A Survey on User Search Goal Inferring and Re-Ranking System

Hemlata Gaikwad[1], P. B. Kumbharkar[2]

[1, 2]Pune University, Department of Computer Engineering, Siddhant College of Engineering, Pune, Maharashtra, India

Abstract: *We know that a non-detachable part of human beings is the internet throughout the world. But as the internet called as an ocean of information so that each and every information we can get from the internet. To infer the user search goals through user profiles an excellent effort have made by many researchers, knowledge of user searching or history of user searching and pattern but their failed the most of the techniques as it's not that over the internet the same contents or documents there will always try to search by the user. To find the location specific queries another technique to guess the user goals made use of user location and answer them. Thus for the user goal search we are going to analyze all the implemented algorithms.*

Keywords: Hidden Web Crawler, Query Optimization, Search engines, Metadata, document frequency

1. Introduction

In many websites for finding the user need the search engine are widely used. As there are short in size queries i.e. normally two or three words. But an ambiguous results gives by these queries. For the user's expectations these result does not exactly matches. Many times different search result produced by different search engine. So that non useful results arises and to satisfy the user's expectations those are fail. Therefore we have proposed a user search goal inferring system to match the relevant search result with user's needs. In this we are treating the user's need as a cluster. This will be very useful to improve the performance of search engine. We can able to redesign the result by grouping the needs of the user at different time. The user need can assigned by a word on which the clustering will be done. The ranking of result are depending upon the clustering. Many methods were invented to make searching more effective like classification of query, recognition of search results, and session limit detection for better searching. However, since the number of different clicked URLs of a query may be small this method has limitations. By the search engine other works analyse the search results returned when a query is submitted.

Therefore, to search engines for issuing the queries there is no standard or optimal way. To adequately train machine learning approaches there focuses on classifying documents by most text classification research which contain enough terms. In that web queries are short the task of classifying web queries is different, very few inherent features providing. Therefore, as features most approaches use the documents retrieved to classify it by a query.

For example, the user has entered a query „phoenix" in Google search engine. Basically for phoenix as a bird it should produce the results. But it is displaying the result of a shopping mall in pune. The expected result is found to user but it is not ranked as a first result. Many times user have to search for many pages of search results to find his need. Every time user wanted to submit query „phoenix" instead of bird it will firstly shows the result of mall.

Figure 1: Variation In output of query „phoenix" submitted by user.

2. Research Elaborations

A. Query Recommendation Using Query Logs in search Engines

For a search engine in this paper author proposed a scheme which given a query submitted and a list of queries which are related are suggests. On previously issued queries the related queries are depends, and for the search engine to tune or redirecting the search process which can be issued by the user. Based on a process of query clustering the method proposed in which there are identified groups of semantically similar queries. The content of historical preferences of users are uses by the clustering process which are registered in the search engine's query log. The method not only ranks the related queries, but also discovers them according to a relevance criterion.
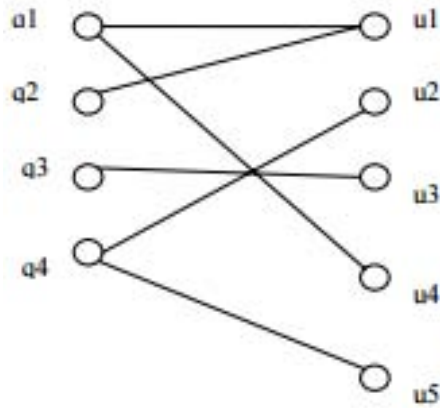
Figure 2: Query – URL representation as a bipartite graph

Finally, we show with experiments which effectiveness of the method over the query log of a search engine [1].

B. Varying Approaches to Topical Web Query classification

Because of the promise it offers in improving retrieval effectiveness and efficiency, recent interest has drawn by topical classification of web queries. However, whether classification is performed before or after much of this promise depends on it the query is used to retrieve documents. In classification of query: classification effectiveness of pre versus post-retrieval and from classified queries versus bridging a classifier trained using a document taxonomy the effect of training are examine two previously unaddressed issues explicitly. A document taxonomy categories are mapped by Bridging classifiers to provide sufficient training data onto those of a query classification problem. Author find that in F1 score from manually classified queries training classifiers explicitly outperforms the bridged classifier by 48%. Also, there only the query terms using a pre-retrieval classifier performs merely 11% worse than the bridged classifier from retrieved documents which requires snippets [2].

C. Context-Aware Query Suggestion by Mining Click-Through

A novel context-aware query suggestion approach is proposed by author in this paper which is in two steps. In the offine model-learning step, queries are summarized into concepts to address data sparseness by clustering a click-through bipartite. Then, as the query suggestion model, a concept sequence suffix tree is constructed from session data.
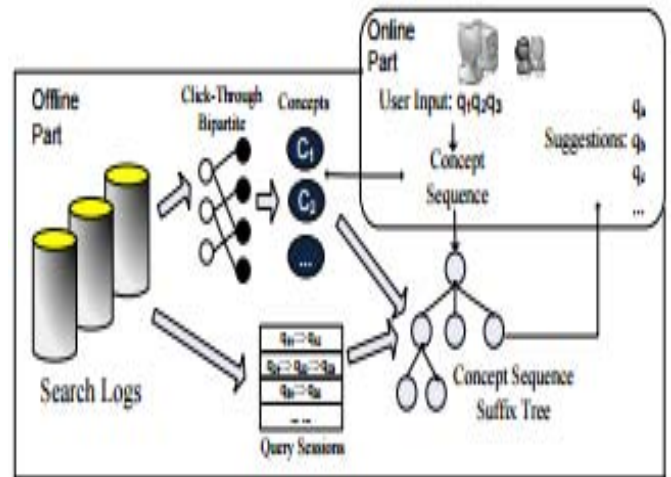

Figure 3. Our framework approach

A user's search context is captured in the online query suggestion step by mapping the query sequence submitted to a sequence of concepts by the user. This approach was test by author on a commercial search engine"s large-scale search log of which containing 2:6 billion clicks, 840 million query sessions, and 1:8 billion search queries. In both coverage and quality of suggestions this approach outperforms two baseline methods [3].

D. Bringing order to the Web: automatically categorizing search results

Figure 4: Presenting web pages within category structure

For organizing the Web search results into hierarchical categories author developed a user interface. For classifying the arbitrary search results automatically there were used the text classification algorithms into an existing category

structure on-the-fly. Our new category interface compared by a user study with the typical ranked list interface of search results.

In objective and subjective measures the category interface is superior, both was the study. There much better the subjects liked the category interface than the list interface, and that was organized into categories at finding information they were 50% faster. In categories of interest focus on items rather than having to browse sequentially through all the results [4].
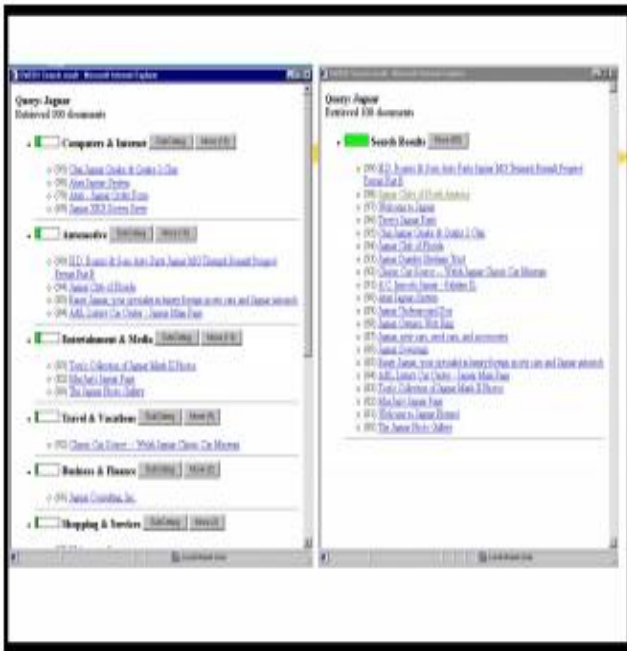

Figure 5: Category vs. List Interface

E. Relevant Term Suggestion in Interactive Web Search Based on Contextual Information in Query Session Logs
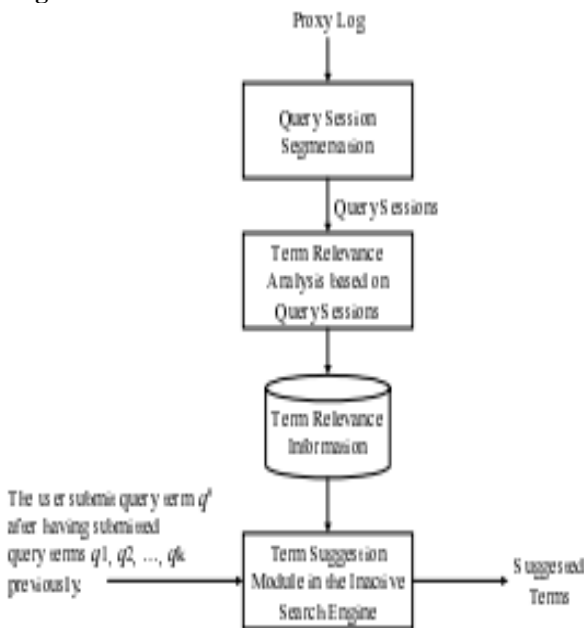
Figure 6: System framework

For extraction of relevant term and suggestion of term a new, effective log-based approach present by author. Using this approach, for a user query the relevant terms suggested rather than in the retrieved documents in similar query session from search engine logs which are those that co-occur.

About two million query transactions which are submitted to the search engines in Taiwan, using a proxy server log the proposed approach was tested. Organized and highly relevant terms can provide by the proposed approach, and in a user's query session can exploit the contextual information to make more effective suggestions [5].

F. Evaluating Retrieval Performance Using Click through Data

For evaluating the quality of retrieval functions in this paper, a new method proposed. By experts or explicit user feedback unlike traditional methods which relevance judgments require, on click through data is it depends. Since at very low cost the click through data can be collected and without overhead for the user this is a key advantage.

Figure 7: Example for query "support vector machine".

As evaluation with traditional relevance judgments the method gives the same results under mild assumptions which shown in a theoretical analysis. The assumptions are indeed justified are verifies by an empirical analysis and in a WWW retrieval study that conclusive results leads by that new method [6].

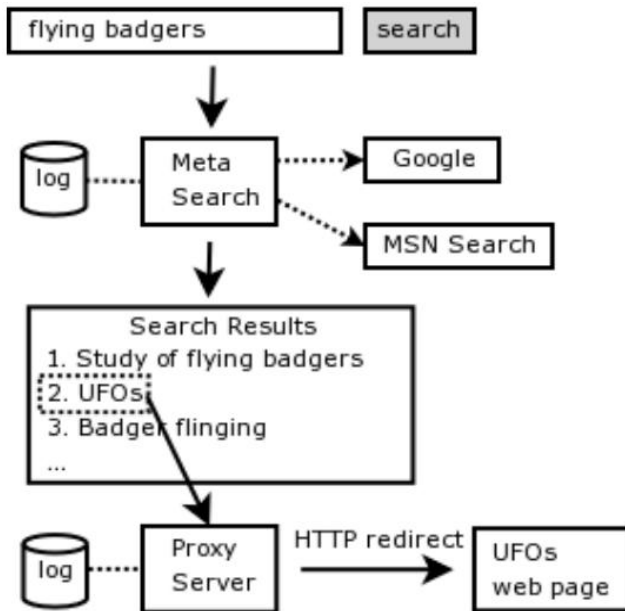### G. Optimizing Search Engines Using Clickthrough Data.

Figure 8: Search Procedure

For automatically optimizing the retrieval quality of search engines an approach is presents in this paper for using click through data. Intuitively, in the ranking there should present relevant documents high by a good information retrieval system, with their following below the less relevant documents. While from examples exist previous approaches to learning retrieval functions, from relevance judgments the training data generated by experts typically they require it. This makes them difficult and expensive to apply.

Figure 9: Ranking presented for the query "support vector machine". Marked in bold are the links the user clicked on.

To develop a method that utilizes clickthrough data for training is the goal of this paper, the users clicked on in the presented ranking with the log of links namely the query-log of the search engine in connection. At very low cost in abundance and can be recorded such clickthrough data is available. For learning retrieval functions this paper presents a method taking an approach of Support Vector Machine

(SVM). In a risk minimization framework this method is shown to be well-founded from a theoretical perspective. Furthermore even for large sets of queries and features, it is shown to be feasible. In a controlled experiment the theoretical results are verified [7].

H. Accurately Interpreting Clickthrough Data as Implicit Feedback

From clickthrough data in WWW search the implicit feedback generated is examines in this paper. Against manual relevance judgments using eyetracking as well as comparing implicit feedback analyzing the users" decision process, It can be concluded that clicks are informative but biased.

Figure 10: Questions used in the study.

While as absolute relevance judgments difficult this makes the interpretation of clicks, we show that from clicks are reasonably accurate the relative preferences derived on average.

3. Conclusion

In this paper using search results or clicked URLs, to analysed to infer user search needs we have introduce feedback sessions. As user implicit feedbacks both the clicked URLs and the unclicked ones are considered before the last click and to construct feedback sessions it taken into account. Here we have maintain the sequence of most relevant search results to represent need of user. We have used the concept of pseudo documents to design the feedback sessions. This concept will make the searching easy to user. And it is producing most relevant results.

References

[1] 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.

Paper ID: NOV151668
2111

[2] 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.

[3] 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.

[4] 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.

[5] C.-K Huang, L.-F Chien, and Y.-J Oyang, "Relevant Term Suggestion in Interactive Web Search Based on Contextual Information in Query Session Logs," J. Am. Soc. for Information Science and Technology, vol. 54, no. 7, pp. 638-649, 2003.

[6] T. Joachims, "Evaluating Retrieval Performance Using Clickthrough Data," Text Mining, J. Franke, G. Nakhaeizadeh, and I. Renz, eds., pp. 79-96, Physica/Springer Verlag, 2003.

[7] 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.

[8] 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.

Author Profile

Hemlata Gaikwad received her B.E. degree in Computer Science and Engineering from Pune University (Maharashtra India) in 2011, and Pursuing M.E. degree in Computer Engineering from Siddhant College of Engineering Pune (India).

Prashant Kumbharkar received his B.E. degree in Computer Science and Engineering from Amravati University(Maharashtra India) in 1994, his M.E. degree in Computer Engineering from Bharati Vidyapeeth Pune (India) in 2005 and pursuing Ph.D. Degree in Computer Science Engineering at JJT University, Rajasthan. He has been working as an Associate Professor and Head of Department in the Department of Computer Engineering at Siddhant College of Engineering Pune, India since June 2008.He has 20 Yrs of experience in the field of education. He has published various papers in National and International journals. His research interests include computer networks, particularly in network optimization, Network Security, Unix Programming and Databases.