Improved Framework for Location Aware Keyword Query Suggestion

Ms. Nikhila A
^M. Tech student ,Dept of CSE  Thejus Engineering College ,Vellarakkad

Abstract: A user who has issued a keyword query to a search engine. Using that keywords search engine retrieves the documents, the suggested keyword queries to the user, which are semantically relevant to the original query and they have as results documents that correspond to objects near the users location. For this purpose, here propose a weighted keyword-document graph which captures semantic and distance between queries and documents. Then, use the graph to suggest queries that are near in terms of graph distance to the original queries. Partition based algorithm is used to make the system more scalable, using the suggested keywords the system again perform keyword routing . In keyword routing ,system removes all the stop words from the sentence . Then construct an candidate query graph , to find the minimum distance between the elements. Minimum spanning tree is used to find the distance between the elements. Then convert the result into SPARQL query, using this technique the system removes all the unwanted documents from the list, only give the correct documents to the user which is semantically relevant to the user original query and also nearer to the user location. Using this method system efficiency is increased.

Keywords: Query suggestion, Keyword routing, Spatial databases.

1. Introduction

Data mining is the process of discovering meaningful data from large amounts of data stored in repositories using different technologies and techniques. Today, organizations are generating huge and growing amounts of data in different formats and different databases. Data mining is the process of analyzing data and finding useful information patterns, associations, or relationships from it. For better decision making, the large amount data collected from different resources require proper methods for extracting knowledge from the databases.

With the rapid growth of data on the web more and more people rely on the search engine for exploiting the information they need. A search engine is a software program or script available through the internet that searches documents and files for keywords and returns the results of any files containing those keywords. Search engines essentially act as a filters for the wealth of information available on the internet. They allow users to quickly and easily find information of genuine interest or value to them, without the need to wade through numerous irrelevant web pages.

In figure 1.1 illustrate when User enter the keyword that in which they want to search for a particular document, the web server send the query to the index server. Index servers provides the pages which contains the word that match with query. The query travels to the document server which retrieve the stored documents. Then search result return to the user within a second this is the simple working flow diagram of search engine. Recently, most search engines using bag-of-words model to respond to a users query, which matches keywords between the query and web documents . However the drawbacks of this model become increasingly prominent. The inherent ambiguity of natural language makes the search engine cannot find out the documents that meet the users need. The average length of queries submitted to search engines is only 2 to 3 words, which make it difficult to speculate the meaning of the queries. Features of search engine. Advanced site search engine complete customization, Scheduled re-indexing, content monitoring, No fixed page limit, Ease of use, Indexing of password-protected pages.

![Figure 1.1 working flow of search engine](image-url)

Location-aware keyword query suggestion(LKS),that the suggested queries retrieve documents not only related to the user information needs but also located near the user location .A
Spatial Keyword query is an approach of searching qualified spatial objects by considering both the query requester’s location and user specified keywords. Taking both spatial and keyword requirements into account, the goal of a spatial keyword query is to efficiently find results that satisfy all the conditions of a search. Searching is a common activity happening in data mining. This motivated to develop methods to retrieve spatial objects. A Spatial Keyword query is an approach of searching qualified spatial objects by considering both the query requester’s location and user specified keywords. Taking both spatial and keyword requirements into account, the goal of a spatial keyword query is to efficiently find results that satisfy all the conditions of a search. This motivated to develop methods to retrieve spatial objects. A spatial object consists of objects associated with spatial features. In other words, spatial objects involve spatial data along with longitude and latitude of location. The importance of spatial databases is reflected by the convenience of modeling entities of reality in a geometric manner. However, existing keyword suggestion techniques do not consider the locations of the users and the query results. Users often have difficulties in expressing their web search needs they may not know the keywords. After submitting a keyword query, the user may not be satisfied with the results.

2. Literature Survey

Beeferman et.al.[1] propose a collective agglomeration that embody the common clicked URLs. Collective agglomeration algorithmic rule accustomed establish connected queries and URLs for clustering cluster of queries that are similar in an repetitious approach. The queries within the same cluster are used as suggestions for each other. the standard of the question suggestions was evaluated by the click-through rate on the live Lycos program. However, this technique has high procedure value and can’t rescale to massive information.

U.Ozertem et.al.[2] Learning to rank considered the task of suggesting related queries to users after they issue their initial query to a web search engine and proposed a machine learning method to learn the probability that a user may find a follow-up query both useful and relevant, given his initial query. The method is based on a machine learning model which enables the system to generalize queries that have never occurred in the logs as well. The model is trained on co-occurrences mined from the search logs, with novel utility and relevance models, and the machine learning step is done without any labeled data by human judges. The learning step allows system to generalize from the past observations and generate query suggestions that are beyond the past co-occurred queries. estimating a scoring function that measures how useful and relevant is a follow-up query to a given query. estimate this score by a probabilistic utility function that relies on the query co-occurrence. The scores are used as the target values in machine learning model. This is a discriminating advantage of the method and it saves the costly and time consuming human labeling process. This model enables us to rank the suggestion candidates for a given a query, and eliminate the irrelevant and useless ones.

Jiang et.al. [5]: In this a query search procedure constructs queries that rank the document high enough for user to see it; from this set of queries the suggestions is given.

R. Zhong et.al.[6] The location-aware instant search problem, which returns users location-aware answers as users type in queries letter by letter. the main challenge is to achieve high interactive speed. a novel index structure, prxregiontree(calledPR-Tree),to efficiently support location aware instant search. PR-Tree is a tree-based index structure which seamlessly integrates the textual description and spatial information to index the spatial data. Using the PRTree, develop efficient algorithms to support single prx queries and multi-keyword queries.

Ji-RongWen et.al.[8] Introduced query clustering approach using content words and user feedback, combining content and feedback similarity approach so it is efficient but its difficult to set parameters for linear combination of two similarity metrics.

3. Existing System

Keyword suggestion in web search helps users to access relevant information without having
to know how to precisely express their queries. Existing keyword suggestion techniques do not consider the locations of the users and the query results; i.e., the spatial proximity of a user to the retrieved results is not taken as a factor in the recommendation. A baseline algorithm extended from algorithm BCA is introduced to solve the problem. Then, we proposed a partition-based algorithm (PA) which computes the scores of the candidate keyword queries at the partition level and utilizes a lazy mechanism to greatly reduce the computational cost. The performance of the proposed algorithms is low

4. Proposed System

Design the improved Location-aware Keyword query Suggestion framework, for suggestions relevant to the user's information needs that also retrieve relevant documents close to the query issuer's location. Location-aware Keyword query Suggestion (LKS) framework constructs an initial keyword-document graph (KD-graph). This directed weighted bipartite graph between Documents and Keyword queries captures the semantics and textual relevance between the keyword query and document nodes, i.e., the first criterion of location-aware suggestion. partition algorithm which will divide the keyword queries and documents in the KD-Graph into groups. To improve the performance of the system here introduce the keyword routing mechanism. In this case, all the meta-data which we get from the link there we perform the term mapping process. Term mapping process is removing stop words from the sentence, including connectives and store the remaining part in a list. During query graph construction, then find the minimum distance between the elements using minimum spanning tree algorithm. By doing this, can improve the performance the system.

Algorithm

Step 1: Do term mapping.

1.a) Remove stop words from the sentence, including connectives and store the remaining part in a list L(A).

1.b) For each element "a" in L(A) do,

   Find a list of direct mapping of "a" and store it in a list L(B)

   Find a list of substring mapping of "a" and store it in a list L(C)

Step 2: Apply query graph construction for L(B) and L(C).

During query graph construction, find the distance between the elements of L(A) in L(B). (For this use minimum spanning tree algorithm)

Do the same in L(C).

Step 3: If the distance is zero, ignore otherwise,

3.a) For each L(B), sort the distance and the elements in the lower order of their distance

3.b) Apply the same for list L(C)

Step 4: Convert the result into a SPARQL query (It is a Comma Separated Value query) which is in the form of Name-Value pair.

Step 5: Store and print the first "n" elements from L(B) and print the first "m" elements from L(C) if necessary. (Necessary condition m<<n).

4.1 Modules

A. User aware module

This is the module the user can be authenticated whether the user is valid user or not before that the user wants to register first. In registration the user have to give username, password, mail id. If the user is valid the user enters into the application.

B. Domain Specific Search

This module is the simple search engine, when user given a keyword, the system find out the related documents corresponding to the keyword.

C. Location aware keyword query suggestion

In this module the suggestion of a searching query will be display depending upon the latitude and longitude of the user. The location of the particular place will also display in a Google map.

D. Keyword routing

Term mapping: In our project maintain a stop word file which contain around 200 commonly seen stop words. The token obtained from the previous step is compared with stop word file. When a match occur corresponding token is removed otherwise it is given as the input to the next stage. Stop word list includes commonly used adjectives, connectives, verbs and certain other words. All the stop-words are removed from the Meta data. Split keyword set: removing all the stop words from the meta data, then the meta data contain only the basic keyword set, these keywords are combine with each other and then retrieve the relevant document. Candidate query graph: this graph is used to find out the minimum distance between each keyword query in the document, minimum distance document is selected and passed to the user. This document must satisfy both the condition, it is more nearer to the user location also semantically relevant to the initial keyword query.

4. Conclusion

This paper proposed an improved LKS framework providing keyword query suggestions that are relevant to the user information needs at the same time can retrieve relevant documents near the user location. Proposed a partition based algorithm which computes the scores of the candidate keyword queries at the partition level and utilizes a lazy
mechanism to greatly reduce the computational cost. Here proposed an keyword routing to increase the performance of the system also retrieve the most relevant documents. Keyword routing system removes all the stop words from the sentence. Then construct an candidate query graph to find the minimum distance between the elements. Minimum spanning tree is used to find the distance between the elements. Then convert the result into SPARQL query. Using this technique the system removes all the unwanted documents from the list, only give the correct documents to the user which is semantically relevant to the user original query and also nearer to the user location. Using this method system efficiency is increased.

5. References


[7] Shuyao Qi, Dingming Wu, Nikos Mamoulis,(2015)“Location Aware Keyword Query Suggestion Based on Document Proximity”,In IEEE Transactions on Knowledge and Data Engineering , pp. 82-97.

