



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

Review on Analyzing User Search Goals by Clustering Pseudo-Document

Ujwala Ninawe¹, Prof.M.R.Wanjari²

M.Tech Student, Department of CSE, Shri Ramdeobaba College of Engineering & Management, Nagpur, India¹

Assistant Professor, Department of CSE, Shri Ramdeobaba College of Engineering & Management, Nagpur, India²

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 since many ambiguous queries may get covered by 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 goal of the user to get any information at the time of searching the result is unknown to the browser. The ultimate aim 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. Recognized goal results are shown to the user. For the query, click through log is maintained. Using click through log feedback session is created. By which pseudo-documents are generated. Pseudo-document depicts with goal texts for clustering which is then used to infer user search goals. Different properties and different structures that various clustering algorithms supports will get observed and the time required for each one will be compared. The inference and analysis of user search goal helps in improving search engine relevance and user knowledge.

KEYWORDS: User search goal; feedback sessions; pseudo-documents; clustering

I. INTRODUCTION

Web is a large, informative, which supplies abundant information. Nowadays Internet is widely used by users to satisfy cluster of different information needs. However, it may be possible that an ambiguous query or topic submitted by the user to the search engine application is unable to satisfy user information needs, because different users might have different information needs on diverse aspects upon 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 can be helpful in improving relevancy of the search engine and user required knowledge. So it becomes necessary to collect 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 typical user behavior in a web site is not random; rather, it is driven by the user information goal. That is, the user makes traversal decisions looking for information that would satisfy his or her need. The key challenge is to understand the relationship between content-organization and user information need on one hand with the user behavior on the other.

There are many advantages for the evaluation and analysis of user search goals. Some of them are as follows:

- Web search results can be reorganized according to user search goals by combining very similar search results with same information need. This can be useful to other users who have different search goals to find easily what they actually require.
- Reranking web search results to satisfy users' goal.

II. LITERATURE REVIEW

[1] A New Algorithm for Inferring User Search Goals with Feedback Sessions:

In this paper unsupervised clustering technique has been used. At first the phrases are extracted from the documents, which are then clustered called as candidate cluster. And then final clusters are created by merging all these candidate clusters. At last the clustering algorithm is applied to get the clustering of candidate cluster.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

[2] Agglomerative Clustering of a Search Engine by Query Log:

This paper has proposed a method for data mining. Data has been collected using a search engine by taking the user log. A cluster of same type of queries and URLs similar to particular queries gets discovered. In every log, user fires 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 is viewed by bipartite graph, having vertices on one side and URL's corresponding to the respective queries are 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 queries cannot be accurately estimated by analyzing the contents (e.g. keywords). In this paper, user logs (recording user document clicks) are taken as a supplement. A clustering method is proposed that suggests if two queries retrieved from the same or similar document clicks, they are similar. An analysis of the clustering results concludes 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 is evaluated for result organization with the help of logs of a commercial search engine like Google, Yahoo. This method is 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 are used for making term suggestions extracts key terms which gets co-occur from the retrieved relevant documents. This paper evaluates effective log-based approach for performing relevant term extraction and term suggestion. 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 may Many relevant terms that are conceptually related but do not frequently occur in the same retrieved document can, therefore, be accurately identified 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 is 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.

[7] Supporting Informational Web Search with Interactive Explorations

The effectiveness of search engines may impact on Internet users. Nevertheless, depending on different user goals in web search, it is noticed that users occasionally get lost in the large number of search results when trying to explore them. In this paper, Author has proposed to support users to conduct such informational web searches of explorations. Proposed approach is feasible in improving usability of current search engines.

[8] Context-Aware Query Suggestion by Mining Click-Through and Session Data

Query suggestion plays an important role in improving the usability of search engines. Proposed methods can make meaningful query suggestions by mining query patterns from search logs, none of them are context-aware queries. In this paper, they proposed context-aware query suggestion approach in two steps. In the offline model learning step, to address data sparseness, queries are summarized into concepts by clustering a click-through bipartite. Then, from session data a concept sequence suffix tree is constructed as the query suggestion model. In the online query suggestion step, a user's search context is captured by mapping the query sequence submitted by the user to a sequence of concepts.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

III. CONCLUSION

User search goal is the information on different aspects of a query that user groups want to obtain [1]. Information need is user's particular desire to obtain information to satisfy his/her need. User search goals are used to consider as the clusters of information needs for a query. The inference and analysis of user search goals have a lot of advantages in improving search engine relevance and user experience. This approach can discover user search goals for some popular 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 online. Thus, users can find what they require conveniently.

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.