

A Major Project Report On

**MODIFIED
WEIGHTED PAGERANK ALGORITHM**

Submitted in partial fulfilment of the requirements
for the award of the degree of

**MASTER OF TECHNOLOGY
IN
SOFTWARE ENGINEERING**

By

Ravinder Kumar

(Roll No. 2K13/SWE/16)

Under the guidance of

Dr. Kapil Sharma

Associate Professor

Department of Computer Science & Engineering

Delhi Technological University, Delhi



Department of Computer Science & Engineering

Delhi Technological University, Delhi

2013-2015



DELHI TECHNOLOGICAL UNIVERSITY
CERTIFICATE

This is to certify that the project report entitled **MODIFIED WEIGHTED PAGERANK ALGORITHM** is a bonafide record of work carried out by Ravinder Kumar (2K13/SWE/16) under my guidance and supervision, during the academic session 2013-2015 in partial fulfilment of the requirement for the degree of Master of Technology in Software Engineering from Delhi Technological University, Delhi.

To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University/Institute for the award of any Degree or Diploma.

Dr. Kapil Sharma
Associate Professor
Department of Computer Science & Engineering
Delhi Technological University
Delhi



DELHI TECHNOLOGICAL UNIVERSITY

ACKNOWLEDGEMENT

I feel immense pleasure to express my heartfelt gratitude to **Dr. Kapil Sharma** for his constant and consistent inspiring guidance and utmost co-operation at every stage which culminated in successful completion of my research work.

I also would like to thank the faculty of Software Engineering Department, DTU for their kind advice and help from time to time.

I owe my profound gratitude to my family which has been a constant source of inspiration and support.

Ravinder Kumar

Roll No. 2K13/SWE/16

TABLE OF CONTENTS

| | |
|--|---------------|
| <i>Certificate</i> | <i>I</i> |
| <i>Acknowledgement</i> | <i>II</i> |
| <i>Table Of Contents</i> | <i>III</i> |
| <i>List of Figure(s)</i> | <i>V</i> |
| <i>List of Tables(s)</i> | <i>VI</i> |
| <i>Abstract</i> | <i>VII</i> |
| CHAPTER 1: | |
| INTRODUCTION | Error! |
| Bookmark not defined. | |
| 1.1. AN OVERVIEW..... | 1 |
| 1.2. MOTIVATION OF THE WORK..... | 5 |
| 1.3. GOALS OF THE THESIS..... | 5 |
| 1.4. ORGANIZATION OF THE THESIS..... | 6 |
| CHAPTER 2: LITERATURE SURVEY | 7 |
| CHAPTER 3: PAGE RANKING ALGORITHMS : A SURVEY | 9 |
| 3.1. WEB MINING..... | 9 |
| 3.2. VARIOUS PAGE RANKING ALGORITHMS..... | 12 |
| 3.2.1. PAGE RANK ALGORITHM..... | 12 |
| 3.2.2. WEIGHTED PAGE RANK ALGORITHM..... | 16 |
| 3.2.5. PAGE RANK USING VISITS OF LINKS..... | 20 |

| | |
|--|-----------|
| 3.2.6. WIGHTEDPAGE RANK USING VISITS OF LINKS..... | 23 |
| CHAPTER 4: PROPOSED WORK..... | 28 |
| 4.1. ALGORITHM TO CALCULATE $EWPR_{VOL}$ | 29 |
| 4.2. BENEFITS OF PROPOSED METHOD | 30 |
| CHAPTER 5: RESULTS AND IMPLEMENTATION | 32 |
| CONCLUSIONS AND FUTURE WORK | 37 |
| REFERENCES | 38 |

List of Figure(s)

| Figure No | Description | Page No |
|-----------|--|---------|
| 1.1 | Working of a Crawler..... | 2 |
| 1.2 | Working of an Indexer..... | 2 |
| 1.3 | Working of a Searcher..... | 3 |
| 1.4 | Working of a Search Engine..... | 4 |
| 3.1 | Process of Web Mining..... | 9 |
| 3.2 | Categories of Web Mining..... | 9 |
| 3.3 | Structure of a Web Graph..... | 11 |
| 3.4 | Working of Page Rank Algorithm..... | 12 |
| 3.5 | A Web Graph..... | 14 |
| 3.6 | Algorithm of Weighted Page Rank Algorithm..... | 18 |
| 3.7 | Algorithm of PageRank using VOL..... | 21 |
| 3.8 | A Web Graph with VOL..... | 22 |
| 3.9 | Algorithm of WPR_{VOL} | 25 |
| 4.1 | Algorithm to calculate $EWPR_{VOL}$ | 30 |
| 5.1 | Weighted Page Rank Values..... | 33 |
| 5.2 | Weighted Page Rank Values using VOL..... | 34 |
| 5.3 | Proposed Algorithm values..... | 35 |

5.4 Comparison of WPR, WPR using VOL, Proposed Algorithm.....36

List of Table(s)

| Table No | Description | Page No |
|-----------------|--|----------------|
| 3.1 | Values of Page Ranks using PageRank Algorithm..... | 15 |
| 3.2 | Values of Page Ranks using Weighted PageRank Algorithm..... | 19 |
| 3.3: | Values of Page Ranks using PageRank using VOL Algorithm..... | 22 |
| 3.4: | Values of Page Ranks using Weighted PageRank using VOL Algorithm.. | 26 |
| 3.5 | Comparison of Page Ranking Algorithms..... | 27 |

ABSTRACT

There are billions of web pages available on the World Wide Web (WWW). So there are lots of search results corresponding to a user's query out of which only some are relevant. The relevancy of a web page is calculated by search engines using page ranking algorithms. Most of the page ranking algorithm use web structure mining and web content mining to calculate the relevancy of a web page. In this thesis, we provide an extension to standard Weighted PageRank algorithm by combining web structure mining with web usage mining. The proposed method takes into account the importance of both the number of visits of inlinks and outlinks of the pages and distributes rank scores based on the popularity of the pages. So, the resultant pages are displayed on the basis of user browsing behavior.

Keywords: World Wide Web, Search Engine, Web mining, Inlinks, Outlinks.

