IMPROVE THE PERFORMANCE OF USER SEARCH GOALS USING FEEDBACK SESSION

D.Kavitha[1], K.M.Subramanian[1], and Dr.K.Venkatachalam[2]

[1]Computer Science and Engineering, Erode Sengunthar Engineering College,
Anna University Chennai, Tamilnadu, India

[2]Electrical Communication and Engineering, Vellalar College of Engineering and Technology,
Anna University Chennai, Tamilnadu, India

ABSTRACT: Search engine is one of the most important applications in today's internet. Users collect required information through the search engine in the internet. Analyzing user search goals are essential to provide best result for which the user looks for in the internet. In existing system, various techniques such as Feedback session, goal text, Pseudo-documents restructuring search result based on term frequency are used to infer user search goals. Existing search results based on term frequency (keywords) which may display unwanted results. In proposed system "Classified Average Precision (CAP)" algorithm is used to understand user search goals efficiently and evaluate the performance of inferring user search goals. Phrase search is performed in proposed system instead of keyword search. Initially Noun Phrase of user query is framed using natural language processing. Framed noun phrases are searched in webpages available in Internet. Term frequency of each noun phrase is found in Pseudo document i.e., finding number of webpages a particular noun phrase is occurred. Based on term frequency, place the webpage/document which contain only the above noun phrases at top link. Here user needs is highlighted and provides a user friendly search engine. Performance of inferring user search goal is evaluated using a new CAP algorithm.

KEYWORDS: Classified Average Precision (Cap), User Search, Feedback Session.

1 INTRODUCTION

In web search applications, queries are submitted to search engines to represent the information needs of users mostly queries may not represent user's specific information exactly. User desire to obtain information to satisfy the need exactly. In order to achieve the user's desire web search result can be restructured to provide user search information at top among result patterns displayed various steps involved in acquiring user's goal are first restructure web search results based on grouping the results of some goal together. Second representing some phrases frequently used in grouped search result. Last step reranking search results that containing different user search goal. Various people attempts to infer user goals using various techniques like predefining the queries into two classes (i.e.) product intent and job indent, analyzing the clicked URL's directly from user click through logs to organize search results.

User search goal represent for what user looks for in the internet. The feedback session is defined as the series of both clicked and unclicked URLs and ends with the last URL that was clicked in a session from user click through logs. Since feedback session represent both clicked and unclicked URLs. Feedback session provides idea to cluster user care about and does not care about. Instead of goal text the phrase search is performed. Based on the feedback session construct the pseudo document for analyzing the accurate result. This pseudo document consist of phrases for each URLs present in the feedback session. This is called as enriched URLs. The enriched URLs are clustered and form a pseudo document. Clustering is the process of grouping the data into classes or clusters. So that objects within a cluster have a high similarity in comparison to one another but are very dissimilar to object in other clusters. After constructing the pseudo document the web search results are restructured and CAP algorithm is used to evaluate the performance of restructured search results.

The main contribution of this paper as follows:

- Clustering feedback sessions is more efficient than clustering search results or clicked URLs directly.
- Combination of enriched URLs in the feedback session to form a pseudo document which reflect information needs of users.
- To achieving the user search goal efficiently and provide user friendly environment in web search application.

Search engine displays only user wanted information. Unwanted or related search results are placed at last. User search goal is achieved from top search results of the proposed system. Noisy data is completed eliminated by the proposed policies. Main objective is to provide user searching information in first fewer links of the search result. In existing system, keyword search was performed. In proposed, phrase search is performed. Phrase search provides efficiency and better search result.

The rest of the paper is organized as follows: The related work can be presented in section2. The system architecture is presented in section3. The proposed system and the evaluation is based on restructuring web search results is presented in section4. Implementation of proposed system is depicted in section5 and concludes the paper in section6.

2   RELATED WORK

Efficient algorithm is used for finding user friendly environment in web search application using data mining concepts. In order to finding the automatic goal identification based on human subject study by using user-click behavior and anchor link distribution, Zheng Lu et al. [2]. In [3] Barbara Poblete et al. to achieve better results using non-supervised tasks such as clustering and labeling.

In [4] clustering search results is an efficient way to organize search results using commercial search engine log data. In [5] Rosie Jones et al. generating a new query to replace a users original search query using query pair algorithm. In [6] Xiao Li et al. increasing the amount of training data based on click graph using semi-supervised click graphs.

In [7] Steven M.Beitzed et al. web query classification is used to improving retrival effectiveness and efficiency. In [8] Daxin Jiang et al. query suggestion by mining query patterns from search logs using query suggestion step and concept sequence suffix. In [9] Dou Shen et al. web query classification aims to classify web users queries, which are often short and ambiguous, into a set of target categories. In [10] Rosie Jones et al. analysis of typical timeouts used to divide query streams into sessions, and demonstration that they are less than optimal for this task

3   SYSTEM ARCHITECTURE

In this section, we represent the system architecture for user search goal evaluation. Initially we enter the user query. System architecture of the proposed system is depicted using the above figure Fig.1. Initially user enter query for which he/she search for in the Internet. Original search results will be displayed. Feedback session is formed from the original search results to monitor user logs. Next, noun phrase is generated using natural language processing. Pseudo documents contain noun phrase along with its term frequency. Based on term frequency, original search results is re-structured which provides user search goal on top links. Finally performance of original and re-structured search results is analyzed.
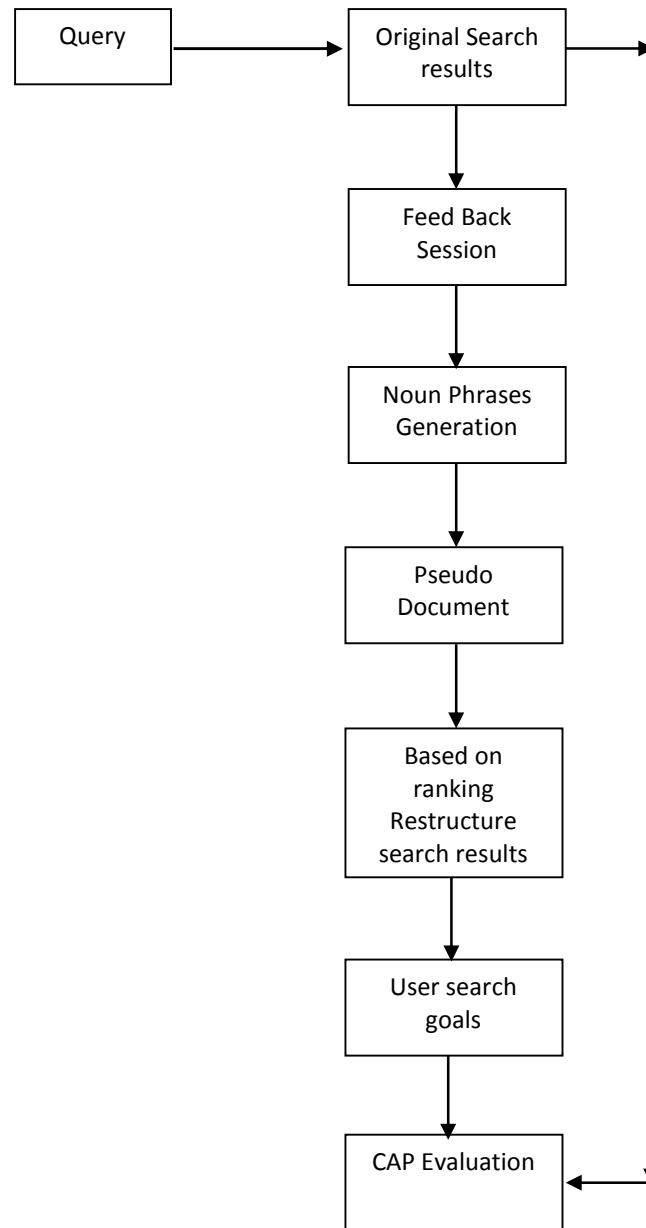
Fig. 1.   Architecture for User Search Goal Evaluation

4   PROPOSED SYSTEM

User's search goal is inferred efficiently by displaying his/her required information at top links among search results. Phrase search is performed instead of keyword search. Noun phrase for user query is framed using natural language processing. Newly framed nun phrases are searched in webpages available in Internet and web page containing exact phrase is placed at top link. Proposed system consists of variety of modules. First module is Construction of Feedback Session in which session is formed to indentify user search goal. Second module is Noun Phrase generation in which noun phrase of user entered query is generated. Third module is Clustering the Pseudo Documents which consists of noun phrase with its term frequency stating a particular phrase present in how many documents. Fourth module is Re-ranking the web search results in which search result containing user goal is placed in top links. Last module is Performance Analysis, in this module performance of existing system search results and proposed system search results is observed.

4.1 CONSTRUCTION OF FEEDBACK SESSION

Feedback session is considered by grouping set of URL's based on user click through logs. Feedback session consists of both clicked and unclicked URL's. In which clicked URL's tells about "what user care for". Unclicked URL's tells about "what user does not care about". Feedback session is used to focus on user's search goal. Feedback session provides user logs, using which usage of a particular document is determined.

Each feedback session can tell what a user requires and what he/she does not care about. Moreover, there are plenty of diverse feedback sessions in user click-through logs. Therefore, for inferring user search goals, it is more efficient to analyze the feedback sessions than to analyze the search results or clicked URLs directly. Multiple feedback sessions are considered to provide efficient information regarding user search goals.

Search results	Click sequence
www.thesun.co.uk/	0
www.nineplanets.org/sol.html	1
www.solarviews.com/eng/sun.htm	2
en.wikipedia.org/wiki/Sun	0
www.thesunmagazine.org/	0
www.space.com/sun/	0
en.wikipedia.org/wiki/The_Sun_(newspaper)	3
imagine.gsfc.nasa.gov/docs/science/know_l1/sun.html	0
www.nasa.gov/worldbook/sun_worldbook.html	0
www.enchantedlearning.com/subjects/astronomy/sun/	0

Fig. 2.   Feedback Sessions

Fig. 2. A feedback session in a single session. "0" in click sequence means "unclicked." All the 10 URLs construct a single session. The URLs in the rectangular box construct a feedback session.

4.2 NOUN PHRASE GENERATION

Noun phrase for user enter query is framed using natural language processing. Natural language processing is a technique used to indentify noun, verb, object etc in a sentence. Here user query is considered as sentence and it is analyzed to indentify the meaning of user query. Natural language separate noun, verb etc in a query and analyze the meaning of the user query. After analyzing the query, phrases are framed as following example. Phrases is an incomplete sentence which contain noun and verb. Synonyms for keywords in user entered query is also found and phrases are framed using those synonyms. Main purpose of finding synonyms is to cover user search goals efficiently. An example for noun phrase generation for a sample query is given below

Query: What is use of Internet?

Phrases:

      i.   Internet is used for
      ii.   Uses of Internet
      iii.   Benefits of Internet
      iv.   Advantages of Internet
      v.   Functions of Internet
      vi.   Purpose of Internet etc.,

4.3 CLUSTERING THE PSEUDO DOCUMENTS

Clustering groups user search content. K-mean algorithm is used to cluster the Pseudo documents. Since feedback sessions vary a lot for different click-throughs and queries, it is unsuitable to directly use feedback sessions for inferring user search goals. Some representation method is needed to describe feedback sessions in a more efficient and coherent way. There can be many kinds of feature representations of feedback sessions.

In proposed system the Pseudo document contains phrases framed along with its respective term frequency. Term frequency states, how many documents contain a specific phrase. Each and every noun phrase generated is searched thoroughly in web document available in World Wide Web. By searching, pick the web documents which contain above stated phrases exactly.

Term Frequency represents number of document a particular phrases occurs in various document is calculated

PSEUDO DOCUMENT FOR GENERATED PHRASES

PHRASES	NO-OF-DOCUMENT
1. Uses of Computer	10
2. Uses of a Computer	0
3. Uses Computer	2
4. Use of Computer	2
5. Use of a Computer	0
6. Purpose of computer	0
7. Purpose of a computer	0
8. Importance of computer	0
9. Importance of a computer	0
10. Computer use	2
11. Computer uses	1
12. Computer is used	2
13. Computer purpose	0
14. Computer Importance	0

Fig. 3.   Pseudo document

Figure 3, depicts the Pseudo document which consists of noun phrases and number of documents in which a particular noun phrases occurs exactly.

4.4 RE-RANKING THE WEB SEARCH RESULTS

Since search engines always return millions of search results, it is necessary to organize them to make it easier for users to find out what they want. In existing system mostly clicked URL is placed at top links, Restructuring web search results is an application of inferring user search goals. This paper introduce how to restructure web search results by inferred user search goals at first.

Based on Term frequency of the phrase in pseudo document re-structuring is performed. Phrase which occurs in maximum number of document is considered first. In that, Document containing exact phrase is displayed first. Likewise all the documents are re-structured accordingly. Related links are displayed at last.

4.5   PERFORMANCE ANALYSIS

In this module, performance of existing system and proposed system in measured. In existing system user search goal is inferred. It just states what user search for and what user does not search for. Hence existing system may display unwanted or noisy search results. In proposed system, user search goal is provided exactly at first link among the search results. Based on availability of user search goal at top link, efficiency is stated. User required information can be gathered with in some top search results.
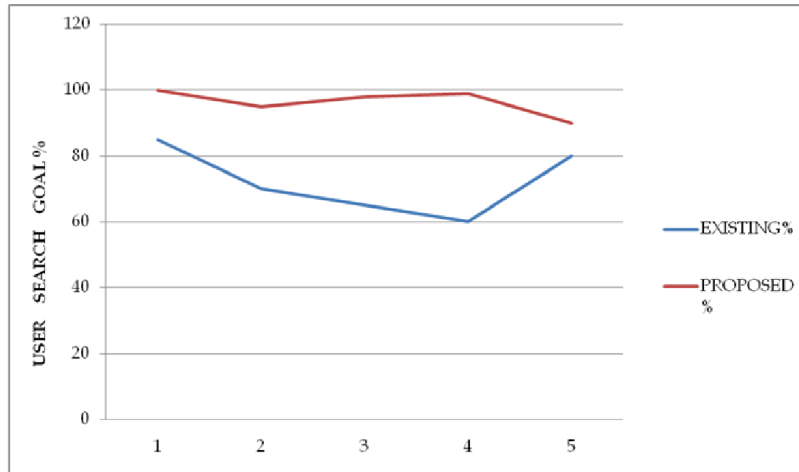

Fig. 4.   Performance Analysis

5   IMPLEMENTATION

Proposed system has been implemented using PHP & MySQL. PHP is used at front end and MySQL is used at back end to store various webpages from the Internet. Google Search Engine is considered as existing system, which displays users search goal in elegant manner. Usually Google search engine displays often clicked webpages at top link which consists of keywords in user query. Wikipedia content links mostly displayed at top links in Google search engine. Users goal may be obtained from forth coming links. User search goal at top links cannot be assured and noisy data may be displayed.

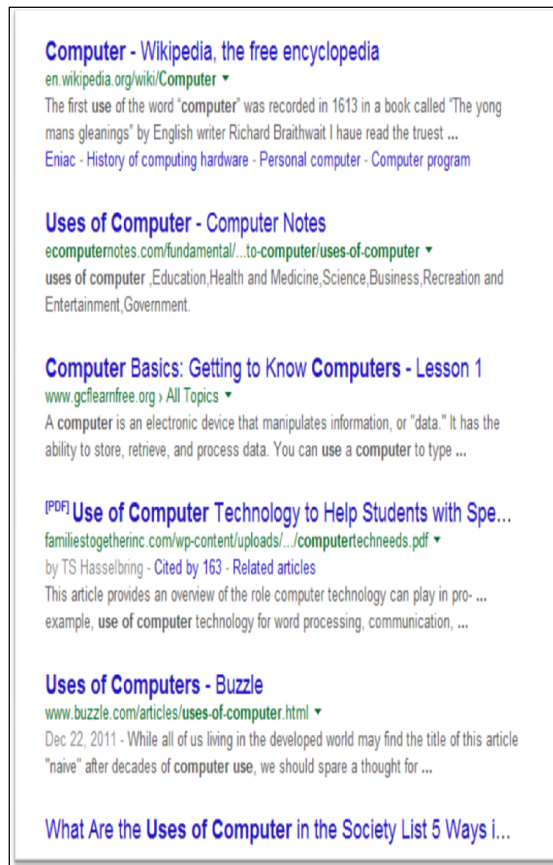Search results of Google search engine is depicted below.

Fig. 5.   Existing Search results

Proposed search engine provides user search goal at top links. This can be achieved by techniques like Feedback session, Noun Phrase, Pseudo documents and Restructuring search results based on phrase frequency. As stated about modules in Section 4, proposed has been designed. Feedback session provides user click logs. Noun phrase generation provides noun phrase of user entered query. Pseudo document provides frequency of noun phrases. Number of documents a particular noun phrase is occurred.

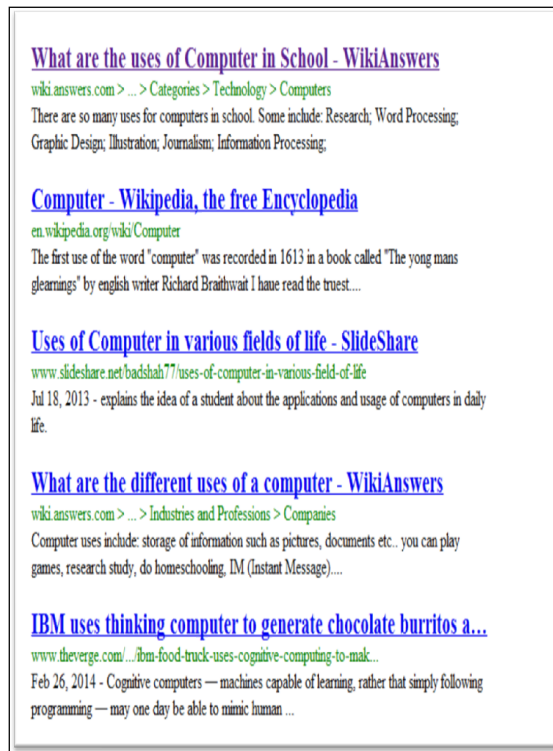Search results of Proposed search engine is depicted below.

Fig. 6.   Restructured Results

Figure 6, depicts the search result of the proposed system. In this, uses of computer in schools was placed in first link. All the displaying links contain only the uses of computer, not the general information about the computer. Hence users search goal is achieved efficiently. Proposed search engine lists only information needed by the user.

6   CONCLUSION

In this paper, 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 analyzed 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. Proposed system is used to infer user search goal efficiently. Existing system is used to categorize about what user care and does not care about. Re-structuring is performed in such a way. In proposed system, phrase search is performed for user query. Searching for phrase provides user requirement exactly. During Re-structuring Webpage containing exact phrase can be placed at top link. Unwanted or noisy search results will not be displayed in the proposed system. Thus user goal is inferred efficiently and exactly.

REFERENCES

[1]   Beitzel. S, Jensen. E, Chowdhury. A, and Frieder. O, (2007), "Varying Approaches to Topical Web Query Classification," Proc. 30th Ann.Int'l ACM SIGIR Conf. Research and Development (SIGIR '07), pp. 783-784.
[2]   Baeza-Yates. R, Hurtado. C, and Mendoza. M, (2004), "Query Recommendation Using Query Logs in Search Engines," Proc. Int'l Conf. Current Trends in Database Technology (EDBT'04), pp. 588-596.
[3]   Beeferman. D and Berger. A, (2000), "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.

[4]   Cao. H, Jiang. D, Pei. J, He. Q, Liao. Z, Chen. E, and Li. H, (2008), "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.

[5]   Chen. H and Dumais. S, (2000), "Bringing Order to the Web: Automatically Categorizing Search Results," Proc. SIGCHI Conf. Human Factors in Computing Systems (SIGCHI '00), pp. 145-152.

[6]   Huang C.K, Chien. L.-F, and Oyang. Y.-J, (2003), "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, issue no. 7, pp. 638-649.

[7]   Jones. R, Rey. B, Madani. O, and Greiner.W, (2006), "Generating Query Substitutions," Proc. 15th Int'l Conf. World Wide Web (WWW '06), pp. 387-396.

[8]   Joachims. T, Granka. L, Pang. B, Hembrooke. H, and Gay. G, (2005), "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.

[9]   Joachims. T, (2003), "Evaluating Retrieval Performance Using Clickthrough Data," Text Mining, J. Franke, G. Nakhaeizadeh, and I. Renz, eds., Physica/Springer Verlag,   pp. 79-96.

[10]   Jones. R and Klinkner .K.L, (2008), "Beyond the Session Timeout: Automatic Hierarchical Segmentation of Search Topics in Query Logs," Proc. 17th ACM Conf. Information and Knowledge Management (CIKM '08), pp. 699-708.

[11]   Lee. U, Liu. Z, and Cho.J, (2005), "Automatic Identification of User Goals in Web Search," Proc. 14th Int'l Conf. World Wide Web (WWW '05), pp. 391-400.

[12]   Li. X, Wang. Y.-Y, and Acero. A, (2008), "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.

[13]   Poblete. B and Ricardo. B.-Y, (2008), "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.

[14]   Shen. D, Sun. J, Yang. Q, and Chen. Z, (2006), "Building Bridges for Web Query Classification," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '06), pp. 131-138.

[15]   Wang. X and Zhai. C.-X, (2007), "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.

[16]   Zheng. Lu, Hongyuan Zha, Xiaokang Yang, (2006), "Automatic Query Log Identification of User Goals in Web Search" Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR' 06), pp. 135-148.