

## A New Algorithm for Inferring User Search Goals with Feedback Sessions

Bhavesh Pandya\*, Charmi Chaniyara\*\*, Darshan Sanghavi\*\*\*, Krutarth Majithia\*\*\*\*

\*(Assistant professor, Department of Information Technology, St. Francis Institute of Technology, Mumbai University, Maharashtra, India)

\*\* (Assistant professor, Mumbai University, Maharashtra, India)

\*\*\* (Student, Department of Information Technology, St. Francis Institute of Technology, Mumbai University, Maharashtra, India)

\*\*\*\* (Student, Department of Information Technology, St. Francis Institute of Technology, Mumbai University, Maharashtra, India)

### ABSTRACT

When different users may have different search goals when they submit it to a search engine. The inference and analysis of user search goals can be very useful in improving search engine relevance and user experience. The Novel approach to infer user search goals by analyzing search engine query logs. Once the User entered the query, the Resultant URLs will be filtered and the Pseudo-Documents are generated. Once the Pseudo documents are generated the Server will apply the Clustering Mechanism to URL's. So that the URLs are listed as different categories. Feedback sessions are constructed from user click-through logs and can efficiently reflect the information needs of user. Second, we propose a novel approach to generate pseudo documents to better represents the feedback sessions for clustering. Finally we proposed new criterion "Classified Average Precision (CAP)" to evaluate the performance of inferring user search goals. Experimental results are presented using user click-through logs from a commercial search engine to validate the effectiveness of our proposed methods. Third, the distributions of user search goals can also be useful in applications such as re ranking web search results that contain different user search goals.

**Keywords-** Classified Average Precision, Data Contents, Presentation Styles, Average Precision, Integrated Interface Schema, Uniform Resource Locator, Mining Spanning Tree.

### I. INTRODUCTION

• For a broad-topic and ambiguous query, different users may have different search goals when they submit it to a search engine. The inference and analysis of user search goals can be very useful in improving search engine relevance and user experience. In this paper, we propose a novel approach to infer user search goals by analyzing search engine query logs. First, we propose a framework to discover different user search goals for a query by clustering the proposed feedback sessions. Feedback sessions are constructed from user click-through logs and can efficiently reflect the information needs of users. Second, we propose a novel approach to generate pseudo-documents to better represent the feedback sessions for clustering. Finally, we propose a new criterion "Classified Average Precision (CAP)" to evaluate the performance of inferring user search goals. Experimental results are presented using user click-through logs from a commercial search engine to validate the effectiveness of our proposed methods.

• Our framework consists of two parts divided by the dashed line. In the upper part, all the feedback sessions of a query are first extracted from user click-through logs and mapped to pseudo-documents. Then, user search goals are inferred by clustering these pseudo-documents and depicted with some keywords. Since we do not know the exact number of user search goals in advance, several different values are tried and the optimal value will be determined by the feedback from the bottom part. In the bottom part, the original search results are restructured based on the user search goals inferred from the upper part. Then, we evaluate the performance of restructuring search 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 upper part..

[1] This paper highlights " A New Algorithm for Inferring User Search Goals with Feedback Sessions ".For a broad-topic and ambiguous query, different users may have different search goals when they submit it to a search engine. The inference and

analysis of user search goals can be very useful in improving search engine relevance and user experience. In this paper, we propose a novel approach to infer user search goals by analyzing search engine query logs. First, we propose a framework to discover different user search goals for a query by clustering the proposed feedback sessions. Feedback sessions are constructed from user click-through logs and can efficiently reflect the information needs of users. Second, we propose a novel approach to generate pseudo-documents to better represent the feedback sessions for clustering. Finally, we propose a new criterion "Classified Average Precision (CAP)" to evaluate the performance of inferring user search goals. Experimental results are presented using user click-through logs from a commercial search engine to validate the effectiveness of our proposed methods.

[2] This paper highlights Using Feedback Sessions for Inferring User Search Goals. Identifying or inferring user's search goal from given query is a difficult job as search engines allow users to specify queries simply as a list of keywords which may refer to broad topics, to technical terminology, or even to proper nouns that can be used to guide the search process to the relevant collection of documents. 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.

[3] This paper "A New Algorithm for Inferring User Search Goals with Feedback Sessions". When different users may have different search goals when they submit it to a search engine. The inference and analysis of user search goals can be very useful in improving search engine relevance and user experience. The Novel approach to infer user search goals by analyzing search engine query logs. Once the User entered the query, the Resultant URLs will be filtered and the Pseudo-Documents are generated. Once the Pseudo documents are generated the Server will apply the Clustering Mechanism to URL's. Feedback sessions are constructed from user click-through logs and can efficiently reflect the information needs of user. Second, we propose a novel approach to generate pseudo documents to better represents the feedback sessions for clustering. Finally we proposed new criterion "Classified Average Precision (CAP)" to evaluate the performance of inferring user search goals. Experimental results are presented using user click-through logs from a commercial search engine to validate the effectiveness of our proposed methods. Third, the distributions of user search goals can also

be useful in applications such as re ranking web search results that contain different user search goals.

## II. Other Existing System

We define user search goals as the information on different aspects of a query that user groups want to obtain. Information need is a user's particular desire to obtain information to satisfy his/her need. User search goals can be considered as the clusters of information needs for a query. The inference and analysis of user search goals can have a lot of advantages in improving search engine relevance and user experience.

### 2.1 Problems on existing system:

- What users care about varies a lot for different queries, finding suitable predefined search goal classes is very difficult and impractical.
- 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.
- Only identifies whether a pair of queries belongs to the same goal or mission and does not care what the goal is in detail.

## III. System Architecture

The implementation stage involves careful planning, investigation of the existing system and its constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

### 3.1 Interaction Model:

#### Client-driven interventions:

Client-driven interventions are the means to protect customers from unreliable services. For example, services that miss deadlines or do not respond at all for a longer time are replaced by other more reliable services in future discovery operations.

#### Provider-driven interventions

Provider-driven interventions are desired and initiated by the service owners to shield themselves from malicious clients. For instance, requests of clients performing a denial of service attack by sending multiple requests in relatively short intervals are blocked.

### 3.2 Modules:

- Preprocessing modules.
- Re-formatting web log file.
- Clustering of users.

- Association rule.
- Recommendation

**Step 1: Preprocessing Module**

In the preprocessing step, there are two major tasks: re-format the log file and retrieve web pages to local space. Log file re-formatting involves revising the log file to an appropriate format for further steps. Web page retrieving involves reading the URL address part of the re-formatted log file, retrieving the Web pages accordingly, and storing them to local space.

**Step 2: Re-Formatting Web Log File**

Each tuple of the log file is a record for one request from a certain user. For example, Figure 2 shows an IIS log file entry, as viewed in a text editor and Table 1 lists and describes the fields used in this work.

Y.Y.Y.Y, -, 03/20/07, 7:55:20, W3SVC2, SERVER, X.X.X.X, 4502, 163, 3223, 200, 0, GET, /Logo.gif, -,

**Step 3: Clustering**

In clustering the data is been clustered. Where the required data is been clustered into one cluster and unwanted data into one cluster. In web document clustering the web pages are preprocessed during clustering the document have multiple topics. It clears the html, xml tags from web pages. Eliminating all punctuations, integrating the Reformatted log files and web document clustering. In user clustering it extract the use query and display the clustering table. Then integration of user cluster with web document clustering.

**Step 4: Association Rule**

Mining association rules among items in a large database is one of the most important problems of data mining. The task is to find interesting association or correlation relationships among a large set of data items. The typical rule mined from database is formatted as follows:

$$X \rightarrow Y [Support, Confidence]$$

It means the presence of item X leads to the presence of item Y, with [Support]% occurrence of [X,Y] in the whole database, and [Confidence]% occurrence of [Y] in set of records where [X] occurred.

**Step 5: Recommendation**

**K-mean Algorithm**

K-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a

cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid.

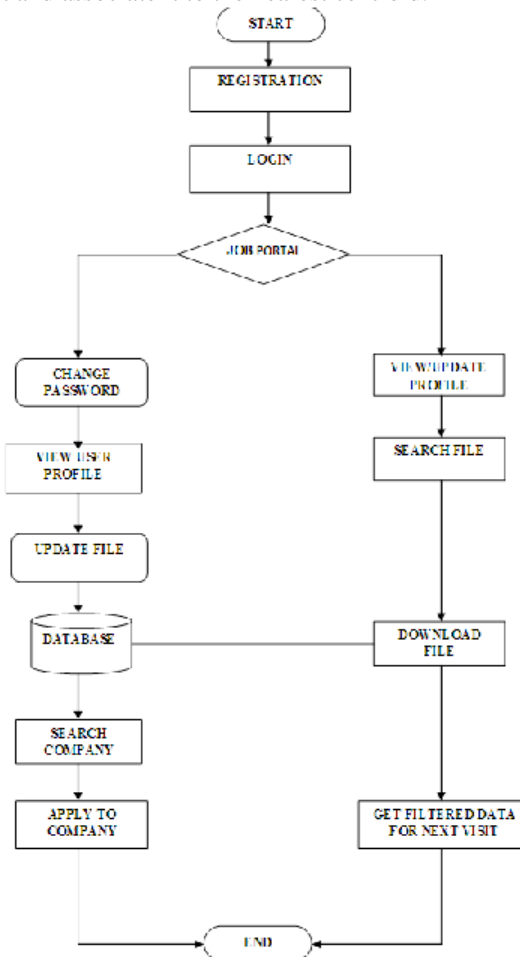


Figure 1 Flowchart

**IV. Conclusion**

In this Project, a novel approach has been proposed to infer user search goals for a query by clustering its feedback sessions represented by pseudo-documents. First, we introduce feedback sessions to be analysed to infer user search goals rather than search results or clicked URLs. Both the clicked URLs and the unclicked ones before the last click are considered as user implicit feedbacks and taken into account to construct feedback sessions. Therefore, feedback sessions can reflect user information needs more efficiently. Second, we map feedback sessions to pseudo-documents to approximate goal texts in user minds. The pseudo-documents can enrich the URLs with additional textual contents including the titles and snippets. Based on these pseudo-documents, user search goals can then be discovered and depicted with some keywords. Finally, a new criterion CAP is formulated

to evaluate the performance of user search goal inference. Experimental results on user click-through logs from a commercial search engine demonstrate the effectiveness of our proposed methods. The complexity of our approach is low and our approach can be used in reality easily. For each query, the running time depends on the number of feedback sessions. Therefore, the running time is usually short. In reality, our approach can discover user search goals for some popular queries offline at first. Then, when users submit one of the queries, the search engine can return the results that are categorized into different groups according to user search goals online. Thus, users can find what they want conveniently.

## REFERENCES

- [1] Zheng Lu, Student Member, IEEE, Hongyuan Zha, Xiaokang Yang, Senior Member, IEEE, Weiyao Lin, Member, IEEE, and Zhaohui Zheng
- [2] International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 12, December 2013
- [3] A New Algorithm for Inferring User Search Goals with Feedback Session J. Shobana, Asst. Professor in SRM University, Department of Computer Applications, Chennai