



INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

DERIVING USER SEARCH OBJECTIVES USING INNOVATIVE ALGORITHM

Ms. Nayan Narayan Sonawane *, Prof. Mr. V. V. Bag

* Computer Science and Engineering, NK Orchid College of Engineering & Technology, Maharashtra, India

ABSTRACT

For a broad-topic and questionable query, different users at different time may have different search objectives when user submit a query to a search engine. Deriving and analyzing user search objectives for a given query can be very useful in improving search engines purpose and user experience.

In Developed system query is reformed using location of user and type of query (content based or location based). Query reformed is then passed to multiple search engines will give better and accurate results. Unique results from multiple search engines are extracted which are considered for generating feedback session. User click-through logs efficiently reflect the user's information need. Both clicked and unclicked links from feedback session are considered. Pseudo document is created to represent the feedback session generated. For a query different user search objectives are derived by clustering the pseudo documents. After clustering pseudo documents original unique results will get clustered into groups of User Search Objectives. Retrieved clusters represent different user search objectives for a given query; hence user can get what he/she wants conveniently.

KEYWORDS: User search objectives, Feedback Sessions, Pseudo-documents, Restructuring and Reranking search results, Clustering Search Results

INTRODUCTION

In internet search applications, queries which are submitted to search engines show the information needs of users. Nevertheless generally queries might not specifically depict users specific information needs since many questionable queries might cover a vast topic and different users at different time might need to get information on different forms once they submit the same query. For this case, when the query "galaxy" is submitted to a search engine, few users need to find information of our galaxy, whereas few others need to search out information on galaxy phones of Samsung. Accordingly, it is required and likely to derive completely different user search objectives in internet search applications. User search objectives are nothing but the information on different forms of a query that user want to derive. Motivation behind obtaining information is to satisfy his/her information need. The interpretation and analysis of user search objectives can have lot of advantages in improving search engine importance and user experience.

A few benefits of deriving user search objectives are; system can restructure web search results according to user search objectives by grouping the search results with the same search objective; hence, end users with different search objectives can easily find what they need actually. User search objectives depicted by some keywords can be used in query recommendation. Query suggestions can help users to form their own queries more accurately. Click-through logs are useful in applications such as reranking of web search results.

Aim of Developed system is to derive user search objectives for a query and depict each objective with some keywords. When user submits a query it will get reformed and then forwarded to multiple search engines. Unique results out of n. number of search results from multiple search engines are used for generating feedback session; it is defined as the combination of both clicked and unclicked links from these unique search results. Pseudo-documents are created which better reflect user's information needs in turn reflect feedback session. Developed System then perform clustering of these pseudo documents to derive user search objectives and depict them with some keywords. The grouping of different user search objectives can be obtained efficiently after clustering of feedback sessions. Hence system derives the number of user search objectives for a submitted query.

RELATED WORK

User's motivation behind information need is important for search engines to better provide user. Developed system overcomes the drawbacks of existing systems and tries to boost the search engine performance and give results as

[http:// www.ijesrt.com](http://www.ijesrt.com)

© *International Journal of Engineering Sciences & Research Technology*

per user's objectives. In the work done by [1] finding the adequate search engine is difficult because every search engine will have their own format for finding the best search results. So the best results may not be retrieved in a single search engine. The search results which are retrieved are not suitable and hence it is not consistent. Clustering of feedback session is done using conventional clustering method. Some works evaluate the search results returned by the search engine to exploit different query forms.[7],[13] Nonetheless, query forms without user feedback have restrictions to improve search engine importance. Some works consider user feedback and evaluate the different clicked URLs of a query in user click-through logs directly; number of different clicked URLs of a query might be not big enough to get ideal results. Other works introduce search goals to detect session boundary [11] Nonetheless, their method only finds if a pair of queries belongs to the same goal and does not care about what the goal is in real. Utilization of user click-through logs is to obtain user implicit feedback when learning ranking functions in information retrieval. Works on how to use implicit feedback to improve the retrieval quality[8],[9],[10]. There are some related works concentrates on organizing the search results.[6],[8],[13]

THE ARCHITECTURE

Following Figure 1.1 shows the System Architecture. From above given introduction and existing related work, in which several techniques are being developed to address various problems faced in deriving user search objectives in search engines but even after various proposed methods the problem of deriving user search objectives was not resolved and therefore there was a need of designing search engine which gives most accurate search result and thereby overcoming the drawbacks of existing systems. User search objectives which are derived and analysed can be very useful in improving search engines accuracy and user experience. In order to address various problems system is developed in which instead of providing many search results, developed application only aims at displaying top re-ranked results according to user search objectives and feedback sessions. Thus, developed systems application is to derive user search objectives, reform query activity tracks users GPS location for analyzing search queries submitted by user, Type of query is derived according to this user entered query can be either content based or location based. Multiple search engines are used for accuracy of results, feedback session is represented by pseudo documents. Reranking of search results and clustering of pseudo documents is performed to find user search objectives. Developed system consists of following major activities:

1. Analyse and Reform Query:

When User Submits A Query, developed System will analyze query, users location is tracked to find type of query. Click-through logs of user are also checked to understand particular user's need of information. Query submitted by user is reformed according to type of query, preferences of user and his/her previous search queries. Reformed Query will be then forwarded to multiple search engine APIs.

2. Multiple Search Engines

Reformed Query is submitted to multiple search engines where each search engine will display N. number of results, out of that developed system will extract top results from them. Top results are more accurate for finding user search objectives.

3. Display Unique Search Result

Top extracted results from search engines are then sorted for finding unique single bunch of results for query and then present it to user. User now can see the unique results from search engines; user can now click on any link which will generate feedback session.

4. Generate Feedback Session

Feedback sessions are constructed from user click-through logs and can efficiently reflect user's information need. Feedback session contains all the clicked links by user and unclicked links from unique search results displayed to user. In developed system feedback session is generated by considering all results displayed in unique results.

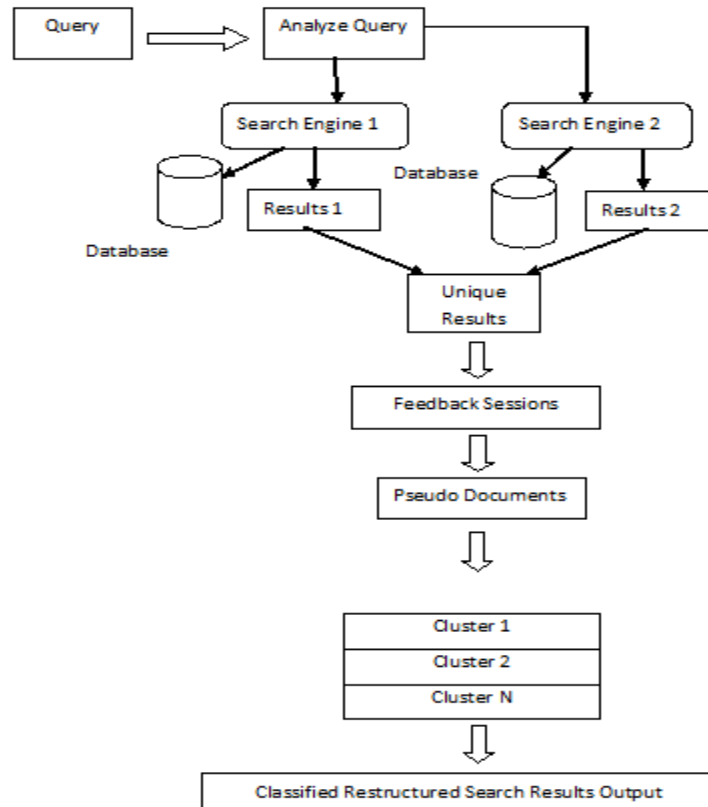


Figure 1.1 The Overall System Architecture

5. Creating Pseudo Documents

Then developed system generates pseudo-documents to better represent the feedback sessions. A Pseudo document contains the words from search results title and snippet.

6. Reranking

Reranking of Search results is done on the basis of click through logs of user, gives the top rated search result link. Click Sequences are used for finding ranking of results. Reranking is useful in finding users objective. It will give what information user is interested in.

7. Clustering Pseudo Documents

Finally the clustering of pseudo documents is done which will give clusters of user search objectives. By analyzing contents of Search results and Clusters, developed System will Classify and Rearrange search results. Search engine can return the results that are categorized into different groups according to user search objectives online. Thus, users can find what they want conveniently.

ALGORITHMS**Algorithm of Developed System is as follows:****Input:**

1. Search String
2. Users Feedback in terms of Clicks.

Output: Clustered Search Results according to User Search Objectives

- 1: Input the search query.
- 2: Reforming input query submitted.
 - i system finds user location information
 - ii Previously stored pseudo document history of input query checked by system for giving query suggestions
 - iii Find Type of Query = Location Based or Content Based
- 3: To give best results to user reformed query is passed to multiple search engines.
- 4: To get single bunch of unique search result set, top 20 search results are extracted from Multiple Search Engines, then sorted and merged together and finally these unique result set displayed to user.
- 5: When user clicks any link then order of click is stored by system and then it updates click through data of user
- 6: Unclicked links of unique result are also needed to generate feedback session.
- 7: Pseudo document for each link is created using title and snippet information of link.
- 8: End results are derived depending on user search objectives clustering of pseudo documents is performed; restructured and reranked results are displayed to user in the form of clusters of user search objectives.
- 9: End of Algorithm

Algorithm to Reform Query**Input:** Search String submitted by User**Output:** Query Reformed

1. Define the set and subset of Abstract Keywords .
Example: Keywords = {Earth, graffiti, India, . . . } And Earth is set of three elements, e1, e2, e3 where e1 = {Google, map, Wikipedia, planet}, e2={planet, solar, system, nine planet}, e3={nasa, science, gov, nine planet}
2. If submitted query found any Keywords, then check other query words for corresponding subset of elements
Example: If submitted query have word Earth, then check for remaining words of query against subset of earth.
3. If any match found between query words and subset words, then initiate the variable array of these subset words
 $SW[] = \{Google, map, Wikipedia, planet\}$ where SW = subset words
4. Now, check pseudo documents (feedback session database) for related query submitted in previous sessions with SW
5. If found, with that query and with SW, restructure Qsub
 $Qrst = Qsub \cup SW \cup (PQ + PD)$ (Pseudo documents and related previous query)
6. If Qtyp = LST, then append user location to end of query
7. If Qsub have words like 'near', then append location with word 'from'
8. If not, append location with 'at' $Qref = Qrst + LOCATION(User)$ (string concation)
9. Return Qref

Algorithm for Sorting and Merging of search results

Input: Reformed Query

Output: Unique Search Results from Multiple Search Engines

1. Save output of third party API search engine under datasets, 'yahoo' (YDS = Yahoo Dataset), 'google' (GDS = Google Dataset), 'bing' (BDS = Bing Dataset)
2. Declare SRDS[RW,RPT] (Server Result Dataset), a two dimensional array where RW=result row and RPT=Repeated Counter
3. Now, take row from GDS (dataset), compare with for each row of YDS and BDS
4. If match found with both, set RPT=3 and insert into SRDS for RPT=3 at last position of array
5. Else set RPT=2 and insert into SRDS for RPT=3 at last position of array
6. Else RPT =3 and insert into SRDS for RPT=3 at last position of array
7. Repeat steps 3 to 6 for each row in GDS
8. Declare SRDSsort array
9. Now, for each row of SRDS, make comparison with Pseudo Documents(Feedback session database)
10. If match found, check corresponding click sequence, ranking
11. According to value in click sequence and ranking, assign row from SRDS to SRDSsort array at top of all records, in between records or at last (Same method two dimension array was used here)
12. Repeat steps 9 to 11 for each row in SRDS
13. After completion, SRDSsort have search result set arranged in sequence having, most visited(clicked) result at top
14. Update search_result database with SRDSsort having entities like search result row, rank, search provider
15. Send top predefined number of rows from SRDSsort to user GUI with ID, ID represents corresponding details in search_result database

Algorithm to generate pseudo documents.**Input: Unique Search Results****Output: Pseudo Documents**

1. Initiate the click sequence counter of two dimension CS[ID, Seq] where ID is search row id and Seq is the order of click sequence
2. Whenever user clicks on any link to visit search result page, user get navigate to search page in new window while data was sent to server
3. Now with that data, CS is set by value ID and Seq, where Seq=Seq(last)+1
4. The rank was incremented by one for each click on same search result set row
5. The feedback session database was get updated by result of step 3 and 4
6. The words from result set row submitted back to server by clicking are extracted by removing irrelevant words like adjectives and nouns and compared against keywords
7. If match found more than 50%, then keywords subset are updated with words from result set row

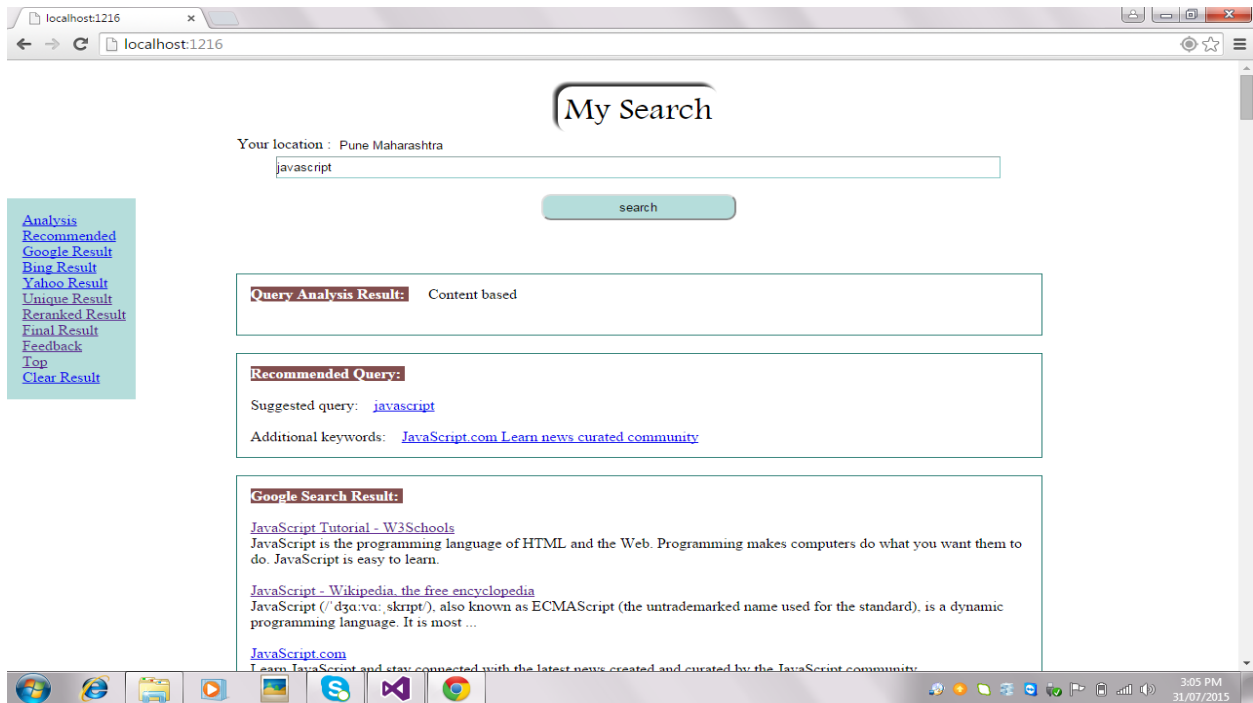
EXPERIMENTAL SETUP

When user submits a query, system will reform the original query by analyzing query and finding type of query (content or location based) to understand particular user's need of information. The click-through of particular user helps system in displaying most used result, used in Reranking of search results. After reforming query, system will forward this query to multiple search engines, where each search engine will display number of search results, out of which developed system will extract top search results and sort them for finding unique single set of results for query and then present them to user.

A framework is used to discover different user search objectives for a query by clustering feedback sessions. Feedback sessions contain both clicked and unclicked links from unique results displayed; reflect the information needs of users. The developed system then generates pseudo-documents which better represent the feedback sessions. Clustering of pseudo documents is performed to find the clusters of user search objectives. The grouping of different user search objectives can be done efficiently after clustering. Unique search results from multiple search engines will then categorized into different groups as said previously. Hence system derives the number of user search objectives for a submitted query.

The result of the system consists of following screenshots. Screenshot 1 shows how the search engine looks like. In given textbox user can enter his/her search string i.e. query. Here user has entered query "javascript" and clicked "search" button, in turns server will display following results. Query Analysis Result gives the 'type of query', here 'javascript' is a 'content based' query.

Recommended Query gives the additional words from feedback sessions generated through click through logs. When user will click on the additional words then automatically the search engine will fire a new query which contains original query entered by user plus additional keywords from Recommended Query.

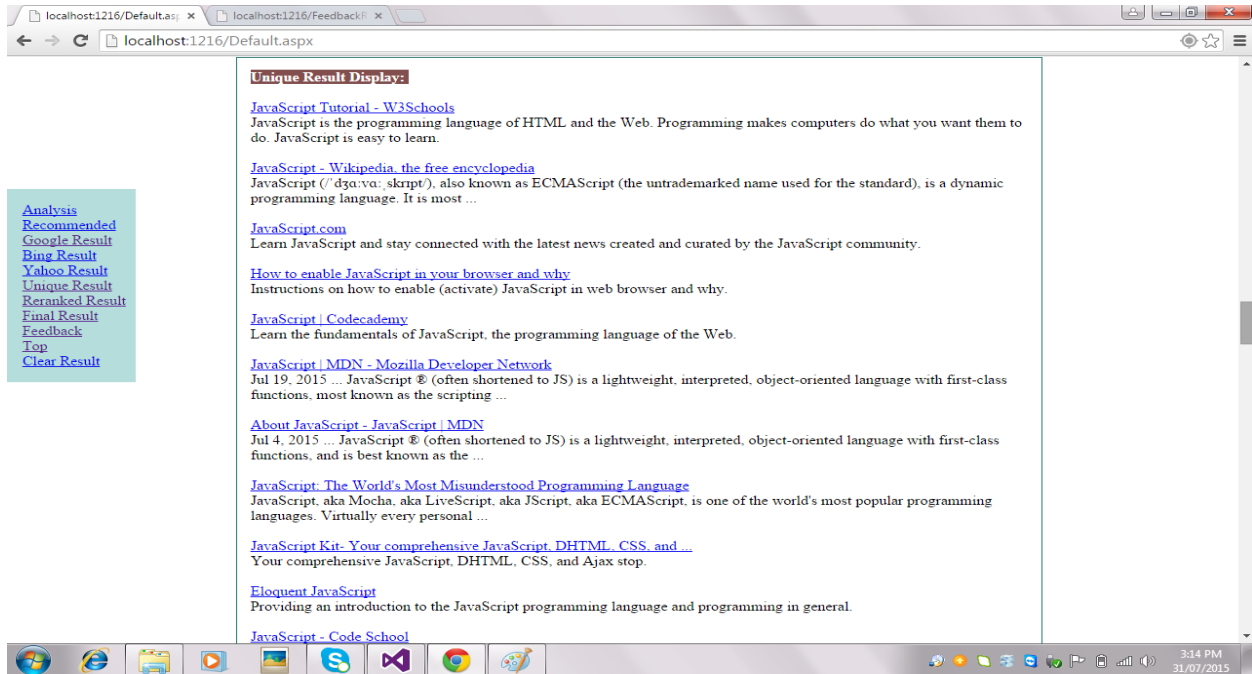


Screenshot 1 'javascript query'

Query entered by user is forwarded to the Multiple Search Engines after user clicks 'search' button. Google, Bing and Yahoo Search Engines are used for increasing the accuracy of search results and in turn the accuracy of user search objective. As shown in above screenshot it is displaying the results extracted from Google API for query 'javascript'. It displays total of 20 results from Google API. As the 'javascript' query is entered by user it is giving the top results for the same query. System has also extracted the results from Bing Search Engine API can be seen by clicking on 'Bing Result' tab. For Yahoo search engine, system has extracted the results from first page of yahoo results. Which are displayed in search engine interface, using 'Yahoo Results' tab user will be able to see the results. Tabs given on Left side of page will drag a control to respective results of activities. 'Feedback' will open the new web page displaying the contents of tables Pseudo Documents and Click sequences. 'Top' will move control to top of current web page. 'Clear Result' will clear all the results so that user can enter new query.

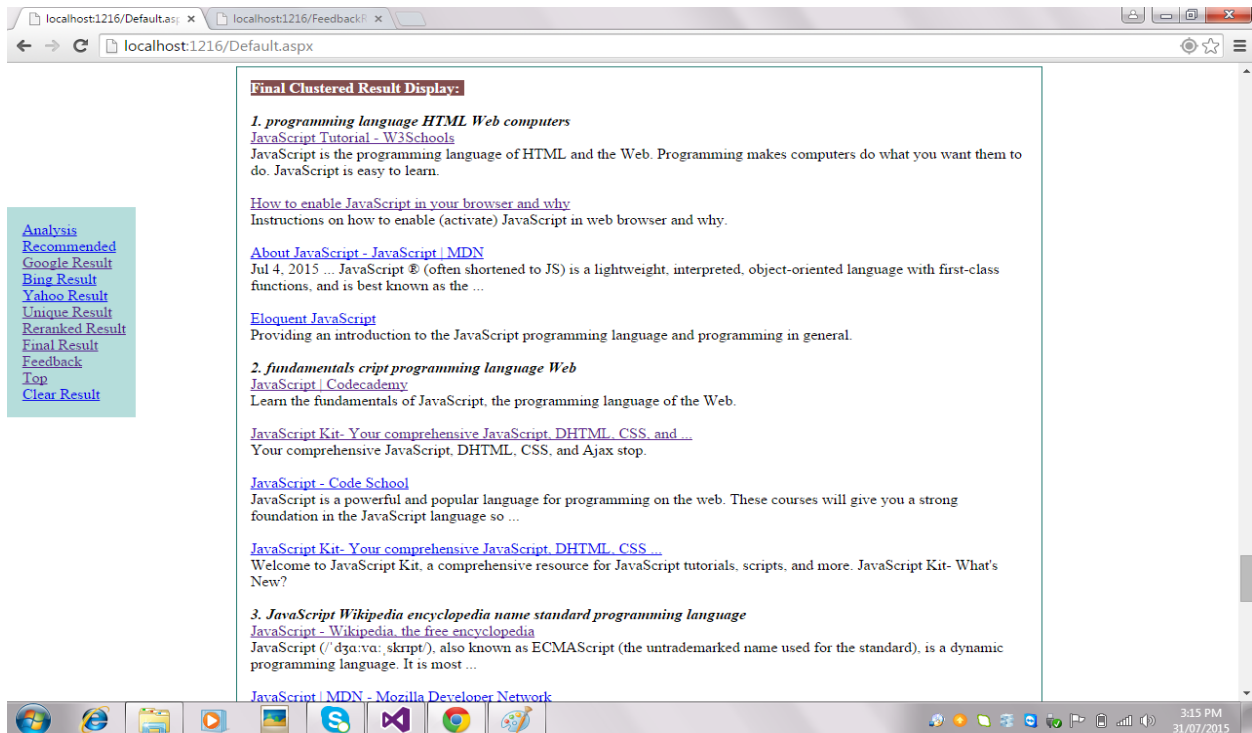
'Unique Result' display (Screenshot 2) gives the single bunch of sorted results from all above multiple search engines. Results are sorted to maintain the consistency. Duplicate results are distracted from final unique resultset. Now User have the final top rated results for the query 'javascript', here user can click on the links which are relevant to his/her objective of search. After user clicks any link, in new tab that particular page will get opened. So on for any number of clicks.

Finally Clustering is performed on the Pseudo Documents which will Derive the "User Search Objectives" for the query 'javascript'.



Screenshot 2 'Unique Result Display'

Screenshot 3 gives the first 3 'User Search Objectives' respectively. Search Result Links which fall under particular cluster (Group) will give information related to that particular 'User Search Objective'. Hence it will give a Faster, Accurate and More Precise Search Results from the point of view of "User Search Objective"

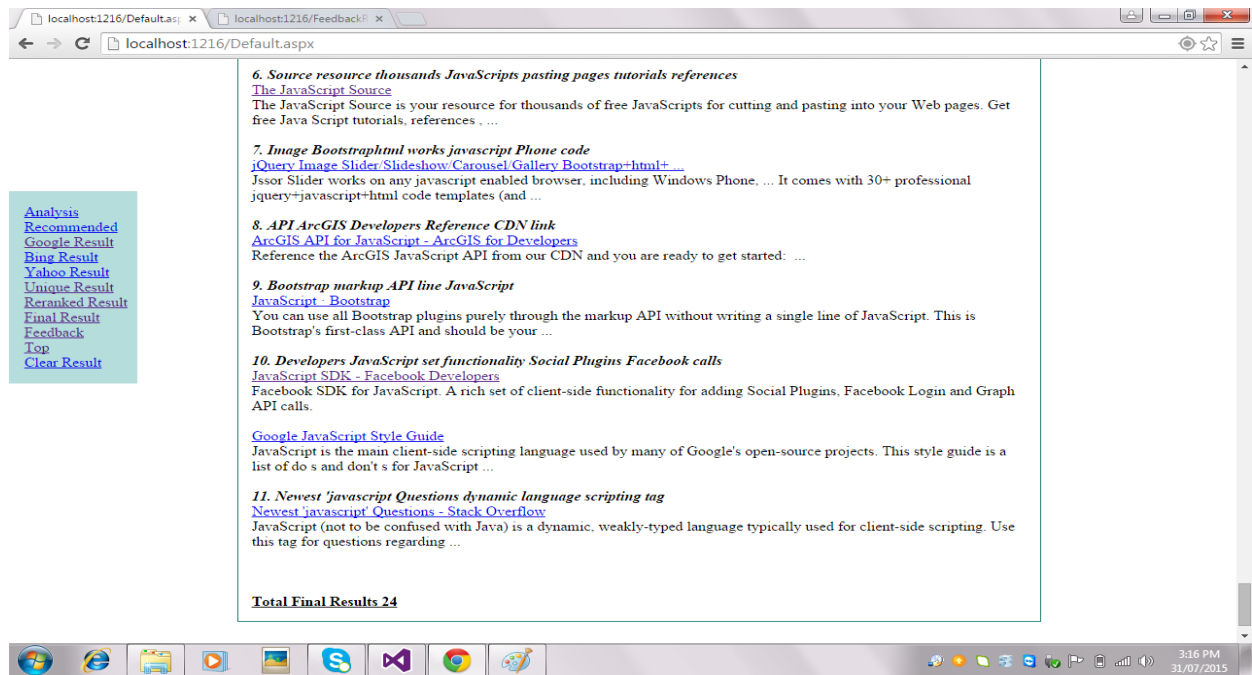


Screenshot 3 'Final Clustered Result Display'



Screenshot 4 ‘Final Clustered Result Display’

Screenshot 4 and Screenshot 5 gives the remaining Clusters of “User Search Objective”. Total 11 “User Search Objectives” for 24 “Unique and Re-ranked Search Results”



Screenshot 5 ‘Final Clustered Result Display’

CONCLUSION

System uses an Innovative Algorithm for deriving ‘User Search Objectives’. It is required and likely to derive ‘User Search Objectives’. The information on different forms of a query that user groups need to find is nothing but user search objectives. Advantages of deriving and analysing ‘User Search Objectives’ in improving search engine importance and user experience includes developed system can be able to restructure web results according to user search objectives by grouping the search results with same search objective; therefore users with different search objectives can find what they actually want easily. User search objectives depicted by some keywords can be used in query recommendation. Reranking of web search results can also be performed using click-through logs of user. It also helps in tracking user search behaviour. When a user submits query, the system returns the results that are categorized into different groups according to user search objectives. Hence system derives user search objectives for a query.

REFERENCES

- [1] “A New Algorithm for Inferring User Search Goals with Feedback Sessions” Zheng Lu, Student Member, IEEE, Hongyuan Zha, Xiaokang Yang, Senior Member, IEEE, Weiyao Lin, Member, IEEE, and Zhaohui Zheng
- [2] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. ACM Press, 1999.512 IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 3, MARCH 2013
- [3] R. Baeza-Yates, C. Hurtado, and M. Mendoza, “Query Recommendation Using Query Logs in Search Engines,” Proc. Int’l Conf. Current Trends in Database Technology (EDBT ’04), pp. 588-596, 2004.
- [4] D. Beeferman and A. Berger, “Agglomerative Clustering of a Search Engine Query Log,” Proc. Sixth ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining (SIGKDD ’00), pp. 407-416, 2000.
- [5] S. Beitzel, E. Jensen, A. Chowdhury, and O. Frieder, “Varying Approaches to Topical Web Query Classification,” Proc. 30th Ann. Int’l ACM SIGIR Conf. Research and Development (SIGIR ’07), pp. 783-784, 2007.
- [6] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li, “Context-Aware Query Suggestion by Mining Click-Through,” Proc. 14th ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining (SIGKDD ’08), pp. 875-883, 2008.
- [7] H. Chen and S. Dumais, “Bringing Order to the Web: Automatically Categorizing Search Results,” Proc. SIGCHI Conf. Human Factors in Computing Systems (SIGCHI ’00), pp. 145-152, 2000.
- [8] T. Joachims, “Evaluating Retrieval Performance Using Click-through Data,” Text Mining, J. Franke, G. Nakhaeizadeh, and I. Renz, eds., pp. 79-96, Physica/Springer Verlag, 2003.
- [9] T. Joachims, “Optimizing Search Engines Using Clickthrough Data,” Proc. Eighth ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining (SIGKDD ’02), pp. 133-142, 2002.
- [10] T. Joachims, L. Granka, B. Pang, H. Hembrooke, and G. Gay, “Accurately Interpreting Clickthrough Data as Implicit Feedback,” Proc. 28th Ann. Int’l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR ’05), pp. 154-161, 2005.
- [11] R. Jones and K.L. Klinkner, “Beyond the Session Timeout: Automatic Hierarchical Segmentation of Search Topics in Query Logs,” Proc. 17th ACM Conf. Information and Knowledge Management (CIKM ’08), pp. 699-708, 2008.
- [12] X. Wang and C.-X. Zhai, “Learn from Web Search Logs to Organize Search Results,” Proc. 30th Ann. Int’l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR ’07), pp. 87-94, 2007.
- [13] H.-J. Zeng, Q.-C. He, Z. Chen, W.-Y. Ma, and J. Ma, “Learning to Cluster Web Search Results,” Proc. 27th Ann. Int’l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR ’04), pp. 210-217, 2004.