

Literature Review: Analyzing and Organizing User Search Queries

Chandana Nighut¹, Ashish Gurav², Anjaneya Naik³.

^{1,2,3}Department of Computer Engineering,
Atharva College of Engineering,
University of Mumbai, India

Abstract— A broad topic may have wide interpretations for a particular query, Search goals differ from user to user when they submit a particular query to the search engine .The understanding of users search goals is very useful in improving the user experience .

Primarily, user specific search goals are recognized by the clustering of the feedback sessions. User's click through logs help to build the feedback sessions which efficiently reflect the user's search goals. Secondly, an Incremental approach to create pseudo-documents for better representations of feedback session can be used for clustering. Experimental results will be shown by the user's click through logs from a commercial search engine to show the accurate search results entered by the user.

Keywords— User search goals, feedback sessions, pseudo-documents.

I. INTRODUCTION

User search history becomes more useful and reliable by organizing query search history and displaying the correct search results for particular search query using data mining concepts of association rule mining and combining it with click through data algorithm thus giving an incremental results which will be dynamic and user specific. User search goals are used for search results. Need for information is a user's particular desire to obtain accurate information to satisfy his/her need. User search goals can be considered as the clusters of information needed for query analysis and inference are important part of the user search history program and plays a major role while organizing data. A Web is a collection of interrelated files on one or more Web servers. Data mining is a mining technique which is used to extract knowledge from user search history. Web mining and Data mining are used for complete organization of data and is also used for better analysis of data for accurate results. Feedback session contains all the information for search results and then the data is clustered for reviewing the data and feedback clustering is performed. It is also used for analyzing the user search history and give organic results to users.

II. LITERATURE REVIEW

[1] Summarizes a new novel approach to solve users search goals by analysing user search logs. It solves this problem by clustering proposed feedback sessions. It uses both clicked and unclicked urls for constructing it as a feedback sessions. Last clicked urls are considered for

feedback sessions, which are then mapped to pseudo documents to predict the search goals. User search goals are different for different users. It also uses "Classified Average Precision" (CAP) for performance evaluation. For Example A user may be interested in searching about Mango and his goal would be to know more about it as a fruit or else he may be interested to gather information about the mango brand which is also a fashion brand. So a search query mango has ambiguous search results. It is very important to display accurate search results which matches user's search goals. A search query result which doesn't matches according to user's search goals is time consuming for users.

The clustering problem as a salient phrase ranking problem.[2] proposes a method such that it predicts search results when one submits their user search query. It is based on clustering process where similar queries are grouped together and are displayed to users having a similar search goal. Existing search engines such as Google, often return a long list of search results, ranked by their relevance. Search engine users have to go through the list and examine to identify their required results.

[3] proposes a technique for collection of similar queries and websites. Here each cluster contains both user query search and also its links associated with the queries .It does not directly uses the actual content for queries but it uses content which are correlated to queries from click through data. These data are then used to assist users for search results. The performance of a search engine is affected by multiple factors. Accuracy is very important factor that can affect the utility of a search engine significantly.

[4] suggests a query searching technique where users are suggested with queries based on mining query patterns. Query suggestion are used by mapping query by users to the desired concepts. These query suggestions are then suggested to users. A good retrieval system should be the first priority rather than the system which gives us with inaccurate results.

It is better to have good search results which satisfies user search goals rather than have inaccurate search results for the query .In this paper we proposed a method that utilizes click-through data for training, namely the query-log of the search engine.

Most of the search engines give results using a flat rank list which is good for unambiguous queries but queries

which have multiple outcomes cannot use flat rank list. [5] proposes a method where the search results are organized using users feedback, on this feedback text processing is done to create a pseudo document and on these clustering is performed using fuzzy k means clustering also preference is given to most visited links that occur at the top. A user interface is developed that organizes the search results into hierarchical structure here a Support Vector Machine classifier is built using web pages which are manually classified. This model is used to classify new web pages on-the-fly. The advantage of this approach is to assist user in focusing on the task relevant information. This system allows user to browse and choose the categories and to view the categories according to their own search relevance.

Conventional methods make use of co-occurring keywords which are based on the rank list[5]. These approaches face difficulties in extracting the terms that are conceptually related to the query but do not occur frequently. [6] presents a log-based approach to extract the terms, here search engine's logs are considered to suggest the relevant terms of the search query that co-appears in the query logs, unlike the conventional methods the suggested terms are re-organized incrementally in this system.

Query recommendation through pseudo-document feedback is costly and leads to different results. [7] gives a scale for evaluating query recommendation, they have built a model for recommending query by taking into consideration the query-user pair and by fitting it to the model of user judgment which improves the relevance of the query generated even further.

[8] Checks if the feedback generated from the clickthrough data is reliable or not, analyzes the user's decisions through eye tracking which is used to compare feedback through user's judgments, while the clickthrough data is sometimes biased hence they take the average relevance of the clicked. The results from [8] indicate that the users clicking decisions are based on correctness of the search results but the decision is also based on the trust they have on the search engine. In order to remove the confusion improve quality of results machine learning method is used for pair wise preferences to properly interpret the implicit feedback.

[9] gives a method to automatically improve the quality of retrieving of the search engines using clickthrough data. As the training data can be obtained from clickthrough data in the form of relative preference, [9] derives an algorithm for learning a ranking function. In this paper a Support Vector approach, is used and the resulting training problem is tractable even for a large number of queries. Furthermore, there are many situations where the goal of learning is ranking based which shows that algorithm is not limited to meta-search engine this is the working of most recommender problems.

The automatic identification of the users goal for a query. [10] Conducts study on a human subject showing that 60% of the queries have predictable goals irrespective of the user and for the next 40% of the queries search engines employ simple techniques to handle them

separately. The techniques which are used in this paper is the combination of user-click behaviour and anchor-link distribution.

III. METHODOLOGY

Working starts with registration of the user, login id and password is given to the user post registration. Every user has a unique login id to save user specific preferences for better accuracy.

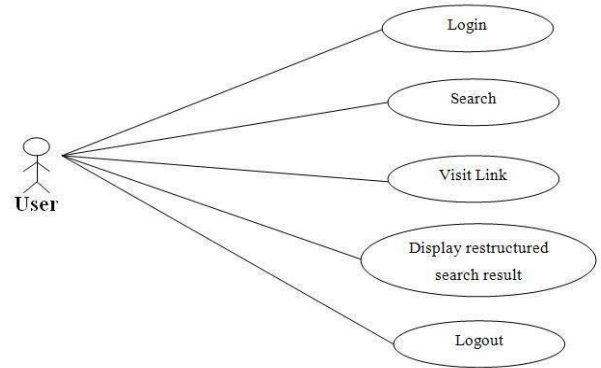


Figure 1: Use Case Diagram

The methodology is as follows,

1] When the user submits a query to the search engine all the feedback sessions of a query are first extracted from user click-through logs and mapped to pseudo-documents.

Later we try to understand the user search goals by clustering the aforementioned pseudo-documents and are depicted with some keywords. Since the exact number of user search goals are not known in advance, the second part will help us to determine the optimal value by trying several different values.

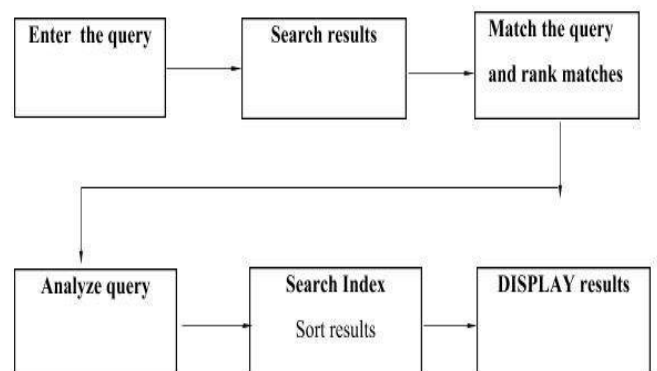


Figure 2: Control Flow Diagram

2] The original search results are restructured based on the user search goals inferred from the first part. Then, evaluation of performance is done through restructuring search results by Incremental Algorithm. The evaluation result will be used as the feedback to select the optimal number of user search goals in the first part.

IV. DISCUSSION

[1] Summarizes a new novel approach to solve users search goals by analysing user search logs. It solves this problem by clustering the feedback sessions. It uses both clicked and unclicked urls for constructing it, as a feedback session. Last clicked urls are considered for feedback sessions, which are then mapped to pseudo documents to predict the search goals. [2] The similar queries are grouped together and are displayed to users having a similar search goal. To improve the accuracy even further a technique for collection of similar queries and websites [3] is used. Here each cluster contains both user query search and also its links, associated with the queries.

Then a query searching technique [4] is used where users are suggested with queries based on data mining query patterns.

Then the search results are organized using users feedback. [5] text processing is done to create a pseudo document and on these clustering is performed using fuzzy k clustering and preferences is given to most visited links that users click on. These approaches face difficulties in extracting the terms that are conceptually related to the query but do not occur frequently. Hence a log-based approach [6] to extract the terms, It focuses on search engine's logs to suggest the relevant terms of the search query that co-appears in the query logs. Based on these method the search results are shown to the user.

V. CONCLUSION

The major contributions as follows: is to understand different users search goals for a particular query when they submit at a search engine. By clustering the feedback sessions, the system states that clustering of the feedback sessions is more useful than clustering of clicked urls. Further, the distributions of different users search goals for a particular query can be obtained after feedback sessions are clustered.

VI. FUTURE SCOPE.

Clustering can be done while considering different parameters like feedback sessions, Clicked URL's, But a pseudo document can be created by combining the enriched URLs of click through data and of the feedback sessions. It can successfully reflect the need of a user when a particular query is submitted. Thus, it can tell what the user search goals are in detail.

REFERENCES

- [1] Zheng Lu, Student Member, IEEE, Hongyuan Zha, Xiaokang Yang, Senior Member, IEEE, Weiyao Lin, Member, IEEE, and Zhaohui Zheng – "A New Algorithm for Inferring User Search Goals with Feedback Sessions"- IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 3, MARCH 2013.
- [2] Baeza-Yates. R, 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.
- [3] Beeferman. D 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.
- [4] Cao. H, 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.
- [5] Chen. H 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.
- [6] Huang C.K, 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.
- [7] Jones. R, B. Rey, O. Madani, and W. Greiner, "Generating Query Substitutions," Proc. 15th Int'l Conf. World Wide Web (WWW '06), pp. 387-396, 2006.
- [8] Joachims. T, 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.
- [9] T. Joachims. Optimizing search engines using clickthrough data. In Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD), 2002.
- [10] Lee. U, Z. Liu, and J. Cho, "Automatic Identification of User Goals in Web Search," Proc. 14th Int'l Conf. World Wide Web (WWW '05), pp. 391-400, 2005.