to extract structure information from the web and focuses on hyperlink structure.

In this Paper, we use both the breadth first search and depth first search techniques in Web Structure Mining. Breadth First Search is a searching technique which uses QUEUE (First in First out) data structure to be implemented. It starts at the initial node and then searching is applied in downward on a tree structure in breadth wise order. And Depth First Search uses STACK (Last in First out) which starts as initial node and applied downward on a tree structure in depth wise order.

This paper work has the following objectives:-

- To implement Web Structure Mining using Breadth First Search technique to download link.

- To implement Web Structure Mining using Depth First Search technique to download link.
- To count the total no. Of links crawled and analyze the performance

## 2. RELATED WORK

Web Mining has many applications in different research fields such as artificial intelligence and information retrieval or information extraction. Web contain large amount of data, So, We need to collect the web document together before mining. And the web documents should be retrieved and represented into some format for easy processing.

We can view the Web as directed graph in which document represents the nodes and hyperlink represents the edges. Web Structure Mining uses this process to analyze the edges and node structure of web sites. It considers web pages as an object in World Wide Web. [9]Web Structure Mining has two popular algorithms which are HITS and Page Rank. To find the importance of web pages, both algorithms focus on link structure of the web. Structural information about sites and web documents is provided by Web Structure Mining. Depending on kind of Structure information, Web Structure Mining can be divided into two types which are Hyperlinks (connects location of web pages either in different web pages known as inter-document hyperlink or in same web pages known as intra-document hyperlink) and Document Structure.

There are basically two main problems that are faced by any web user. The first is an irrelevant search results and the second problem is not able to index large amount of information available on the web. Due to large amount of information available on the web, these two problems are solved by Web Structure Mining

## 3. IMPROVEMENT

This work has the following improvement over the existing work: -

In existing work, when we implement the Web Structure Mining with Breadth First Search technique, then sometimes it does not extract all the links. If the crawler finds the weight of any page zero, then it stops extracting the links and gives the output without extracting all other links. So user is not able to find the complete information because all links are not extracted by the algorithm. But in this paper, when we use BFS, then if the crawler finds the weight of a page is zero, then it will not stop extracting the links, but it moves to the page having the lowest weights and extract the other links also. So, it saves the time and effort of a web user.

#### 4. IMPLEMENTATION RESULT AND ANALYSIS

The user has to enter the URL into the Graphical User Interface as shown in the Fig. 1.

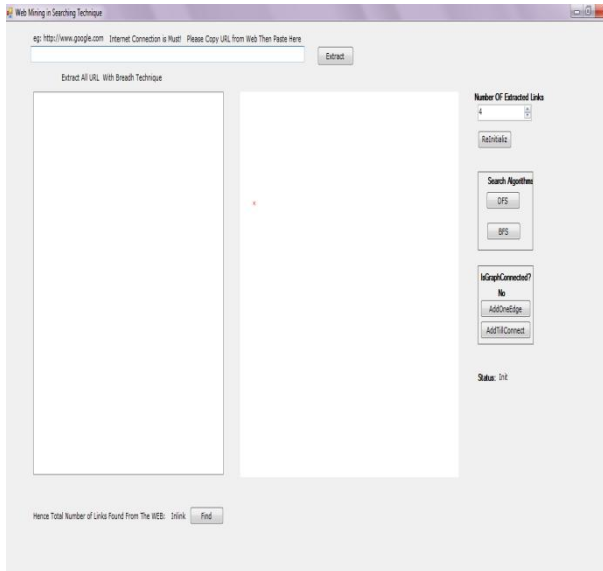


Fig. 1. Graphical user interface for entering the URL

We can enter any URL starting with http://. In this paper, we use URL as https://www.google.co.in/ and then click on extract. We have the list of URL as shown in the Fig. 2.

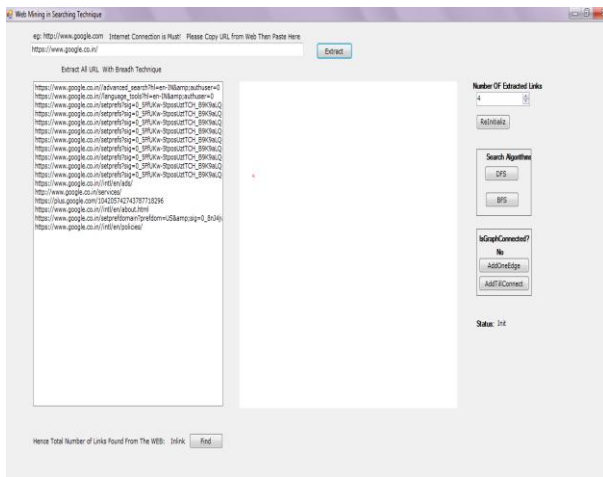


Fig. 2. Extracted links of a given URL

Now we can calculate the total number of link found from the Web. To do so, we have to click on find and we have the total no. of links as shown in Fig. 3. We have 17 links of URL “http://google.co.in?”.

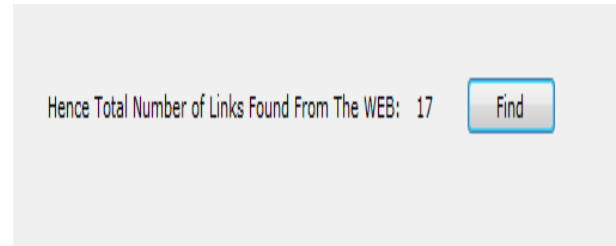


Fig. 3. Total no. of links

Now, we can draw the graphs for Breadth First Search and Depth First Search with their weights. The graph for Breadth First Search is shown in the Fig. 4.

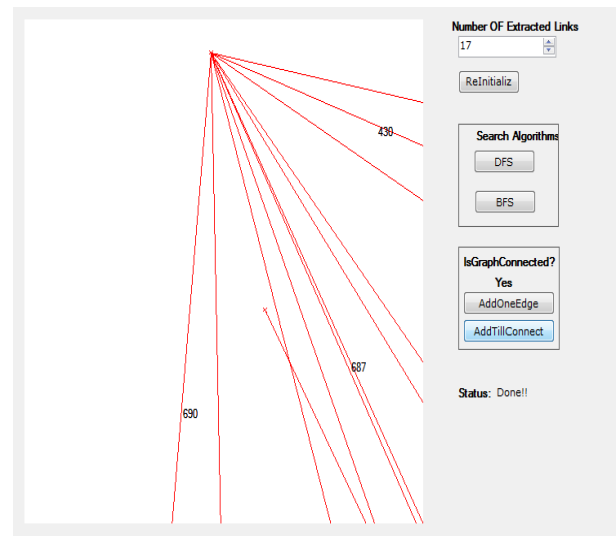


Fig. 4. Graph for Breadth First Search

The Graph for Depth First Search is shown in the Fig. 5.

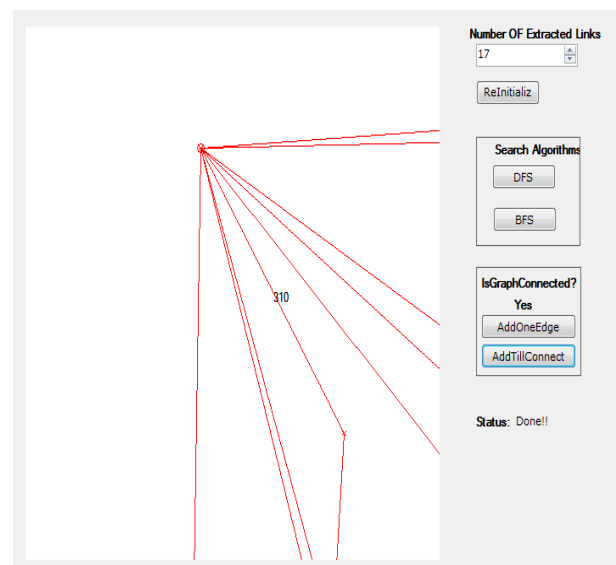


Fig. 5. Graph for Depth First Search



When we compare the performance of Web Structure Mining with these both techniques, then we analyze that Web Structure Mining with Breadth First Search gives the better performance as compared to Web Structure Mining with Depth First Search because it takes less time to extract all the links as compared to DFS.

## 5. CONCLUSION

This paper deals with Web Structure Mining with Breadth First Search and Depth First Search for extracting the links. The work in this paper analyzes the extraction of web link information using .Net Framework and standard interface. We have tested this work with different URL successfully. And we conclude that it saves the time for user to search for their query.

## 6. ACKNOWLEDGEMENT

The authors thank to Dr. Deepti Gaur for her technical support and the reviewers for their valuable suggestion so that we can improve this topic.

## 7. REFERENCES

- [1]. Mrs. V. Sujatha and Dr. Punithavalli, , September – 2011 “A Study of Web Navigation Pattern Using Clustering Algorithm in Web Log Files,” *International Journal of Scientific & Engineering Research*, Vol.- 2, Issue – 9.
- [2]. Ms Kiruthika M, Mr. Rahul Jadhav and Ms Dipa Dixit, 2011 “ Pattern Discovery Using Association Rules,” *International Journal of Advanced Computer Science and Applications*, Vol. – 2, No.- 12.
- [3]. Brijendra Singh and Hemant Kumar Singh, 2010, “Web Data Mining Research: A Survey,” .
- [4]. Prof. Marathe Dagadu Mitharam, February – 2012, “ Preprocessing in Web Usage Mining,” *International Journal of Scientific & Engineering Research*, Vol.- 3, Issue – 2, ISSN 2229-5518.
- [5]. Cooley R, Mobasher B, and Srivastava J. , 1999 , “Data preparation for Mining World Wide Web Browsing Patterns”. *Journal of Knowledge and Information Systems.*, 1: 5-32.
- [6]. Kosala R, and Blockeel H, 2000, “Web Mining Research : A Survey. *ACM SIGKDD Explorations*.
- [7] K. Pol, N. Patil, S. Patankar, and C. Das, 2008, "A Survey on Web Content Mining and Extraction of Structured and Semistructured Data", *First International Conference on Emerging Trends in Engineering and Technology, ICETET, Nagpur, India*, pp. 543-546.
- [8] V. Sathiyamoorthi, V. M. Bhaskaran, Nov. 2009, “Data Preparation Techniques for Web Usage Mining in World Wide Web – An Approach”, *International Journal of Recent trends in Engineering*, Vol 2, No 4.
- [9]. Jaideep Srivastava, Robert Cooley, Mukund Deshpande and Pang-Ning Tan, Jan.-2000, “Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data,’ *SIGKDD Explorations*, Vol.- 1, Issue – 2.
- [10]. S.Ajoudanian, and M.D. Jazi, 2009, “Deep Web Content Mining”, *World Academy of Science, Engineering and Technology*, 49.