



# **iJRASET**

International Journal For Research in  
Applied Science and Engineering Technology



---

# **INTERNATIONAL JOURNAL FOR RESEARCH**

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 3**

**Issue: II**

**Month of publication: February 2015**

**DOI:**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# Deducing User Seek Goals with Opinion Array Using Fuzzy Self Design Algorithm

Yuvashree A.K<sup>#1</sup>, Venkatakrishnan .M<sup>\*2</sup><sup>#1,\*2</sup>Department of Information Technology,<sup>#1,\*2</sup>Sri Venkateswara College of Engineering, Pennalur, Sriperumbudur, Pin-602117, India.

**Abstract-** For an ambiguous query, based on new ideas many users may have different search goals when it come to a search engine. In improving search engine relevance and user experience is based on the inference and analysis of user search goals. In this paper, we propose a new approach to infer user search goals by deeply analysing search engine query logs. First, this paper proposes a user search goals by clustering the proposed opinion sessions by a query. Opinion sessions are constructed from user click-through logs it can efficiently reflect the information needs of users. Second, it proposes a new approach to generate pseudo-documents to better represent the clustering by opinion sessions. Finally, in this paper proposed a new criterion based on "Classified Average Precision (CAP)". This is to evaluate the performance of inferring user search goals. Our proposed method is an experimental result that presented using user click-through logs from a commercial search engine to validate the effectiveness.

## I. INTRODUCTION

A new approach to capture user search goals in image search by pseudo-images method which is extracted in user click-through logs to reflect user information needs by mining single sessions. Hence, this also proposes a new evaluation criterion to determine for a query by the number of user search goals. An experimental result is the effective of the proposed method by demonstration. Capturing users future search actions has many potential applications they are query recommendation, advertisement arrangement, and web page re-ranking so on. Predicting users' future queries and URL clicks based on global users' query logs and their current access behaviours in this paper. This explore various features from queries and clicked URLs in the users' current search sessions, use them for prediction and select similar intents from query logs. Due to an intent shift problem in search sessions, this paper discusses more effects on the prediction, what representations are more accurate to represent users' intents, how to retrieved similar intents affect the prediction and how the intent similarity is measured. Performance evaluation shows that our proposed method outperforms the existing method and also shows that the developed model demonstrates the effectiveness of the Personalized Image Search. Search Engine retrieves significant and essential information from the web based on the query term given by the user. Many papers exploit the Social annotations and Novel Framework for considering the user query relevance and user specific-topic to learn personalized image search. The proposed framework contains two techniques: Utility and Prediction model for social annotations. We introduce a Hit Matrix technique for user query relevance and preference into the specific topic space. Evaluating the Performance shows the proposed method outperforms the existing method and also shows that the developed model demonstrates the effectiveness of the Personalized Image Search. The significant and essential information from the web based on the query term given by the user is retrieved by Search Engine. The query log file contains log entry for every request given by the user to the search engine that is maintained in the system desktop or in the proxy server. The performance of the search engine is improved by the process of mining the query log file. The similarity between the query keywords and URLs in its first phase is a proposed algorithm that the query log file to discover. In the second phase, similarity measure generated from the first phase is combined with the query cluster and the URL cluster is created. Based on their previous search histories and click through data the clusters recommend query to the user to frame their future queries. The user about the URL selection for the future queries is combined similarity based approach.

## II. SYSTEM DESCRIPTION

In this paper, Depicting each goal with some keywords automatically we aim at discovering the number of diverse user search goals for a query. We first propose a new approach to infer user search goals for a query by clustering our proposed opinion sessions. Then, we propose a new optimization method to map opinion sessions to pseudo-documents which can efficiently reflect user information needs. Depict them with some keywords; at last we cluster these pseudo documents to infer user search goals. The proposed opinion session ends with the last URL that was clicked in a single session we propose this new criterion "Classified Average Precision" to evaluate the restructure results and consists of both clicked and unclicked URLs. Based on the proposed criterion, the best cluster number also describes the method to select. We first propose a new approach to infer user search goals for a query by clustering our proposed opinion sessions. Then, we propose a new optimization method to map

## International Journal for Research in Applied Science & Engineering Technology (IJRASET)

opinion sessions to pseudo-documents which can efficiently reflect user information needs.

- A. We propose a framework to infer different user search goals for a query by clustering opinion sessions.
- B. We demonstrate that clustering opinion sessions is more efficient than clustering search results or clicked URLs directly.
- C. Moreover, the clusters distributions of different user search goals can be obtained conveniently after opinion sessions.
- D. We propose a new optimization method which can effectively reflect the information need of a user to combine the enriched URLs in an opinion session to form a pseudo-document. Thus, we can tell what the user search goals are in detail.
- E. We propose a new criterion CAP to evaluate the performance of user search goal inference based on restructuring web search results. Thus, we can determine the query for number of user search goals.

### III. METHODOLOGY

#### A. Ambiguous Query

Queries are submitted to search engines to represent the information needs of users. Since many ambiguous queries may cover a broad topic and different users may want to get information on different aspects when they submit the same query. Hence, sometimes queries may not exactly represent users' specific information needs. For example, when the query "the sun" is submitted to a search engine, while some others want to learn the natural knowledge of the sun and some users want to locate the homepage of a United Kingdom newspaper.

#### B. Restructure web search results

We need to restructure web search results according to user search goals by grouping the search results with the same search goal users with different search goals can easily find what they want. User search goals represented by some keywords can be utilized in query recommendation. Web search results that contain different user search goals by distributing of user search goals can also be useful in applications such as re-ranking. Due to its usefulness, many works about user search goals analysis have been investigated. They can be classified into three classes: query classification, search result reorganization, and session boundary detection.

#### C. Opinion Sessions

The opinion session ends with the last URL that was clicked in a single session we propose this new criterion "Classified Average Precision" to evaluate the restructure results and consists of both clicked and unclicked URLs. All the URLs have been scanned and evaluated by users is motivated that before the last click. So, besides the clicked URLs, part of the user opinions the unclicked ones before the last click should be done. Opinion session can tell what a user requires and what he/she does not care about. Moreover, there are plenty of diverse opinion sessions in user click-through logs. Hence, it is more efficient to analyse the opinion sessions than to analyse the search results or clicked URLs directly, for inferring user search goals.

#### D. Pseudo document

In this paper, we need to map opinion session to pseudo documents User Search goals. There are two steps to build the pseudo-document. One is representing the URLs in the opinion session. URL in a opinion session is represented by a small text paragraph that consists of its title and snippet. Then, some textual processes are implemented, such as transforming all the letters to lower cases, stemming and removing stop words to those text paragraphs. Another one is Forming pseudo-document based on URL representations. To obtain the an opinion session for feature representation , we propose an optimization method to combine both clicked and un-clicked URLs in the opinion session.

$$w_k(x) = \frac{1}{\sum_j \left( \frac{d(\text{center}_k, x)}{d(\text{center}_j, x)} \right)^{2/(m-1)}} .$$

The opinion session ends with the last URL that was clicked in a single session we propose this new criterion "Classified Average Precision" to evaluate the restructure results and consists of both clicked and unclicked URLs. All the URLs have been scanned and evaluated by users is motivated that before the last click.

#### E. User Search Goals

We cluster pseudo-documents by FCM clustering which is simple and effective. Since we do not know the appropriate number of user search goals for each query, we set number of clusters to be five different values and perform clustering based on these five values, respectively. After clustering all the pseudo-documents, one user search goal is considered for each cluster. The

# International Journal for Research in Applied Science & Engineering Technology (IJRASET)

centre point of a cluster is computed as the average of the vectors of all the pseudo-documents in the cluster.

## IV. SYSTEM DESIGN

### A. Architecture Diagram

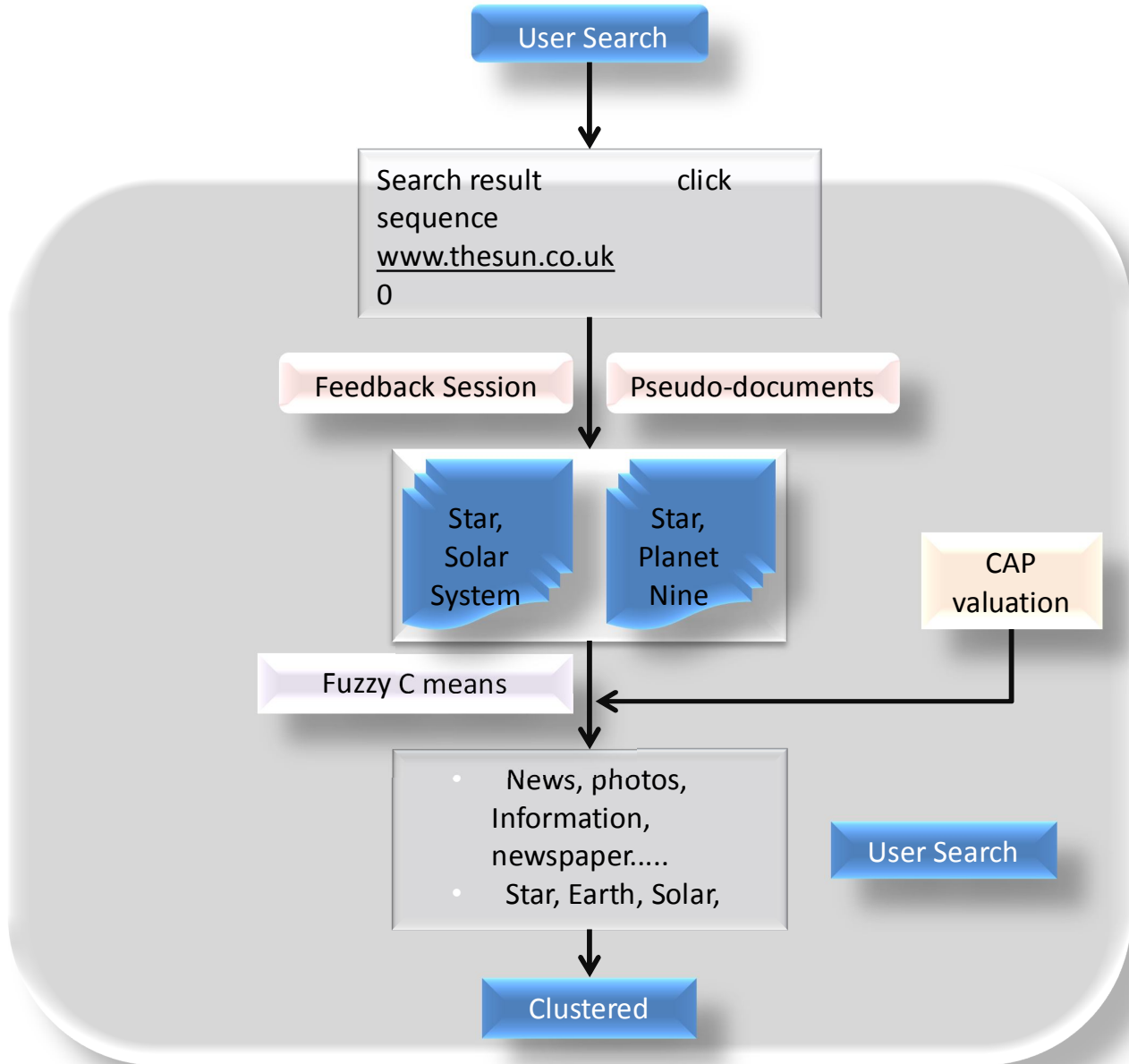


Fig.1 Architecture Diagram

The Architecture Diagram is given in the Fig.1, the flow is from the user search to the cluster. In-between the fuzzy means is used.

### B. Block Diagram

## International Journal for Research in Applied Science & Engineering Technology (IJRASET)

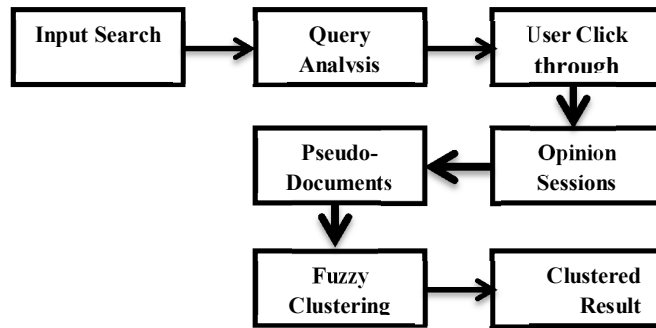


Fig.2 Block Diagram

The Fig.2 shows the block diagram for the system, where the input search is given to the query analysis, after the query is analysed the user should click through the opinion Session and by sending it to Pseudo-Documents and the Fuzzy Clustering is made, so that the clustered result is obtained.

### V. MODULES

#### A. Login

- 1) In computer security, a log in or log on is to identifying and authenticating the user referring to credentials presented by the user is the process by which individual access to a computer system is controlled.
- 2) A user can log in to a system to obtain access, and when log out or log off the access is no longer needed. To close off one's access the log out is done in a computer system after having previously logged in.

#### B. User Search Logs

The user enters the queries to the search engine. The queries are maintained as a log and the results will be produced based on the keywords. The search goals for a query and depicting each goal with some keywords automatically. The user's queries are saved.

#### C. Opinion Sessions

The opinion sessions is defined as the array of both clicked and un clicked URLs and ends with the last URL that was clicked in a session from user click-through logs. Then we map the opinion sessions to pseudo-documents which can effectively reflect user information needs. We combine the enriched URL's in an opinion sessions to form a pseudo document. The opinion session is based on a single session .and also it can be extended to the whole session. So besides the clicked URLs, the UN clicked ones before the last click should be a part of the user opinions. For inferring user search goals it is more efficient to analyse the opinion sessions than to analyse the search results or clicked URLs directly.

#### D. Pseudo Documents

The search log will be represented as 0 in the click sequence. The binary vector is used to represent the opinion sessions 1 as clicked and 0 as UN clicked.

Steps to build pseudo documents:

- 1) Represent the URL in the opinion session.

It extracts the titles and snippets of the returned URL's from the opinion sessions. Each URL is represented as a small text paragraph then some textual process is implemented as text paragraphs such as transforming all the letters to lower case, stemming and removing stop words.

- 2) Forming pseudo documents according to URL representations: Process of combining both clicked and un clicked URL's in the opinion sessions.

#### E. Clustering The Pseudo Documents

The Pseudo documents are clustered into K means clustering. It performs clustering based on the five values. The terms with the

## International Journal for Research in Applied Science & Engineering Technology (IJRASET)

highest values in the centre points are used as the keywords to depict user search goals. The clustering is the process on a term-weight based on vector representation of queries, of the clicked URLs for the query. Similar queries may not share query-terms but they do share terms in the documents selected by the users.

Thus we avoid the problems of comparing and clustering sparse collection of vectors in which similar queries are difficult to find a problem that appears in previous works on clustering. So we do rank the suggested queries based on two criteria's:

- 3) The similarity of the queries to input query (the query submitted to the search engine).
- 4) The support which measures how much the answers of the query have attracted the user's attention.

### *F. Final Restructured Results*

The results are restructured based on the evaluation of web search goals. This method is called CAP (Classified Average Precision). Search engines will return millions of search results so it is necessary to organize them to make it easier for users to find what they want. The user search goals are represented as the vectors. So we perform categorization by choosing the smallest distance between the URL vector and user-search-goal vectors. By this way the results can be restructured according to the inferred user search goals.

## VI. CONCLUSION

In this paper, a new approach has been proposed to infer user search goals for a query by clustering its opinion sessions represented by pseudo documents. First, we introduce opinion sessions to be analysed to infer user search goals which is much better than search results or clicked URLs. Both the clicked URLs and the UN clicked ones before the last click are considered and taken into account to construct opinion sessions as user implicit opinions. Therefore, opinion sessions can reflect user information needs more efficiently. Second, we map opinion sessions to pseudo documents to approximate goal texts in user minds. The pseudo documents which can enrich the URLs including the titles and snippets with additional textual contents. Based on these pseudo documents, user search goals can then be depicted and discovered with some keywords. Finally, a new criterion CAP is created to evaluate the performance of user search goal inference. On user click through logs from a commercial search engine demonstrate the effectiveness of our proposed methods as an experimental results.

## REFERENCES

- [1] H. Chen, M. Lin, and Y. Wei, "Novel Association Measure Using Web search with Double checking", Proc. 21<sup>st</sup> Int'l Conf. Computational Linguistics and 44<sup>th</sup> Ann. Meeting of the Assoc. for Computational Linguistics (COLING/ACL '06), pp. 1009-1016, 2006.
- [2] 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.
- [3] 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.
- [4] S. Beitzel, E. Jensen, A. Chowdhury, and O. Froder, "Varying Approaches to Topical Web Query Classification", Proc. 30<sup>th</sup> Ann. Int'l ACM SIGIR Conf. Research and Development (SIGIR'07), pp. 783-784, 2007.
- [5] H. Cao, D. Jian, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li, "Context-Aware Query Suggestion by Mining Click-Through", Proc. 14<sup>th</sup> ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD'08), pp. 875-883, 2008.
- [6] 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.
- [7] C. K. Huang, L. F. Chien and Y. J. Ovang, "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.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)