Optimization of Search Information on Cloud

N. P. Joshi
IT Dept.WIT, Solapur

R. B. Kulkarni, Ph.D
CSE Dept.WIT, Solapur

ABSTRACT
Cloud Computing, another buzzword for internet based computing, is used to deliver services on the internet by sharing computing resources. Amongst a large number of services web service is the most important one. To get access to the vast amount of information, we use search engine. It is often the case where users are starved for the correct information, and observed effectiveness is quite low. Our proposed work explains a method to improve effectiveness of Internet information by constructing a cloud of search engines. The method uses query processing and result optimization returned by multiple search engines. A combination of different term-weighting schemes is used to improve rank updation process which in turn improves information retrieval. The search results are submitted to an optimizer where they are re-ranked by updating the existing pagerank values. These optimized search results are returned to the user..

Keywords
Cloud Computing, Rank Updater

1. INTRODUCTION
Cloud Computing is next generation computation, which gives everything as per the user needs. Lots of information is already available on the web and the updating is still in progress. With such a huge amount of information it is not easy to pick up the desired stuff every time. The softwares such as search engines can't also help to resolve the issue completely. The need of a methodology providing the right stuff throughout is emerging continuously. It is with the changing scenario, the World Wide Web is becoming an important information source in every walk of life. The information obtained from the web is not only used in everyday life but also is frequently referenced in scientific publications .The popular search engines are used for retrieving information but , the task of finding relevant information related to a specific query or topic is becoming increasingly difficult. Therefore users usually browse only the first or perhaps the second page of the search results retrieved by the search engine. This problem mainly occurs due to the query not being properly framed by the user. As a result of this, ranking is a difficult and an important task to obtain the exact match for the query given by the user[2]. Thus we propose an algorithm for re-ranking of results returned by multiple search engines . The re-ranking of the pages is based on term weighting scheme.

This paper aims at to optimize search information of multiple search engine with the use of re-ranking approach which uses clustering which is important task.

To execute task we used query logs to store query information.The paper is organized as section II describes about related work, section III tells about proposed work , section IV concludes the paper and section V is with some references.

2. LITERATURE REVIEW
Yubao Zang [1] has proposed an idea to improve effectiveness of Web search results by query processing and optimizing the results.Query processing is done with scheme which integrates different locations and identifies query terms

.The term weighting scheme is very important as per as ranking is concern.

A K Sharma , Neha Aggrawal[2] proposed an approach to optimize the results which is based on learning historical query logs and it considers the users information needs. And the paper aims to optimize the results by giving more relevant information on the top.

Ulrich Guntzer , Werner kibling[3] has proposed an algorithm to combine multi-feature results for efficient top ranked results. In this algorithm mainly concerns with weighted queries and multimedia data such as retrieval of images .To handle similarity of multimedia data, storage of data is the important task. It very difficult to treat with similarity measure because it changes with user needs. This algorithm is adaptation to previous optimization query model as multi –data is involved.

3. METHODOLOGY
In our proposed method , steps given below are followed which are mentioned in fig.1

1.User sends query to the interface of search engine.

2 .Query processor sends it to Query logs and then it send it to query optimizer to optimize the query.

3. After optimization repeated results are removed then the result is sent to the rank updater.

4. Rank Updater gives results back to the optimizer and final results are returned to the interface.
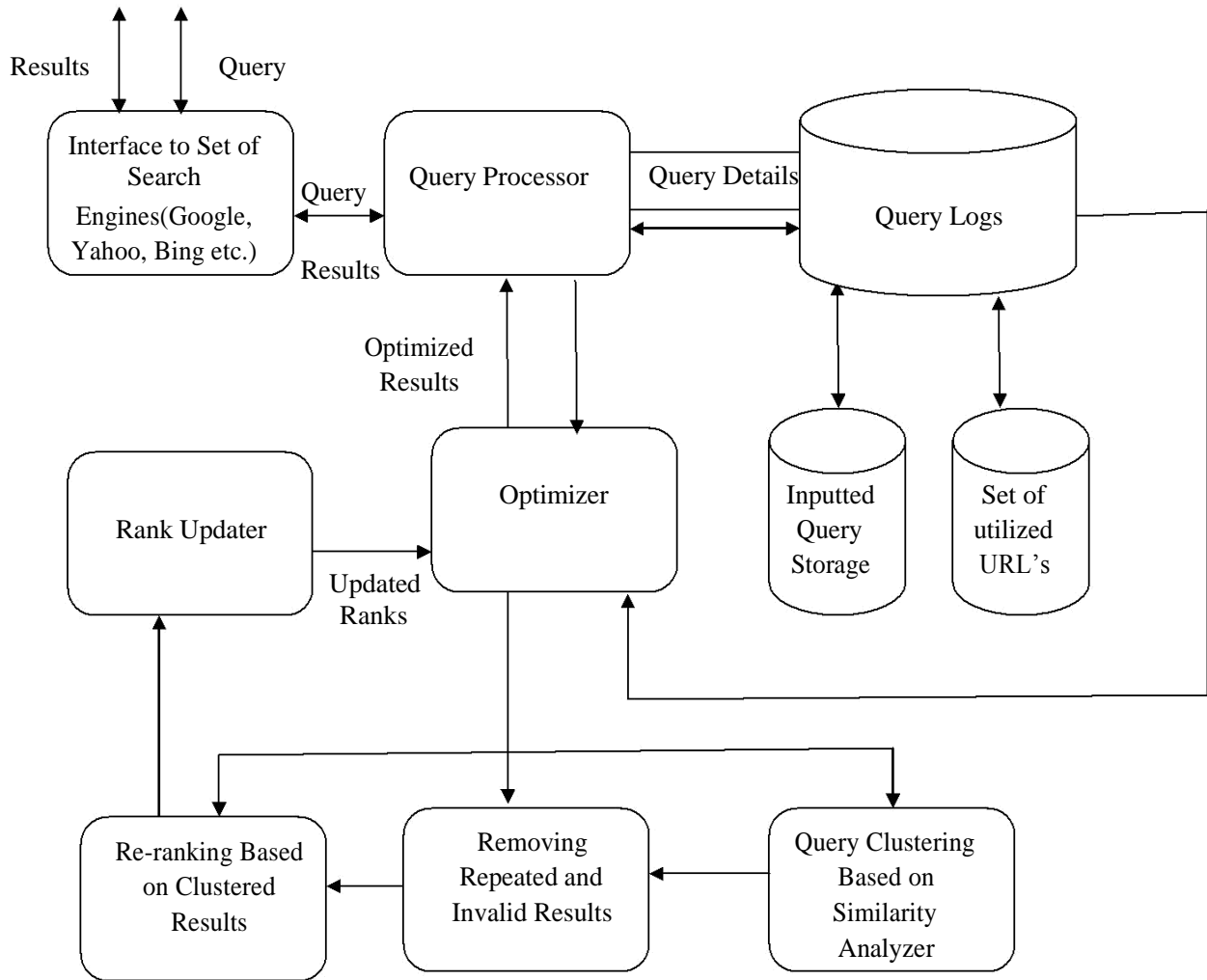
Fig. Proposed system architecture

Important modules of above fig.1 are discussed below: to the optimizer that optimizes the query to have the effective utilization during result generation.

3.1 Set of Search Engines:

When it comes to search the information on the web, the first thing that comes in to mind is the search engine which helps us to do so. User uses different search engines such as Google, Yahoo!, Bing etc. This module of the figure represents the set of such a search engines and takes the inputs to search on the web. Such an input is called as query. This query is then forwarded to the query processor to process it in a required compatible processing format.

3.2 Query Processor:

For every keyword entered by the user, it is very much necessary to process the query and then execute it to generate the desired results. The query processer module takes the input as user entered query from the set of search engines. Also, it sends that processed query along with details to the query log, so that the same kind of query can be handled quickly on the next appearance. Also, it sends the same details

3.3 Query Logs:

These logs are maintained for the betterment of the time efficiency of the result generation. The logs stores query details such as the processed query, the keywords defined in it, user_id of the user, access time, page preferences of the user, storage of links visited etc. These query logs are finally utilized for the process of optimization, where the similar type of query if arrived can be served with the accurate and desired results, in no time. Log Records(query, user_id, access time)

Table 1.The storage of query is given below in the table :

Query Number	Query id	Query
1	101	Sachin tendulkar
2	102	Rahul dravid

3.4 Query Optimizer, Clustering and Rank Updation

This unit is very much important in this entire process. Here, it takes input from query processor i.e. processed query, query logs i.e. log records of similar types of queries[3]. Then it gives those inputs to similarity analyzer, where it finds the similar types of results. Based on those results the queries are assigned to the clusters formed. This process is called as query clustering. Here, there might be the case that, a single query can be assigned to more than one cluster. For avoiding this situation a removal of repeated results is very necessary, which can be done by deleting the repeated queries in the clusters. The repetition can be judged by comparing the URL's of the two same queries, and one of them can be removed if the URL's are same. Then, the process of re-ranking of the clustered results is carried out and the new ranks are given to the web pages. There are more such methods of doing this such as weighting scheme, entire web-page score, partial web- page score etc .In this we are using term weighting scheme. These new ranks are then forwarded to the rank updater, where it confirms the new ranks and perform the updation successfully. The optimizer then revert back the results of the inputted query with respect to newly ranked pages, to the query processor, which in turn revert back it to the set of search engines. The search engines then display that most desired and intended results to the user.

4 Conclusion

Though a World Wide Web is a huge collection of information, it is a kind of labyrinth. The paper here discusses different approaches to fulfill such a need. It first discusses the technology of result optimization, where it collects the results of user queried stuff from various search engines. Then it separates the most relative and the irrelative results by using techniques such as duplication avoidance, URL uniformity etc. Then based on these results it updates the ranks of the relative web pages and bringing the pages with the highest ranks in front of user as the desired stuff. The paper further discusses another technique called query logs. In this, user submitted queries are stored in the query logs along with other details such as user_id, web page preferences, access time etc. Based

on these details, the similarity patterns are formed and the query clustering is done, with the visited links of the users. The re-ranking is then done for the web pages and most visited, clustered and the high ranked relative web pages are returned as the desired stuff according to the inputted query. The techniques discussed in this paper helps a lot to a user to reach up to a desired web information resource, resulting in, a nearly accurate search.

REFERENCES

[1] aeyong Kang and Kwang Mong Sim (2011) "A Cloud Portal with a Cloud Service Search Engine" International Conference on Information andIntelligent Computing IPCSIT vol.18

[2] A.K. Shanna, Neha Aggarwal "Web Search Result Optimization by Mining the Search Engine Query Logs" ©201 0 IEEE

[3] Ulrich G¨untzer , Wolf-Tilo Balke, Werner Kießling "Optimizing Multi-Feature Queries for Image Databases" Proceedings of the 26th VLDB Conference,Cairo, Egypt, 2000.

[4] Yubao Zhang "Result optimization returned by multiple Chinesesearch engines based on XML" Institute of Mechanical and Information Engineering, Zhejiang Textile and Fashion College

[5] Jaeyong Kang, Kwang Mong Sim "Cloudle: An Ontology-Enhanced Cloud Service Search Engine" Web Information Systems Engineering – WISE 2010 Workshops Lecture Notes in Computer Science Volume 6724, 2011, pp 416-427.

[6] Nikolaos Nanas , Victoria Uren , Anne De Roeck "A Comparative Study of Term Weighting Methods for Information Filtering

[7] A. Arasu, 1. Cho, H. Garcia-Molina, A. Paepcke, and S.Raghavan, "Searching the Web," ACM Transactions on InternetTechnology, Vol. I, No. I, pp. 97-101, 2001.