# Deduce User Search Progression with Feedback Session

B.Bazeer Ahamed[1] and T.Ramkumar[2]

[1] *Faculty in computer Science & Engineering, Sathyabama University, Chennai,Tamil Nadu,India*

[2] *Department of Computer Applications, AVC College of Engineering, Mayiladuthurai,Tamil Nadu,India*

**Abstract**

The web is a medium for accessing a great variety of information stored in various locations. As data on the web grows rapidly it leads to several problems such as increased difficulty of finding relevant information. When a user submits a query to the search engine, it must be able to retrieve information according to the user's intention. But search engine retrieves the list of pages ranked based on that similarity to the query. Sometimes the results are not according to users interests, because many relevant terms may be absent from queries and words may be ambiguous. Therefore, the results produced by the search engine are not satisfactory to fulfill the user query request. In order to solve this ambiguity, the proposed work is to discover the number of diverse user search goals for a query and represent each goal with some keywords automatically.

**Keywords** Clustering; HITS; Restructuring search results; Classified Typical Meticulousness; Feedback session.

## 1 Introduction

With the fast growth of the Web, a user can obtain abundant information easily by submitting a query to a search engine. Many existing search engines use keyword matching as the search mechanism, which usually causes the situation that a large number of non-relevant documents containing query terms are founded out, and the user will make strenuous efforts to browse these non-relevant documents. Thus, it is not simple to find out the real user goal from such short queries.

Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. These tools can include statistical models, mathematical algorithms and machine learning methods (algorithms that improve their performance automatically through experience such as neural networks or decision trees). Consequently, data mining consists of more than collecting and managing data, it also includes analysis and prediction. Data mining can be performed on data represented in quantitative, textual or multimedia forms. Data mining applications can use a variety of parameters to examine the data. They include association (patterns where one

event is connected to another event such as purchasing a pen and purchasing paper), sequence or path analysis (patterns where one event leads to another event such as birth of a child and purchasing dress), classification (identification of new patterns such as coincidences between duct tape purchases and plastic sheeting purchases), clustering (finding and visually documenting groups of previously unknown facts, such as geographic location and brand preferences) and forecasting (discovering patterns from which one can make reasonable predictions regarding future activities, such as prediction that people who join an athletic club may take exercise classes).

Information Retrieval (IR) is essentially a matter of deciding which documents in a collection should be retrieved to satisfy a user's need for information. The user's information need is represented by a query or profile, and contains one or more search terms, plus some additional informations. Hence, the retrieval decision is made by comparing the terms of the query with the index terms appearing in the document itself. The decision may be binary (retrieve/reject), or it may involve estimating the degree of relevance that the document has to the query. A stemming algorithm is a process of linguistic normalization, in which the variant forms of a word are reduced to a common form. It is important to appreciate that we use stemming with the intention of improving the performance of Information Retrieval systems [1]. Unfortunately, the words that appear in documents and in queries often have many morphological variants. Thus, pairs of terms such as "computing" and "computation" will not be recognized as equivalent without some form of natural language processing (NLP).

Improved Hypertext-Induced Topic Selection (HITS) algorithm is a very popular and effective algorithm to rank documents based on the link information among a set of documents. The algorithm presumes that a good hub is a document that points to many others, and a good authority is a document that many documents point to. Hubs and authorities exhibit a mutually reinforcing relationship: a better hub points at many good authorities, and a better authority is pointed to by many good hubs. To run the algorithm, we need to collect a base set, including a root set and its neighborhood, the in- and out-links of a document in the root set [2]. Because the HITS algorithm ranks documents only depending on the in-degree and out-degree of links, it will cause problems in some cases. For example, a) mutually reinforcing relationships between hosts and b) topic drift. Both problems can be solved or alleviated by adding weights to documents. The first problem can be solved by giving the documents from the same host much less weight, and the second problem can be alleviated by adding weights to edges based on text in the documents or their anchors. The simple modification to the HITS algorithm for the first problem achieves a remarkable better precision, while further precision can be obtained by adding content analysis [3]. Clustering is a

common descriptive task where one seeks to identify a finite set of categories or clusters to describe the data. Clustering is the process of identification of classes, also called clusters or groups, for a set of objects whose classes are unknown. The objects are so clustered that the intra class similarities are nearly maximized and the interclass similarities are minimized based on some criteria defined on the attributes of objects. Once the clusters are decided, the objects are labeled with their corresponding clusters, and common features of the objects in a cluster are summarized to form the class description. Agglomerative Hierarchical clustering techniques is used to cluster the pseudo - documents since Agglomerative Hierarchical Clustering is a classical clustering algorithm, originating analogically like k-means from the statistics domain. The main advantage of AHC is to create descriptions of clusters, removes redundant descriptions and attaching cluster to another one whose description is a subset of its description [4]. Improved HITS algorithm is used to rank cluster results relevant for a particular topic. The web results are restructured. Finally, we introduce user editable browser that allow the user to perform editing operations such as deletion and emphasis while browsing the search results.

The rest of the paper is organized as follows. Section 2 reviews various techniques for effective inferring user search goals and Generalized Feedback Session Techniques. Deduce User search progression is presented in Section 3. Experiments Measures in Section 4.Section 5 concludes the paper and shows our future directions on this topic.

## 2 Literature Review

### *2.1 Identifying User Goals from Web Search Results*

Yao-Sheng Chang et al propose a novel probabilistic inference model which effectively employs syntactic features to discover a variety of confined user goals by utilizing Web search results.[5] On the basis of analyzing the user goals in the viewpoint of Natural Language Processing (NLP) process. Assume that the user goal should be expressed with the form of a hidden sentence in his/her mind. In general, a typical sentence includes a subject (S), a verb (V), and an object (O). Also, assume that the subject of the hidden sentence in user mind is the user himself/herself and the combined pair of the verb and object is called VO - pair. On the basis of VO - pairs, potential user goal can be represented.

For example when users submit a query “Michael Jackson”, predict that the hidden sentence in the user mind is “I want to download Michael Jacksons music,” and the potential user goal is the VO - pairs “download music” (verb + object). The object can be regarded as the noun after the verb [5]. The user senses his own sentence according to his own mindset but awareness to the mechanism is much different.

### *2.2 An Overview of Personalization in Web Search*

The above mentioned technique has some drawbacks as follows: More challenges to adopt VO-pair classes to certain languages and time Consuming to identify VO-pair. In order to overcome these drawbacks Indu Chawla introduces Web search personalization algorithms to improve the Web search experience by using an individuals data e.g. user's domain of interest, preferences, query history, browser history etc . Using these factors they extract the results that are the most relevant to that individual. Personalization can be broadly categorized in two types: context oriented and individual oriented [6]. Context oriented personalization include factors like the nature of information available, the information currently being examined, the applications in use, when, and so on. Individual oriented personalization uses user interests, query history, browser history, pages visited etc [7].

Even though the Web search personalization algorithm improve the Web search experience by using an individual's data, but has some drawbacks as follows [8]: In Context oriented personalization expecting searchers to provide context information explicitly as part of their search is not ideal. Many users are simply unwilling to provide this type of additional information and even asking for it can lead to frustration certainly asking the user for anything close to personal information is liable to alienate many users because of privacy concerns [9]. And in the second option where the user can choose from among the categories provided, the problem is to know about the users' interest for displaying the categories to the user. Moreover, searchers often do not have enough knowledge available to them to explicitly express such context information even if they were inclined to do so [10]; In Individual oriented personalization a single user profile or model can contain a too large variety of different topics so that new queries can be incorrectly biased and also users are becoming more concerned about threats to privacy in the online environment.

### *2.3 Query recommendation using query logs in search engines*

Baeza et al [11] present an algorithm to recommend related queries to a query submitted to a search engine. The related queries are based in previously issued queries, and can be issued by the user to the search engine to tune or redirect the search process. The method proposed is based on a query clustering process in which groups of semantically similar queries are identified. The clustering process uses the content of historical preferences of users registered in the query log of the search engine [12]. The method not only discovers the related queries, but also ranks them according to a relevance criterion. Ranking the queries according to two criteria: First one by the similarity of the queries to the input query (query submitted to the search engine).trail by the next one by the measures, how much the answers of the query have attracted the attention of users.The combination

of measures (a) and (b) defines the interest of a recommended query.

### *2.4 Automatic Identification Of User Goals In Web Search*

Uichin Lee et al [13] study whether and how to identify the user goal automatically without any explicit feedback from the user. Two types of features for the goal identification task,one is Past user-click behavior and another one is Anchor-link distribution.

First feature is based on the intuition that the user's goal for a given query may be learned from how users in the past have interacted with the returned results for this query. If the goal of a query is navigational, then in the past users should have mostly clicked on a single Website corresponding to the one they have in mind. On the other hand, if the goal is informational, in the past users should have clicked on many results related to the query. Thus by observing how the results for a particular query have been clicked so far, and easily tell whether the current user who issues that query has a navigational or an informational goal.

The Anchor link distribution is used to find the destination of the links with the same anchor text as the query. For example, for a navigational query pub med, a single authoritative Website exists (which is www.ncbi.nlm.nih.gov)[9]. As a result, if extract all the HTML links with the anchor text pub med, to find that a dominating portion of these links point to that single Website. On the other hand, for an informational query hidden markov model, because of lack of a single authoritative site, can expect that the links with the anchor text hidden markov model point to a number of different destinations.

### *2.5 Learn From Web Search Logs To Organize Search Results*

Xuanhui Wang et al propose a different strategy for partitioning search results, which addresses these two deficiencies through imposing a user-oriented partitioning of the search results.[10] Learn "interesting aspects" of similar topics from search logs and organize search results based on these "interesting aspects". Finally Generate more meaningful cluster labels using past query words entered by users.

To know what the users are really interested in given this query, first retrieve its past similar queries in preprocessed history data collection. Based on the similarity scores, we rank all the documents in history data set. The top ranked documents provide us a working set to learn the aspects that users are usually interested in. Each document in history data set corresponds to a past query, and thus the top ranked documents correspond to q's related past queries.

### *2.6 Query-Sets: Using Implicit Feedback And Query Patterns To Organize Web Documents*

Barbara Poblete et al present a new document representation model based on implicit user feedback obtained from search engine queries.[14] The main objective

of this model is to achieve better results in non-supervised tasks, such as clustering and labeling, through the incorporation of usage data obtained from search engine queries. This type of model allows user to discover the motivations of users when visiting a certain document. The query document model reduces the feature space dimensions considerably, because the number of terms in the query vocabulary is smaller than that of the entire website collection. This model is very similar to the vector model, with the only difference that instead of using a weighted set of keywords as vector features, we will use a weighted set of query terms.

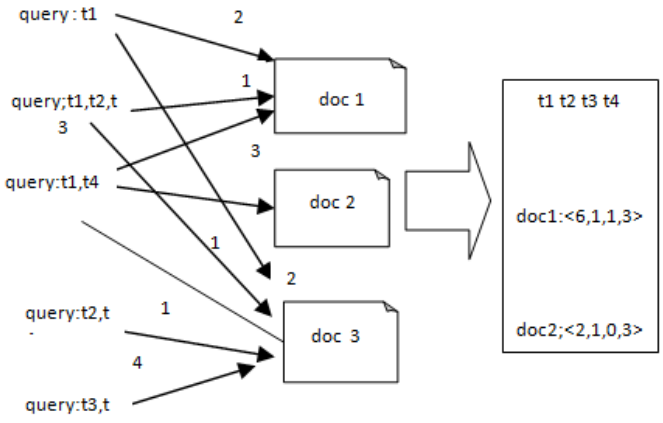


**Fig.** 1 Example of the Query Document Representation.

*2.7 User Intent Based Searching*

Haibo Yu et al propose a user intent based searching mechanism (UIBS )in order to enable precise discovery of a Web site (for navigational searching) as well as aggregating of their information (for informational searching) and performing further activities (for transactional searching).[15] This mechanism includes three main components: a Web site capability description interface for the explicit description of a Web site's capability, a search interface that enables explicitly describing user's search intent and requirements and their relevant responses, and a query engine which provides relevant search results based on user's intent and requirements. In a UIBS enabled user, a UIBS client should be installed which can issue UIBS search requests based on user inputs or other sources of user intents, preferences and query [9]. The client should also process the received UIBS results, and presents the processed result to the user or other applications. It explicitly describes the general information of the Web site and the links to published content as well as Web services that the Web site can provide. The

WSCD works like a site map and presents all information that the Web site wants to publish.

### *2.8 Generalized Feedback Session Techniques*

Zheng Lu et al extended the previously discussed algorithm in an effective manner. He focused on user search goals by organizing search results by aspect learned from user click through logs.[16, 17]

In the feedback sessions related to the given query will be extracted from user click- through logs. 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. The clicked URLs tell what users require and the unclicked URLs reflect what users do not care about [18]. Then, map each feedback sessions to pseudo-documents using keywords which can efficiently reflect user information needs.

By using k- means clustering techniques to cluster the pseudo-documents to infer user search goals [19]. Finally restructure the web search results inferring user search goals. Fig.2 represents the data flow diagram for the existing system.

But there is some confines that we are focused in obtainable System i.e. Generate noisy and redundant search results problem and Cluster labels generated are not informative enough to allow the user to identify the right search result. Do not generate the search results if the query has been entered for the first time.

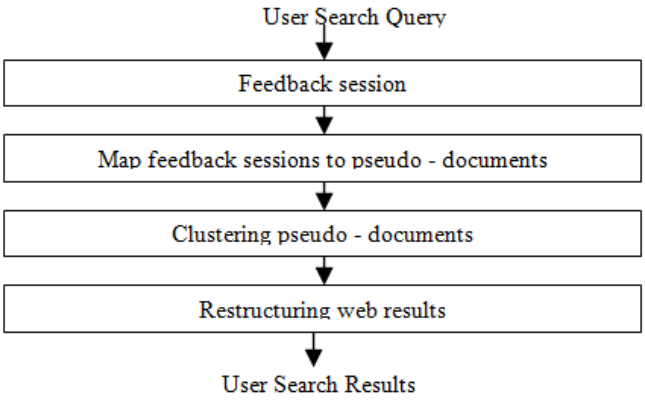


**Fig.** 2 Dataflow diagram for the Generalized Feedback Session Techniques.

## 3 Deduce User Search Progession

With the fast growth of the Web, a user can obtain abundant information easily by submitting a query to a search engine. Many existing search engines use keyword matching as the search mechanism, which usually causes the situation that

a large number of non-relevant documents containing query terms are founded out, and the user will make strenuous efforts to browse these non-relevant documents. Thus, it is not simple to find out the real user goal from such short queries. A variety of ranking algorithms have been proposed and used in many Web search engines. People search for information using search engines. Search engines, however, cannot always return good ranked search results which satisfy user's search intentions adequately. Hence, it is difficult to recognize users' search intentions just by analyzing their input queries. The search results sometimes do not correspond to the user's search intentions because of this diversity. In this case, the user must check the search results sequentially until he obtains sufficient information from the linked pages from the page containing the search results.

The query contains parts of user general verbal communication and special characters which are not required for analysis as they do not truly reflect the relevance of a search result. If this query is used for analysis, it may give inconsistent and inaccurate results. Therefore the user query will be pre-processed to identify the root words. The feedback sessions has been introduced to infer user search goals for a query. Then, map each feedback sessions to pseudo-documents using keywords which can efficiently reflect user information needs and relate to the given query will be extracted from user click- through logs. 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. The clicked URLs tell what users require and the unclicked URLs reflect what users do not care about. In order to apply the evaluation method to large-scale data, the single sessions in user click-through logs are used to minimize manual work. Because from user click-through logs, we can get implicit relevance feedbacks, namely "clicked" means relevant and "unclicked" means irrelevant. A possible evaluation criterion is the typical meticulousness (TM) which evaluates according to user implicit feedbacks. TM is the average of precisions computed at the point of each relevant document in the ranked sequence However, TM is not suitable for evaluating the restructured or clustered searching results. Therefore we introduce Classified Typical Meticulousness (CTM) System has been initiated to evaluate the performance of the restructured web search results. Where the CTM of the class including more clicks namely confer is calculated, CTM selects the TM of the class that user is interested in (i.e., with the most clicks/confer).

Then, map each feedback sessions to pseudo-documents using keywords which can efficiently reflect user information needs. K-mean clustering is too simple and it performs poorly for large set of data. And also it requires prior knowledge of number of clusters to be generated. Therefore, instead of using k-mean clustering we had used Agglomerative Hierarchical Clustering to cluster the pseudo - documents.

## 4 Experiments And Measures

The query submitted by the user contains parts of speech and special characters which are not required for analysis as they do not truly reflect the relevance of a search result. If this query is used for analysis, it may give inconsistent and inaccurate results. Therefore, the user query will be pre-processed to identify the root words.

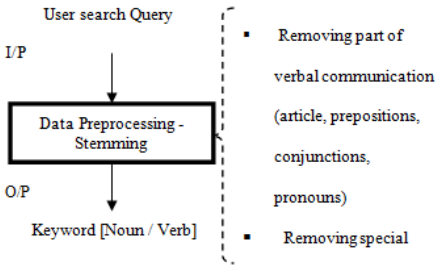


**Fig.** 3 Data Preprocessing

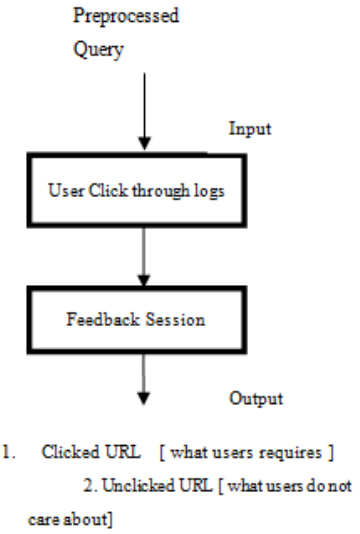


**Fig.** 4 General Feedback Session progression.

Feedback session consists of both clicked and unclicked URLs and ends with the last URL that was clicked in a single session. The clicked URLs tell what users require and the unclicked URLs reflect what users do not care about. Table

1 shows a feedback session with lists of 10 search results of the given query "computer", where "0" in the click sequence represents the "unclicked URL" and remaining represents "clicked URL".

The pseudo-document can be used to infer user search goals. The mapping of feedback sessions into a pseudo-document includes two steps. They are as follows: First enrich the URLs with additional textual contents by extracting the titles and snippets of the returned URLs appearing in the feedback session. In this way, each URL in a feedback session is represented by a small text paragraph that consists of its title and snippet. Then, some textual processes are implemented to those text paragraphs, such as transforming all the letters to lowercases, stemming and removing stop words.

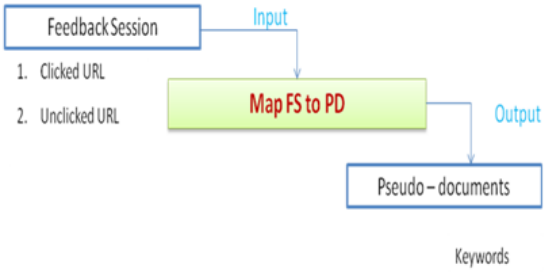


**Fig.** 5 Map Feedback Sessions to Pseudo-Documents.

Improved HITS (Hyperlink - Induced Topic Search) algorithm is applied to the clusters generated by Agglomerative Hierarchical Clustering to rank cluster results relevant for a particular topic. Ranking is done by assigning relevance weight to the cluster results. Restructure the web search results based on result generated by Improved HITS algorithm.

Each webpage will be treated as a node, with hyperlinks treated as directed links from one node to another. Each node i is assigned an authority score a(i) and hub score h(i). Given a directed graph, the authority and hub score is defined as follows:

a(i) = The sum of the hub scores of the nodes pointing to node i.

h(i) = The sum of the authority scores of the nodes that node i is pointing to.

The higher a node's authority/hub score is, the better authority/hub is. This makes intuitive sense; a node is a good authority if good hubs are pointing to it, and a node is a good hub if it is pointing to good authorities. These authority and hub updates can be described through matrix notation. Suppose that we define matrices A, U and V as follows:A = the adjacency matrix of the graph. That is A(ij) = 1 if node i has a directed link to node j, and 0 otherwise.

U = a column matrix containing the hub score of all of the nodes.

V = a column matrix containing the authority score of all of the nodes.
With these definitions, it follows that for any iteration k:

$$U(k) = A * V(k) and V(k) = transpose(A) * U(k)$$

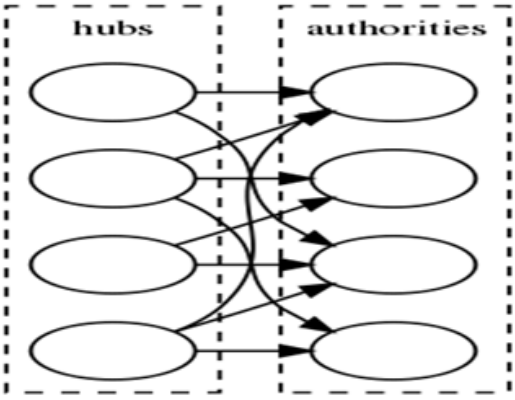


**Fig.** 6 Hubs are pages that link to authorities.

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. Restructuring web search results is an application of inferring user search goals. The inferred user search goals are represented by the vectors and the feature representation of each URL in the search results can be computed. Then, categorize each URL into a cluster centered by the inferred search goals. Perform categorization by choosing the smallest distance between the URL vector and user-search-goal vectors. By this way, the search results can be restructured according to the inferred user search goals.
The adjacency matrix of the graph

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} A^t = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}$$

Assume the initial hub weight vector

$$u = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

Compute the authority weight vector by:

$$v = A^t \cdot u = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} \bullet \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$$

Then, the updated hub weight

$$v = A \cdot u = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \bullet \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}$$

Editable Browser for Selecting the Results & Re - ranking the result based on user editing strategies. We enhanced an editing operation such as deletion and emphasis that can be employed while users are browsing Web search results. This system enables users to edit any portion of the page of Web search results at any time while searching. Our system detects user's search intentions from the editing operation. Then our work propagates user's search intentions based on their editing operation to all of the Web search results. This system guesses the user's search intentions, for example assuming, "This user does not want this kind of the result", if the user deletes a part of the search results, or "This user wants this kind of the results more", if the user emphasizes a part of the search results. After guessing the user's search intentions, our system re-ranks the search results according to the intention and shows the re-ranked results to the user. In this way, the user can easily obtain optimized search results.

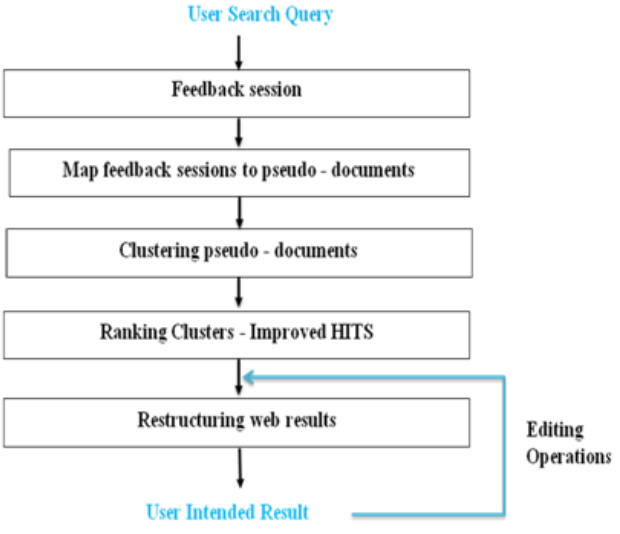


**Fig.** 7 Editing Browser Processing.


**Fig.** 8 Restructured Outputs with Editable Options.


**Fig.** 9 Restructured Result after Deletion Operation.

| Anon ID | Query | Query Time | Item Rank | URL |
|---|---|---|---|---|
| 7869 | Sun | 2014:05:17 10:57:05 | 1 | http://solarsystem.nasa.gov/planets/profile.cfm?Object=Sun |
| 1608 | Sun | 2014:05:17 10:58:13 | 2 | http://science.nationalgeographic.co.in/science/space/solar-system/sun-article/ |
| 8795 | Sun | null | 1 | http://sunnewsonline.com/new/ |
| 3786 | Sun | null | 4 | https://www.google.co.in/search?q=sun&newwindow=1&tbm=isch&tbo=u&source=univ&sa=X&ei=EptfUo-6JpDtrQf0poHIDA&ved=0CIwBEIke |
| 2820 | Sun | null | 4 | http://www.holidayphilippinesblog.com/filipino-culture-food/pinoy-power-breakfast/attachment/sunrise/ |
| 7005 | Moon | null | 5 | http://en.wikipedia.org/wiki/Moon |
| 8321 | Moon | null | 1 | https://solarsystem.nasa.gov/planets/profile.cfm?Object=Moon |
| 1807 | Moon | null | 4 | http://moon.com/ |
| 7422 | college | null | 1 | http://www.pabcet.com/ |
| 8503 | sun | null | 2 | http://www.vancouversun.com/index.html |
| 6470 | moon | null | 1 | http://www.space.com/55-earths-moon-formation-composition-and-orbit.html |
| 5325 | moon | null | 2 | http://www.bibliotecapleyades.net/luna/esp_luna_16.htm |

**Fig.** 10 Feedback session for the Query "The Sun".

User editable browser has been introduced to perform editing operations such as deletion and emphasis while browsing the search results. When the user deletes a part of the search result, the system degrades search results which include the deleted term or sentence. When the user emphasizes a part of the search result, the system upgrades the search results which include the emphasized term or sentence. The system re-ranks the search results according to the user intention and shows the re-ranked results to the user.

*Deletion Operation:*

Deletion is an operation that indicates what types of search results the user does not want to obtain from the system. consider the sample class shown in the Fig.7 , if the user does not want the $36rd$ link the user uses the deletion operation to remove the link from the generated search result.

*Performance Evaluation Based on Restructured Web Search Results*

Each URL in the click session is categorized into one class, we introduce the Typical Meticulousness(TM) process for calculating the user click through log. TM will always be the highest value namely 1 no matter whether users have so many search goals or not. Therefore, there should be a risk to avoid classifying search results into too many classes by error. So, we can further extend TM by introducing the above Risk and propose a new criterion called "classified TM" it is calculated by

$$CTM = (\frac{1}{y}\sum_{y}^{r=1}\frac{Rr}{r}) \times (1 - d_{ij}/C_Y^2)^x$$

Where Y is the number of relevant (or clicked) documents in the retrieved ones, r is the rank, Rr is the number of relevant retrieved documents of rank r and x is used to adjust the influence of Risk on CTM which can be learned from training data. We select 10 queries and empirically decide the number of user search goals of these queries. Then, we cluster the feedback sessions and restructure the search results with inferred user search goals. We tune the parameter x to make

CTM the height. Based on the above process, the optimal x is from 0.6 to 0.8 for the 10 queries the mean and the variances of the optimal x are 0.697 and 0.005, respectively. Thus, we set x to be 0.7.

**Table** 1 User Editable Browser Comparisons with other methods for 100 ambiguous queries

| Method | Probabilistic inference model | personalizati on algorithms | Query recommender algorithm | identification of user goals | Generalized feedback session | user editable browser |
|---|---|---|---|---|---|---|
| I | 0.7124 | 0.787 | 0.755 | 0.562 | 0.7173 | 0.8055 |
| II | 0.7911 | 0.625 | 0.584 | 0.742 | 0.8031 | 0.8875 |
| III | 0.801 | 0.749 | 0.611 | 0.632 | 0.9895 | 1 |
| IV | 0.5391 | 0.6245 | 0.512 | 0.654 | 0.6646 | 0.6787 |
| V | 0.632 | 0.745 | 0.741 | 0.524 | 0.7845 | 0.833 |

Sample Manual Calculation for Single Class Result:

$1.CTM = 1/3(1/1 + 2/3 + 3/4) * (1-0)^{0.7} = 0.8055$

$2.CTM = 1/4(1/1 + 2/2 + 3/4 + 4/5) * (1-0)^{0.7} = 0.8875$

$3.CTM = 1/2(1/1 + 2/2) * (1-0)^{0.7} = 1$

$4.CTM = 1/5(1/1 + 2/3 + 3/5 + 4/7 + 5/9) * (1-0)^{0.7} = 0.6787$

$5.CTM = 1/2(1/1 + 2/3) * (1-0)^{0.7} = 0.833$

In order to demonstrate that when inferring user search goals, clustering our proposed feedback sessions are more efficient than other clustering search results and clicked URLs directly In order to further compare our method with the existing method, we test the 100 most ambiguous queries such as “Apple”, “The Sun”, “car”, “mobile”, “college”, “earth” and so on. CTM has the highest mean average which is significantly higher than existing method.

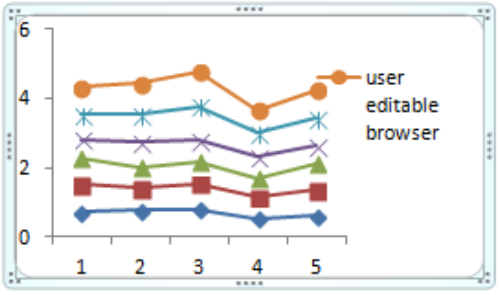


**Fig.** 11 CTM VS. Mean average of other methods.

## 5 Conclusion

This paper aims to discover the number of diverse user search goals for a given query and predict with some keywords automatically. First, the given user query is pre-processed to find the root word. Then, a feedback session has been intro-

duced to infer user search goals for a query. 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. Each feedback sessions are mapped to pseudo-documents using keywords which can efficiently reflect user information needs. By using Agglomerative Hierarchical clustering techniques to cluster the pseudo - documents. Improved HITS (Hyperlink - Induced Topic Search) algorithm is used to rank cluster results relevant for a particular topic. We introduced Classified Typical Meticulousness(TM) process for calculating the user click through log, the web results are restructured. Finally, user editable browser has been introduced allow the user to perform editing operations such as deletion and emphasis while browsing the search results. This paper can be extended to support query recommendation there by suggesting queries that can helps the user to form queries more precisely.

**Acknowledgement**

The authors would like to express their sincere thanks to the editorial board members and anonymous reviewers for providing their valuable comments.

The first author would like to express their thanks to management and other authorities for providing an feasible environment to carry out the research work successfully.

**References**

[1] Beeferman. D and Berger. A (2000), "Agglomerative clustering Of a search engine query log", *ACM SIGKDD Proc. Sixth Int'l Conf. Knowledge Discovery and Data Mining(SIGKDD '00)*, pp.407-416.

[2] Joachims .T Text Mining, J. Franke, G. Nakhaeizadeh and Renz, eds.(2003), "Evaluating retrieval performance using click-through data", *Physica/Springer Verlag*, pp.79-96.

[3] 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.

[4] Joachims .T,(2002), "Optimizing search engines using click through data,",Proc. *Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '02)*, pp.133-142.

[5] Yao-Sheng Chang, Kuan-Yu He, Scott Yu and Wen-Hsiang Lu (2006), "Identifying User Goals From Web Search Results",*In IEEE.*

[6] Indu Chawla,(2010), "An overview of personalization in web search", *In IEEE.*

[7] 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.

[8] Chen .Hand 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.

[9] Zeng .H.-J, He. Q.-C, Chen .Z, Ma. W.-Y and Ma. J.(2004), "Learning to cluster web search results",Proc. $27^{th}$ Ann. *Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '04)*, pp.210-217.

[10] Shital C. Patil, Prof. R.R. Keole, (2013), "Web usage mining and webcontent mining C a combine approach for enhancing search result delivery", Vol.3, No.10.

[11] Baeza-Yates .R, Hurtado. C and Mendoza. M, (2004), "Query recommendation using query logs in search engines", *The Journal of supercomputing(EDBT '04)*, pp.588-596.

[12] Joachims .T,"Optimizing search engines using click through data".

[13] 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.

[14] 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.

[15] Haibo Yu , Tsunenori Mine and Makoto Amamiya(2011), "Towards user intent based searching", *2011 International Joint Conference of IEEE TrustCom-11/IEEE ICESS-11/FCST-11.*

[16] 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.

[17] Zheng Lu, Hongyuan Zha, Xiaokang Yang, Weiyao Lin and Zhaohui Zhengr, (2013), "A new algorithm for inferring user search goals with feedback sessions,", *IEEE Transactions on Knowledge and Data Engineering*,Vol.25, No.3.

[18] Takehiro Yamamoto, Satoshi Nakmura and Kutsumi Tanaka, "An editable browser for reranking web search results".

[19] 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. *or Information Science and Technology*, Vol.54, No.7, pp. 638-649.

**Corresponding author**
B.Bazeer Ahamed can be contacted at:bazeerahamed@gmail.com