International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 06 | June-2016 www.irjet.net p-ISSN: 2395-0072

Analysis of User Search Goals For Search Results

Ujwala Ninawe[1], Prof.M.R.Wanjari[2]

[1] *M.Tech Student*
CSE Department, Shri Ramdeobaba College Of Engineering & Management,Nagpur,India

[2] *Assistant Professor, Department of computer science and engineering*
Shri Ramdeobaba College Of Engineering & Management,Nagpur,India

Abstract:

In web browsers, queries are submitted to search engines to satisfy the information needs of users. Sometimes it is very difficult to get the exact information for the queries because of many ambiguous queries may covered a broad topic and different users may be wanted to get information on different aspects when they submit the same query on search engine. The User's goal to get specific information at the time of searching the result is unknown to the web browser. Thus the ultimate goal of searching process is to first recognize the users' goal. Based on the user's URL click pattern on the search engines result page, Inferring user search Goal. Restructured goal results will be shown to the user. For the ambiguous query when it is submitted to the search engine, User click through log is maintained. Feedback session is created by taking feedback from multiple users based on their interest. Feedbacks are then map to generate pseudo-documents indicating goal text or keywords may come to users mind. Pseudo-documents depicts with goal texts for clustering which is then used to infer user search goals. The inference and analysis of user search goal helps in the improvement of search engine relevance and user knowledge.

Keywords: *User search goal, feedback sessions, pseudo-documents, clustering*

1. INTRODUCTION

Extraction of data from the data set and convert into more understandable format, concept of data mining is used. The data mining technique is used in many fields like in the study of human genetics, educational research, electrical power engineering etc. Web is a large, informative, dynamic, which supplies abundant information for the query. However, it might possible that an ambiguous query or topic fired by the user to the search engine application is unable to recognize user information needs. Users might have different information needs on diverse aspects on submission of same query or topic to search engine. Hence it is necessary and important to capture the user search goals as a cluster of information for the query. The evaluation and depiction of user search goals help to improve relevancy of the search engine and user required knowledge**.** So it is necessary to capture the different user goal and also retrieve the efficient information for a required query. Capturing different user's search goal related to Information requires the normal query based information retrieval. The user's behavior in a web site is not random; rather, it is driven by the user which satisfies user information goal. The user makes traverse decisions looking at information that would satisfy his or her need.

2. LITERATURE REVIEW

[1] A New Algorithm for Inferring User Search Goals with Feedback Sessions:
In this paper unsupervised clustering technique has been used. Author aims at discovering the number of diverse user search goals for a query and then depicting each goal with some vague keywords automatically. At first the phrases are extracted from the documents, which are then clustered as candidate cluster. And then final clusters are created by merging all these candidate clusters by using clustering algorithm

[2] Agglomerative Clustering of a Search Engine by Query Log:
Data has collected using a search engine by taking the user click log for query. A cluster of similar queries and URLs matching queries gets discovered. In every log, user submits a query to particular search engine and URL selected among the retrieved data by the user is collected from the search engine. The data set has beeen viewed by bipartite graph, having vertices on one side and URL's corresponding to the respective queries on the other side of graph. In this paper, agglomerative clustering algorithm is used to get the similar clusters for the particular query.

[3] Query Clustering Using User Logs:
Given the different forms of queries and user intentions to get the needed information, the similarity of query might not be accurately estimated by analyzing the contents (e.g. keywords). In this paper, user's logs (recording user document clicks) has taken as a supplement. A clustering method which has been proposed, suggests if two queries retrieved from the same or similar document clicks, then they might be similar. An analysis of the clustering result reveals that this clustering strategy can combine the similar queries together more effectively than using keywords alone.

[4] Learn from Web Search Logs to Organize Search Results:
Method used in this paper has been used to evaluate for result organization with the help of logs of a commercial search engine like Google, Yahoo. This method is then compared with ranking of default search engine and clustering of search results is performed. The advantage of this method is for the effectiveness of results in improving search utility and the labels which are generated using past query words are more readable and useful than those are generated using traditional clustering approaches.

[5] Relevant term suggestion in interactive web search based on contextual information in query session logs:
Conventional approaches used for making term suggestions that extract key terms which gets co-occur from the retrieved relevant documents. In this paper, evaluation effective to log-based approach for performing relevant term extraction and suggestions is considered. By using this approach, the relevant terms suggested queries are those that found in similar query sessions of search engine logs in response to users' original queries, rather than from retrieved documents. There might be Many relevant terms that relates with each other conceptually but do not frequently occur in the same retrieved document, therefore, can helps in accurately identifying and presents the information to the user.

[6] Personalized Concept Based Clustering of Search Engine Queries:
In this paper, they proposed a new personalized concept based clustering technique. This technique has been used to obtain personalized query suggestions for individual users based on their conceptual profiles. These techniques make use of click through data and the concept relationship graph mined from web-snippets, both of which can be captured at the back end and as such do not add extra burden to users. An adapted agglomerative clustering algorithm is used for finding queries that are conceptually close to one another.

3. PROPOSED APPROACH

Feedback sessions are considered as user implicit feedback and will be using an optimization method to combine both clicked and unclicked URLs in feedback sessions. It helps to find out what users really require and what they do not care. An approach is proposed to infer user search goals for a query by clustering its feedback sessions represented by pseudo-documents. Pseudo documents represent some goal text or vague keywords that might come in users mind. Then applying clustering Algorithms to cluster the different pseudo documents to uniquely identify clusters. Finally, restructure the web search results are displayed to the user.

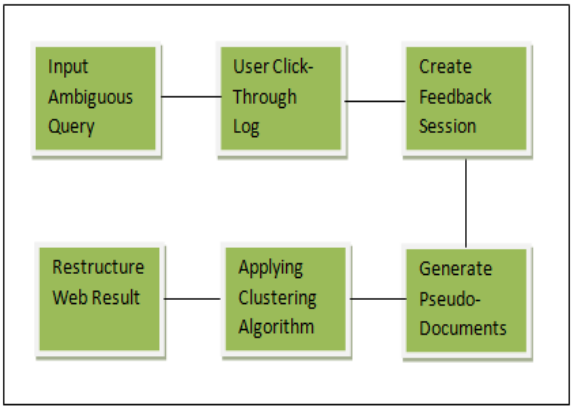

Fig1: General Approach

3.1. User Click-Through Log:

When the user enters the query to the search engine, the query is recorded as a log and then the results will get produced based on the entered keywords.

3.2. Feedback Sessions:

The feedback sessions is defined as "The series of both clicked and unclicked URLs and ends with the last URL that was clicked in a session from user click-through logs" [1]. It is required that before the last click, all the URLs should have been scanned and then evaluated by users. In addition to the clicked URLs, the unclicked ones before the last click should be considered as user feedbacks.

3.3 Pseudo Documents:

Mapping of the feedback sessions to pseudo-documents helps in depicting user information needs effectively. Feedback sessions are used to form a pseudo document. The feedback sessions may vary for different clicks through logs and queries submitted by the user. It is not suitable to use the feedback sessions directly. Instead some method id is needed to represent the feedbacks to make pseudo document in a more efficient way. The search log will be represented in the form of 0's and 1's. To identify the feedback sessions in binary vector format, 1 is used as clicked and 0 as unclicked. But this representation is not appropriate to give the proper contents of user search goal. We introduce pseudo-documents to approximate goal texts. Thus pseudo-documents can be used to infer user search goals. Following are the two main Steps to build pseudo documents:

a. Represent the URL in the feedback session: Titles and snippets for each URL's that appears in feedback session are first extracted. Each URL is represented as a small text paragraph. Some textual process is implemented to text paragraphs such as converting all the letters to lower case, performing stemming and removing stop words. Finally, Term Frequency-Inverse Document Frequency (TF-IDF) vector is used each URLs title and snippets.

b. Forming pseudo documents based on URL representations: This step obtains the feature

representation of a feedback session. It combines both clicked and unclicked URL's in the feedback sessions. Feature representation of a feedback session indicates the importance of terms in feedback session. It reflects and identifies what users actually want and what they do not care about. It is used to approximate the goal texts in users' mind.

3.3. Clustering Pseudo Documents:

The Pseudo documents are clustered using clustering algorithms. Since exact numbers of clusters or groups are unknown for each query, we set k clusters. The terms with the highest values in the center point are used as the keywords in order to depict user search goals. The clustering is the process based on TF-IDF vector representation of queries, obtained from the aggregation of the term-weight vectors of the clicked URLs for the query.

3.4. Final Restructured Results:

The results are restructured based on the evaluation of web search goals. Search engine will return millions of search results to the end user. So it is necessary to organize them to make it easier for users to find what they require. The user search goals are represented as the vectors. So we perform categorization by choosing the smallest distance between the URL vector and user search goal vectors. The results can be restructured according to the inferred user search goals.

4. RESULT

For the ambiguous query, the system gives organized result to the user depending upon the clusters. This helps user to achieve desired goal in order to save search time. Following table shows the ambiguous query depicting user search goals with some keywords.

Query	Keywords to depict user goal
Cricket	Cricket, book, insect
	Mobile, game, live, player, team
Sun	Tv, Newspaper, daily, sun
	Star, Planet, Solar, contain
Apple	Tree, fruit
	Store, ipad, iphone, job, online

Table 1: Abstracted Keywords Used to Depict User Search Goals for Some Ambiguous Queries

5. CONCLUSION

User search goal is defined as the cluster of information on different aspects of a query that user groups want to obtain in order to obtain the useful information [1]. Information need represents a user's particular desire to obtain information that would satisfy his/her need. Advantages of the inference and analysis of user search goals are improving search engine relevance and user experience. This approach helps in discovering user search goals for some popular ambiguous queries. When users submit one of the queries, the search engine will return the results which are then categorized into different groups according to user search goals. Thus, users can find what they require conveniently in order to achieve the desired goal related to the query.

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, 2013.
2. 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.

3. J.-R Wen, J.-Y Nie, and H.-J Zhang, "Clustering User Queries of a Search Engine," Proc. Tenth Int'l Conf. World Wide Web (WWW '01), pp. 162-168, 2001.
4. 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.
5. 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.
6. Kenneth Wai-Ting Leung, Wilfred Ng, and Dik Lun Lee, "Personalized Concept-Based Clustering of Search Engine Queries," IEEE transactions on knowledge and data engineering, vol. 20, no. 11, november 2008.
7. Tai-Wei Chen, Ping-Lin Chang and Wei-Guang Teng. "Supporting Informational Web Search with Interactive Explorations," IEEE International Conference on Signal Image Technology and Internet Based Systems 2008.
8. Huanhuan Cao, Daxin Jiang, Jian Pei, Qi He, Zhen Liao, Enhong Chen, Hang Li, "Context-Aware Query Suggestion by Mining Click-Through and Session Data ," *KDD'08,* August 24–27, 2008.
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.