

Phrase Prioritization Algorithm With Supporting Data Structure And Its Implementation

Sachin Kumar, Krishna Prasad K

Abstract: Search Engines are the most common data on the World Wide Web Although Search Engine Optimization deals with growing visibility of search results web pages, this study work focuses on the search algorithm that will provide the most appropriate, relevant and needed records with the anticipated search query. A comparative research of the intersection algorithm and its variants provided the requirement for the fresh algorithm. CRANTOP database is a conventional WWW database that not only provides a database of documents and queries, but also provides appropriate queries for real appropriate documents. The study was conducted on that database, available in the Appendix, and the analysis is based on the values of precision, recall and F-measure. The proposal algorithm, which preference of the search queries, is at the core of the study job. This algorithm indicates fetching phrases methodology and then looking for all possible sentences in order to increase the recall value. In this respect, the most significant problem is the use of such sparse matrix representation, which promotes the effective search of sentences into the papers. For doing it, sparse matrix representation was suggested, which contains Columns corresponding to the papers, making searching for any possible sentence effective. Rows Linked Representations preserve the vibrant nature of papers and insert and delete phrases from papers. In addition, emphasis was placed on general words and query words into the dictionary, resulting in precision search outcomes. Synonyms were also regarded for document retrieval, thereby enhancing comprehension.

Index Terms: Phrase prioritization, Algorithm, Data structure, CRANTOP database, SEO, Search Engine, SPARSE

1. INTRODUCTION

SEO is study of methods & methodologies of search engine outcomes designed to influence or enhance the visibility of internet papers. SEO Theory, an informal discipline, is a natural outcome of the extensive interest in SEO and exercise. Despite significance in search engine in the internet, very little scholarly study has performed on them. It usually operates on delivering "spider" and "internet crawler" returning to everyone files this might discover. Based on its word content, each return document is read and indexed. The indices are developed using an algorithm to return outcomes for query only in most instances. The vertical search engines are websites that concentrate on specific subjects and can be specifically linked to those subjects. The word "vertical" derives from the concept that of horizontally of subjects, there are areas to search, we find vertically within a narrow group of concern. Peters S. et.al [2011] only one that includes content particular to certain consumers can be described as a Search Engines. They currently dominate today's online search industry, but there is a increasing popularity of unique search engines for niche markets. LookSmart is one of the largest unique engines at the moment, while VSE are not new, their enhanced popularity is what has changed. There is a typical 31.9 percent search failure rate among company users on broad-based search engines.

2 LITERATURE review

Through a thorough review of KM literature and the findings of researchers resources, technology and other factors in the organization being surveyed on the data collection. Changiz V. et.al [2016] This compiles literature highlighting possible references to CSFs for projects implementing ERP. There is a deep knowledge of different CSFs that other scientists have already recognized, "content assessment" is used. For literature, four phases of content analysis,. Shahin D.et. al[2009] Nearly all organisations that seek to satisfy market and customer dynamic requirements and plan to integrate innovations into their products and services. The new strategy can be used to determine and prioritize development variables based on the point of view of clients, while traditional methods were based solely on the point of view of technicians and developers..Arash S. et. al [2013] Search engines have developed to be more secure to the requirements of individual consumers, including personal preferences and place of the person..Zoe C. [2016].This was done among university learners and questionnaires distributed at least once a month to learners in the library service. A model created for different libraries was the foundation of the questionnaire. The features / quality elements with fuzzy methods were prioritized to satisfy student and scholarly libraries demands. In improving the use of this strategy to academic libraries, books are expected to be able to meet customer requirements and improvements. Reza J. et. al [2009] This offers a clear and straightforward way to evaluate and manage a portfolio of IT projects. The ultimate objective is a significant project decrease. There is widespread use of the term "alignment of company and technology objectives." But how do you "align" these intangibles? And what is meant by "alignment" when it comes to IT projects? Most businesses are struggling with these questions The true issue is that on a distinct project basis they are dealing with the problems. The use of one or more of the methods mentioned here has given clear, definable advantages. The rates of failure are very good. Vin D.[2005] RDM Development of policies and services has developed a complicated collection of LAN within and beyond organisations involving various professional groups. Data sharing is regarded a major policy and service activity, but its prominence can be ascribed to big study funders through the

- Sachin Kumar is currently working in Hierank Business School, Noida Sector-62 as Associate Professor & HOD-IT(BCA) which is affiliated to CCS University, India. Email:sachin.suryan2005@gmail.com
- Co-Author Krishna Prasad K, is currently working in College of Computer and Information Science Srinivas University, Mangalore, Karnataka, India as Associate Professor.

adoption of positions. Rosie H. et.al [2015] Today, organizations and people need a extensive policy on virus protection to address the increasing threats posed by Internet computer viruses. Because of the variety of ideas used in XML resources, selecting a right query is not simple. It's about capturing the concepts based on the all data collection. It enables very good selection of concepts, usually when there are many different types of information in a structured XML document. Keng H. G. et. al[2015]. On the results that internet users often depend highly on the search engine ranking without paying much attention to the web pages ' credibility. In addition, SEO shows outcomes in different scenario other than may promote credibility assessments of searchers. It offers perspectives for scientists, search engine developers, teachers, and learners. Kemmerer Y et.al [2012]. Social search definition is these search engines that analyzed and characterized by its particular characteristics. Social search and social search engine comparison summarize the many methods in which individuals can search the internet. Although various attempts have been made in the past to define social search, it presents an argument that unifies some current definitions and is the other social interpretation of social expectations. It is a comprehensive review of the various social search engine kinds. Burghardt M et.al[2012].

3 RELATED WORK

On queries, which were selected from the conventional crantop database, comparative analysis of very easy intersection algorithms was performed. More sophisticated algorithms are in use, but by displaying the computed values of Precision, Recall and F-Score, the easy algorithms were chosen to emphasize the need for a fresh algorithm. These values are very small, showing the need to create a new document search concept. A novel algorithm was suggested based on Phrase Prioritization. The phrases have been given emphasis and priority, so that the user who types more than one word can be considered together and the phrases that contain the documents have a higher priority, which is usually the case. It has been suggested a data structure that supports the sentence, which is the connected application of a sparse matrix that maps the nature of the connection between papers and dictionary phrases. Furthermore, emphasizing the need to categorize words in particular and general terms is also a very fresh and useful concept, the common words that compare them in context with those papers A dictionary is a collection of different documents The software like MS-word or MS-Excel will convert sentences, sentences, stop words and paragraphs into a single word and put them sort alphabetically by the sort command in MS-Excel. By doing so, a dictionary of words should be formed, and this is a method of representing links. The page must be placed vertically and the records should be horizontal. Here are just some of the primary extra stuff to do.

1) The study would like to determine the set of terms and provide retrieval, i.e., spelling tolerant and word selection

inconsistent. 2) Searching for compounds or sentences referring to a notion such as ' operating software systems ' is often helpful. It may also want to do proximity queries The index must be expanded to capture the closeness of terms in papers to respond to such queries. 3) A Boolean model records only word or lack, but often the study would like to collect proof, providing more weight to papers that have a word many times than those containing only one time. To do this, the study requires data on the term frequency in the list of posts. 4) Boolean queries only obtain a set of matching records, but the study is usually an efficient technique of ordering (or "ranking") the findings returned. This needs a system to determine how well a document score is encapsulated. Christopher D. M. et.al[2005].

4 SPARSE REPRESENTATION OF THE MATRIX

Here as the study enters a fresh stage, the study introduces the sparse matrix representation idea. The one with many entries as zero is a sparse matrix.

Example

If there are three documents.

Doc 1:- This is a unique representation of complete range of literature.

Doc 2:- A sops and powders developing organization have many retailers.

Doc 3:- All retailers have one or more queue.

It is its sparse representation of the matrix which is pictorially showing below.

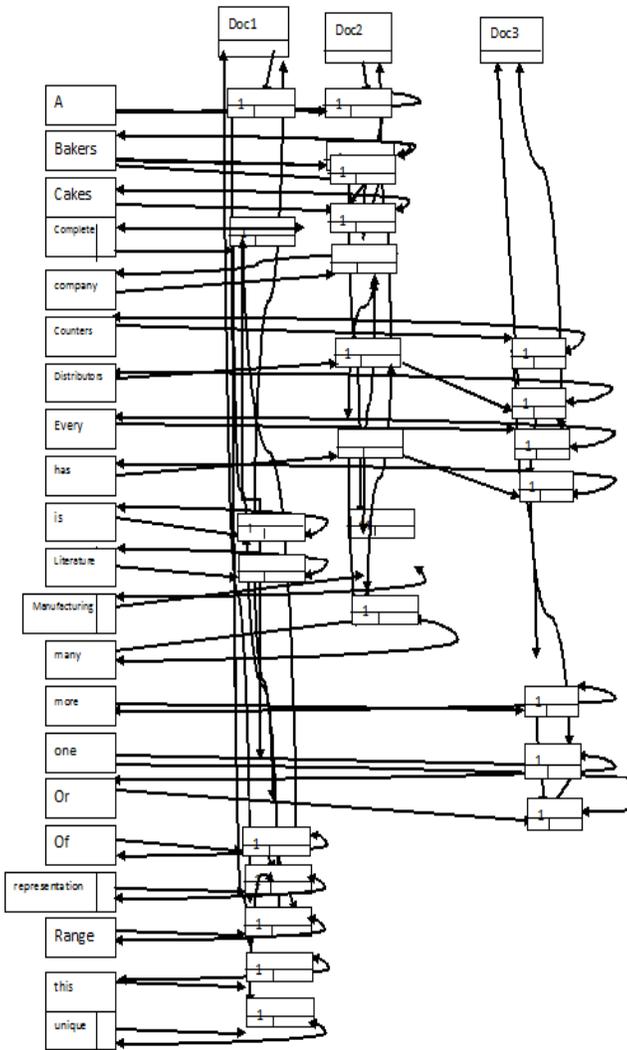
5 ALGORITHMS POINT

1. Break Query
2. Phrase Queries
3. Two –Words Phrases
4. Three –Words Phrases
5. Remove Stop words
6. Convert Remaining Into Root Words
7. Classify Into Specific And General Words
8. Synonyms
9. Intersection With Specific Words

Demonstration of Phrase Prioritization algorithm has been found in the Crantop database. This demonstration proves that more relevant documents will be retrieved as compared to the results. Linked Representation of the Sparse Matrix formed for three documents only showing the simple collection of documents.

6 PROPOSED WORK

It will do the information retrieval better in comparison with other work in the future. By using appearing related algorithm and by showing good results of precision and recall and f-measure the related algorithm can be used as a better tool of Information Retrieval. A tool will be practically implemented in Python language or Matlab.



REFERENCES

- [1] Changiz Valmohammadi, Ali Ghassemi, (2016) "Identification and prioritization of the barriers of knowledge management implementation using fuzzy analytical network process: A case study of the Iranian context", *VINE Journal of Information and Knowledge Management Systems*, Vol. 46 Issue: 3, pp.319-337, <https://doi.org/10.1108/VJKMS-08-2015-0046>
- [2] Shahin Dezdar, Ainin Sulaiman, (2009) "Successful enterprise resource planning implementation: taxonomy of critical factors", *Industrial Management & Data Systems*, Vol. 109 Issue: 8, pp.1037-1052, <https://doi.org/10.1108/02635570910991283>
- [3] Arash Shahin, Ensiyeh Bakhshi, (2013) "Prioritization of innovation factors by the integration of concurrent function deployment and P diagram with a case study in Sepahan Industry Group", *Journal of Manufacturing Technology Management*, Vol. 24 Issue: 6, pp.952-971, <https://doi.org/10.1108/JMTM-04-2011-0039>
- [4] Zoe Chao, (2016) "From query analysis to user information needs: a study of campus map searches", *Library Hi Tech*, Vol. 34 Issue: 1, pp.104-129, <https://doi.org/10.1108/LHT-12-2014-0110>
- [5] Reza Jamali, Hossein SayyadiTooranloo, (2009) "Prioritizing academic library service quality indicators using fuzzy approach: Case study: libraries of Ferdowsi University", *Library Management*, Vol. 30 Issue: 4/5, pp.319-333.
- [6] Vin D'Amico, (2005) "Manage your IT projects like an investment portfolio", *Handbook of Business Strategy*, Vol. 6 Issue: 1, pp.251-255
- [7] Rosie Higman, Stephen Pinfield, (2015) "Research data management and openness: The role of data sharing in developing institutional policies and practices", *Program: electronic library and information systems*, Vol. 49 Issue: 4, pp.364-381.
- [8] Farrokh Mamaghani, (2002) "Evaluation and selection of an antivirus and content filtering software", *Information Management & Computer Security*, Vol. 10 Issue: 1, pp.28-32.
- [9] KengHoon Gan, KeatKeongPhang, (2015) "Finding target and constraint concepts for XML query construction", *International Journal of Web Information Systems*, Vol. 11 Issue: 4, pp.468-490
- [10] Yvonne Kammerer, Peter Gerjets, (2012), Chapter 10 How Search Engine Users Evaluate and Select Web Search Results: The Impact of the Search Engine Interface on Credibility Assessments, in Dirk Lewandowski (ed.) *Web Search Engine Research (Library and Information Science, Volume 4)* Emerald Group Publishing Limited, pp.251 - 279
- [11] Manuel Burghardt, Markus Heckner, Christian Wolff, (2012), Chapter 2 The Many Ways of Searching the Web Together: A Comparison of Social Search Engines, in Dirk Lewandowski (ed.) *Web Search Engine Research (Library and Information Science, Volume 4)* Emerald Group Publishing Limited, pp.19 - 46
- [12] Dirk Ahlers, (2012), Chapter 3 Local Web Search Examined, in Dirk Lewandowski (ed.) *Web Search Engine Research (Library and Information Science, Volume 4)* Emerald Group Publishing Limited, pp.47 - 78
- [13] Kin Fun Li, Yali Wang, Wei Yu, (2012), Chapter 7 Personalised Search Engine Evaluation: Methodologies and Metrics, in Dirk Lewandowski (ed.) *Web Search Engine Research (Library and Information Science, Volume 4)* Emerald Group Publishing Limited, pp.163 - 202
- [14] Sarah E. Crudge, Frances C. Johnson, (2007) "Using the repertory grid and laddering technique to determine the user's evaluative model of search engines", *Journal of Documentation*, Vol. 63 Issue: 2, pp.259-280, <https://doi.org/10.1108/00220410710737213>
- [15] Elaine Menard, NoufKhashman, (2014) "Image retrieval behaviours: users are leading the way to a new bilingual search interface", *Library Hi Tech*, Vol. 32 Issue: 1, pp.50-68, <https://doi.org/10.1108/LHT-06-2013-0067>
- [16] Jurriaan Souer, Dirk-Jan Joor, Remko Helms, Sjaak Brinkemper, (2011) "Identifying commonalities in web content management system engineering", *International Journal of Web Information Systems*, Vol. 7 Issue: 3, pp.292-308, <https://doi.org/10.1108/17440081111165901>
- [17] Yanbo Ru, Ellis Horowitz, (2007) "Automated

- classification of HTML forms on e-commerce web sites", *Online Information Review*, Vol. 31 Issue: 4, pp.451-466, <https://doi.org/10.1108/14684520710780412>
- [18] Brendan Luyt, Dion Goh, Chei Sian Lee, (2009) "Searching locally: a comparison of Yehey! and Google", *Online Information Review*, Vol. 33 Issue: 3, pp.499-510, <https://doi.org/10.1108/14684520910969925>
- [19] Wendy L. Currie, Xinkun Wang, Vishanth Weerakkody, (2004) "Developing Web services using the Microsoft.Net platform: technical and business challenges", *Journal of Enterprise Information Management*, Vol. 17 Issue: 5, pp.335-350, <https://doi.org/10.1108/17410390410560973>
- [20] Anita Elberse, , Patrick Barwise, , Kathy Hammond, (2002), The impact of the internet on horizontal and vertical competition: Market efficiency and value chain reconfiguration, in Michael R. Baye (ed.) *The Economics of the Internet and E-commerce (Advances in Applied Microeconomics, Volume 11)* Emerald Group Publishing Limited, pp.1 – 27
- [21] Sybille Peters, Claus-Peter Ruckemann, Wolfgang Sander-Beuermann(2011)," A NEW APPROACH TOWARDS VERTICAL SEARCH ENGINES Intelligent Focused Crawling and Multilingual Semantic Techniques", pp.251 – 279.
- [22] Kevin Curran, Jude Mc Glinchey(2007)," Vertical Search Engines" *ITB Journal*, pp 22-27.
- [23] Ayush Jain(2013)," The Role and Importance of Search Engine and Search Engine Optimization", *IJETTCS*, ISSN 2278-6856, pp 99-102.
- [24] RAHUL TELANG, UDAY RAJAN, AND TRIDAS MUKHOPADHYAY(2004)," The Market Structure for Internet Search Engines", *Journal of Management Information Systems / Fall 2004*, Vol. 21, No. 2, pp. 137–160.
- [25] Mayank Trivedi(2009)," A study of search engines for health sciences" , *International Journal of Library and Information Science* Vol. 1(5) pp. 069-073.
- [26] Manoj M, Elizabeth Jacob(2008)," Information Retrieval on Internet using Meta- Search Engines: A Review" ,*Journal of Scientific & Industrial Research*,pp 739-746.
- [27] Yue Hou-guang, Zhang Ling, Meng Fan-jun, Songhong-hao(2014),"Research and Implementation of a Vertical Search Engine in the Financial Domain", *International Journal of u-and -e Service, Science and Technology*, pp 117-128.
- [28] Hui Wang, Jianzhuo Yan, Liying Fang, Xinqing Shi, Hairui Luo(2013)," The Design and Implementation of Vertical Search Engine Based on Nutch", The authors - Published by Atlantis Press 1430-1441.
- [29] George Almpandis, Constantine Kotropoulos, and Ioannis Pitas(2009), "Focused Crawling Using Latent Semantic Indexing – An Application for Vertical Search Engines", pp.47 – 78.