



# Restructure Search Results for Efficient Web Search

Addlin Shinney R<sup>1</sup>, Saravana Kumar T<sup>2</sup> and Roslinmary M<sup>3</sup>

M.TECH / IT, Dr.Sivanthi Aditanar College of Engineering, Tiruchendur-628215, Tamilnadu, India<sup>1,3</sup>

Asst. Prof / IT, Dr.Sivanthi Aditanar College of Engineering, Tiruchendur-628215, Tamilnadu, India<sup>2</sup>

**ABSTRACT:** Nowadays every individual are accessing web to find out the needed information. Usage of web rapidly increases and for that the search engine results should be reorganized for easier user interaction for web search user has different goals for different query while submitting into the search engine. Inference of user goal for a query helps to improve search engine relevance by analyzing query logs. In this paper a framework is proposed to identify different user search goal for a query by clustering filter session. Filter session is generated based on user clicked logs that reflect user needs. Feedback contains both clicked and un-clicked URL's Second, propose a new method to generate precise text to appropriate representation of filter session for clustering. Mapping filters session to precise text to find goal text in user mind. Finally propose a novel Criterion "Assorted Average Precision" to determine the performance of inferring user search goals.

**KEYWORDS:** User search goal, Filter session, precise text, restructuring search results, Assorted Average Precision.

## I. INTRODUCTION

In web application, user has to submit their query to the search engine. Search engine list out the results related to that particular query. Different user's wanted to get different aspects of information for a same query. For example, user submits query as "apple", some user want to view about the apple product and some user want to know about the nutrients of the fruit apple. So it is important to know the user search goal for a query. User search goal is defined as an information need of the user .Information needed by the user is based on their desire and that should satisfy their need. Lot of advantages is there for analyzing user search goal .Search goals are represented by the cluster of information needs. Advantages are summed as follows. First, reorganizing search results based on the user search goal and similar search results are placed on the same cluster; by this type of representation user can easily find the information they want. Second, Keywords are used to represent different user search goals. Keywords are used for query recommendation and the indicated query helps the user to find their queries more precisely. Third, re-ranking web search results based on the distribution of user search goals for a query. The overall methodology of our project is given below.

Sum up of our work as follows:

- We propose a framework to find out the user search goal for a query. This is done by clustering filter session. This method is more efficient than clustering web search results and clicked URLs. After filter sessions are clustered we can obtain distribution of different user search goals.
- We propose a new method to combine the URLs in the filter session. Based on that pseudo- document are generated which exactly reflect the information need of the user.
- We propose an AAP to examine the performance of user search goal. Thus we can determine different user search goals for a single query.

**International Journal of Innovative Research in Computer and Communication Engineering**

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 1, March 2014

**Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)**

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6<sup>th</sup> & 7<sup>th</sup> March 2014

The rest of the paper organized as follows: The model of our approach is in section 2. The filter session and its representation in section 3. Section 4 find user goal by clustering precise text. Section 5 reviews several related works. Section 6 concludes the paper.

**II. MODEL OF OUR APPROACH**

Fig 1 shows the model of our work. First the original web search result. To reorganize the resulting search result first the filter session are taken from user clicked logs and then map to precise text. Depicting each with a keyword and finally based on that resulting web pages are restructured.

**III. FILTER SESSION AND ITS REPRESENTATION**

In the section, describes the filter session and precise text to represent filter sessions. This precise text will exactly helps to predict the information need of the user.

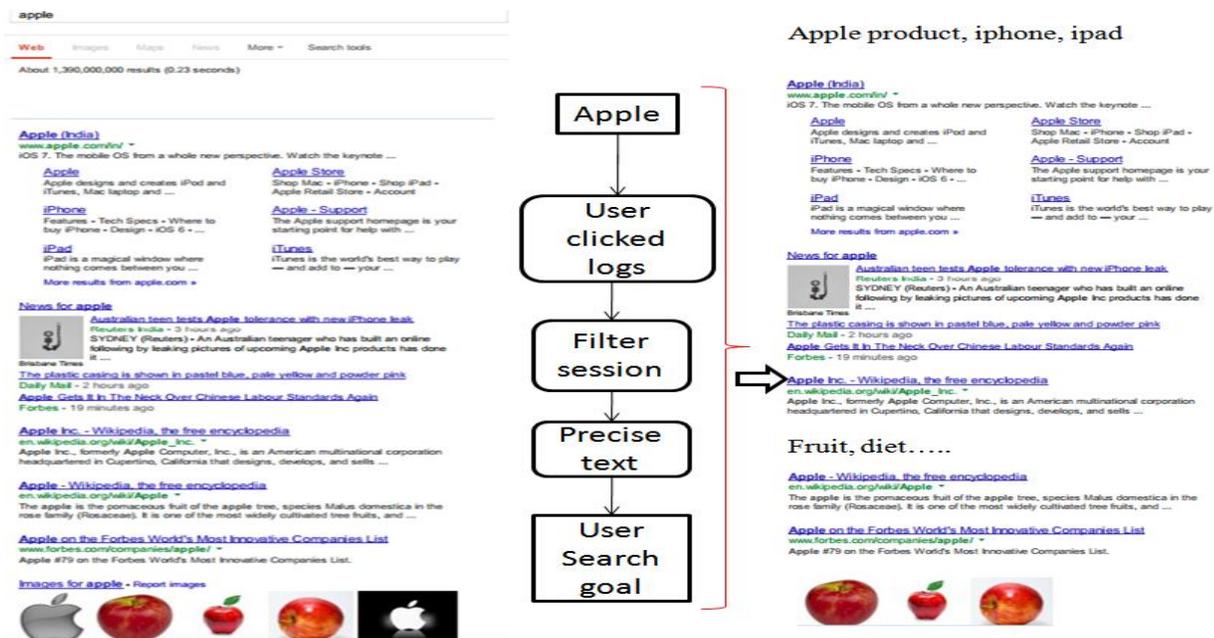


Fig 1. The model of our approach

**3.1 Filter Sessions**

Basically, a session is a series of consecutive queries to satisfy single information need. In this paper we try to infer search goal for a specific query. Single session have a single query is introduced. Meanwhile the filter session we considered is a single session and further we can extend this paper to the whole session. The filter session contains both clicked and un clicked URLs and finish the session with the last URL clicked in the single session. Before the final click all URLs are scanned by the user. So we consider both the clicked and un-clicked URLs before the last click on the

**International Journal of Innovative Research in Computer and Communication Engineering**

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 1, March 2014

**Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)**

**Organized by**

**Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6<sup>th</sup> & 7<sup>th</sup> March 2014**

session. Fig.2 illustrates the single filter session. The left section shows the 8 search results for the query “apple”. The right section shows the click sequence and “0” shows un-clicked. From the fig, single session has 8 URLs and only 6 URLs are considered for filter session. The filter session is indicated by the red color rectangular box. The six URLs consist of 3 clicked and 3 un-clicked URLs. Generally user will scan all the URLs of the search result web page and reasonably the 3 URL in the rectangular box also evaluated by the user.

| SEARCH RESULTS                  | CLICK SEQUENCE |
|---------------------------------|----------------|
| www.apple.com/in/               | 0              |
| en.wikipedia.org/wiki/apple_Inc | 1              |
| en.wikipedia.org/wiki/apple     | 0              |
| www.forbes.com/companies/apple  | 2              |
| www.fruits.com/apple            | 0              |
| www.bloomberg/quote/apple       | 3              |
| bgr.com/apple                   | 0              |
| Techcrunch.com/apple            | 0              |

**Fig. 2. A filter session for a single session**

Each filter session that reflects user needs and user don’t care about the information. There is plenty of filter session in user logs. This method is more efficient to infer user search goal compared with search logs.

**3.2 Generate Precise Text**

Based on user click logs and queries, filter session varies a lot. So inferring user search goal only by referring filter session is not suitable. Some representation method is needed to represent filter session in a coherent way. To represent filter session there are many ways.

| SEARCH RESULTS                  | Click Frequency | Binary Vector |
|---------------------------------|-----------------|---------------|
| www.apple.com/in/               | 0               | 0             |
| en.wikipedia.org/wiki/apple_Inc | 1               | 1             |
| en.wikipedia.org/wiki/apple     | 0               | 0             |
| www.forbes.com/companies/apple  | 2               | 1             |
| www.fruits.com/apple            | 0               | 0             |
| www.bloomberg/quote/apple       | 3               | 1             |

**Fig. 3. The binary vector method**

Fig. 3. Shows the binary vector representation of URLs returned for the query “apple”. For binary vector clicked URLs are represented by “1” and un-clicked URLs are denoted by “0”. This filter session has the binary vector [010101]. Different filter session has different number of URLs and based on that the filter session also gets changed.

Binary vector is not so efficient to predict user search goals. So it is not suitable to use binary vector representation and we need to introduce new method to represent filter session.

For a single query, user has a vague representation of keywords. Using the keyword to check whether the retrieved document satisfy their needs. Key words are named as “target text”. These target texts reflect the information need of the user but not done in explicit way. So we introduce a precise text that helps to identify user information. A new method is introduced to map the filter session to precise text. It includes two steps to build the precise text showed in Fig.4.

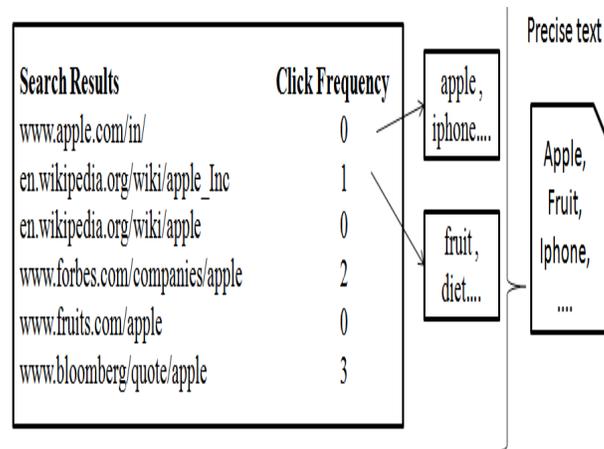


Fig.4.Mapping filter session

They are described in the following way:

**1. Presenting the URLs in the filter session.** In the first step we first select the URLs and enrich it by adding text content that are taken from the filter session. In the same way each URLs in the filter session are described by small text summary. That summary contains both titles and snippet. Then we have to apply some text process to the text summary. The text processes are transforming all the letters to lowercases, stemming and removing stop words. The title and snippet represented by Inverse Document Frequency. The feature representation of the URLs in the filter session and for the title, snippets assign weights. For weights of snippets is set to 1 initially. Then we specify titles are more significant than the snippets. Therefore we assign weight of the title is 2 more than the snippets.

**2. Generating precise text based on URL representation.** To find the feature representation of the filter session considers both clicked and un-clicked URLs. While doing search user skip some URLs because that is similar to that of the previous one. In those situations the un-clicked URLs wrongly reduce the weight of the precise text. Our method solve the problem in three cases: The first case is the ideal case that one term appears in all the clicked URLs and not appear in the un-clicked URLs. In the case people won't prefer the un-clicked URLs, because it doesn't have important information. The second case is defined as general case. In this case the term appears in clicked and subset of un-clicked URLs. User skips this because of duplication. Skipping doesn't affect this case. The third class is the bad case. In this case the term appears in both the clicked and all the un-clicked URLs. People skip because of duplication, hence this method assign reasonable weight. Then we find the user search goal based on the precise text. Precise text is discovered next we find the different goal text for each query. Based on these 3 cases we assign weights for the title and snippets present in the precise text.

#### IV. EVALUATE WEB SEARCH RESULTS

This type of evaluation is a big problem because user goals are not yet predicted. Correct numbers of clusters are not determined yet. The filter information is needed to find the best cluster number. So we have to develop a new metrics to evaluate the performance. First user goal has to be determined and based on that the web pages are reorganized. If user goal evaluated correctly then we can easily restructure the resulting web page. It is the one of the application finding user search goal. A method Assorted Average Precision (APP) helps to find the performance of web search results. We also describe the new method to find the correct number of clusters.

##### 4.1 Evaluation Measure

Single session is considered to minimize the manual work. Because from user clicked logs we get relevant and irrelevant feedback. The clicked URLs are relevant and un-clicked URLs are irrelevant. An evaluation based on the user

**International Journal of Innovative Research in Computer and Communication Engineering**

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 1, March 2014

**Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)**

**Organized by**

**Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6<sup>th</sup> & 7<sup>th</sup> March 2014**

feedback, ranked the relevant document. This is also an unsatisfactory, so we have to avoid the risk while classifying search results.

**V. RESULTS**

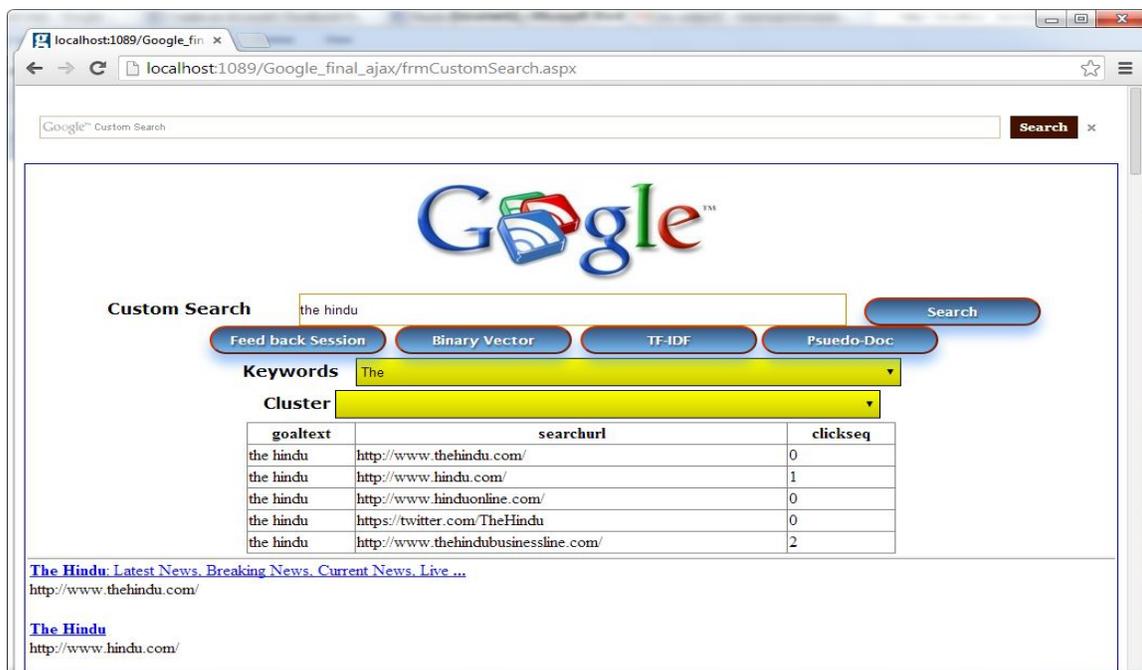


Figure 5.1 Generating Filter Session for Multiple Clicks

**International Journal of Innovative Research in Computer and Communication Engineering**

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 1, March 2014

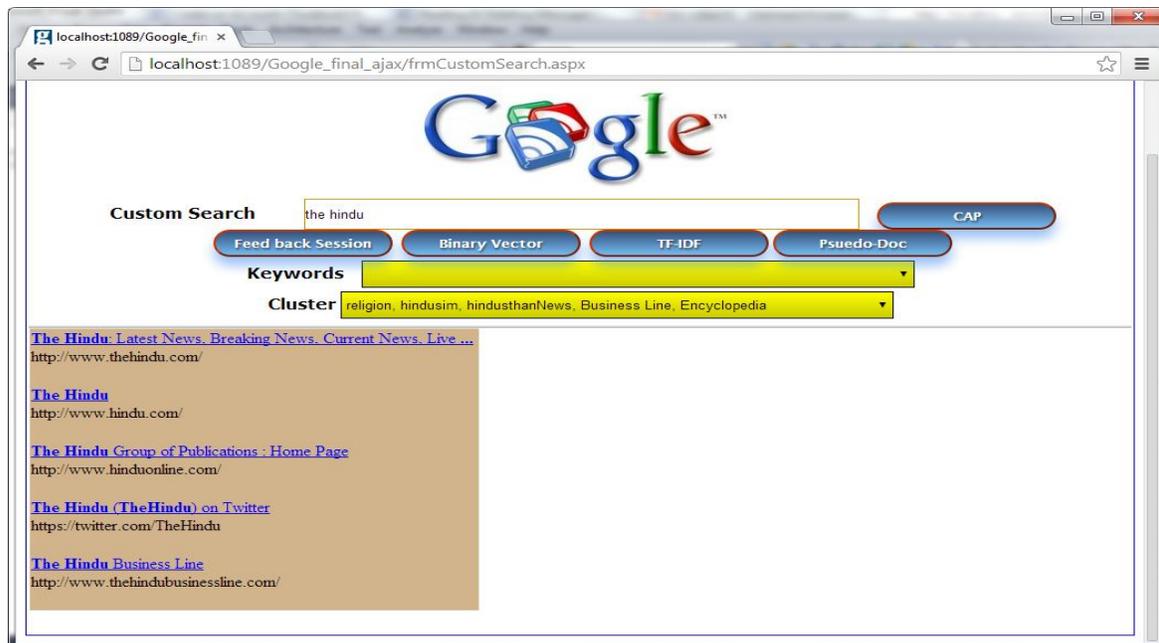
**Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)****Organized by****Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6<sup>th</sup> & 7<sup>th</sup> March 2014**

Figure 5.2 Assorted Average Precision

**VI. CONCLUSION**

In this paper, a new approach is introduced to find user search goal. That is done by clustering filter session represented by precise text. First we introduced a filter session to find the user search goal. For the case of considering search results or user-logs our method is efficient. Comparing with all the methods we explained in the above section our method effectively does restructuring search result web pages. This filter session effectively reflect user needs. Second, we map filter session to precise text to appropriate target text in user mind. The precise text is formed by adding some title and snippets. Based on that user target text is identified and depicting each of them with a keyword. Finally the reorganized search results are evaluated for its performance.

**REFERENCES**

- [1] Allen, R. B., Two digital library interfaces that exploit hierarchical structure. In Proceedings of DAGS95: Electronic Publishing and the Information Superhighway (1995).
- [2] Chakrabarti, S., Dom, B., Agrawal, R., and Raghavan, P. Scalable feature selection, classification and signature generation for organizing large text databases into hierarchical topic taxonomies. The VLDB Journal 7,(1998)..
- [3] T.Joachims. Optimizing search engines using clickthrough data. In Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD), 2002.
- [4] M. Just and P. Carpenter. A theory of reading: From eye fixations to comprehension. Psychological Review, 87:329–354, 1980.
- [5] B.J. Jansen and U. Pooch. A review of Web searching studies and a framework for future research. J. of the American Society of Information Science and Technology, 52(3):235 { 246, 2001.
- [6] D. Beeferman and A. Berger. Agglomerative clustering of a search engine query log. In Proceedings of ACM SIGKDD '00, 2000.
- [7] M. Pasca and B.-V Durme, "What You Seek Is what You Get: Extraction of Class Attributes from Query Logs," Proc. 20th Int'l Joint Conf. Artificial Intelligence (IJCAI '07), pp. 2832-2837, 2007.