

INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATION & MANAGEMENT

I
J
R
C
M



A Monthly Double-Blind Peer Reviewed (Refereed/Juried) Open Access International e-Journal - Included in the International Serial Directories

Indexed & Listed at:

Ulrich's Periodicals Directory ©, ProQuest, U.S.A., EBSCO Publishing, U.S.A., Cabell's Directories of Publishing Opportunities, U.S.A.

Open J-Gate, India [link of the same is duly available at Inlibnet of University Grants Commission (U.G.C.)].

Index Copernicus Publishers Panel, Poland with IC Value of 5.09 & number of libraries all around the world.

Circulated all over the world & Google has verified that scholars of more than 3770 Cities in 175 countries/territories are visiting our journal on regular basis.

Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA

<http://ijrcm.org.in/>

CONTENTS

Sr. No.	TITLE & NAME OF THE AUTHOR (S)	Page No.
1.	USING CYBER PEDAGOGY (WIBEKI/01/2014) MODEL TO INITIATE MULTILITERACIES AND PROMOTE A VIRTUAL CLASSROOM: A PILOT STUDY <i>WILLIAM NKOMO, BERTHA KARIMBIKA & KITSO MOLEFE</i>	1
2.	THE RIGHT TO HEALTH – A CONSTITUTIONAL VIEW <i>HIRANMAYA NANDA & DR. JAYADEV PATI</i>	11
3.	FINANCIAL PERFORMANCE OF SELECT PRIVATE SECTOR BANKS USING CAMEL APPROACH <i>DR. H N SHIVAPRASAD</i>	14
4.	A COMPARATIVE STUDY OF SELECTED EQUITY DIVERSIFIED SCHEMES IN MUTUAL FUND <i>DR. VIJAY H. VYAS</i>	24
5.	THE INFLUENCE OF INTELLIGENT TRANSPORTATION SPACES IN INTELLIGENT TRANSPORTATION SYSTEM <i>KALAISELVI S, SANGEETHALAKSHMI G & SIVASANKARI A</i>	33
6.	A STUDY ON THE SOCIO-ECONOMIC CHARACTERISTICS OF INTERNET BANKING ADOPTERS IN CHENNAI METROPOLITAN CITY WITH REFERENCE TO INDIAN BANK <i>P.SARAVANAN & P.SRIDHARAN</i>	37
7.	COMPARATIVE STUDY OF NEW RAPID BUSINESS PROCESS MODEL WITH EXISTING MODEL BPMN AND UML-AD <i>AMIT LAXMIDAS VADERA & DR. YOGESH R. GHODASARA</i>	42
8.	A DETAILED STUDY ON QUALITY OF SERVICE IN COMPUTER NETWORKS <i>HARIPRIYA N, SANGEETHALAKSHMI G & SIVASANKARI A</i>	48
9.	TATA GROUP AND CSR: AN EXEMPLARY CASE REVIEW <i>KOMAL CHAUDHARY</i>	52
10.	THREE DIMENSIONAL HEALING: BENEFITS FROM THE WELLNESS <i>DR. VANDANA DESWAL</i>	55
11.	EMOTIONAL INTELLIGENCE AND JOB PERFORMANCE IN SERVICE INDUSTRY <i>PREETI BHASKAR</i>	60
12.	AN OVERVIEW OF THE BANKING INDUSTRY IN INDIA <i>DR. SHILPAN D. VYAS & PARINA S. VYAS</i>	66
13.	COUNTERFEITING GOODS IN GULF BUSINESS: ANY ECONOMIC IMPACT? <i>DR. THRESIAMMA VARGHESE & KARIMA AL. QARTOOPI</i>	74
14.	GREEN MARKETING: AN INDIAN EXPERIENCE <i>KANCHAN SEHRAWAT, AMOGH TALAN, DR. A. K. TYAGI & GAURAV TALAN</i>	77
15.	ROLE OF RBI AND GOVERNMENT OF INDIA TOWARDS FINANCIAL INCLUSION OF THE RURAL POOR: ISSUES AND SUGGESTIONS <i>MANOHAR LAMANI & SANGANAGOUDA PATIL</i>	81
16.	CORPORATE SOCIAL RESPONSIBILITY: REGULATION AND ITS SURVEILLANCE <i>RACHANA VISHWAKARMA</i>	85
17.	PAGE RANK ALGORITHMS BASED ON WEB CONTENT MINING AND WEB STRUCTURE MINING <i>N.KANCHANA</i>	90
18.	WEB CONTENT MANAGEMENT SYSTEM: COMPONENTS AND SECURITY <i>OMOSEBI, PAUL ADEOYE & OLORUNLEKE, FEHINTOLUWA E.</i>	93
19.	DETERMINANTS AND PROSPECTS OF ECONOMIC GROWTH IN ETHIOPIA <i>HABTAMU NIGATU ELALA</i>	96
20.	HIGHLY SECURED LOSSLESS IMAGE CRYPTOGRAPHY ALGORITHM BASED ON HAAR WAVELET TRANSFORM <i>MAHIMA GUPTA</i>	105
	REQUEST FOR FEEDBACK & DISCLAIMER	108

CHIEF PATRON

PROF. K. K. AGGARWAL

Chairman, Malaviya National Institute of Technology, Jaipur
(An institute of National Importance & fully funded by Ministry of Human Resource Development, Government of India)
Chancellor, K. R. Mangalam University, Gurgaon
Chancellor, Lingaya's University, Faridabad
Founder Vice-Chancellor (1998-2008), Guru Gobind Singh Indraprastha University, Delhi
Ex. Pro Vice-Chancellor, Guru Jambheshwar University, Hisar

FOUNDER PATRON

LATE SH. RAM BHAJAN AGGARWAL

Former State Minister for Home & Tourism, Government of Haryana
Former Vice-President, Dadri Education Society, Charkhi Dadri
Former President, Chinar Syntex Ltd. (Textile Mills), Bhiwani

CO-ORDINATOR

DR. SAMBHAV GARG

Faculty, Shree Ram Institute of Business & Management, Urjani

ADVISORS

DR. PRIYA RANJAN TRIVEDI

Chancellor, The Global Open University, Nagaland

PROF. M. S. SENAM RAJU

Director A. C. D., School of Management Studies, I.G.N.O.U., New Delhi

PROF. S. L. MAHANDRU

Principal (Retd.), Maharaja Agrasen College, Jagadhri

EDITOR

PROF. R. K. SHARMA

Professor, Bharti Vidyapeeth University Institute of Management & Research, New Delhi

EDITORIAL ADVISORY BOARD

DR. RAJESH MODI

Faculty, Yanbul Industrial College, Kingdom of Saudi Arabia

PROF. PARVEEN KUMAR

Director, M.C.A., Meerut Institute of Engineering & Technology, Meerut, U. P.

PROF. H. R. SHARMA

Director, Chhatrapati Shivaji Institute of Technology, Durg, C.G.

PROF. MANOHAR LAL

Director & Chairman, School of Information & Computer Sciences, I.G.N.O.U., New Delhi

PROF. ANIL K. SAINI

Chairperson (CRC), Guru Gobind Singh I. P. University, Delhi

PROF. R. K. CHOUDHARY

Director, Asia Pacific Institute of Information Technology, Panipat

DR. ASHWANI KUSH

Head, Computer Science, University College, Kurukshetra University, Kurukshetra

DR. BHARAT BHUSHAN

Head, Department of Computer Science & Applications, Guru Nanak Khalsa College, Yamunanagar

DR. VIJAYPAL SINGH DHAKA

Dean (Academics), Rajasthan Institute of Engineering & Technology, Jaipur

DR. SAMBHAVNA

Faculty, I.I.T.M., Delhi

DR. MOHINDER CHAND

Associate Professor, Kurukshetra University, Kurukshetra

DR. MOHENDER KUMAR GUPTA

Associate Professor, P.J.L.N. Government College, Faridabad

DR. SAMBHAV GARG

Faculty, Shree Ram Institute of Business & Management, Urjani

DR. SHIVAKUMAR DEENE

Asst. Professor, Dept. of Commerce, School of Business Studies, Central University of Karnataka, Gulbarga

DR. BHAVET

Faculty, Shree Ram Institute of Business & Management, Urjani

ASSOCIATE EDITORS

PROF. ABHAY BANSAL

Head, Department of Information Technology, Amity School of Engineering & Technology, Amity University, Noida

PROF. NAWAB ALI KHAN

Department of Commerce, Aligarh Muslim University, Aligarh, U.P.

ASHISH CHOPRA

Sr. Lecturer, Doon Valley Institute of Engineering & Technology, Karnal

TECHNICAL ADVISOR

AMITA

Faculty, Government M. S., Mohali

FINANCIAL ADVISORS

DICKIN GOYAL

Advocate & Tax Adviser, Panchkula

NEENA

Investment Consultant, Chambaghat, Solan, Himachal Pradesh

LEGAL ADVISORS

JITENDER S. CHAHAL

Advocate, Punjab & Haryana High Court, Chandigarh U.T.

CHANDER BHUSHAN SHARMA

Advocate & Consultant, District Courts, Yamunanagar at Jagadhri

SUPERINTENDENT

SURENDER KUMAR POONIA

CALL FOR MANUSCRIPTS

We invite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the areas of Computer Science & Applications; Commerce; Business; Finance; Marketing; Human Resource Management; General Management; Banking; Economics; Tourism Administration & Management; Education; Law; Library & Information Science; Defence & Strategic Studies; Electronic Science; Corporate Governance; Industrial Relations; and emerging paradigms in allied subjects like Accounting; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Rural Economics; Co-operation; Demography; Development Planning; Development Studies; Applied Economics; Development Economics; Business Economics; Monetary Policy; Public Policy Economics; Real Estate; Regional Economics; Political Science; Continuing Education; Labour Welfare; Philosophy; Psychology; Sociology; Tax Accounting; Advertising & Promotion Management; Management Information Systems (MIS); Business Law; Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labour Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; International Relations; Human Rights & Duties; Public Administration; Population Studies; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism & Hospitality; Transportation Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic; Web Design and emerging paradigms in allied subjects.

Anybody can submit the **soft copy** of unpublished novel; original; empirical and high quality **research work/manuscript** **anytime** in **M.S. Word format** after preparing the same as per our **GUIDELINES FOR SUBMISSION**; at our email address i.e. infoijrcm@gmail.com or online by clicking the link **online submission** as given on our website ([FOR ONLINE SUBMISSION, CLICK HERE](#)).

GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION:**

DATED: _____

THE EDITOR
IJRCM

Subject: SUBMISSION OF MANUSCRIPT IN THE AREA OF

(e.g. Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)

DEAR SIR/MADAM

Please find my submission of manuscript entitled '_____ ' for possible publication in your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published elsewhere in any language fully or partly, nor is it under review for publication elsewhere.

I affirm that all the author (s) have seen and agreed to the submitted version of the manuscript and their inclusion of name (s) as co-author (s).

Also, if my/our manuscript is accepted, I/We agree to comply with the formalities as given on the website of the journal & you are free to publish our contribution in any of your journals.

NAME OF CORRESPONDING AUTHOR:

Designation:
Affiliation with full address, contact numbers & Pin Code:
Residential address with Pin Code:
Mobile Number (s):
Landline Number (s):
E-mail Address:
Alternate E-mail Address:

NOTES:

- a) The whole manuscript is required to be in **ONE MS WORD FILE** only (pdf. version is liable to be rejected without any consideration), which will start from the covering letter, inside the manuscript.
- b) The sender is required to mention the following in the **SUBJECT COLUMN** of the mail:
New Manuscript for Review in the area of (Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)
- c) There is no need to give any text in the body of mail, except the cases where the author wishes to give any specific message w.r.t. to the manuscript.
- d) The total size of the file containing the manuscript is required to be below **500 KB**.
- e) Abstract alone will not be considered for review, and the author is required to submit the complete manuscript in the first instance.
- f) The journal gives acknowledgement w.r.t. the receipt of every email and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of manuscript, within two days of submission, the corresponding author is required to demand for the same by sending separate mail to the journal.

2. **MANUSCRIPT TITLE:** The title of the paper should be in a 12 point Calibri Font. It should be bold typed, centered and fully capitalised.

3. **AUTHOR NAME (S) & AFFILIATIONS:** The author (s) **full name, designation, affiliation (s), address, mobile/landline numbers, and email/alternate email address** should be in italic & 11-point Calibri Font. It must be centered underneath the title.

4. **ABSTRACT:** Abstract should be in fully italicized text, not exceeding 250 words. The abstract must be informative and explain the background, aims, methods, results & conclusion in a single para. Abbreviations must be mentioned in full.

5. **KEYWORDS:** Abstract must be followed by a list of keywords, subject to the maximum of five. These should be arranged in alphabetic order separated by commas and full stops at the end.
6. **MANUSCRIPT:** Manuscript must be in **BRITISH ENGLISH** prepared on a standard A4 size **PORTRAIT SETTING PAPER**. It must be prepared on a single space and single column with 1" margin set for top, bottom, left and right. It should be typed in 8 point Calibri Font with page numbers at the bottom and centre of every page. It should be free from grammatical, spelling and punctuation errors and must be thoroughly edited.
7. **HEADINGS:** All the headings should be in a 10 point Calibri Font. These must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.
8. **SUB-HEADINGS:** All the sub-headings should be in a 8 point Calibri Font. These must be bold-faced, aligned left and fully capitalised.
9. **MAIN TEXT:** The main text should follow the following sequence:

INTRODUCTION**REVIEW OF LITERATURE****NEED/IMPORTANCE OF THE STUDY****STATEMENT OF THE PROBLEM****OBJECTIVES****HYPOTHESES****RESEARCH METHODOLOGY****RESULTS & DISCUSSION****FINDINGS****RECOMMENDATIONS/SUGGESTIONS****CONCLUSIONS****SCOPE FOR FURTHER RESEARCH****ACKNOWLEDGMENTS****REFERENCES****APPENDIX/ANNEXURE**

It should be in a 8 point Calibri Font, single spaced and justified. The manuscript should preferably not exceed **5000 WORDS**.

10. **FIGURES & TABLES:** These should be simple, crystal clear, centered, separately numbered & self explained, and **titles must be above the table/figure. Sources of data should be mentioned below the table/figure.** It should be ensured that the tables/figures are referred to from the main text.
11. **EQUATIONS:** These should be consecutively numbered in parentheses, horizontally centered with equation number placed at the right.
12. **REFERENCES:** The list of all references should be alphabetically arranged. The author (s) should mention only the actually utilised references in the preparation of manuscript and they are supposed to follow **Harvard Style of Referencing**. The author (s) are supposed to follow the references as per the following:
 - All works cited in the text (including sources for tables and figures) should be listed alphabetically.
 - Use (ed.) for one editor, and (ed.s) for multiple editors.
 - When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order.
 - Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
 - The title of books and journals should be in italics. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
 - For titles in a language other than English, provide an English translation in parentheses.
 - The location of endnotes within the text should be indicated by superscript numbers.

PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:**BOOKS**

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

CONTRIBUTIONS TO BOOKS

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

JOURNAL AND OTHER ARTICLES

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

CONFERENCE PAPERS

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19-22 June.

UNPUBLISHED DISSERTATIONS AND THESES

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, Kurukshetra University, Kurukshetra.

ONLINE RESOURCES

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

WEBSITES

- Garg, Bhavet (2011): Towards a New Natural Gas Policy, Political Weekly, Viewed on January 01, 2012 <http://epw.in/user/viewabstract.jsp>

PAGE RANK ALGORITHMS BASED ON WEB CONTENT MINING AND WEB STRUCTURE MINING

N.KANCHANA
ASST. PROFESSOR & RESEARCH SCHOLAR
SCHOOL OF INFORMATION TECHNOLOGY & SCIENCE
DR.G.R.DAMODARAN COLLEGE OF SCIENCE
COIMBATORE

ABSTRACT

The Web is a place where we can find lot of information. The main aim of the web site is to provide relevant information to the users to satisfy their needs. A major challenge in web mining research is to find relevant document or pages by neglecting the noise and outliers. In this paper we compare different type of page rank algorithms that supports web content mining and web structure mining.

KEYWORDS

Web Mining, Page Rank, Web Content Mining, Web Structure Mining, Outliers.

I. INTRODUCTION

Web Mining is one of the application of data mining techniques to extract knowledge from web content. The WWW is a place in which the information and contents expand in size and complexity. Retrieving the required web page on the WWW efficiently and easily is becoming a major challenge. The Huge amount of information becomes very difficult for the users to find, extract, filter or evaluate the relevant web page. This challenge rises to find some technique to solve these problems. Most of the popular search engine such as Google, yahoo, Amazon etc are famous because of their crawling and ranking methods. These search engines download, index and store millions of web pages. So, web content mining and web structure mining becomes very important for effective information retrieval. Existing web mining algorithms do not consider documents containing outliers. Generally, Outlier mining is dedicated to finding data objects, which differ significantly from the rest of the data. Outlier mining has been extensively studied in statistics and recently data mining. However, exploring the Web for outliers has received very little attention in the mining community. Web content outliers are documents with 'varying contents' compared to similar Web documents taken from the same domain. Mining Web content outliers may lead to the identification of competitors and emerging business patterns in electronic commerce. This paper proposes Signed approach page rank algorithm for mining Web content outliers using a domain dictionary.

II. WEB MINING

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 site etc.

A. Categories of Web Mining

Web Mining is broadly divided into three distinct categories, Web Content Mining, Web Structure Mining, and Web Usage Mining.

1) *Web Content Mining*: Web Content Mining is the process of extracting useful information from the contents of web page. The web page is unstructured and contains a mix of text, images, audio, video or structures records such as tables and lists. Web Content Mining is concerned with the retrieval of information from WWW into more structure form and indexing the information to retrieve it quickly. The technologies that are normally used in web content mining are Natural Language Processing and Information Retrieval.

2) *Web Structure Mining*: Web Structure Mining is the process of discovering structure information from the WWW and finds the similarity and relationships between different web sites. It is the processes by which we can also find the link structure of hyper links at inter document level and the intra document level. Page Rank also falls in to this category.

3) *Web Usage Mining*: Web Usage Mining is the process of data mining techniques to predict the user behavior while the user interacts with the web. Usage data captures the identity or origin of web users along with their browsing behavior at a web site. It extracts data stored in server access logs, referrer logs, agent logs, client-side cookies, user profile and meta data. It also uses the secondary data on the web.

III. KEY CONCEPTS**A. Outlier Mining**

Outliers are the data that obviously deviate from others, disobey the general mode or behavior of data and disaccord with other existing data. Web content outliers are such as noise, irrelevant and redundant page from the web documents. Also, Web content outliers mining can be used to determine pages with entirely different contents from their parent web sites.

B. Page Rank

Ranking is important as it helps the users to look for "Quality" pages that are relevant to the query. Page Rank is a numeric value that represents how important a page is on the web. It is the Google's method of measuring a page's "important" when all other factors such as Title tag and Keywords are taken into account, Google uses page rank to adjust results so that more "Important" pages move up in the results page of a user search result display.

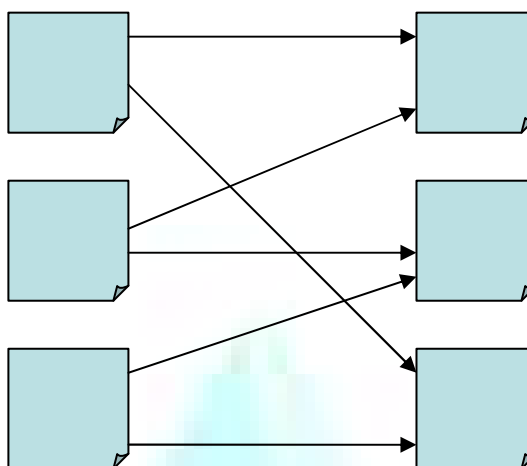
IV. PAGE RANK ALGORITHMS

With the increasing number of web pages and users on the web, the number of queries submitted to the search engines are also increasing rapidly. Therefore, the search engines needs to be more efficient in its process. Web mining techniques are employed by the search engines to extract relevant documents from the web database and provide the necessary information to the users. The search engines become very successful and popular if they use efficient ranking mechanism. Page ranking algorithms are used by the search engines to present the search results by considering the relevance, importance and content score and web mining techniques to order them according to the user interest. Some ranking algorithms depend only on the link structure of the documents i.e. their popularity scores (Web Structure Mining), whereas others look for the actual content in the documents (Web Content Mining), while some use a combination of both i.e. they use content of the document as well as the link structure to assign a rank value for a given document. If the search results are not displayed according to the user interest then the search engine will loose its popularity. So the ranking algorithms become very important. There are number of algorithms proposed based on Web Content Mining and Web Structure Mining. Three important algorithms HITS (Hyper-link Induced Topic Search), Weighted Page Content Rank, and Signed Approach Page Rank algorithm.

A.HITS

Klienberg gives two forms of web pages called as hubs and authorities. Hubs are the pages that act as resource lists. Authorities are pages having important contents. A good hub page is a page which is pointing to many authoritative pages on that content and a good authority page is a page which is pointed by many good hub pages on the same content. A page may be a good hub and a good authority at the same time[8,9]. The HITS algorithm treats WWW as directed graph $G(V,E)$, where V is a set of vertices representing pages and E is set of edges corresponds to link. Figure 1 shows the hubs and authorities in web .

FIG. 1 HUBS AND AUTHORITIES



Hubs Authorities

It has two steps:

1. Sampling Step:- In this step a set of relevant pages for the given query are collected.
 2. Iterative Step:- In this step Hubs and Authorities are found using the output of sampling step.
- Following expressions (1,2) are used to calculate the weight of Hub (H_p) and the weight of Authority (A_p).

$$H_p = \sum_{q \in I(p)} A_q \dots\dots\dots(1)$$

$$A_p = \sum_{q \in B(p)} H_q \dots\dots\dots(2)$$

where H_q is Hub Score of a page, A_p is authority score of a page, $I(p)$ is set of reference pages of page p and $B(p)$ is set of referrer pages of page p, the authority weight of a page is proportional to the sum of hub weights of pages that link to it. Similarly a hub of a page is proportional to the sum of authority weights of pages that it links to.

1) Constraints with HITS Algorithm

Following are some constraints of HITS algorithm

- Hubs and authorities: It is not easy to distinguish between hubs and authorities because many sites are hubs as well as authorities.
- Topic drift: Sometime HITS may not produce the most relevant documents to the user queries because of equivalent weights.
- Automatically generated links: HITS gives equal importance for automatically generated links which may not have relevant topics for the user query.
- Efficiency: HITS algorithm is not efficient in real time.

HITS was used in a prototype search engine called Clever for an IBM research project. Because of the above constraints HITS could not be implemented in a real time search engine.

B. Weighted Page Content Rank

Weighted Page Content Rank Algorithm (WPCR) is a proposed page ranking algorithm which is used to give a sorted order to the web pages returned by a search engine in response to a user query. WPCR is a numerical value based on which the web pages are given an order. This algorithm employs web structure mining as well as web content mining techniques. Web structure mining is used to calculate the importance of the page and web content mining is used to find how much relevant a page is? Importance here means the popularity of the page i.e. how many pages are pointing to or are referred by this particular page. It can be calculated based on the number of in links and out links of the page. Relevancy means matching of the page with the fired query. If a page is maximally matched to the query, that becomes more relevant.

Algorithm: WPCR calculator

Input: Page P, In link and Out link Weights of all back links of P, Query Q, d (damping factor).

Output: Rank score

Step 1: Relevance calculation:

- a) Find all meaningful word strings of Q (say N)
- b) Find whether the N strings are occurring in P or not?
- Z= Sum of frequencies of all N strings.
- c) S= Set of the maximum possible strings occurring in P.
- d) X= Sum of frequencies of strings in S.
- e) Content Weight (CW)= X/Z
- f) C= No. of query terms in P
- g) D= No. of all query terms of Q while ignoring stop words.
- h) Probability Weight (PW)= C/D

Step 2: Rank calculation:

- a) Find all back links of P (say set B).
- b) $PR(P)=(1-d)+d (PR(T1)/C(T1)+.....PR(Tn)/C(Tn))$
- c) Output PR(P) i.e. the Rank score

Using the Weighted Page Content Rank can get relevant and important pages easily as it employs web structure mining and web content mining. The input parameters used in Page Rank are Backlinks, Weighted Page Rank uses Backlinks and Forward Links as Input Parameter and Weighted Page Content Rank uses Backlinks, Forward Link and Content as Input Parameters.

C. Signed Approach Page Rank Algorithm

In the proposed Algorithm, web documents are extracted from the search engines by giving query by the user to the web. Then the obtained web documents D is preprocessed, i.e., stop words, stem words and except text other data such as hyperlinks, sound, images etc are removed. The output is a set of documents with white-spaced separated words and it is indexed in two dimensional format (i,j), where 'i' represent web pages and 'j' represent words. Therefore, first word from first web page is indexed as (1,1), second word from the first page is indexed as (1,2) etc,. The domain dictionary is arranged in such a way that, all 1-letter word will be indexed first, followed by 2-letter words, then 3-letter words similarly up to 15-letters word which is a very reasonable upper bounds for number of characters in a word. Each page is mined individually to detect relevant and irrelevant documents using signed approach. Finally, a relevant web document is obtained which contains required information catering to the user needs.

The proposed algorithm explores the advantages of full word matching and signed approach using organized domain dictionary where the indexing is done based on the length of the word. First, the input web document is preprocessed and separated into white spaced words. The full word profile for the document is generated in matrix form (i.e., $w_{1,4}$ - represents 4th word in 1st page). Then the j^{th} word from i^{th} page is taken and its length is calculated (i.e., $|W_{ji}|$) and depending on the number of characters, the respective index on the domain dictionary is searched. Using Binary search If the word (w_{ji}) is found in the dictionary, then positive count is incremented by one else negative count is incremented by one. This process is carried out for all words in that web page. Finally, positive count is compared with the negative count to check the relevancy of that web page. If the positive count is less than the negative count, then that page is irrelevant, otherwise it is considered as more relevant and rank is calculated for that page.

Algorithm : Signed Approach Page Rank

Input: Domain Dictionary, Web Document D, In link and Out link Weights of all back links of P.

Output: Relevant Pages, Rank Score and Irrelevant Pages.

OtherVariable:Pos_count, Neg_count

```

a) Extract the input web document D after preprocessing. Where
b) D contains Pages P1,P2,.....Pn.
c) Read the contents of web pages P1 to Pn
d) Generate full word profile in to matrix for Wij (Where i represent Page and j represent Word)
for ( i=1;i<=n;i++)
{
Pos_count=0; Neg_count=0;
for(j=1;j<=m;j++)
{
Using BINARY SEARCH find W(i,j) word exists in dictionary
If so { Pos_count++;}
else { Neg_count++; }
if (Pos_count++>=Neg_count)
{
i) Print Pi as relevant web page ;
ii) Find all back links of Pi (say set B).
iii) Calculate the Rank using the formula PR(Pi)
=(1-d)+d(PR(T1)/C(T1)+.....PR(Tn)/C(Tn))
iv) Output PR(Pi) i.e. the Rank score } else Print Pi as irrelevant web page.

```

This signed approach ensure that memory space, search time and run time gets reduced by using domain dictionary, Binary search, and rank calculation than other approaches for checking the relevancy of the web documents. As the efficiency of web content is increased, the quality of the search engines also gets increased. This method is very simple to implement. The proposed method is can be used by business personals to keep track of all the positive and negative aspects related to their business.

V. COMPARISON OF VARIOUS PAGE RANK ALGORITHMS

Table1 shows the difference between above three algorithms

TABLE I: COMPARISON OF HITS, WEIGHTED PAGE RANK AND SIGNED APPROACH PAGE RANK

Content	HITS	Weighted Page Content Rank	Signed Approach Page Rank Algorithm
Mining Technique used	WSM and WCM	WSM and WCM	WSM and WCM
Complexity	<O(log N)	<O(log n)	<O(log n)
Working Procedure	Computes scores of n highly relevant pages on the fly	Gives sorted order to the web pages returned by a search engine as a numerical value in response to a user query	Gives sorted order to the web pages returned by a search engine as a numerical value in response to a user query and domain dictionary
Input/output parameters	Back links, Forward links and content	Backlinks, forward links and content	Backlinks, forward links and content
Advantages	Relatively small, some times rich in relevant pages about the query	It provide important information and relevancy about a given query by using web structure and web content mining	It provide important information and relevancy about a given query by using web structure and web content mining
Search Engine	IBM Search engine Clever	Research Model	Research Model
Limitations	Topic drift and efficiency problem	No limitation best as comparison to Page Rank and Weighted Page Rank	Needs to work along with page rank and best as compared to Weighted page content

VI. CONCLUSION

Web mining is the Data Mining technique that automatically discovers or extracts the information from web documents. HITS, Weighted Page Rank algorithms, and Signed approach page rank are used in Web Structure Mining to rank the relevant pages. In this paper HITS and Weighted Page Rank algorithms may not produces the required relevant documents easily, but in the new algorithm Singed approach page rank user can get relevant and important pages easily as it employs web structure mining and web content mining. As part of the future work is to carry out performance analysis for Singed approach page rank and to work on finding required relevant and important pages more easily and quickly using any mathematical tools for mining the web content.

REFERENCES

1. Ashutosh Kumar Singh, Ravi Kumar P, "A Comparative study of Page Ranking Algorithms for Information Rerieval" International Journal of Electrical and Computer Engineering 4:7 2009.
2. G. Poonkuzhali, K.Thiagarajan, K.Sarukesi and G.V.Uma, "Singed Approach for Mining Web Content Outliers", World Academy of Science, Engineering and Technology 56,2009.
3. Rekha Jain, Dr G.N.Purohit, "Page Ranking Algorithms for Web Mining," International Journal of Computer application,Vol 13, Jan 2011.
4. Tamanna Bhatia, "Link Analysis Algorithms for Web Mining," International Journal of Computer application,Vol 2, June 2011.
5. Wenpu Xing and Ali Ghorbani, "Weighted Page Rank Algorithm", Proceedings of the Second Annual Conference on Communication Networks and Services Research (CNSR '04), IEEE, 2004.

REQUEST FOR FEEDBACK

Dear Readers

At the very outset, International Journal of Research in Computer Application & Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you to supply your critical comments and suggestions about the material published in this issue as well as on the journal as a whole, on our E-mail infoijrcm@gmail.com for further improvements in the interest of research.

If you have any queries please feel free to contact us on our E-mail infoijrcm@gmail.com.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

Sd/-
Co-ordinator

DISCLAIMER

The information and opinions presented in the Journal reflect the views of the authors and not of the Journal or its Editorial Board or the Publishers/Editors. Publication does not constitute endorsement by the journal. Neither the Journal nor its publishers/Editors/Editorial Board nor anyone else involved in creating, producing or delivering the journal or the materials contained therein, assumes any liability or responsibility for the accuracy, completeness, or usefulness of any information provided in the journal, nor shall they be liable for any direct, indirect, incidental, special, consequential or punitive damages arising out of the use of information/material contained in the journal. The journal, nor its publishers/Editors/Editorial Board, nor any other party involved in the preparation of material contained in the journal represents or warrants that the information contained herein is in every respect accurate or complete, and they are not responsible for any errors or omissions or for the results obtained from the use of such material. Readers are encouraged to confirm the information contained herein with other sources. The responsibility of the contents and the opinions expressed in this journal is exclusively of the author (s) concerned.

ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

Our Other Journals

