

# Web Graph based query suggestion for Text and Image

S.DHIVYA<sup>1</sup>, T.A.SANGEETHA<sup>2</sup>

Assistant Professor, Kongu Arts and Science College, Erode, Tamil Nadu, India<sup>1,2</sup>

**Abstract:** The way of getting different and explosive growth of Web information effectively is more critical. The first challenge is that it is not easy to recommend latent semantically relevant results to users. In the Query Suggestion, there are several outstanding issues that can potentially degrade the quality of the recommendations. The first one is the ambiguity which commonly exists in the natural language. Queries containing ambiguous terms may confuse the algorithms which do not satisfy the information needs of users. This research cope with the abbreviation based Query suggestion to provide the recommendation for user query, based on the short term query word. It also proposes K-means clustering Algorithm for generating user group based query matrix. In previous search query words are also taken into query suggestion calculation. In this previous query based suggestion method not consider the Zero Visit count URLs. But in normal find query suggestion also consider the Zero visit count URLs. The Graph query suggestion construct the Bipartite directed graph, based on the given queries and the access URLs. In addition Find Query Suggestion is constructed for domain wise. In addition, to add the recommendation user details play an important role in online query suggestion calculation. The view recommendation user URL access details to provide the user URL access details, it is used to the user can easily identify which one is the best URL for the given query. At run time user URL access matrix to provide all users URL access details and user group based query matrix which is used to give the user access details by group wise. In addition the image can be searched based on the suggested query words. In this proposed method the images can be added under the category wise and image search results returns all the images which is added under the category.

**Keywords:** visit count, query suggestion, URL, image, short term query word.

## 1. INTRODUCTION

Data Mining is the Process of finding relevant and useful information from databases. Although data mining is still in its infancy, companies in a wide range of industries - including retail, finance, health care, manufacturing transportation, and aerospace - are already using data mining tools and techniques to take advantage of historical data. By using pattern recognition technologies and statistical and mathematical techniques to sift through warehoused information, data mining helps analysts recognize significant facts, relationships, trends, patterns, exceptions. Clustering is the task of grouping a set of objects. Objects in the same group are called as a cluster. It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics. The partition clustering techniques partition the databases into a predefined number of clusters. They attempt to determine k partitions that optimize a certain criterion function. The partition clustering algorithms are of two types: K-means algorithms and K-medoid algorithms.

The Hierarchical Clustering techniques do a sequence of partitions, in which each partition is nested into the next partition in the sequence. It creates a hierarchy of clusters from small to big or big to small. In this dissertation using the concept of partition Clustering for generating group based query matrix which is used to give the user access details by group wise. K-Means is a simple learning algorithm for clustering analysis. The goal of K-Means algorithm is to find the best division of n entities in k

groups, so that the total distance between the group's members and its corresponding centroid, representative of the group, is minimized. Formally, the goal is to partition the n entities into k sets  $S_i$ ,  $i=1, 2, \dots, k$  in order to minimize the within-cluster sum of squares. In k-Means algorithm, only find a local optimum, and is commonly run multiple times with different random initializations. Variations of k-means often include such optimizations as choosing the best of multiple runs, but also restricting the centroid to members of the data set, choosing medians, choosing the initial centers less randomly. The algorithm prefers clusters of approximately similar size, as they will always assign an object to the nearest centroid. This often leads to incorrectly cut borders in between of clusters

The Web-Mining is the process of discover useful information or knowledge from data on the web. Based on the primary kinds of data used in the mining process. Web content mining extracts useful information from page contents. Web content consists of several types of data such as textual, image, audio, video, metadata, as well as hyperlinks. Most of the efforts on the web content mining are concentrated on the text and hypertext contents. The textual parts of the web content data consist of unstructured data such as free texts, semi-structured data such as HTML documents, and more structured data such as in the tables or database-generated HTML pages. Most of the web content data is unstructured, free text data. The techniques of text mining can be directly employed for web content mining. Web structure mining discovers knowledge from the hyperlink structure of the Web. Web usage mining extract knowledge from the usage patterns.

## 2. LITERATURE SURVEY

### 2.1 Improving Web Search Rank

According to Eugene Agichtein, Eric Brill and Susan Dumais the incorporating user behavior data can significantly improve ordering of top results in real web search setting [1], and examined alternatives for incorporating feedback into the ranking process and explore the contributions of user feedback compared to other common web search features. The reported results of a large scale evaluation over 3,000 queries and 12 million user interactions with a popular web search engine and showed that incorporating implicit feedback can augment other features, improving the accuracy of a competitive web search ranking algorithms by as much as 31% relative to the original performance. Millions of users interact with search engines daily and issue queries, follow some of the links in the results, click on ads, spend time on pages, reformulate their queries, and perform other actions. These interactions can serve as a valuable source of information for tuning and improving web search result ranking and can compliment more costly explicit judgments. Implicit relevance feedback for ranking and personalization has become an active area of research. In "Optimizing Search Engines Using Click through Data", according to Joachims exploring implicit feedback in controlled environments has shown the value of incorporating implicit feedback into the ranking process [2]. The motivation for that work is to understand how implicit feedback can be used in a large-scale operational environment to improve retrieval. It compares and compliment evidence from page content, anchor text, or link-based features such as in links or Page Rank. It is intuitive that user interactions with the web search engine should reveal at least some information that could be used for ranking, estimating user preferences in real web search settings is a challenging problem. This paper explores whether implicit feedback can be helpful in realistic environments, where user feedback can be noisy (or adversarial) and a web search engine already uses hundreds of features and is heavily tuned. To this end, explore different approaches for ranking web search results using real user behavior obtained as part of normal interactions with the web search engine.

The specific contributions of that paper include:

1. Analysis of alternatives for incorporating user behavior into web search ranking.
2. An application of a robust implicit feedback model derived from mining millions of user interactions with a major web search engine.
3. A large scale evaluation over real user queries and search results, showing significant improvements derived from incorporating user feedback.

### 2.2 Query Expansion by Mining User Logs

According to Ji-Rong Wen, Jian-Yun Nie, and Wei-Ying Ma, in recent years, the user has been witnessing the explosive growth of information on the World Wide Web [3]. People are relying more and more on the Web for their diverse needs of information. The Web is an information hotpot where innumerable authors have created and are creating their Web sites independently. There is an acute

requirement for search engine technology to help users exploit such an extremely valuable resource. Despite the fact that keywords are not always good descriptors of contents, most existing search engines still rely solely on keyword-matching to determine the answers. Users usually describe their information needs by a few keywords in their queries, which are likely to be different from those index terms of the documents on the Web. This problem is general in Information Retrieval (IR) systems and has been documented before the popularization of the Web: New or intermittent users often use the wrong words and fail to get the actions or information [4]. As a consequence, in many cases, the documents returned by search engines are not relevant to the user information need. This raises a fundamental problem of term mismatch in information retrieval, which is also one of the key factors that affect the precision of the search engines. Very short queries submitted to search engines on the Web amplify this problem. Many important words or terms may be missing from the queries.

To solve this problem, investigated the utilization of query expansion techniques to help users formulate better queries. Query expansion involves supplementing the original query with additional words and phrases. There are two key aspects in any query expansion technique: the source from which expansion terms are selected and the method to weight and integrate expansion terms. It not says the concept of abbreviation based query suggestion. In this dissertation the abbreviation based Query suggestion to provide the recommendation for user query suggestion based on the short term query word and it also improve the search result.

### 2.3 Collaborative Filtering

According to Collaborative filtering is valuable in e-commerce and for direct recommendations for music, movies, news etc [5]. But today's systems use centralized databases and have several disadvantages, including privacy risks. It move toward ubiquitous computing, there is a great potential for individuals to share all kinds of information about places and things to do, see and buy, but the privacy risks are severe. It introduced a peer-to-peer protocol for collaborative filtering which protects the privacy of individual data. A second contribution of this paper is a new collaborative filtering algorithm based on factor analysis which appears to be the most accurate method for CF to date. The new algorithm has other advantages in speed and storage over previous algorithms. It is based on a careful probabilistic model of user choice, and on a probabilistically sound approach to dealing with missing data and experiments on several test datasets show that the algorithm is more accurate than previously reported methods, and the improvements increase with the sparseness of the dataset.

Finally, factor analysis with privacy is applicable to other kinds of statistical analyses of survey or questionnaire data scientists (e.g. web surveys or questionnaires). Work builds on the paper [6] that introduced collaborative filtering with privacy. That paper described a protocol for encrypting data and aggregating it and publishing the result. The first contribution of the this paper is to

introduce the privacy-preserving CF scheme to the SI-GIR community, and give a simplified description of the protocol. The second contribution is a new collaborative filtering scheme preserving privacy which also appears to be the most accurate CF scheme developed so far. The third contribution is a set of experiments using the new method, characterizing its accuracy, speed and storage requirements. The genesis of this work is the disturbing privacy affronts that have followed from widespread computerization and ecommerce. The fragility of the digital marketplace has put many e-companies in bankruptcy. Finally, as collaborative filtering techniques become more widespread and constitute an important part of the process of diffusion of innovations through society [7]. Diffusion in real societies relies on both homophilous diffusion where recommendations come from others like the inquirer, and heterophilous diffusion, where the inquirer explicitly seeks recommendations from individuals not like them (typically from experts or earlier adopters). Today's collaborative filtering systems support homophilous diffusion well, but offer no plausible version of heterophily. One goal of this work is to support recommendations from designated communities, such as students, gardeners, surfers, antique collectors, elder citizens etc. To design a system that would allow nonmembers of those communities to gain recommendations from them, just as in natural heterophilous diffusion, this pushes two of their design criteria: (i) the protocol should protect the privacy of individuals in the community while allowing individuals both inside and outside the community to gain recommendations from it (ii) a peer to peer design is very desirable so that it is as easy as possible for communities to construct electronic CF groups, even for communities with limited resources (e.g. ability to afford, configure and run a server).

#### **2.4 Query Suggestion**

According to Huanhuan Cao<sup>1</sup> Daxin Jiang Jian Pei Qi He Zhen Liao Enhong Chen<sup>1</sup> Hang L, Query suggestion plays an important role in improving the usability of search engines [8]. Although some recently proposed methods can make meaningful query suggestions by mining query patterns from search logs, none of them are context-aware. It do not take into account the immediately preceding queries as context in query suggestion. The use of novel context-aware query suggestion approach which is in two steps. In the online model-learning step, to address data sparseness, queries are summarized into concepts by clustering a click-through bipartite. Then, from session data a concept sequence suffix tree is constructed as the query suggestion model. In the online query suggestion step, a user's search context is captured by mapping the query sequence submitted by the user to a sequence of concepts. By looking up the context in the concept sequence suffix tree and approach suggests queries to the user in a context-aware manner. In this test the approach on a large-scale search log of a commercial search engine containing 1 : 8 billion search queries, 2 : 6 billion clicks, and 840 million query sessions. The experimental results

clearly show that the approach out performs two baseline methods in both coverage and quality of suggestions. The effectiveness of information retrieval from the web largely depends on whether users can issue queries to search engines, which properly describe their information needs. Writing queries is never easy, because usually queries are short (one or two words on average) [9] and words are am-biguous [10]. To make the problem even more complicated, different search engines may respond differently to the same query. Therefore, there is no "standard" or "optimal" way to issue queries to search engines, and it is well recognized that query formulation is a bottleneck issue in the usability of search engines. Recently, most commercial search engines such as Google, Yahoo!, Live Search, Ask, and Baidu provide query suggestions to improve usability. That is, by guessing a user's search intent, a search engine suggests queries which may better reflect the user's information need.

A commonly used query suggestion method [11] is to find similar queries in search logs and use those queries as suggestions for each other. Another approach [12] mines pairs of queries which are adjacent or co-occur in the same query sessions. Although the existing methods may suggest go odd queries in some cases, none of them are context-aware - they do not take into account the immediately preceding queries as context in query suggestion. It not says the concept of general frame work method for query suggestion. In this research that the framework is built upon the heat diffusion on both undirected graphs and directed graphs. It has several advantages which can be utilized to many recommendation tasks on the Web.

### **3. PROBLEM FORMULATION**

At present, the redundant search pages and irrelevant web links are not eliminated. Expansions of abbreviated query such as (OS for Operating System) are not considered. Chart based analysis is not provided for viewing search result quality. To overcome the problem, an application is required and it should be capable of improving the query suggestion. When designing a framework for recommendations on the Web, we still face several challenges that need to be addressed. The first challenge is that it is not easy to recommend latent semantically relevant results to users. Take Query Suggestion as an example; there are several outstanding issues that can potentially degrade the quality of the recommendations. Queries containing ambiguous terms may confuse the algorithms which do not satisfy the information needs of users. The user perform a search is because of they have little or even no knowledge about the topic they are searching. In order to find satisfactory answers, users have to rephrase their queries constantly. The adoption of personalization will not only filter out irrelevant information to a person, but also provide more specific information that is increasingly relevant to a person's interest.

The phase of overcome the problem in search result, the framework is constructed based on the heat diffusion on

both undirected graphs and directed graphs. The graph is used for the user can easily identify whether the given query is linked with how many number of URLs. The Bipartite directed Graph is generated based on the value of visit count. The problem of improving the search result, proposes the method of previous search query words are also taken into query suggestion calculation. To add the recommendation user details play an important role in online query suggestion calculation. The view recommendation user URL access details to provide the user URL access details, it is used to the user can easily identify which one is the best URL for our query suggestion. In this proposed method the images can be added under the category wise and image search results returns all the images which is added under the category.

#### 4. RESEARCH METHODOLOGY

##### 4.1 Online Generation of the Recommendation Model

The recommendation model collect the user recommendation details and which provide the efficient URL details for the given query. Let  $vd[r\ n]$  represent the given query term  $r$  is compared with number of  $n$  user recommendation details and return the best URL based on the visit count.

Require: a set  $S$  of user sessions recorded in the query log.  
 Ensure: an inverted index  $vdindex$  built over the virtual documents obtained from satisfactory sessions.

```

1: for all  $r \in S$  do
2: if  $sessionIsSatisfactory(r)$  then
3:  $s \leftarrow getTermsFromSession(r)$ 
4:  $vd[r\ n] \leftarrow s$  {given current satisfactory session  $r$ , the occurrences of all the query-terms in  $r$  are added to the (initially empty) body of the virtual document associated to final query  $r$ .}
5: end if
6: end for
7:  $vdindex \leftarrow buildInvertedIndex(vd)$  { build the inverted file indexing all the obtained virtual documents.}
  
```

##### 4.2 Improved query selection by query search (iqsqs) (previous query based search)

The previous search query words are also taken into query suggestion calculation and it's improving the search result. The given query words are compared with previous search query. The query matched with previous query then only provides best Result based on the visit count.

Input: the user's query,  $Q_0$ , the set  $RefDocs$  of reference documents for  $Q_0$ , a number  $N \geq 20$ , and a number  $K$

Output: a set  $QS$  containing  $K$  query suggestions

```

1: // Term Selection
2: for all  $d \in RefDocs$  do
3: quickly find a set,  $F_d$ , of up to 20 terms in  $d$  that are likely to be useful in constructing queries with high coverage
4: score each term in  $F_d$  according to its coverage when combined with  $Q_0$  to form a query
5: sort  $F_d$  (highest scoring term first) and delete all but the first  $N$  terms
6: end for
7: // Query Suggestion Candidate Generation
  
```

```

8: initialize  $QSC$ , the set of query suggestion candidates, to be empty.
  
```

```

9: for all  $d \in RefDocs$  do
  
```

```

10: generate a set,  $QSC_d$ , of queries built from terms in  $F_d$ 
  
```

```

11:  $QSC \leftarrow QSC \cup QSC_d$ 
  
```

```

12: end for
  
```

```

13: // Greedy Selection of final Query Suggestions
  
```

```

14: initialize  $QS$  to be empty.
  
```

```

15: for  $i = 1$  to  $K$  do
  
```

```

16: add to  $QS$  the query  $qs \in QSC$  that most increases MCC (break ties to maximally increase MEC)
  
```

```

17: remove  $qs$  from  $QSC$ 
  
```

```

18: end for
  
```

```

19: Return  $QS$ 
  
```

##### 4.3 Query suggestion by query search (qsqs)(find query suggestion)

Input: a set  $RefDocs$  of reference documents and a number  $K$

Output: a set  $QS$  containing  $K$  query suggestions

```

1: // Lexical Alias Search
  
```

```

2: for all  $d \in RefDocs$  do
  
```

```

3: use Query Search to find a lexical alias for  $d$ ,  $LAd$ , based on  $d$ 's title terms and most frequent terms
  
```

```

4: end for
  
```

```

5: // Query Suggestion Candidate Search
  
```

```

6: initialize  $QSC$ , the set of query suggestion candidates, to be empty.
  
```

```

7: for all  $d \in RefDocs$  do
  
```

```

8: use Query Search to find the set,  $QSC_d$ , of minimal subqueries of  $LAd$  that cover  $d$ 
  
```

```

9:  $QSC \leftarrow QSC \cup QSC_d$ 
  
```

```

10: end for
  
```

```

11: // Greedy Selection of final Query Suggestions
  
```

```

12: initialize  $QS$  to be empty.
  
```

```

13: for  $i = 1$  to  $K$  do
  
```

```

14: add to  $QS$  the query  $qs \in QSC$  that most increases MCC (break ties to maximally increase MEC)
  
```

```

15: remove  $qs$  from  $QSC$ 
  
```

```

16: end for
  
```

```

17: Return  $QS$ 
  
```

##### 4.4 Suggestion Retrieval (image retrieval)

The image can be searched based on the suggested query words.

Require: the head  $r_0\ t_j$  of length  $t$  of the current user session  $r_0$ , and the recommendation model  $vdindex$ .

Ensure the set  $R$  of top- $k$  scored recommendations for the given query.

```

1:  $s \leftarrow getTermsFromSession(r_0\ t_j)$ 
  
```

```

2:  $D \leftarrow getMatchingVirtualDocuments(vdindex, s)$ 
  
```

```

3:  $R \leftarrow new\ heap(k)$ 
  
```

```

4: for all  $d \in D$  do
  
```

```

5:  $shortcut \leftarrow getTitle(d)$ 
  
```

```

6:  $R.insert(shortcut, a\ BM25(s, d) + b\ freq(shortcut))$ 
  
```

```

7: end for
  
```

```

8: return  $R$ .
  
```

##### 4.5 K-Means clustering algorithm

It is the problem of finding homogeneous groups of data points in a given data set. Each of these groups is called a cluster and can be defined as a region in which the density of objects is locally higher than in other regions.



The given data set  $X = \{x_1, \dots, x_n\}$ ,  $x_n \in R_d$ . The M-clustering problem aims at partitioning this data set into M disjoint subsets (clusters)  $C_1, \dots, C_M$ , such that a clustering criterion is optimized. The most widely used clustering criterion is the sum of the squared Euclidean distances between each data point  $X_i$  and the centroid  $m_k$  (cluster center) of the subset  $C_k$  which contains  $x_i$ . This criterion is called clustering error and depends on the cluster centers  $M_1, \dots, M_M$ :

$$E(m_1, \dots, m_M) = \sum_{i=1}^N \sum_{k=1}^M I(x_i \in C_k) \|x_i - m_k\|^2;$$

where  $I(X) = 1$  if  $X$  is true and 0 otherwise.

The k-means algorithm finds locally optimal solutions with respect to the clustering error. It is a fast iterative algorithm that has been used in many clustering applications. It is a point-based clustering method that starts with the cluster centers initially placed at arbitrary positions and proceeds by moving at each step the cluster centers in order to minimize the clustering error. K-Means is a simple learning algorithm for clustering analysis. The goal of K-Means algorithm is to find the best division of  $n$  entities in  $k$  groups, so that the total distance between the group's members and its corresponding centroid, representative of the group, is minimized. Formally, the goal is to partition the  $n$  entities into  $k$  sets  $S_i$ ,  $i=1, 2, \dots, k$  in order to minimize the within-cluster sum of squares.

In k-Means algorithm, only find a local optimum, and is commonly run multiple times with different random initializations. Variations of k-means often include such optimizations as choosing the best of multiple runs, but also restricting the centroid to members of the data set, choosing medians, choosing the initial centers less randomly. The algorithm prefers clusters of approximately similar size, as they will always assign an object to the nearest centroid. This often leads to incorrectly cut borders in between of clusters.

## 5. PROPOSED SYSTEM

### 5.1 Find Query Suggestion

In this dissertation "WEB GRAPH BASED QUERY SUGGESTION FOR TEXT AND IMAGE", is the process of finding the relevant result for the given query, based on the user access visit count. Several URL has some unwanted data and the user doesn't know the best URL for the query suggestion. It is a best method for find out the best URL for the given query, based on URL visit count. This method is constructed for retrieve the text and images.

#### 5.1.1 Text

In Existing methodology there is no concept of abbreviated based query suggestion for retrieve the data from the web. For example Expansions of abbreviated query such as (OS for Operating System) are not considered. It proposes the method of abbreviated based query suggestion. In this proposed method user no need to type full query word, the user can type only abbreviated word. Based on the abbreviated word the information can be retrieved.

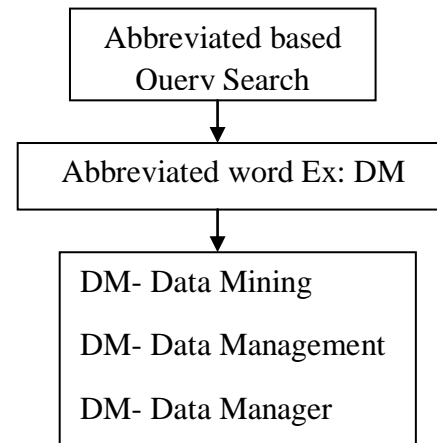


Fig. 1 Abbreviated Based Query Suggestion

In this abbreviated based query suggestion improve the search result time. The user need not to know the full explanation of query word. For example the user can type the abbreviated word as DM means the DM related words are displayed and the user can choose the relevant query and retrieve the relevant result.

The Bipartite Undirected and directed graph can be constructed, based on Queries, URLs and URL visit count. In adding the search word illustrate the concept of add the search word for query suggestion and also include the option of delete the query word. After adding the search word the URL can be added for respective search word.

In the Find Query Suggestion concept, once the user can enter the query after that what are the URLs contain that query terms can be listed. The URLs can be displayed based on the user visit count. Once the user click the URL the visit count will be increment by one.

In the previous query based suggestion, previous search query words are also taken into query suggestion calculation. In this previous query based suggestion method not consider the Zero Visit count URLs. But in normal find query suggestion also consider the Zero visit count URLs. The Graph query suggestion construct the Bipartite directed graph, based on the given queries and the access URLs. In addition Find Query Suggestion is constructed for domain wise.

#### 5.1.2 Image

The images can be searched based on the Query word. In this proposed method the images can be added under the category wise. In the existing methodology the images can be searched based on the query name. Example the user can type the word computer means, the computer images are retrieved as a result. In this proposed method the user can type the query word computer means, it returns what are all the images stored under the category of computer.

In this method the images can be searched based on suggested keywords. The abbreviated based query suggestion also included for image search and also included visit count methodology.

Once the user type the query word means the related words and their visit count details are listed. Based on the visit count the images can be searched and return the result.

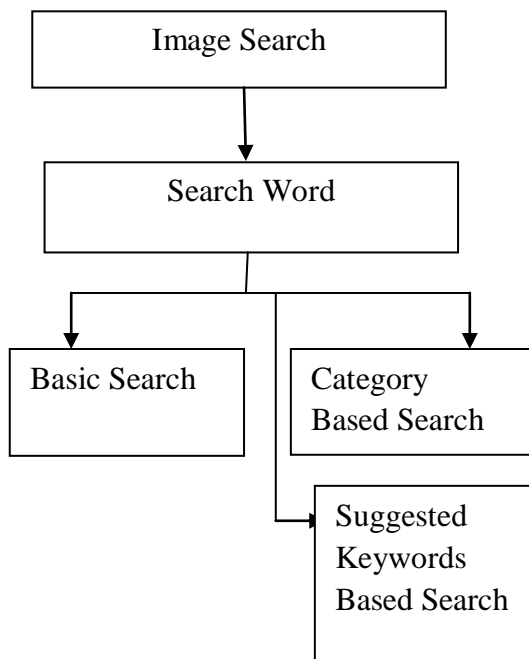


Fig. 2 Image Search

### 5.2 User URL access details

In this concept include user access details such as user-id, visit time, URL, visit count, Name of the user and user professional. In the Link Recommendation Provide the Recommendation link based on the user. In query based recommendation provide the URLs , based on visit count. The URL Access Matrix provides all user URL visit count details. The Group based query matrix provides the user URL access details based on user group based.

### 6. RESEARCH FINDINGS

1. Query Suggestion is a technique widely employed by commercial search engines to provide related queries to users' information need.
2. Abbreviation based Query suggestion can ensure to provide the result for user query suggestion based on the short term query word.
3. Data record the activities of Web users, which reflect their interests and the latent semantic relationships between users and queries as well as queries and clicked Web documents.
4. Recommendation algorithm not only suggests queries which are literally similar to the test queries, but also provides latent semantically relevant recommendations.
5. The heat diffusion process is providing the general framework to construct the directed and undirected bipartite graph. The Web graphs for the recommendation tasks suggest the query suggestions to the user goal.
6. The previous search query words are also taken into query suggestion calculation and it's improving the search result.
7. The method of adding recommendation user details gets the user details and stores it in data set and which is used

to provide information to the user based on the stored recommendation in data set.

8. The view recommendation user URL access details to provide the user URL access details, it is used to the user can easily identify which one is the best URL for our query suggestion.

9. Another challenging the user URL access matrix to provide all user users URL access details. The method of k-mean clustering is used to generate user group based query matrix which is used to give the user access details by group wise. In addition the image can be searched based on the suggested query words.

### 7. CONCLUSION AND FUTURE ENHANCEMENT

The existing system does not deal with the query suggestion in the query search page. The visit count suggestion is not provide for the web information's. The top most URLs does not identified. The redundant search page and irrelevant web links are not avoided. The recommendations of query phrase with related link does not to be categorized. It can provide latent semantically relevant results to the original information need. This model provides personalized recommendations for users. In the proposed system, Similarity measure between two queries is also considered. Abbreviation based query suggestion is also considered. Personalized recommendations are given importance. Previous search query words are also taken into query suggestion calculation. Expansions of abbreviated query are considered. The new search mechanism satisfies the user's information needs. Unwanted Web page links are eliminated to better extent. To suggest the users, which query more visited the same URL means it is best one of the URLs in that query. This query suggestion result displays by using visit count value. In addition the image can be searched based on the suggested query words. The improvements can be appended by changing the existing modules or adding new modules. Several areas to be developed in future, so the application must be upgraded for the new ones required and it is possible to modifications according to new requirements and specifications. It deals with the query and URLs which is stored in the database, and it will be displayed the by the use of web browser control. In future the abbreviated query words will developed in domain wise. The undirected graph visit count to be retrieved from the search page result.

### REFERENCES

- [1] E. Agichtein, E. Brill, and S. Dumais, "Improving Web Search Ranking by Incorporating User Behavior Information", SIGIR '07: Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 19-26, 2006.
- [2] T. Joachims, "Optimizing Search Engines Using Clickthrough Data", In Proceedings of the ACM Conference on Knowledge Discovery and Datamining (SIGKDD), 2002.
- [3] Hang Cui, Ji-Rong Wen, Jian-Yun Nie, and Wei-Ying Ma, "Query Expansion by Mining UserLogs", (SIGKDD), 2003.
- [4] G.W. Furnas, T.K. Landauer, L.M. Gomez, and S.T. Dumais, "THE Vocabulary Problem in Human-System Communication", Comm. ACM, vol. 30, no. 11, pp. 964-971, 1987.

- [5] J. Canny, "Collaborative Filtering with Privacy via Factor Analysis", SIGIR '07: Proc. 25th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 238-245, 2002.
- [6] J. Canny. "Collaborative filtering with privacy", In IEEE Security and Privacy Conference (to appear) , 2002.
- [7] E. M. Rogers. "Diffusion of Innovations", Fourth Edition. The Free Press , 1995.
- [8] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li, "Context-Aware Query Suggestion by Mining Click-Through and Session Data", KDD '08: Proc. 14th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 875-883, 2008.
- [9] Wen, J., et al. "Clustering user queries of a search engine", In WWW'01.
- [10] Cui, H., et al. "Probabilistic query expansion using query logs" In WWW'02.
- [11] Baeza-Yates, R.A., et al. "Query recommendation using query logs in search engines", In EDBT'04 .
- [12] Fonseca, B.M., et al. "Concept-based interactive query expansion", In CIKM'05

### BIOGRAPHIES



**Ms. S.Dhivya** received M.C.A Degree from Anna University, Coimbatore and M.Phil Degree from Bharathiar University, Coimbatore, TN, India .She is currently the Assistant Professor in Kongu Arts and Science College, Erode, TN, India. She has 3 years of

teaching experience.

**Ms. T.A. Sangeetha** received M.Sc(CS) Degree from Bharathiar University, Coimbatore and M.Phil Degree from Bharathiar University, Coimbatore, TN, India. She is currently the Assistant Professor in Kongu Arts and Science College, Erode, TN, India. She has 14 years of teaching and 12 years of research experience. She has guided 20 M.Phil students in the area of Computer Science. She has presented papers in National and International Conference.