ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

An Optimized Search Technique By Query Classification

Rohit Agarwal[1], Rajesh Singh[2]

M.Tech Scholar, Dept. of Computer Science, B.S.A College of Engineering & Technology, Mathura, UPTU Lucknow, India

Assistant Professor, Dept. of Computer Science, B.S.A College of Engineering & Technology, Mathura, UPTU Lucknow, India

ABSTRACT: Web Content mining is defined as extraction of required data from distributed databases. Today we are using search engine for extracting the required information on the basis of requirement of user. But search engines that we are using not always provide the actual required information that the user actually want in a first attempt. So we can say that current search engine techniques provide the approximate output to the user.
In this paper, we proposed a query classification technique by which we can improve the user query to make it more small, simple and more understandable. Such improved query will take a small fraction of time to execute over server and provide only the more actual required information to the user. Also we try to improve the pre-existing search engine algorithms to optimize the search result.

KEYWORDS: Web mining, Optimized Search, Query, Classification.

I. INTRODUCTION

As the era of online resources increases, the problems related with web mining increases. The most general or primary problem is to retrieve only the required information at the right time and at right place. However some user can efficiently retrieve the required information using search engine because he/she is familiar with searching techniques using Boolean operators (AND, OR, NOT etc) because such user are directly or indirectly computer literate. But if we concern about the normal user who are not much computer literate can't search the required information at right time. Today mostly user waste too much time in searching small information that he/she actually want due to the current interface of a search engine. But if we want to improve such interface then we have to optimize our user query to make it small and simple [1]. Because it is general phenomena if we know actually what we want then we can find it easily. Similarly, if a search engine knows that exactly what a user want then it can easily search such information over the databases in a small fraction of time.

II. RELATED WORK

In [2] author proposed a hypertext based architecture to perform the search but its architecture was not able to reduce the search time. In [3] author proposed a classification technique which will take less time to perform a search but drawback is that it can performs classification but not the sub classification in terms of database. It will also consume more time to perform a search. In [4] and [5] authors proposed some search rules and graph based data mining that helps in reducing the search time but more than 80% people of world still not aware with such rules as well as internet architecture is also very complex in which graph problems like konnisberg's bridge problem may arise which is still not possible to solve. In [6] author proposed a distributed approach in extracting processed information in small amount of time. His proposed architecture was decreased the network traffics to increase the processing speed. But the drawback is that his architecture needs more than 100 crawlers to perform its task which is too costly to develop or manage.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

III. PROPOSED APPROACH

We classified the user query that he/she feed in search engine on the basis of 'type' and 'extension'. When a large and complex query will be classify by proper categorization as well as sub categorization then such query will become small, simple and more understandable by the existing system to search it. Let us understand it with a general example, if a user want to search a bitmap image(that will take less memory) over a topic "unity" using an existing search system then he/she directly feed his/her topic "unity" in search box and select the type as images. In that case, existing search system will list out all the images related with topic "unity" with extensions bmp, jpeg, png etc and suppose existing search engine will list out 1200 images as shown in following table.

Image Search Topic	Type of Image	No. of Images shown
"Unity"	bmp	220
"Unity"	jpeg	612
"Unity"	others	368
	Total	1200

Table 1.1 Existing Search result (On the basis of assumption)

But the user actually wants the bmp images. It means that out of 1200 only 220 are the actual required information and remaining 980 are useless information for the user. So if we compare then we get that search engine consume more time to search or access over such 1200 images as comparison to 220 images.
But when we classified or sub classified user query on the basis of 'type' and 'extension' then user will feed "unity" in search box and select "images" in 'type' combo box and "bmp" in 'extension' combo box. We also proposed and implement a query classification algorithm in existing search system that will combine the whole user query in backend to make it small, simple and clearer to understand. Now search result will be as follows as shown in following table.

Image Search Topic	Type of Image	No. of Images shown
"Unity"	bmp	220
"Unity"	others	0
	Total	220

Table 1.2 Search result with query classification technique (On the basis of assumption)

So we can get more optimized output after applying such query classification technique in our existing search engine technique.

IV. IMPLEMENTATION ISSUE

To implement such a query classification technique, we have to consider the improvement in the following existing search system area

A. *Interface Consideration*: Firstly we have to improve our preexisting search system user interface to add one more option that is sub category on the basis of 'extension' of required information.
We include one more combo box i.e. sub category in pre existing search system. After improvement our search system will appear as follows

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

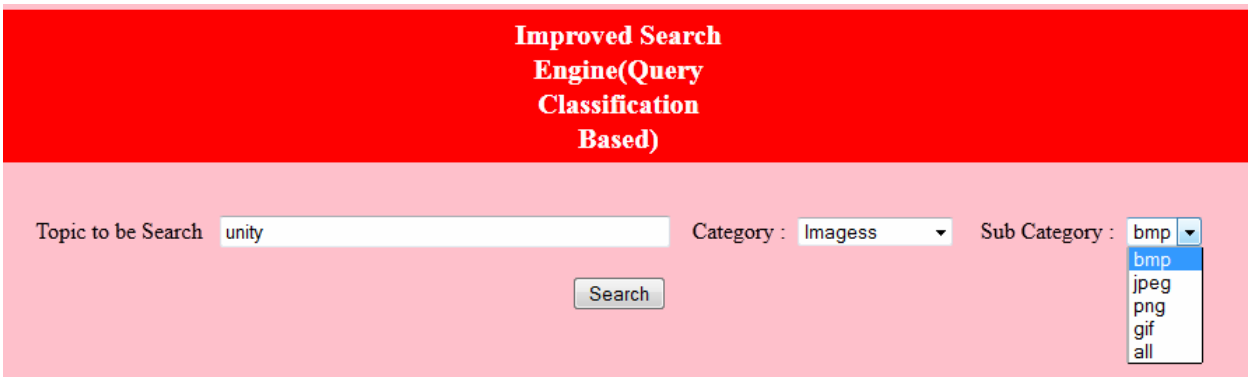

Fig 1.1 Improved Query Classification Technique Based User Interface

B. *Proposed Query Classification Algorithm:* Secondly we have to implement a query classification algorithm by applying a mathematical approach over the query obtained from user interface.
Let we have a query Q that to be search by using existing search system, C is category & S is sub category specified by user then our proposed algorithm are as follows

```
Algo QC ( )
        {
        Dim Q, C, S, Qn
        Q= Input ( )                    // Here Q stored the query feed by user
        C= Input ( )                    // Here C stored the category feed by user
        S= Input ( )                    // Here S stored the sub-category feed by user
if(S!=”All)
  {
            Qn= Q^C^S;                  // Here ^ is Conjunctive AND Operator
// Qn stored Improved Query
   Search (Qn);

  }
else
 {
  Arr[]={S1,S2,…..Sn}           // Here Arr[] I an array that store all subcategories
                                        // related with chosen Category C i.e S1,S2….Sn.
  Loop:: i 0 to n-1 step +1
    {
     Qn=Q^C^S[i]                // Using loop, we perform checking of all subcategories
                                        // with feeded Query Q in Qn one by one.
     Search(Qn)
    }
  }
        }
```

Here, Search(Qn) will performed web search as other search engine did but on Qn that is an improved user query.
In our above proposed algorithm QC [Query Classification], we call two more algorithm i.e. Input () and Search (). Input () will take the input in the form of query from the user and Search () will forward such Query to the web server (crawler) for performing such search.
Then we forward such improved query to the web server and perform searching using indexer and crawler as we have done in our previous searching.
On implementing such query classification technique in our existing search engine technique, our web mining architecture will appear as follows.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

In below architecture, *User* feed *Query* through *Implemented User Interface* that contains user *Query(Q), Category (C)* and extension in the form of *Subcategory (S)* as shown in fig 1.1. then such queries forwarded to search engine server where we implement our proposed algorithm*(Query Classified)* that will make an improved *Query(Qn)* using our *QC()* search algorithm as mentioned above. Then such *Qn* will forward to *Crawler and Spider* by *Search Engine Server* which then search such Query *Qn* over number of distributed databases*(Web Databases) D1,D2….Dn* with the help of *Database Server* and if any related result found that will again forward to *Search Engine Server* and then to *User*. Such result is *Optimized and Improved result* that matches exactly to *User* actual requirement.

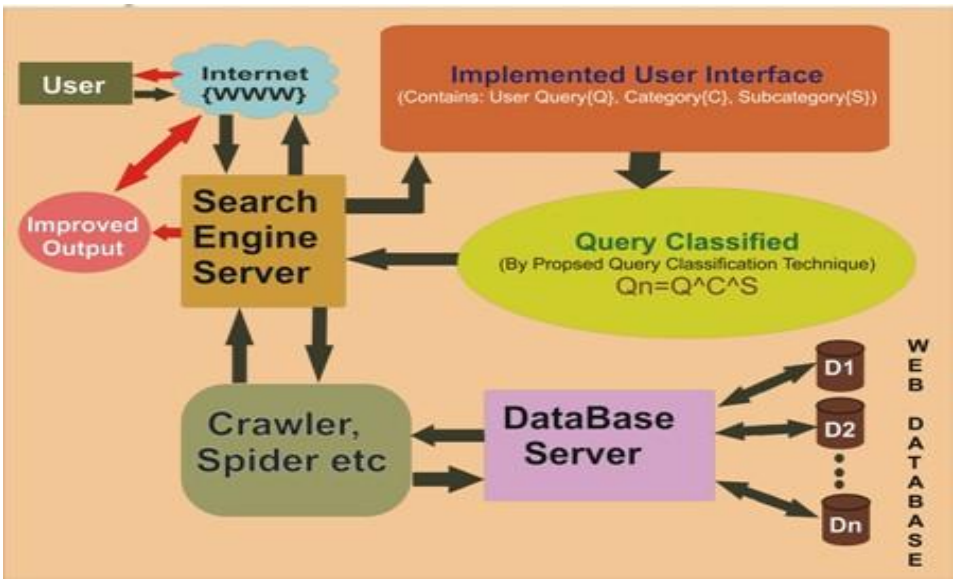

Fig 1.2 Query Classification Based Search Engine Architecture

Our algorithm and implemented interface will help only to make user query more simple and clear to search engine because we know that a search engine will perform better search only when the query feed in it is more clear.

V. SIMULATION RESULT

We implement such optimized search technique by using PHP and WAMP. Such simulation report shows the comparison between search time and required search results using previous search technique as well as our proposed search techniques over distributed databases.

If a user want to search a pdf over a topic “corruption in india” using previous search engine and our proposed search techniques. In fig 1.3 it shows that using previous search technique take more search time as comparison to our proposed search technique. As the comparison results shows that preexisting search techniques found 4 pages in 5 seconds, 18 pages in 10 seconds, 22 pages in 15 seconds, 58 pages in 20 seconds and 80 pages in 25 seconds So, its average result is 37 pages in 15 seconds whereas our proposed search technique found 4 pages in 5 seconds, 23 pages in 10 seconds, 69 pages in 15 seconds, 87 pages in 20 seconds and 119 pages in 25 seconds So, its average result is 60 pages in 15 seconds. Our optimized search results shows that our proposed search technique perform better search as previous search techniques in terms of speed and time.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

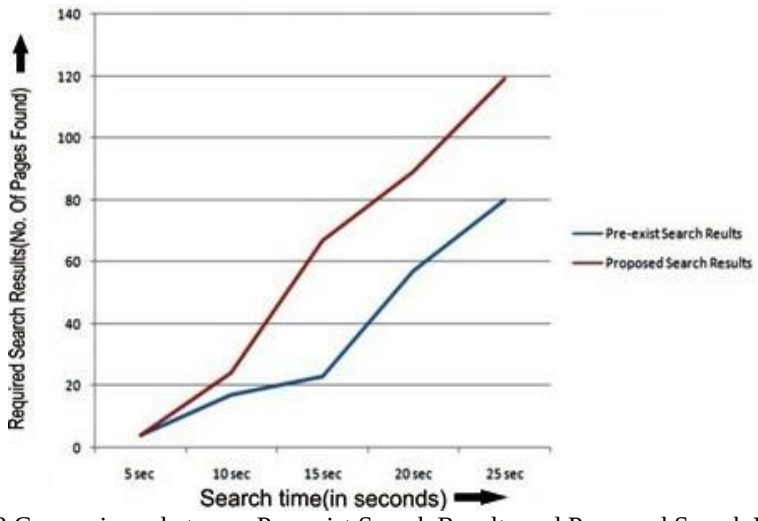

Fig 1.3 Comparisons between Pre-exist Search Results and Proposed Search Results

VI. CONCLUSION

Our present paper discuss about a new query classification technique that will implemented on improving the preexisting search system. Such paper will arises several research issues in such web content mining field or improve in pre existing information retrieval process over search engine.

REFERENCES

1. Debajyoti Mukhopadhyay, Pradipta Biswas; "A Syntactic Classification based Web Page Ranking Algorithm"; 6th Int. workshop on MSPT; 2006.
2. Brin, Sergey; Page, Lawrence; "The Anatomy of a Large-Scale Hypertextual Web Search Engine;" 7th Int. WWW Conf. Proceedings, Brisbane, Australia; April 1998.
3. X. Yin, J. Han, J. Yang, and P. S. Yu. Efficient classification across multiple database relations: Acrossmine approach. *IEEE Trans. Knowledge and Data Engineering*, 18:770{783, 2006.
4. J. Wang and G. Karypis. HARMONY: Efficiently mining the best rules for classification. In *Proc 2005 SIAM Conf. Data Mining (SDM'05)*, pages 205{216, Newport Beach, CA, April 2005.
5. T. Washio and H. Motoda. State of the art of graph-based data mining. *SIGKDD Explorations*, 5:59{68, 2003.
6. R Singh, S K Gupta "Search Engine Optimization - Using Data Mining Approach" International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 2, Issue 9, September 2013, PP 28-32

BIOGRAPHY

Rohit Agarwal is an M.Tech Scholar in Computer Science Department, B.S.A College of Engg. & Tech. Mathura, UPTU Lucknow. He received Bachelor of Technology in 2012 from B.S.A College of Engg. & Tech. Mathura, UPTU Lucknow. His research interests in Search Engine Optimization, Computer Network, Programming etc.

Rajesh Singh is Asst. Prof. in Computer Science Dept. B.S.A College of Engg. & Tech. Mathura, UPTU Lucknow. He received a B.E. in Computer Sci Engineering from Faculty of Engg. & Technology, RBS College, Bichpuri, Agra, in 2000 and M.Tech in Information Technology from Karnataka State University, Karnataka. He had already published two research papers in field of search engine optimization in IJAIEM in 2013 and in IPASJ in 2014. Some of his research interests include the Data Mining, Computer Architecture, human computer interaction, Search Engines, scalability of information access interfaces.