Predicting User Search Intention By Analyzing Semantic Keywords Harvested From Short Texts

Shruti Shridhar Mhatre[1], Dr. Anitha Patil[2]
[1,2]Department of Computer Engineering
[1,2]Pillai's HOC Collage of Engineering & Technology, Rasayani, Maharashtra, India
Email: shrutisarle@gmail.com

Abstract:

It's a known fact that the World Wide Web has become the pinnacle of our daily lives. A never ending volume of information will be found in the deep abyss of the internet; more or less every single idea can be found online. Numerous attempts are, have been made by an equally number of researchers to extract the user behaviour with respect to their search criterions. An understanding of one's online search attributes, search habits, their pattern of search does not quite able to corner down the user goals. Most of the search criterions do not include the data and is therefore prone to ambiguous results. 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. For location related searches user location is necessary to understand the use of location related services. Here we will try to collate and analyse all the relevant techniques and algorithms for the user.

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

1. Introduction

Most of the websites make use of search engine to find what the needs of the user are, as the queries are often short in size, and the queries often are prone to ambiguous results. These result do not exactly match the user's expectations. Also, there are times when different search engine give out different search result. The non-useful results may at times show up at the very top up and could well be used by the user, well these are the ones which cause dubious results. 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 beneficial as it will help improve the overall performance of search engine. Grouping relevant needs of the user at different intervals will help us redesign the end result. Different unique, non-unique needs of the user can be assigned by a word on which the clustering will be done. The ranking of result depend on the clustering. For a search to be more effective, query classification, search result recognition, and techniques to limit sessions were invented. However, since a user may not clink multiple search URL's for a search query this method has a few restrictions. The results from the searches when a query is submitted by the search engine are analyzed.

Therefore, there is no known way or an optimized model for search engines to issue queries. They focus on classifying documents by most text classification research which have adequate terms to sufficiently train machine learning techniques. In those web queries are of limited characters the way to classify those queries is disparate, and inherently very few features

providing. Therefore, to classify it by a query as features most common approaches features use document retrieval.

For example, ideally, in Google search engine if the user enters a query 'phoenix' it should first display the result for phoenix, a symbol of a bird reborn from ashes or the capital of Arizona instead it shows of a shop in the near locality. The result of course is displayed but it's not really the on which the user is looking for or in the top hit list. In fact there are times when the user will eventually will have to browse through many pages to get to the expected, satisfactory result. The very time the user will search for the word phoenix he will be returned the location of the nearest shop instead.

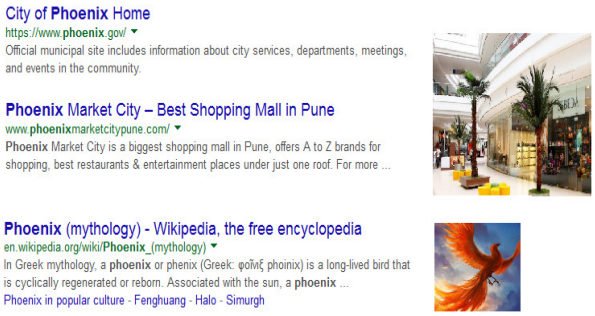


Figure 1. Variation In output of query 'phoenix' submitted by user

2. Literature Survey

2.1. Query Recommendation Using Query Logs in search Engines

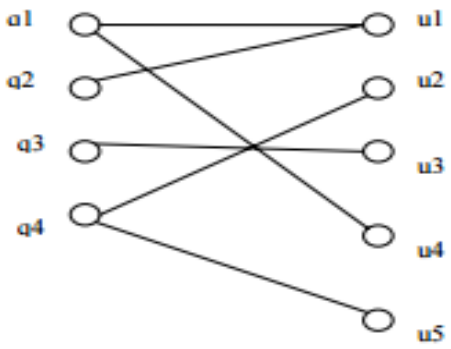


Figure 2. Query – URL representation as a bipartite graph

Another scheme is proposed where once a user submits a query the search engine will in turn put forward a list of suggestions which could be useful to the user and possible related to suggestions, thus related or similar queries depend, and search engine has to tune or direct the search process issued by the end user. The proposed method has identified groups of semantically similar queries based on a process of query clustering. The clustering process will basically store the searches over a period of time and/or which are already present in the sear engines log history. The content of historical preferences of users are used. The method will prioritize the related queries as well as discovers them according to a relevant favorable criterion. Finally, we show with experiments which effectiveness of this method over the query log of a search engine.

2.2. Deviating Approaches to Topical Web Query classification

Off late there has been an interest on topical classification of web queries. Claims are that it offers in improving retrieval effectiveness and efficiency, and because of how it promises to deliver. However, to retrieve documents whether classification is done prior or later, a lot of the promise depends on if the query is used. In classification of query: classification effectiveness of pre against 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. By Bridging classifiers to provide sufficient training data on classification of a query issue a document taxonomy categories are mapped. Author discovers that in F1 score manually classified queries training classifiers clearly performs better than bridged classifier by 48%. Also, the query terms using a pre-retrieval classifier performs merely 11% worse than the bridged classifier from retrieved documents for which snippets are necessary.

2.3. Context based Query Suggestion by Mining Click-Through

Context based query suggestion course is suggested by author, this system consists of two steps. By clustering a click-through binal in the suggested offine model-learning step, to look into data sparseness, queries are summarized into concepts. Then, from session data as the query suggestion model a concept sequence suffix tree is constructed.

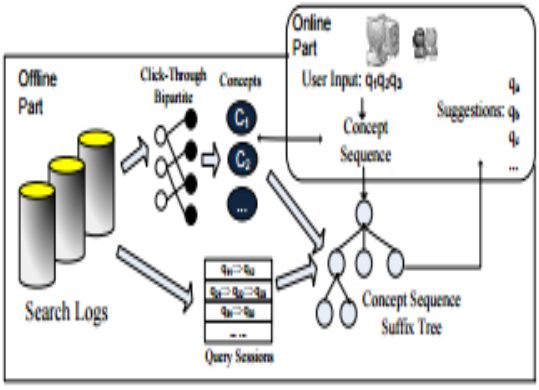


Figure 3. Addressing framework structure

In the online query suggestion step a user's search context is captured by mapping the query sequence submitted to a sequence of concepts. This approach was tested by the author on a commercial search engine's large-scale search log of which easily had approximately 2.6 billion clicks, approximately 850 million query sessions, and 1.8 billion search queries. In both coverage along with quality of suggestions the particular approach easily shoes better results than the previous baseline methods.

2.4. Getting an ordered web: automatically categorize the search results

Figure 4. Web Pages presented as structured category

A user interface to organize the search results into hierarchical categories. They text classification algorithms into a preexisting category structure while the session is in progress for classifying the arbitrary search results automatically. With the characteristic interface of ranked listed search results our new category interface compared by a user study.

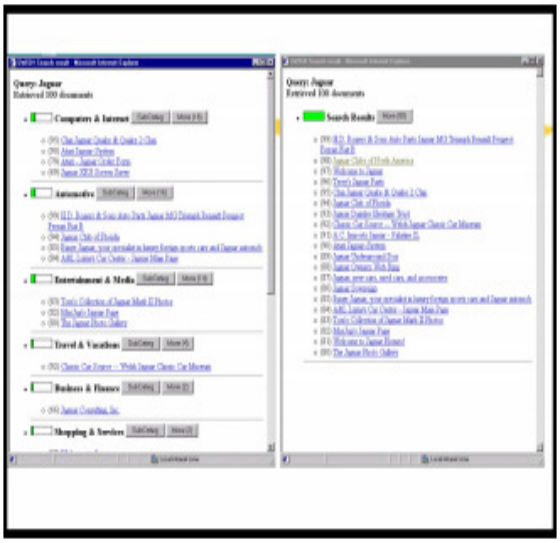


Figure 5. Category interface comparison with List Interface

Study showed that the category interface being superior both in terms objective and subjective measures as compared with the list interface. There much better the subjects liked the category interface, and was well organized into categories at finding information which were 50% faster. To browse sequentially through all the results in categories of interest focus on items rather than having.

2.5. Pertinent word Suggestion Interactive Web Search depending on Contextual Information in Query Session Logs

A new, effective system log approach is put forward by author for extradition of relevant word and suggestion of the same. For a user query the relevant words suggested rather than in the retrieved documents in similar query session from search engine logs which intern are the once which re occur Using this approach, About 2 million query transactions which are fed to a search engines in Taiwan, using a proxy server log the proposed approach was tested. By the approach suggested, Organized and highly relevant terms can provide information to make many such effective suggestions in a user's query session which can exploit the context.

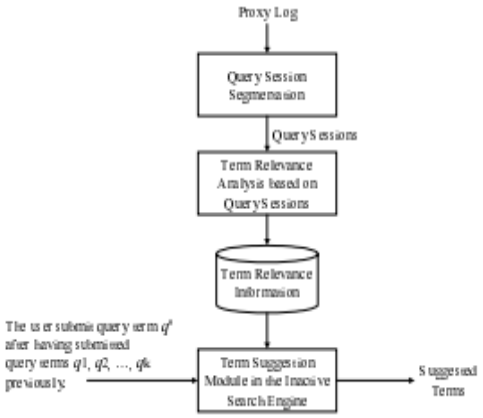


Figure 6. Framework

2.6. Evaluating Retrieval Performance Using Click through Data

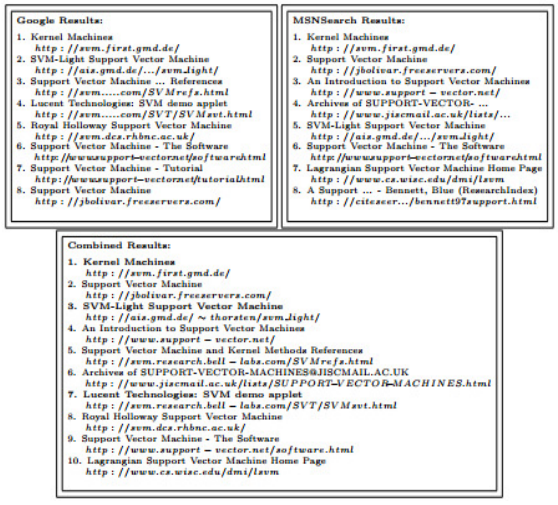


Figure 7. Ex. query "support vector machine"

In this paper, a new method proposed for evaluating the quality of retrieval functions. By experts or explicit user feedback unlike traditional methods which relevance judgments require, on click through data is it depends. Since click through data can be collected at a very low cost and without any risk for the user this is a crucial advantage. The method gives similar results under mild assumptions as evaluation with traditional relevance judgments which shown in a theoretical analysis. Empirically, the assumptions are indeed justified and are verified and in a WWW retrieval study which concludes results leads by that new method.

2.7. Search Engines optimization Using Click-through Data

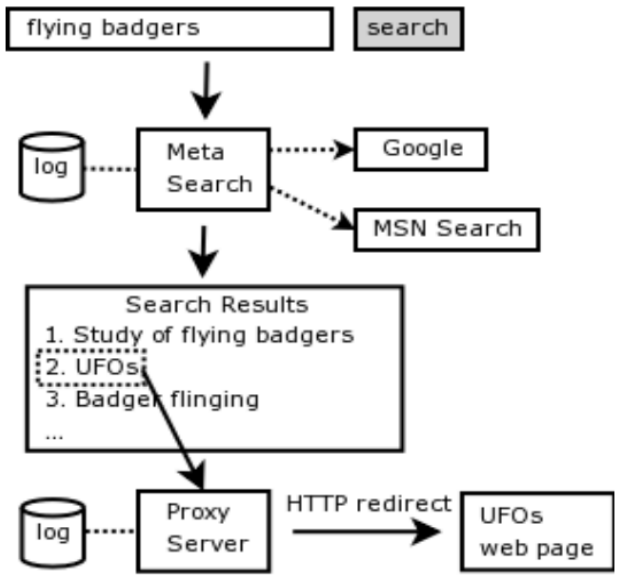


Figure 8. Procedure of a typical search

One approach presented in this paper for using click-through data for automatically boosting the retrieval quality of search engines. Naturally, by a well designed information retrieval system, with their following below the less significant documents in the ranking there should present relevant documents high. While from examples there exists previously tried out approaches to learn retrieval functions, from corresponding judgments. The training data generated by experts typically require it.

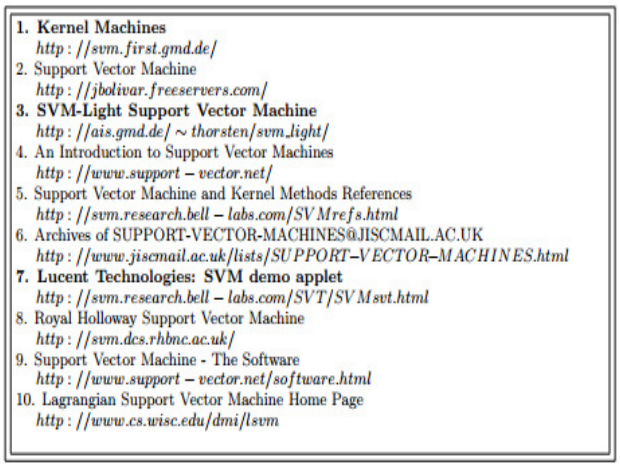
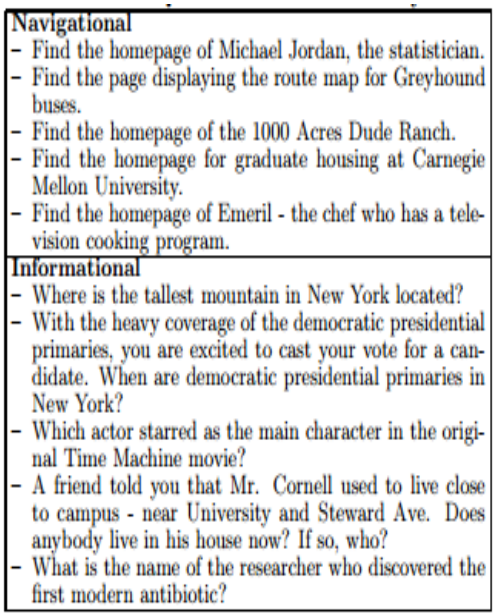

1. **Kernel Machines**
   http : //svm.first.gmd.de/
2. Support Vector Machine
   http : //jbolivar.freeservers.com/
3. **SVM-Light Support Vector Machine**
   http : //ais.gmd.de/ ~ thorsten/svm_light/
4. An Introduction to Support Vector Machines
   http : //www.support – vector.net/
5. Support Vector Machine and Kernel Methods References
   http : //svm.research.bell – labs.com/SVMrefs.html
6. Archives of SUPPORT-VECTOR-MACHINES@JISCMAIL.AC.UK
   http : //www.jiscmail.ac.uk/lists/SUPPORT–VECTOR–MACHINES.html
7. **Lucent Technologies: SVM demo applet**
   http : //svm.research.bell – labs.com/SVT/SVMsvt.html
8. Royal Holloway Support Vector Machine
   http : //svm.dcs.rhbnc.ac.uk/
9. Support Vector Machine - The Software
   http : //www.support – vector.net/software.html
10. Lagrangian Support Vector Machine Home Page
    http : //www.cs.wisc.edu/dmi/lsvm

Figure 9. Ranks for the query "support vector machine". The link in bold is the one user clicked on

The idea of this paper is to develop a method that utilizes click-through data for training. Users clicked on the ranking presented with the log of links i.e. the query-log of the corresponding search engine. At very low cost in abundance and can be recorded such click-through data is available. This paper presents a method taking an approach of Support Vector Machine (SVM) for learning retrieval functions. This method is shown to be well-founded from a theoretical perspective in a minimum risk framework. Furthermore, it proved itself feasible for a large set of queries, feature. The results accumulated in this theoretical were verified in a controlled manner in an experiment.

2.8. Accurately Interpreting Click-through Data as Implicit Feedback


Navigational
- Find the homepage of Michael Jordan, the statistician.
- Find the page displaying the route map for Greyhound buses.
- Find the homepage of the 1000 Acres Dude Ranch.
- Find the homepage for graduate housing at Carnegie Mellon University.
- Find the homepage of Emeril - the chef who has a television cooking program.

Informational
- Where is the tallest mountain in New York located?
- With the heavy coverage of the democratic presidential primaries, you are excited to cast your vote for a candidate. When are democratic presidential primaries in New York?
- Which actor starred as the main character in the original Time Machine movie?
- A friend told you that Mr. Cornell used to live close to campus - near University and Steward Ave. Does anybody live in his house now? If so, who?
- What is the name of the researcher who discovered the first modern antibiotic?

Figure 10. Queries used

In WWW search from click-through data 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 could well be informative but also present with a bias. This in turn makes the interpretation of clicks along with their relevance judgments a bit difficult,. It can be seen that from the from clicks are more or less accurate and on an average the relative preferences can be derived.

3. Conclusion

In this paper we have introduced closed loop sessions using search results or clicked URLs, to analyze to infer user search needs. As user implicit feedbacks the clicked URLs the ones which are unclicked are also are considered before the last click to make a closed loop feedback session. To show the specific need of an end user, systems have logged the sequence of most relevant search results. To design the closed loop feedback sessions the concept of pseudo documents was used. This particular concept will make the searching easy to user with 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.

[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 Click-through 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 Click-through 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 Click-through Data as Implicit Feedback," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '05), pp. 154-161, 2005.

[9] 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.