

ISSN 2395-1621

# An Efficient Approach for Inferring User Search Goal with Feedback Sessions

Sumit Kharche<sup>#1</sup>, Vivek Barkul<sup>#2</sup>, Ashwini Shinde<sup>#3</sup>, Rupali Bhore<sup>#4</sup><sup>1</sup>sumit.kharche@gmail.com<sup>2</sup>barkulvivek5@gmail.com<sup>3</sup>ashwinishinde701@gmail.com<sup>4</sup>rup.211962@gmail.com<sup>#1234</sup> UG Students, Department of Computer Engineering, BSCOER, Narhe Pune.

## ABSTRACT

Information needs of users are represented by queries submitted to search engines and different users have different search goals for a broad topic. Sometimes queries may not exactly represent the user's information needs due to the use of short queries with ambiguous terms. Hence to get the best results it is necessary to capture different user search goals. These user goals may contain information on different aspects of a query that different users want to obtain. Here, feedback sessions are used to discover different user search goals based on series of both clicked and unclicked URL's. The pseudo-documents are generated best representation of the feedback sessions which can reflect need of user objective. Clustering the pseudo documents and each cluster can be considered as one search goal and depict them with some keywords or goal text. Evaluation of clustering is an important problem. So, CAP can be used to evaluate the performance of user search goals inference and restructuring results.

**Keywords—** Search engines, Feedback sessions, pseudo document, CAP (Classified Average Precision).

## ARTICLE INFO

### Article History

Received : 3<sup>rd</sup> April, 2015Received in revised form :  
5<sup>th</sup> April, 2015Accepted : 10<sup>th</sup> April, 2015

### Published online :

11<sup>th</sup> April 2015

## I. INTRODUCTION

In the recent years searching information from the web within time has more importance. The different users have different search objectives when they submit query to the search engine. Sometimes queries may not exactly represent the actual objective of user. As there are so many ambiguous queries and different users may want to get information on different aspects when they submit the same query. In fig 1 it shows that user has entered query "the dawn" which has ambiguous meaning. One can want information of Spacecraft and another may want to locate home page of Pakistan English newspaper. User search objective is the information on different aspects of a query that user groups want to obtain. There are mainly two aspects for improving search engine relevance 1) inference of user search objective and 2) analysis of user search objective

Home - [DAWN.COM](http://www.dawn.com/) - Latest News, Breaking News, Pakistan News ...

<https://www.dawn.com/>

Dawn News - Rs320bn tax relief for .... RSShttp://beta.dawn.com .... The DAWN MEDIA GROUP is not responsible for the content of external sites. Tupermic ...

### E-Paper

DAWN ePaper update time: Pakistan  
1100 PST (0600 GMT ...

### Pakistan

Sindh - KP & FATA - Punjab -  
Balochistan - ...

### Sport

@ICC: RT @Holly\_Ferling: Another  
great win for the ...

### Newspaper

Govt to oppose handover of BB  
jewellery. Zardari's claim to ...

### Dawn

Dawn is the time that marks the beginning of the twilight before sunrise.  
It is recognized by the ...

### Dawn

### Spacecraft

Dawn is a space probe launched by NASA on September  
27, 2007, to study the two most massive ...



Fig.1. Example of different user objectives for” the dawn” query

First, we can restructure web search results obtained according to user search objectives. Restructuring can be done by group in the search results with the same search objective. By using this, we can easily find out users with different search objectives. And also we can find what exactly user want. Second. We are representing user search objectives by some keywords. These keywords can be utilized in query recommendation. By using recommendation, we find out the suggestions about queries which can help users to form their queries more detail. Third, the search objectives are distributed for the use of reranking of documents

There are different kinds of technique for improving the search results and finding the exact users need. We first propose a approach to identify user search objective for a query by clustering our proposed feedback sessions. The feedback session is the series of both clicked and unclicked URLs. For keeping the feedback we have use Click through logs. In the existing feedback system it contains limited number of URLs.

So the number of URLs is increased in the Click through log. For better understanding of user search objective, we propose an optimization method. In this method, feedback sessions are mapped to pseudo-documents. By this mapping process one can efficiently reflect user information needs. Likewise the pseudo documents are clustered using the clustering algorithm. The user search goals are obtained according to the feedback sessions. The restructure result is produced for the user query based on the user search goal. The CAP evaluation can be done for each user search goal and the clustering can be done to find the optimal number of users.

## II. LITERATURE SURVEY

### Pseudo Documents:

By using recommendation, one can find out the suggestions about queries which can help users to form their queries more detail [5] [12] [13].They are representing user search objectives by some keywords. These keywords can be utilized in query recommendation [5] [12] [13].

### Clicked and unclicked URL's:

It consider user goals as “Navigational” and “Informational” .Then they categorize queries into these two classes[9].In this paper define query intents as “Product intent ”and “Job intent” .Then they classify queries into these two defined intents. But finding exact class is very difficult and impractical [4]. So to overcome this problem one has to learn aspects of queries by analyzing the clicked URLs which are directly taken from user click-through logs for the organization of search results [1].

### Clustering:

It shows that clustering of search results is a best way to organize search results [1]. They proposed learning

interesting aspects of a topic from web search logs and generation of cluster labels which has specific meaning [1].

### Restructuring Web Search Results:

Finally the evaluation of clustering is also an important problem. To solve this problem we propose a evaluation criterion classified average precision (CAP).CAP is used to evaluate the performance of the restructured web search result [13].

## III. LIMITATIONS OF EXISTING SYSTEM

1. A what users care about varies a lot for different queries, finding suitable predefined search goal classes is very difficult and impractical.
2. Analyzing the clicked URLs directly from user click-through logs to organize search results. However, this method has limitations since the number of different clicked URLs of a query may be small. Since user feedback is not considered, many noisy search results that are not clicked by any users may be analyzed as well. Therefore, this kind of methods cannot infer user search goals precisely.
3. Only identifies whether a pair of queries belongs to the same goal or mission and does not care what the goal is in detail.

## IV. PROPOSED SYSTEM

We aim at discovering the number of diverse user search goals for a query and depicting each goal with some keywords automatically. We first propose a novel approach to infer user search goals for a query by clustering our proposed feedback sessions. Then, we propose a novel optimization method to map feedback sessions to pseudo-documents which can efficiently reflect user information needs.

### Proposed Framework:

Fig. 1 shows the framework of our approach. The proposed system framework is an enhancement to techniques introduced in [1].The main motive of proposed system is to identify search objective and return search results within few time. In feedback session it keeps more number of URLs as compared with previous method. New framework makes use Feedback session and Hierarchical clustering.

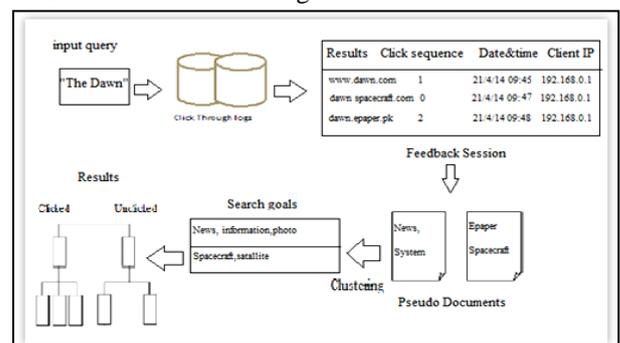


Fig. 2.Proposed System Model (Framework)

### Detail of System Module:

Proposed framework consists of different parts as follows part 1: In the first step queries are submitted to the search engine and all these queries are stored into Click through logs. Different user's has different aspects for searching information. Feedback sessions of a query are first extracted from user click-through logs. In the feedback session different types of information is stored like "Click sequence, Date/Time, IP address".

The feedback session is useful for inferring user search objective. In the proposed system we can consider number of feedback session for analysis. These feedback sessions are mapped to the pseudo-documents. Pseudo-documents consist of keywords which represent the user's information need. By using these pseudo-documents we can easily find out the user search objectives inferred by clustering. Initially we do not know the exact number of user search objective so that several different values are tried. From these values optimal value will be determined by the feedback from the part 2.

Part 2: In this the clustering of the feedback session takes place. One can cluster the retrieved results but in the proposed system feedback sessions are clustered. Clustering of the feedback session is more efficient than the clustering of the retrieved results. Finally Then, we evaluate the performance of retrieved results by our proposed evaluation criterion CAP. And the evaluation result will be used as the feedback to select the optimal number of user search goals in the part 1. In this paper focus is on the feedback sessions, pseudo documents. For identification of user search objective following procedure we have to follow:

#### Feedback Session:

When user submits the query to the search engine we do not know about exactly need of user. So for the identification of user search objective we are maintaining feedback session. Feedback session is the series of queries and some clicked search results. In the previous method it keeps the record of limited URLs. But in the proposed method number of URLs is increased. By using feedback session we can easily identify user search objective. In this we are considering feedback session for only single query. Therefore for the single query feedback session is also single.

The proposed framework consists of feedback session with both clicked and unclicked URLs and they ends with the last URL which was clicked in single session. Before the last click all previous URLs have been scanned and they are evaluated by users. Feedback session consists of 0 which indicates that corresponding URL is unclicked. In feedback session only three URLs are shown. In that one URL is clicked and two URL's are unclicked. Clicked URL indicates what exactly user need. Unclicked URL shows that user doesn't want the information related to this URL. Therefore, for identification of user search objective it is best way to analyze the feedback session then the analyzed search results or clicked URLs.

#### Mapping of Feedback Session to Pseudo document:

There different methods for representation of feedback session. One of the methods is known as "Binary Vector

Method" to represent feedback session. When the query "the sun" submitted to the search engine 0 represents unclicked in the click sequence. For example, binary vector [0110001] can be used to represent feedback session [14]. In that 1 represents "clicked" and 0 represents "unclicked". The binary vector method has disadvantage that doesn't give enough information to identify user search objective.

New method is proposed for representation of feedback session. In this method the feedback session is mapped to pseudo documents. Pseudo documents consist of keyword to determine whether document can satisfy their need. Hence pseudo documents can be used to identify user search Objective.

#### Clustering Pseudo Document:

After mapping the feedback sessions to pseudo documents next task is to form the group of clicked and unclicked URLs. So we have to cluster these URLs into clicked and unclicked clusters. For clustering we use the K-means clustering algorithm which is very easy to implement and use. For clustering k-means algorithm can be used. It is one of the partitioning methods. Clustering refers to grouping of similar objects. In K-means, centroid can be calculated for each cluster. The advantage of k-means algorithm is a large number of variables can be easily computationally faster than compared with other techniques and it produce tighter cluster, especially if the clusters are globular. These are technique implemented in Pseudo-document. The algorithm is as follows:

#### K-means Clustering Algorithm:

1. Randomly choose K data items from X as a initial centroids.
2. Repeat
  - a) Assign each data point to the cluster which has the closest centroid.
  - b) Calculate new cluster centroid.
 Until the convergence criteria is mate.

#### Restructuring Results:

Feedback information is needed to determine the best cluster. Using the pseudo document the keywords or goal text can be generated. Finally, the original search results are reconstructed and the performance can be evaluated using classified average precision. The categorization can be performed by choosing the smallest distance between the user search goals vector and URL vectors. Goals of this method are to evaluate whether User Search Goals are inferred properly or not. CAP (Classified Average Precision) is used to evaluate the performance of restructured web search result. Performance is calculated by

$$CAP = \frac{\text{No. of Relevant URLs}}{\text{No. of Retrived URLs}}$$

#### Algorithm strategy for Proposed Framework:

1. Generate pseudo documents of URL from feedback session.

2. Pseudo document contains all the keywords from given web page's title and description.
3. Also all the stop words and stemming words are removed.
4. Apply K-means clustering algorithm to form a group of relevant keywords from pseudo documents so that each cluster represents one user search objective. Next step is to organize words from one cluster into topic-subtopic hierarchy by using clustering.

**V. RESULT ANALYSIS**

System relies on the feedback of user. Feedback is then converted into pseudo-documents which represents the keywords from the documents. After that the pseudo documents are clustered using the k-means clustering algorithm. Results are evaluated using CAP.

Query	Keywords used to depict user search goals
Apple	Apple Wikipedia
	News
	Official
Sun	Solar Sun
	Sun Wikipedia
	Official
Dawn	News Paper
	Spacecraft
	Dawn Wikipedia
Hindu	Religion
	News
	Hindu Wikipedia

Table 1.1 Keyword depictions of different queries

**VI. CONCLUSION**

Proposed approach is used to infer user search goals by clustering the feedback sessions. Feedback sessions consist of both clicked and unclicked URL's before the last click is considered as user's implicit feedback. Then feedback sessions are mapped to pseudo-documents to approximate goal texts in users mind. These documents enrich URL's with additional contents including titles and snippets. Based on these documents search goals can be depicted with some keyword. Finally, to evaluate the performance of the user search goals CAP is used. By using this method users can find what they want easily.

**REFERENCES**

[1] 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.

[2] B. Poblete and B.-Y Ricardo, "Query-Sets: Using Implicit Feedback and Query Patterns to Organize Web Documents," Proc. 17th Int'l Conf. World Wide Web (WWW '08),pp. 41-50, 2008.

[3] X. Li, Y.-Y Wang, and A. Acero, "Learning Query Intent from Regularized Click Graphs,"Proc. 31st Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '08),pp. 339-346, 2008.

[4] 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.

[5] U. Lee, 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.

[6] H.-J Zeng, Q.-C He, Z. Chen, W.-Y Ma, and J. Ma, "Learning to Cluster Web Search Results," Proc. 27th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '04), pp. 210-217, 2004.

[6] R. Baeza-Yates, C. Hurtdo, 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.

[7] Zheng Lu, Student Member , IEEE , HongyuanZha, Xiaokang Yang, Senior Member, IEEE ,WeiyaoLin,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".

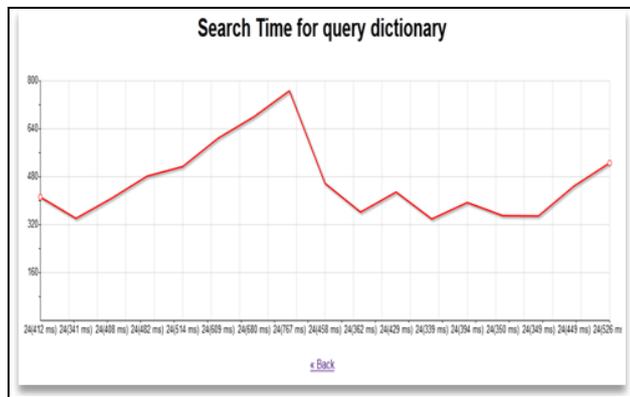


Fig. 3.Graph showing Retrieval time for query

A graph is plotted as number of results versus retrieval time (milliseconds) .The graph shows retrieval time taken for query "dictionary".

The results obtained for user input query can be shown as a bar graph. The clicked count and unclicked count are plotted as sessions versus count for clicked and unclicked count.