

Novel Search Engine Optimization Technique

Khan Saba Shireen¹, Farida Attar², Bamne Naba³, Shaikh Shabnoor⁴

^{1,2,3,4}M.H. Saboo Siddik College of Engineering, Mumbai, India

Abstract: A web search engine is referred to a software system which is used to retrieve information on the World Wide Web. Widely popular search engines for example Google's page rank algorithm aid to gain desire information by making the information obtainable easily through internet. In this research work, a novel approach using weighted technique is introduced to mine the web contents as per the user needs. Web content mining is an area which filters the amount of irrelevant and redundant results by optimizing the searching process. Also along with this technique we will take users feedback to make the results more accurate and relevant as per users need. Results extraction, pre-processing, user's feedback, total relevancy and sorting of results comes under the working of the proposed system. So in this way the drawbacks of the existing systems are overcome by removing the redundant and unwanted data. Thus the proposed system providing with benefits of nicety and relevancy along with the level of excellence in results.

Keywords: Content mining, Feedback, Relevant, Term frequency, Weighted technique.

I. INTRODUCTION

In today's world users are not interested in retrieving large number of documents, but high usefulness of retrieved documents is desired. Search engine makes easier for the user to access any kind of information with the assistance of internet. Search Engine returns list of documents arranged in sequence according to rank of document. For example, Google's page rank algorithm. Page rank algorithm estimates the importance of a website by counting the number and quality of the links. One another example is of yahoo's click popularity, it determine the websites ranking by keeps track of user's click on a website so the website can get close to the top rank. Rank optimization is done to move useful documents upwards in search result list. Here mining of users feedback are beneficial to optimize ranks.

So a novel approach to be developed to work towards ranking content of the web resource based on user query & feedback. In the proposed work, performance of rank optimization architecture is ameliorated so that time of optimization decreases and it returns accurate result. The new approach is introduced to rank the relevant pages based on the content and keywords.

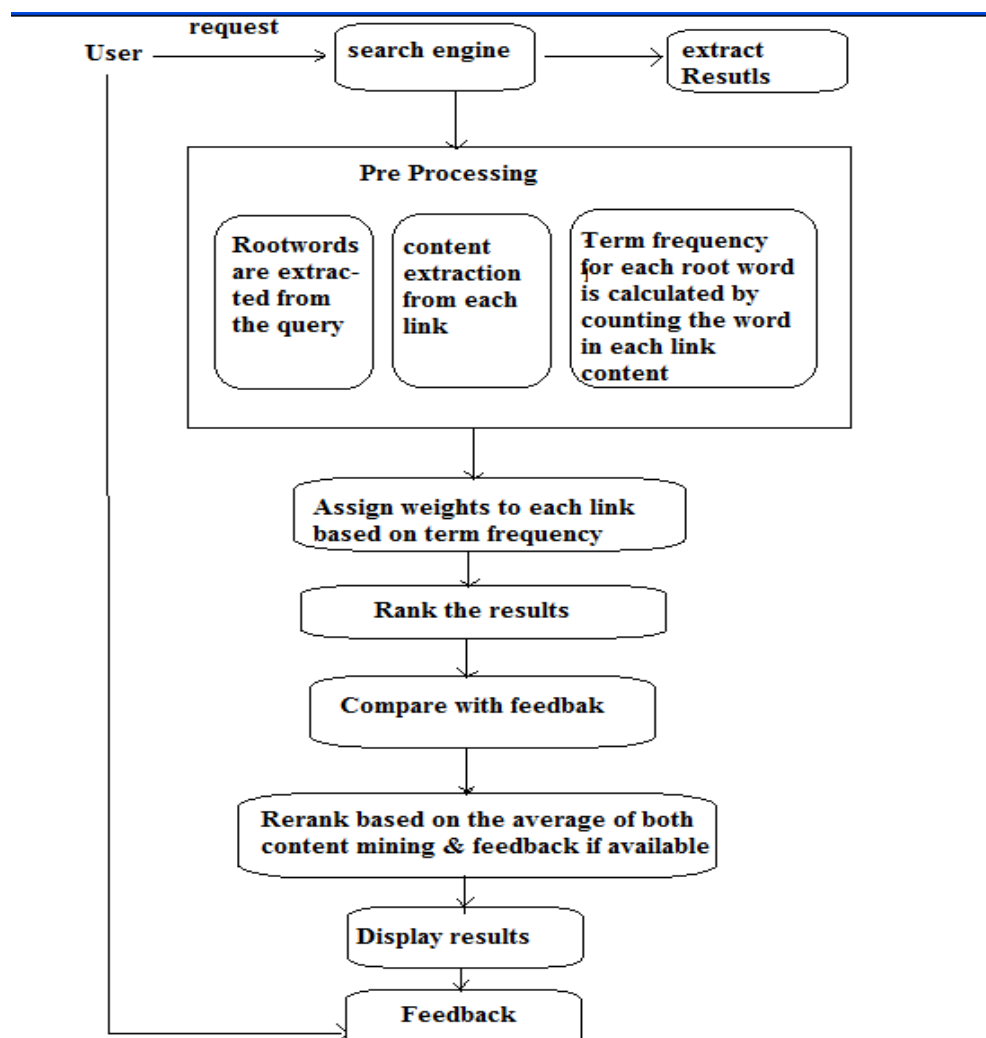
When a user enters a particular query on search engine, the result will be displayed on the basis of content based ranking and optimization done by using the feedback provided by the user stored in the temporal database. If the user finds a specific link serviceable then accordingly the user will rank the website and then the results will be re-ranked based on average of all the feedbacks.

II. RELATED WORK

Many techniques and research works exist for search engine optimization. Paper [1] proposed a method to organize user histories using data mining techniques. In this system, user's account is created and based on its history query grouping is done. Each group consists of similar queries. Using two inputs search logs and list of query grouped reranking algorithm displays the result based on user past behaviour. In paper [2] web pages are added to database which are related to specified domains (i.e. related to our specified Ontologies) and other web pages are discarded. To find the domains in which a Web page belongs to or not relevance scores is calculated of the Web page for all domains, and if the relevance scores of the page are more than predefined scores then we say that the page is specific to these domains under consideration. In paper [3] search engines instead of searching keywords on the web page search meta data for the required information. In this work,

search engine is developed based on this concept. Search engine first searches the pages and then gets the result by searching for the metadata. In the proposed work [4] a new approach is introduced to rank the relevant pages based on the content and keywords. User Query is pre-processed to identify the root words. Every result page keywords and content words are pre-processed and compared against the dictionary. If a match is found then particular weight is awarded to each word. Finally, the total relevancy of the particular link against user request is computed by summarizing all the weights of the keyword and content words. In paper [5] strategy is introduced in which user's profile is constructed automatically from the user's search history and is augmented by a general profile which is extracted automatically from a common category hierarchy. The categories that are likely to be of interest to the user are deduced based on his/her query and the two profiles and these categories are used as a context of the query to improve retrieval effectiveness of web search. In paper [6] novel approach is proposed to infer user search goals by analyzing search engine query logs. Framework is developed to discover different user search goals for a query by clustering the proposed feedback sessions. In this, a new criterion is proposed "Classified Average Precision (CAP)" to evaluate the performance of inferring user search goals. Experimental results are presented using user click-through logs from a commercial search engine to validate the effectiveness of proposed methods.

III. ARCHITECTURAL DESIGN



Based on the user request, search engine results are recruited and stored in the database. After this preprocessing is done in which firstly the root words are identified then content of each link is extracted. Using content mining we will assign weight to each link by computing term frequency according to which rank will be given. The results will then be arranged based on new ranks sorted in descending order of rank and displayed to the user.

If feedback is given then it will be compared against the previous ranks for each link and consequently average rank will be calculated and considered for displaying of results. So now the most pertinent links are found at the top of the list.

IV. WORKING

A. Results extraction:

Queries are being entered in the search box for the purpose of retrieval of information and obtaining the data to the proposed query. User enters their query in search engine for the information they want. According to the query requested appropriate resultant links are extracted from Google & are stored in temporal database. For example user enters query “data mining”, the search box will fetch the links from the Google and store it in the database.

B. Pre-processing:

In Pre-processing the following operations are performed:

- i. From the user query root words are identified and extracted.
- ii. For each link provided and stored in the database, content is retrieved from the links i.e. Content mining is performed on them.
- iii. Content mining performed on the links are used to calculate Term frequency by extracting root words given for the each link.
- iv. Weight is assigned to each link based on term frequency.

C. User's Feedback:

Users are provided with the option to rank the link based on its creditability. Depending on the user likeness which he will find the most useful link he will be given with the option of giving the feedback to that particular link by ranking it out of 5. The most relevant link can be ranked 5 or 4 and the least of importance can be ranked 1 or 2. The given feedback by the user is used to calculate the relevancy of particular link.

D. Total Relevancy:

Ranks obtained from content mining and from user's feedback are used to calculate Total relevancy based on average of weight assigned & ranks stored for each link.

E. Sorting of results:

Results are sorted in descending order of total relevancy calculated. Thus, we get optimized result after sorting. The topmost links in the results are the one which most of the user find useful & the bottom most results are least relevant for the requested user query.

V. ALGORITHM

Step 1: Extraction of search engine results R_j based on user query.

Step 2: Pre-process query & Identify the root words R_{Wi} .

Step 3: For every root word R_{Wi} in each link Term frequency is calculated.

a) Calculate TF_i . (1)

b) Calculate $W_i = \sum TF_i$. (2)

Step 4: Check for feedbacks of user for each link

a) Average feedback for each link is calculated

$F_i = \sum F_j / n$ (where n is total feedbacks available for particular link) (3)

Step 5: Total Relevancy TR_i for particular link is calculated by computing average of F_i & W_i .

Step 6: Repeat step 3 to 5 for all search results R.

Step 7: Sort the results R based on TR_i in descending order.

VI. CONCLUSION

The proposed approach gives better result by optimizing the search engine results. Proposed methodology focuses on content mining to rank the relevancy of web pages and feedback from the users to contribute in the final result. The proposed system provides accuracy and performance in the relevancy of the results obtained. It also provides quality work by reducing and filtering the colossal amount of unwanted data. In future, we also intend to work on providing with the options to users of retrieving the information about images, audio and video along with the additional benefit of searching through images, audio and video. We will also try to accomplish collaboration with more than one search engine.

REFERENCES

- [1] Charanjeet dadiyala and Sulabha patil, "An approach to organize user histories using data mining techniques", International Journal of Advanced Computational Engineering and Networking, ISSN: 2320-2106 Volume- 1, Issue- 7, and Sept-2013.
- [2] Debajyoti Mukhopadhyay and Sukanta Sinha, "A New Approach to Design Graph Based Search Engine for Multiple Domains Using Different Ontologies", Web Intelligence & Distributed Computing Research Lab, Techno India Group West Bengal University of Technology.
- [3] Ritu Khatri Kanwalvir, Singh Dhindsa and Vishal Khatri, "Investigation and Analysis of New Approach of Intelligent Semantic Web Search Engines", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-1, Issue-1, April 2012
- [4] P.Sudhakar, G.Poonkuzhali and R.Kishore Kumar, Member IAENG, "Content Based Ranking for Search Engines".
- [5] Fang Liu, Clement Yu, Weiyi Meng, "Personalized Web Search For Improving Retrieval Effectiveness", University of Illinois at Chicago, Chicago, IL 60607.
- [6] Zheng Lu, Hongyuan Zha and Xiaokang Yang, "A New Algorithm for Inferring User Search Goals with Feedback Sessions", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING VOL. 25, NO. 3, MARCH 2013