International Conference on Recent Innovations in Science, Engineering & Management
R.R. INSTITUTE OF TECHNOLOGY, Bengaluru (India)
10th - 11th May 2019 www.conferenceworld.in ISBN : 978-93-87793-87-3

Implementation of Exploratory Search EnginewithSearch Goal-based QueryRecommendation

Ashwini V. Borkar, Dr. K.A. Waghmare

Dept. of Comp. Sci. and Engg, Govt. College of Engineering,Amravati (Maha),India

Dept. of Comp. Sci. and Engg, Govt. College of Engineering,Amravati (Maha),India

Abstract— *Exploratory search is the term which is effective in terms of scrutinize the searcher's result. In todays advanced technological world, exploratory search is a supreme venture and plays vital role in piercing. Therefore, we had done in-depth perusal and provide a novel approach in exploratory search along with the goal-based query recommendation technique. Firstly, we proposed a search engine with novel approaches for query recommendation and result re-ranking. For query recommendation, search goal shift technique with the help of topic similarity and query similarity is used which is based on machine learning algorithm. In addition, we provide the effectiveness of this novel method and provide an extraordinary way for exploratory search.*

Keywords: Exploratory search, query recommendation, search goal shift, KNN.

I. INTRODUCTION

In today's technological world, all the searchers are under the influence of the advanced technology. Everyone has progressed through science and technology and with the help of innovative way finds the path to explore the search and results. The topic to which we concern is exploratory search, which is grown form the information retrieval and information seeking but has alternatives to the kind of search that has received majority of focus. What is exploratory search? – exploratory search is the activity where user with the unfamiliar domain or goal are requested to seek the information which is destined. It is often, the web searcher desire to be more accurate but ubiquitous results came up and accurate search need to be grown.

To support an exploratory search the search system, need to be more accurate and for that along with the exploratory search the query recommendation technique is used which helps in explore the results. The current query recommendation techniques are mainly focus on the users given information which is far away from the user's desire [1]. To support the exploratory search and query recommendation we came with the effective approach that has suggest the new ideas and novel aspects that needs to be follow to explore the results.

Also, with the rapid increase in the scientific publications, it is much more difficult to the researchers to search particular citations and find the appropriate result. the new way and modern technique need to be established so that it will beneficial for the society of science. In the field of data mining, the KNN classification algorithm plays an important role in the text classification. From the bunch of data, it is difficult to find and understand the related data and to realize which is our concern topic. So in this paper we come to know that the related suggestions plays an important role and the more experiments and research work is needed in this field.

The main focus of this paper is on the query recommendation and search goal shift technique. The work we present here are effective than any other previously mentioned methods.Here, in the proposed system, the relationship between various queries are calculated and put the further evaluation results. These results make the system more attractive and trust worthy. The semantic association and the string similarity provided that makes the system more accurate in providing results. The semantically related keywords from the queries and the permutation combination concept provides the better result to convince the user's desired results.This paper computes the total time required for the calculation of similarity distance between the user's query and the suggested query. It also provides the similarity method by which two queries are related. In this paper we will see is there a search goal shift or not? And if shifted then how much time it takes to shift and provide the suggested query? Meanwhile the paper provides the details working of the provide system and is trustworthy in all activities.

It is quite hard to achieve and provide the desired results to the user by traditional system but if we take a look to the future scope of this experiment, it looks that it will create new magical era in the science world.

In order to avoid the traditional searching and suggesting problems, this new system provides an effective techniques and much more reliable results. Many of the previous works which are done shows the influence of their work on the proposed work system and the technology that has been used currently.

II. LITERATUREREVIEW

To meet the user's demand and looking for the satisfaction of users, the various previous work is appreciable. The topic oriented exploratory search based on an indexing network suggest the theme of semantic association graph. And the concept of query reformulation using wordnet and genetic algorithms makes the research more effective in this area.

A. Topic Oriented

The topic oriented exploratory search [2] which is based on indexing network proposed a new approach and allows the discovery of new associations and knowledge. The work is based on the semantic association graph and popularity of topics. For exploratory search, the expansion of keywords and related topics has been done. On the basis of experimental design, result and evaluation this method shows that the exploratory search can improve user search experience.

B. Concept Understanding Ability

The method focuses on the computation of concept understanding ability [3] based on the knowledge of the user. The concept association model and knowledge model based on the folksonomy were introduced.

C. Random Walk and Topic concepts

Query suggestion method based on random walk and topic concepts (QuS-RWTC) [4] is the method which is used to build the query URL bipartite graph. The transition probability has been computed from the initial and final query. From the preliminary and final query set, the usergets the higher coverage and wider scope for their results.

D. Query Reformulation

The author introduces the approach of wordnet and genetic algorithm for query reformulation [5] which satisfy the user in many ways and provide an efficient result. very simple concept is used in this paper, if the queries are more ambiguous then the result will be more effective.

E. Cuckoo optimized query recommendation

Cuckoo search is actually used to optimize the weight of the graph and for the query flow graph. The cuckoo search method [6] is more efficient than any other methods for the recommendation of the query. It helps in finding the query similarity and for the expansion of query and also for the reformulation of the query.

F. Negative Relevance Feedback

While searching in the scientific areas where focus is on promoting a explore results the negative relevance feedback [7] plays and important role while others look only for the positive response. It is used with the visual interactive intent modeling. It shows that the quality of retrieved information is much better than the positive feedback and better for the exploratory search

III. PROPOSED METHODOLGY

3.1 Query Recommendation

Along with the exploratory search, query recommendation is another part which based on the search goal shift graph that helps to provide better suggestions to the searcher which want the exact desired result from the house of information.

3.2 Search Goal Shift

Another main task of our searching is search goal shift graph which relates for the query recommendation. Search goal shift method improves the result and gives more effective exploratory search. To detect the search goal shift various algorithms are used. Here the goal shift obtained from different algorithms and works efficiently.

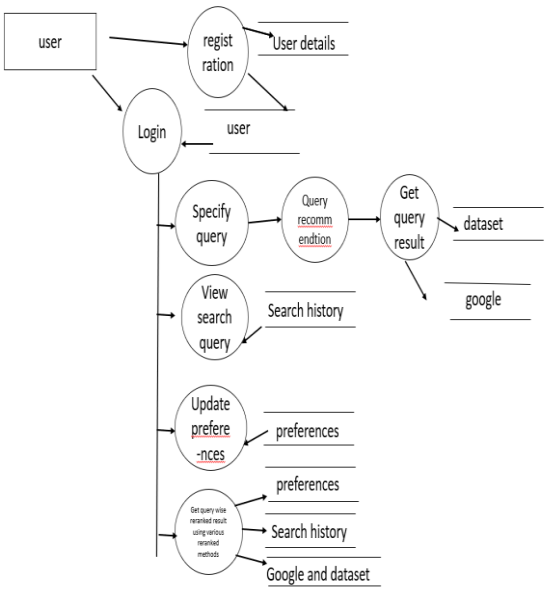


Fig 3.2Data Flow Diagram

3.2.1 Topic similarity

In the topic similarity concept, the sessions are provided to the user. The tokenization is applied on the queries in the session. For that purpose, we maintain a stopword dictionary which will remove the stop-words from the query and topic similarity will be easier.

3.2.2 Semantic similarity

In semantic similarity, On the basis of queries that are collected in one session the semantic similarity is done. Suppose two queries q1 and q2 are collected then in order to detect the similarity we use NLP i.e. Natural Language Processing. As we are using Wordnet to form the semantic relationship, the NLP is necessary.

3.2.3 String similarity

The term string similarity refer as query similarity or Jaccard Index. The Jaccard Index also known as intersection over union. The Jaccard similarity coefficient is a statistic used for comparing the similarity and diversity of sample sets.

The formula to find the Index is:

Jaccard Index = (the number of keywords present in both keywords sets) / (the number of keywords present in either set) * 100

The same formula in notation is: $\mathbf{J(X, Y) = |X \cap Y| / |X \cup Y|}$

In Steps, that's:

1. Count the number of members which are shared between both sets.
2. Count the total number of members in both sets (shared and un-shared).
3. Divide the number of shared members (1) by the total number of members (2).

Multiply the number you found in (3) by 100.

This percentage tells you how similar the two sets are.

- Two sets that share all members would be 100% similar. the closer to 100%, the more similarity (e.g. 90% is more similar than 89%).
- If they share no members, they are 0% similar.
- The midway point — 50% — means that the two sets share half of the members.

- **Search Goal Shift Detection:**

The search goal shift detection is main part of our proposed system which shows us whether the user search for relevant queries so that the suggestion will be accurate or the goal shift in another direction. For that we used jaccard similarity algorithm which shows how the queries are similar to each other so that it concludes the relevant result.

Algorithm 1: Search Goal Detection

1. START
2. Set Input =specify search query
3. Set keyw[]=Keywords_Extraction_using_NLP(input)
4. Set matchingQueries[]=Select queries Matching with keyw[]
5. If matchingQueries[].len=0 then
    a. **Set searchGoalShift=YES**
6. Else
    a. Calculate jaccard index for matchingQueries[]
    b. Set jaccardDist[]=Calculate JaccardDist(matchingQueries[])
    c. If jaccardDist[].len>0 then
        i. set searchGoalShift=NO
    d. else
        i. set semanticQueries=Fetch Queries semantically related to input query
        ii. if semanticQueries.len>0 then
            1. **set searchGoalShift=No**
        iii. else
            1. **searchGoalShift=Yes**
        iv. end if
    e. end if
7. End if

- **Query recommendation based on K - nearest neighbor**

k-nearest neighbor is the non-parametric, lazy learning algorithm. Its purpose is to use a database in which the data points are separated into several classes to predict the classification of new sample points.

As the main goal of our experiment is to provide the query recommendation based on search goal shift which is achieved here with the help of KNN. Here we calculate the Euclidian distance between the query and the dataset. The Euclidian distance is arranged and then calculated. The array is in ascending order.

Algorithm 2 :K- nearest neighbor

1. START
2. Set Input= Search Query
3. Set k=10
4. Set searchQueries[]= fetch related queries from online API
5. For i=0 to searchQueries.len
    a. Set eudist[i]=EUCLIDIAN_DIST(input, searchQueries[i])
6. End For
7. Sort eudist[] in ascending order
8. Select first k queries and display on page
9. END

- **Working of system**

The figure 3.2.2 shows the detail working of the experiment done for the query recommendation and exploratory search. In the developed system, the user has to register themselves and get authorized.

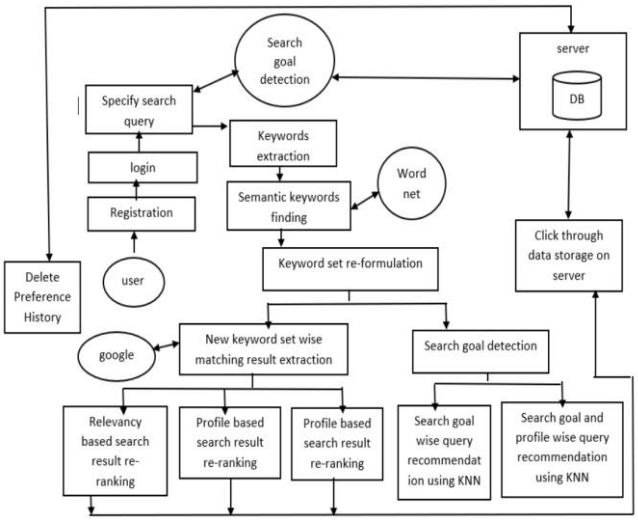


Fig 3.2.2 Block diagram

After registration, the query gets fired. Form the query similar keywords get extracted. The semantic relationship is calculated from the keywords and keywords get reformulated. From the extracted keywords the different types of result re-ranking like relevancy, profile and preference are done. On the other hand, at the

same time of keyword extraction and reformulation, the search goal detection is occurred. The data gets fetched from the google and the searched data is saved to the user's database.

In the proposed system, various functionalities and technical aspects are used which are describe in the section of the implementation. The working of natural language processing (NLP), Wordnet, KNN, jaccard similarity, search goal detection various terms are included in the proposed system which works efficiently to make proposed work more reliable, more competitive and more efficient than the previous one.

EXPERIMENTAL RESULTS AND DISCUSSION

In our experiments, we need a collection of data but as we are not using any real time data, we are going to fetch it from the google with the help of application programming interface (API). Below is the graph of searching time evolution which shows the total time consume for the results of user's query.

- **Evaluation of recommendation result-**

Here, in the fig. 4.1 searching time evaluation shows the total time required to search the suggested query to the specified query. The time is in mili-seconds and it will increase as per the query increased in the sessions.

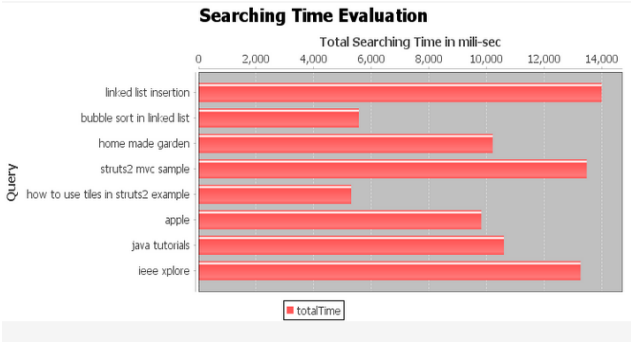
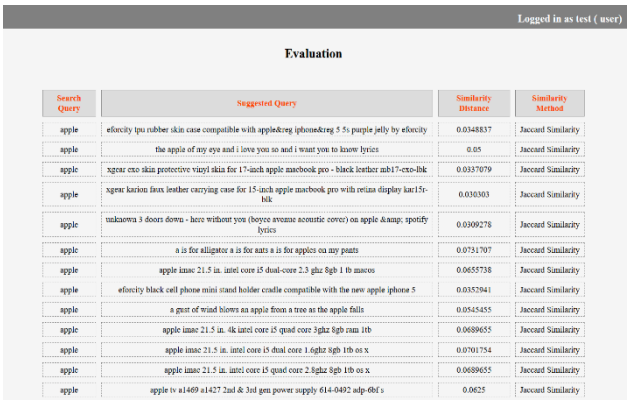


Fig 4.1 Searching Time Evaluation.

As we are using the Euclidian distance, the similarity measures of the distances are estimated in the KNN graph. The result is stated below in the figure which is mentioned in our experiment. Fig. 4.2 shows the evaluation of the search query and suggested query. It also shows the similarity distance and the similarity method by which the queries are related.



Logged in as test (user)

Evaluation

Search Query	Suggested Query	Similarity Distance	Similarity Method
apple	eforcity tpu rubber skin case compatible with apple&reg iphone&reg 5 5s purple jelly by eforcity	0.0348837	Jaccard Similarity
apple	the apple of my eye and i love you so and i want you to know lyrics	0.05	Jaccard Similarity
apple	xgear exo skin protective vinyl skin for 17-inch apple macbook pro - black leather mb17-exo-blk	0.0337079	Jaccard Similarity
apple	xgear karion faux leather carrying case for 15-inch apple macbook pro with retina display kar15r-blk	0.030303	Jaccard Similarity
apple	unknown 3 doors down - here without you (boyce avenue acoustic cover) on apple & spotify - lyrics	0.0309278	Jaccard Similarity
apple	a is for alligator a is for aats a is for apples on my pants	0.0731707	Jaccard Similarity
apple	apple imac 21.5 in. intel core i5 dual-core 2.3 ghz 8gb 1 tb macos	0.0655738	Jaccard Similarity
apple	eforcity black cell phone mini stand holder cradle compatible with the new apple iphone 5	0.0352941	Jaccard Similarity
apple	a gust of wind blows an apple from a tree as the apple falls	0.0545455	Jaccard Similarity
apple	apple imac 21.5 in. 4k intel core i5 quad core 3ghz 8gb ram 1tb	0.0689655	Jaccard Similarity
apple	apple imac 21.5 in. intel core i5 dual core 1.6ghz 8gb 1tb os x	0.0701754	Jaccard Similarity
apple	apple imac 21.5 in. intel core i5 quad core 2.8ghz 8gb 1tb os x	0.0689655	Jaccard Similarity
apple	apple tv a1469 a1427 2nd & 3rd gen power supply 614-0492 adp-6bf s	0.0625	Jaccard Similarity

Fig 4.2. Similarity distance calculation

In Fig. 4.3 the comparison graph for query suggestions shows the result of normal search engine suggestions vs proposed system suggestions. Here we search some queries on normal search engine and proposed system, the given are the suggestions which are provided by the both systems. It is clearly visible that our proposed system provide more suggestions than the previous one.

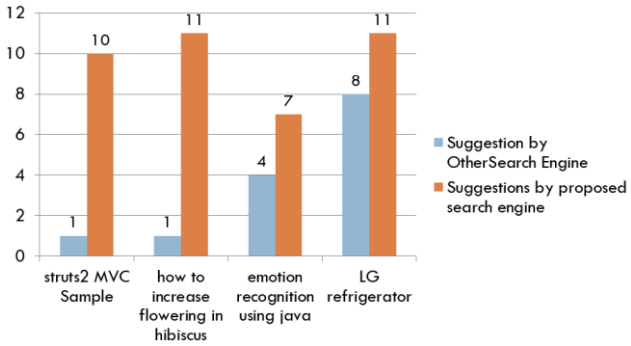


Fig 4.3: Comparison Graph for Query Suggestion

In Fig. 4.4 we took a various queries and shows the search rank obtained by both the systems.

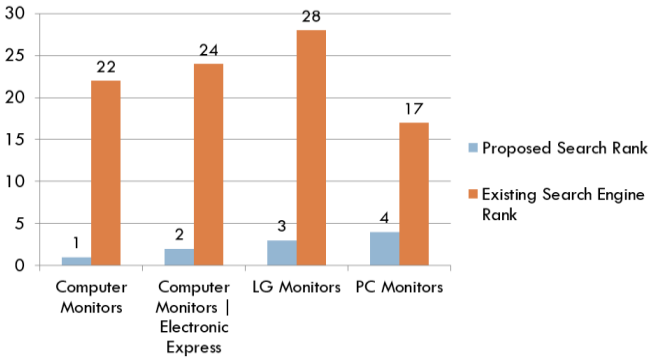


Fig 4.4 : Search Rank Graph for Normal and Proposed Search Engine

IV. CONCLUSIONS

In proposed system, we introduced profile-based query recommendation technique along with search goal wise query recommendation. This may improve the recommendation. We focus on query recommendation as well as searching mechanism. To improve matching search result in case of improper query, we proposed semantic searching system.

V. FUTURE SCOPE

The future scope of the system is to provide better recommendations and the more efficient results to the user. In future, we can explore the results by using more efficient algorithms and techniques. Also, may keywords have synonyms that are not consider here so to provide exact result if synonyms of any query are inserted.

REFERENCES

[1] c. ma and B. Zhang, "A New Query Recommendation Method Supporting Exploratory Search Based on Search Goal Shift Graphs," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 11, pp. 2024-2036, 1 Nov. 2018.

[2] Sun H C, Jiang C J, Ding Z J, et al. "Topic-Oriented Exploratory Search Based on an Indexing Network." *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no.2, pp. 234-247, 2016.

[3] Xinwei Liu, "Estimating concept understanding ability for exploratory search query recommendations," *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, Chengdu, 2017, pp. 1024-1028.

[4] J. Liu, Q. Li, Y. Lin and Y. Li, "A query suggestion method based on random walk and topic concepts," *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, Wuhan, 2017, pp. 251-256.

[5] B. Al-Khateeb, A. J. Al-Kubaisi and S. T. Al-Janabi, "Query reformulation using WordNet and genetic algorithm," *2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT)*, Baghdad, 2017, pp. 91-96.

[6] S. Jagan and S. P. Rajagopalan, "Cuckoo optimized query recommendation in web search," *2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET)*, Chennai, 2017, pp. 1-7.

[7] Peltonen J, Strahl J, Floréen P. "Negative Relevance Feedback for Exploratory Search with Visual Interactive Intent Modeling," in *Proc.* The 22nd ACM International Conference on Intelligent User Interfaces, pp.149-159, 2017.

[8] T. Onoda, T. Yumoto and K. Sumiya, "Extracting and Clustering Related Keywords based on History of Query Frequency," *2008 Second International Symposium on Universal Communication*, Osaka, 2008, pp. 162-166.

[9] Q. Liu, M. Jiang and Z. Chen, "Query Recommendation with TF-IQF Model and Popularity Factor," *2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, Shandong, 2008, pp. 203-207.

[10] Ling Liu, Lin Li and Zhenglu Yang,Masaru Kitsuregawa, "Query-URL Bipartite Based Approach to Personalized Query Recommendation" AAAI'08 Proceedings of the 23rd national conference on Artificial intelligence - Volume 2

[11] Mei Q, Zhou D, Church K. "Query suggestion using hitting time," in *Proc. The 17th ACM conference on Information and knowledge management*, pp. 469-478, 2008.

[12] B. E. Ghali, A. E. Qadi, M. Ouadou and D. Aboutajdine, "Probabilistic Query Expansion method using recommended past user queries," *Second International Conference on the Innovative Computing Technology (INTECH 2012)*, Casablanca, 2012, pp. 406-411.

[13] M. A. Potey, D. A. Patel and P. K. Sinha, "A survey of query log processing techniques and evaluation of web query intent identification," *2013 3rd IEEE International Advance Computing Conference (IACC)*, Ghaziabad, 2013, pp. 1330-1335.

[14] Hassan A, White R W, Dumais S T, et al. "Struggling or explor-ing?: disambiguating long search sessions," in *Proc. The 7th ACM international conference on Web search and data mining*, pp. 53-62, 2014.

[15] J. Zahir, A. El Qadi and S. Mouline, "Access plan recommendation: A clustering -based approach using queries similarity," *2014 Second World Conference on Complex Systems (WCCS)*, Agadir, 2014, pp. 55-60.

[16] S. D. *et al*., "QRGQR: Query Relevance Graph for Query Recommendation," *2015 IEEE Region 10 Symposium*, Ahmedabad, 2015, pp. 78-81.

[17] L. P. Nanni and V. D. Feltrim, "Desire: A Dynamic Approach for Exploratory Search Results Recommendation," *2015 Brazilian Conference on Intelligent Systems (BRACIS)*, Natal, 2015, pp. 288-293.

[18] Bespinyowong R, Chen W, Jagadish H V, et al. "ExRank: an exploratory ranking interface." Proceedings of the VLDB En-dowment, vol. 9, no. 13, pp.1529-1532, 2016.

[19] Boldi P, Bonchi F, Castillo C, et al. "Query suggestions using query-flow graphs," in *Proc. The 2009 workshop on Web Search Click Data*, pp. 56-63, 2009.

[20] R. Umagandhi and A. V. S. Kumar, "Time independent query recommendations from search engine query logs," *International Conference on Software Engineering and Mobile Application Modelling and Development (ICSEMA 2012)*, Chennai, 2012, pp. 1-6.

[21] R. L. Cilibrasi and P. M. B. Vitanyi, "The Google Similarity Distance," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 3, pp. 370-383, March 2007.

[22] Bahareh sarrafzadeh, Edward lank, "improving exploratory search experience through hierarchical knowledge graph" SIGIR 17, august-2017 shinjuku,Tokyo,Japan.