

# Implementation of Intellectual Browser for User Search Goal Discovery and Reordering System

Hemlata Gaikwad<sup>1</sup>, Manisha Darak<sup>2</sup>

<sup>1,2</sup>Pune University, Department of Computer Engineering, Siddhant College of Engineering, Pune, Maharashtra, India

**Abstract:** *Current age is called as Internet age due to immense use of internet in day to day life by billions and trillions of users. Considering the ambiguity of the keywords being used of searches over the internet, every user has a variant goal for searching different contents with similar keyword searches. Thus the user goals and analysis of searched keywords can improve the search engine results retrieving efficiency and can also aid in reducing the latency to obtain intended results quicker by clustering the random results in a structured manner. Now a day, every single search query fired provides sample of ambiguous results as many terms being searched over internet are similar in terms of keywords only but the semantic aspect is totally different. In proposed technique framework will give the data identified with the client objectives. In this paper we have plan a novel structure to find the client destinations by bunching the client seek objectives and after that new way to deal with produce the pseudo record to speak to the grouping viably. Toward the end we have proposed novel methodology.*

**Keywords:** User feedback, Clicked through URLs, Pseudo documents, Clustering, CAP evaluation, Restructuring the search results

## 1. Introduction

The users are used the web search engines to retrieve different required information from internet. If multiple users are search the same keyword with the different motives they will not get appropriate results because that keyword may cover broad topic. Users get the ambiguous results. Most probably the query with minimum keywords makes the problem. They may cover many meanings of information related to it. To identify the exactly we can do the following; firstly we can guide the user by R. Baeza-Yates et.al. [1] query recommendation. Thus these suggested queries are helps the users to build their query appropriate to their need. Secondly to enhance the search engines by clustering the search results according to the same search goal. Thirdly search results are reconstructed depending on the goals of the user. For better information retrieval search engines must be correctly identify the user search goals. User desired information is considered as user search goal.

The analysis of user search goal has many research are examine. Those are divided into the three categories: S. Beitzel et.al. [2] classification of query, reorganization of search result, and detecting the boundary of particular session. The first classification of query is depending on the user's perspective. This classification may have different classes accordingly users needs to identify the class for query which is in his/her mind. For example the 'Informational query', Geographical query etc. also the user can also use the concept of tagging. Tags attached to the particular query to describe the concept in brief for particular query. At the user level classification of particular query is very difficult. Sometime users are not capable of identify the class in which the query belongs. Second, reorganization of the search results H. Chen and S. Dumais [4] of any search engine also analyzed by the feedback of user about clicked URLs from click through logs directly. But it makes the limitation many times if clicked URLs are minimum in numbers. Other one is when direct analysis of search results provided by search engines are performed means many noisy

results without user feedback also analyzed but they cannot infer the user search goal capably. Third, detection of session boundary is nothing but forecasting the goal from hierarchically segmented query logs. It will care about only the queries which related to the same goal, doesn't care about the user need in detail.

For explaining semantic difference suppose the student has fire a query 'pencil' in Google. Here student intention for searching pencil is for writing equipment as a result. But search engine provides the result for software, movie, and documentation for pencil query. For the expected result user needs to go through entire page. It is not displays or ranked first. Every time student got such mixed result after submission of query 'pencil'.

Because of ambiguous queries, an unnecessary web site response takes place and this unwanted results are disappoint the user's necessities. To fulfill user's need efficiently, we have proposed system which understand user search goal for searching relevant results according user expectations. For this, we are thinking the user's need as a cluster. Because of this clustering, performance of proposed system is improved. Proposed system is skilled to reconstruct the results by clustering the expectations of the user on several periods. The user need can allocate with words on which the clustering is performed. The reconstructions of results are based on the clustering.

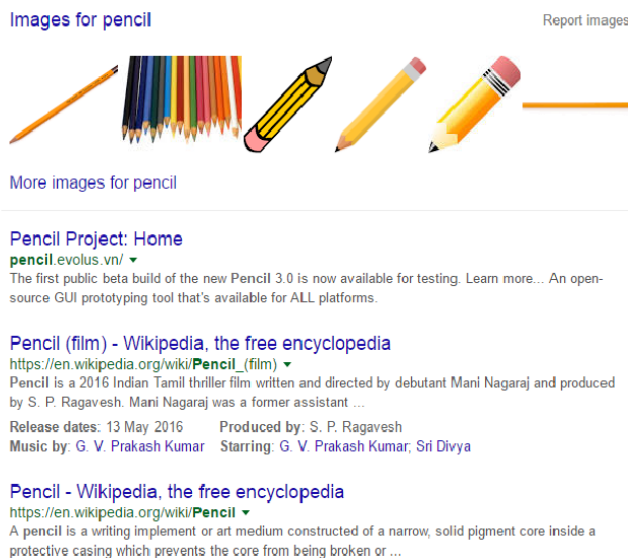


Figure 1: Ambiguous result for 'pencil'

To increase the searching value, various schemes were introduced such as categorization of query, identification or prediction of search results, and session limit detection to enhance seeking. On the other hand, if the amount of various clicked URLs of a user need may be insignificant then these methods have restrictions. In the context of search engine, existing systems only focus on examination the search results returned for submitted query. So the proposed system improves the performance of inferring the user search goals by the use of clustering the feedback sessions and evaluation of the system by CAP method.

This research paper is organized as follows: The literature survey is described in 'Research Elaborations' Section 2. 'System Architecture' is described in Section 3. Workings of different components of system are presented as 'Module Description' in Section 4. The Section 5 shows the results of implemented system. Section 6 describes the analysis implemented by CAP evaluation. And Section 6 concludes the paper.

## 2. Research Elaborations

### A. New Algorithm for Inferring User search Goals using Feedback Sessions.

The work focused on user click feedback only for a single session, as user may search different concepts at different sessions.

### B. User Profile for Personalized Web Search

Uses the users profile information and designation information to infer the user search goal. Only those results are displayed which had relevance with user designation.

### C. PMSE: A Personalized Mobile Search Engine

Due to the importance of location information in mobile search, PMSE classifies these concepts into content concepts and location concepts. In addition user's locations (positioned by GPS) are used to supplement the location concept in PMSE.

### D. Personalized Web Search with Location Preferences

The work just mapped the user query with user location, as most of the user queries are location specific.

### E. A survey on Inferring User Search Goals using Feedback Session

The goal identification requirement can be fulfilled by using pseudo document and feedback session. Click through logs are used to create feedback session. Pseudo document is the revised version of feedback session for clustering the websites. This paper reviews many web mining techniques for inferring the user goals. Query suggestion techniques are examined in this.

## 3. System Architecture

Following figure shows system architecture

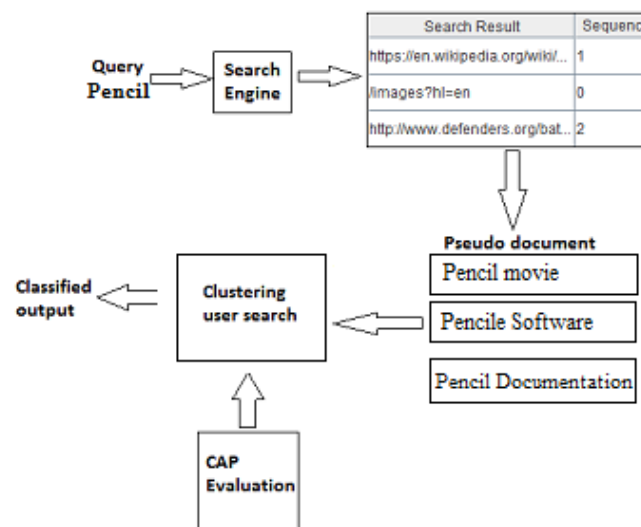


Figure 2: System Architecture

Main components of system architecture are search engine, feedback session, Pseudo documents, clustering, CAP.

## 4. Module Description

### I) Click through Log

The major outcome of the proposed system depends on user feedback for clustering the obtained results. Once the user fires the query, the unstructured results are obtained which needs to be clustered as per user feedback. So URLs are to be clicked on to create the binary vector and record the click sequence for evaluation.

### II) TF - IDF Calculation

Once the clicked and unclicked URLs are recorded for current session, the terms from the URLs are needed to be counted for determining the relevance ratio of the terms to the clicked URLs. So term frequency and Inverse document frequency is needed to be calculated for analysing the term count and further pseudo document creation.

### III) Pseudo Document Creation

Once the term frequency is computed major clustering criteria is to be decided which is done on basis of Higher TF

values obtained for all terms in the documents. Higher 10 TF values are considered as pseudo document tiles.

#### IV) K-Means Clustering

Based on TF IDF values obtained for terms, and the pseudo document cluster titles decided, the URLs containing the same terms are categorized under respective pseudo document. thus all terms and URLs get structured on.

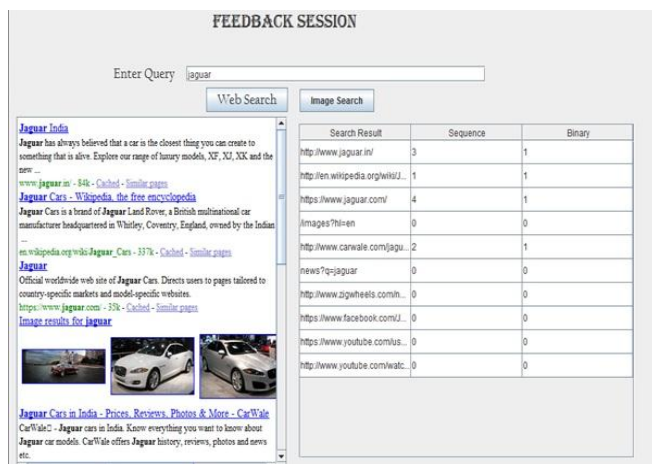
#### V) Re-ranking of search results

After clustering the pseudo document, search results are re-ordered according to higher cluster which matches user need. Performance of the re-ordering assess by CAP evaluation.

#### VI) CAP Evaluation

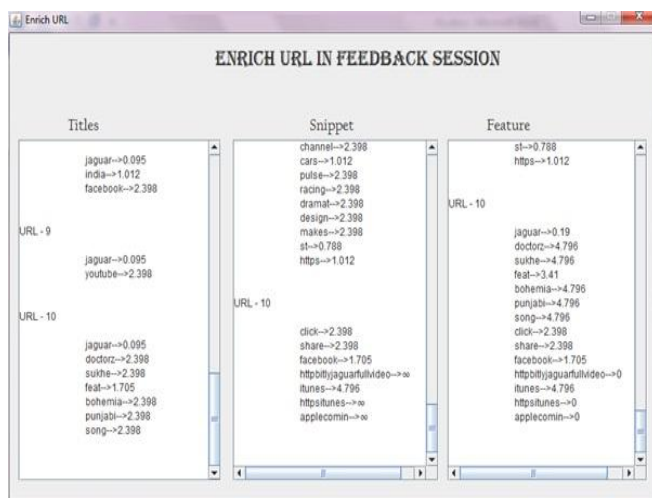
Basically CAP evaluation technique refers click through logs. The clicked URL reflects the relevant results used in calculation of CAP evaluation. CAP evaluation value indicates that cluster of restructured results is relevant to user search goals.

### 5. Results

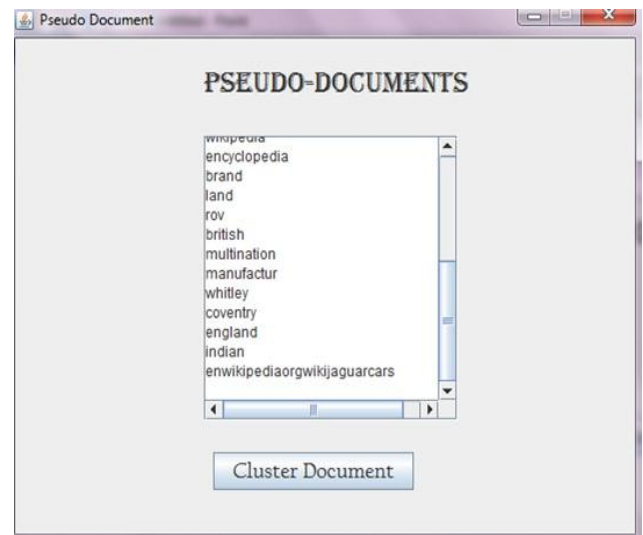


**Figure 3:** Accept the query and feedback from user.

The user enters some keywords as a query to search engine. For example 'jaguar'. It displays the results in one column and other 3 columns are search results which indicate the feedback of user with number sequence and binary vector.



**Figure 4:** Discovery of TF IDF values for text



**Figure 5:** Creation of Pseudo documents

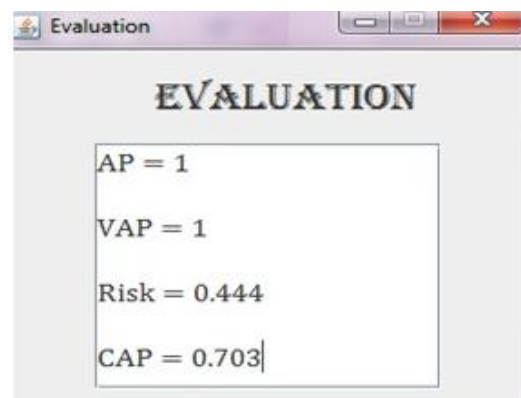


**Figure 6:** Re-ranked web results

### 6. Result Analysis

For comparing performance of proposed system we used CAP evaluation. If CAP value is greater than 0.5 then the restructured results are closer to search goal of user.

As CAP value is greater than 0.5 so proposed system results are acceptable, And closer to user search goal.



**Figure 7:** Performance criteria values

## 7. Conclusion

Internet is a great source of information. The search engines are mainly focus on to fulfill the user requirements in minimum response time. For this purpose the existing system's are consider user's history information, user profile information, location preferences. These methods efficient when users are searching for previously searched data. But if user intensions are variable for each searching time then these systems are not understand the current user intention for searching. Ultimately their performance degrades in terms of identifying user search goals. The proposed system makes use of feedback sessions to collect the user needs instead of focusing only on search results or clicked URLs. Feedback sessions are considers the clicked as well as unclicked URLs before last clicked URL. Pseudo documents make the clear intensions of user search goal. Clustered search results are display to the user after CAP evaluation. CAP evaluation guarantees that given results are reflects user search goals efficiently. So the proposed system has better performance.

## References

- [1] R. Baeza-Yates, 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.
- [2] S. Beitzel, E. Jensen, A. Chowdhury, and O. Frieder, "Varying Approaches to Topical Web Query classification," Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development (SIGIR '07), pp. 783-784, 2007.
- [3] 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.
- [4] H. Chen 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.
- [5] 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.
- [6] T. Joachims, "Evaluating Retrieval Performance Using Clickthrough Data," Text Mining, J. Franke, G. Nakhaeizadeh, and I. Renz, eds., pp. 79-96, Physica/Springer Verlag, 2003.
- [7] 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.
- [8] T. Joachims, 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.

## Author Profile

**Hemlata Gaikwad** received her B.E. degree in Computer Science and Engineering from Pune University (Maharashtra India) in 2011, and Pursuing M.E. degree in Computer Engineering from Siddhant College of Engineering Pune (India).

**Manisha Darak** She has been working as an Associate Professor in the Department of Computer Engineering at Siddhant College of Engineering Pune, India