



An Effective Mobile Search Engine for Geo-Based and Non Geo-Based Keywords Through Personalization

P. Vasavi¹, Dr. A. Govardhan²

¹Software Engineering & JNTU-SIT, India²Software Engineering & JNTU-SIT, India

¹vasavipallepati@gmail.com; ²govardhan_cse@gmail.com

Abstract – PGNMSE is a mobile text query client-server model search engine which takes text query as input and gives geo and non-geo based results based on user preferences. In the geo-based search, it searches for the user specified location or place depending on the keyword. In non-geo based search, it searches for the related data specified by the user. The main aim of this paper is to give appropriate results using the rank filtering algorithm and also using ontology (Entity Relationship). It has two types of searches, one is personalized and another is non-personalized. In this system the client collects and stores the click through data locally to protect privacy. The server is responsible for concept extraction, training and re-ranking of search results.

Keywords— ontology, personalization, ranking, location, content, geo- based, non geo-based, click through data

I. INTRODUCTION

Existing Search engines will give thousands of irrelevant results for a single keyword. To reduce number of search results we have to add extra features to the keyword. If are unable to give appropriate query to the search engine, we may get ambiguous results. If geographical search is enabled then search became local and results would be more relevant.

The responsibility of a mobile search engine is to reduce ambiguity in results with shorter interactions between user and extractor (search engine). Major problem we are facing with earlier search engines is ambiguity in results without user's interest. Another problem of mobile search engines is unable to provide security for user's search results. And moreover the performance of a mobile is limited so complicated tasks cannot be performed. To solve these problems we are introducing (list of algorithms names).

In this mobile search engine the click through data can be divided into two categories one is geographical and another is non geographical. In Geographical category the search results are dependent on GPS (Global Positioning System).Using GPS, location will be traced and local data has first preference. In Non geographical category the search results are dependent on keyword and description of keyword only.

GPS has a vital role in Geographical based mobile search engine, by using this location details can be easily found. For example ,if someone want to search the information about "Universities" then earlier search engines give the list of all universities without considering

the his/her location whereas PGNMSE first locates user ,then gives nearby “Universities” based on ranking algorithms. So that the user can get more relevant information rather than bundles of irrelevant search results.

In this mobile search engine the search results should be produced by combining Geographical and Non geographical preferences into personalization process .In the personalization process for the same keyword it gives different results for different users as long as preferences and interests are not same. For this process we need to maintain the profile of all the users who are registered in the PGNMSE .User profile should contain the details of user location, preferences, and interests.

The Admin can upload data related to both Geographical and Non geographical information. This paper also gives security to the user search data by ranking or number of views count is limited to that user only. Even though users don't want to acknowledge his/her interests and preferences directly, they can be learned from history of user's personal search. Even if the users interest changes, according to that re-ranking is done over the search queries. For example, if someone searches for hotel nearby for once they may look for less costly hotels and some other time they may look for luxurious hotels, it depends on individual's interest only. According to that re-ranking of search result varied from one time to other time.

II. LITERATURE SURVEY

From the first search engine “Archie” onwards so many search engines became popular because of more number of users and bundles of data to be search. Search engines like Google and Yahoo became more popular because of their efficiency in search results, and provide quality content for instant queries. Even these search engines are not convenient for mobile users because user may not have more time and more space while searching for a query. Now a day's no one is stick to one place he/she has to move one place to another. In this situation mobility of search engine is necessary .To achieve users expectations some algorithms are introduced. These algorithms works on the basis of GPS (Global Positioning System) .

Identifying geo-geographical queries in the given keyword/phrase:

Many of the geographical-queries contain at least one geo-keyword; others may be geo or non geo. Thus for each query we decide it as geo based on its geographic content. We may have geo queries with non geo terms.

The queries that user going to search are basically three types.

1. Navigational queries: These queries contain keywords related to a part of website URL, which results to a website on one strike.
2. Informational queries: These are of many categories based on keyword, to know more about something or somebody we use this type of queries. Here to get a perfect answer for your query you need to go through more than one result.
3. Transactional queries: These queries include downloading of resources like documents, media etc

1. Related work:

Mostly all search engines will produce similar results to all users, even different users may search for different information .For example a computer science student want to search for “tree” to know tree concepts in data structures, while a play school kid use the same keyword to draw a tree and a business man searches for Mind tree to know the stock price using the same keyword. Click tools are important in analyzing data for tracking user action on search engine.

This paper will give an approach to re-ranking algorithm which does not need any implicit feedback from the user. Re-ranking is obtained from the collection of click events to train content weight vectors, location weight vectors. Location weight vector is incremented per each visit tracked by GPS. Content weight vector is incremented per each click of user. The number of views of particular search result and search result are sending to search engine for each click. To locate user, GPS is used. GPS gives the latitude and longitude value pairs to train the ranking function in location based search.

2. Modules:

2.1 Ontology:

Ontology consists of relatively generic knowledge that can be reused by different types of application. Click through data can be classified into 2 types.

a. Content ontology:

If a query is repeated to search then that would store in the web snippet as most preferred by the user.

b. Location ontology:

If a given keyword/phrase is present in the predefined web document then the location will be stored as a result. For each query the location related string can be extracted, then it will compare with the predefined stored location and keyword/phrase if it matches it is location ontology.

2.2 User interest profiling and extraction of user preferences:

Earlier search engines will not provide accurate results from abundant search results because those search engines not considering user interests. But proposed system will maintain the profile of user interest and provide a solution for personalization by extracting the profiles of user interests.

2.3 Personalized Re-Ranking functions

There are two types of weight vectors which can decide ranking of web documents in abundant documents. Those are content weight vector and location weight vector. Location weight vector can be calculated through latitude and longitude pair values. These value pairs can be tracked by GPS. So that location weight vector can be incremented.

3 Algorithms:

3.1 Kokono search algorithm:

Location based search method is based on the distance between user's locations and location present in the query. The location of the user can be tracked through their latitude and longitude value pairs. By using these pairs the users nearby location can be found easily. The comparison of kokono search and keyword based search overlooked above 26% of location based information.

3.2 Joachims search algorithm:

This algorithm is used for extraction of click through data by the optimization of quality of search engine. If the search engine retrieves a set of results for a query, and that list of documents d_a, d_b, d_c, \dots in the result set must be viewed by the user in an order of d_a, d_b where d_a 's rank $>$ d_b 's rank.

3.3 OMF profiling:

To build personalize search engine for mobile users, user preferences are organized in ontology based and multi-facet user profile, rank adaptation methods are used for future results.

3.4 SpyNB algorithm:

Spy Naïve Bayes method produce a set of user preferences those are send to RSVM. The set contains positive and negative data sets. Positive data set contains clicked documents and unlabelled documents which are not preferred by the user. SpyNB generates a new data set called negative data set by combining positive and unlabelled data sets. These generated data sets are helpful to find accurate preferences of user.

III. CONCLUSION & FUTURE WORK

Extracting the user preferences on both location and content on the user click through data will be done by PGMSE. It also contains ontology based user preferences, multi- facet user profile and rank adaption methods for future results. This search engine will collect additional data, repeatedly examine the recently obtained data and address privacy issues. For user mobility it includes location based mobile search by using GPS locations, it helps to improve retrieval effectiveness. PGMSE works more precise and efficiently along with the appropriate results.

WWW (World Wide Web) has reached a size where it becoming increasingly challenging to satisfy user information needs. The user looking for search results those should give appropriate and more relevant information. For that user has added the additional phrases to keyword which helps to get accurate and user preferred results. Future work is to identify the user alike queries and their results by query based recommendation procedures. For further enhancement of personalization effectiveness of PGMSE exploit the regular travel patterns and query patterns from click through data and GPS.

REFERENCES

1. Kenneth Wai-Ting Leung, Dik Lun Lee and Wang- Chien Lee, "Personalized Mobile Search Engine", IEEE Transaction on Knowledge and Data Engineering, 2013
2. K.W.-T. Leung, D.L. Lee, and W.-C. Lee, "Personalized Web Search with Location Preferences," IEEE Transactions on Data Mining , 2010
3. S. Yokoji, "Kokono Search: A Location Based Search Engine," Proc. Int'l Conf. World Wide Web (WWW), 2001.
4. The Ranking SVM (Support Vector Machine) algorithm W. Ng, L. Deng, and D.L. Lee, "Mining User Preference Using Spy Voting for Search Engine Personalization," ACM Trans. Internet Technology, vol. 7, no. 4, article 19, 2007.
5. T. Joachims, "Optimizing Search Engines Using Click through Data," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, 2002.
6. Q. Gan, J. Attenberg, A. Markowetz, and T. Suel, "Analysis of Geographic Queries in a Search Engine Log," Proc. First Int'l Workshop Location and the Web (LocWeb), 2008
7. Y.-Y. Chen, T. Suel, and A. Markowetz, "Efficient Query Processing in Geographic Web Search Engines," Proc. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2006.
8. Xu Yingchen, Yang Jing, Gu Junzhong, "An Ontology Based Approach for Mobile Personalized Recommendation", IEEE Transactions on Services Computing, 2009.
9. H. Li, Z. Li, W.-C. Lee, and D.L. Lee, "A Probabilistic Topic-Based Ranking Framework for Location-Sensitive Domain Information Retrieval," Proc. International ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2009.
10. Shen X., Tan B., and Zhai C. Privacy protection in personalized search. SIGIR Forum, 41(1):4–17, 2007.

BIOGRAPHY

P. Vasavi pursuing M.Tech in the stream Software Engineering in JNTUH-School of Information Technology. This work is supported and made under the esteemed guidance of Dr. A. Govardhan is presently a Professor of Computer Science & Engineering & Director at School of Information Technology and Executive Council Member, Jawaharlal Nehru Technological University Hyderabad (JNTUH), India. He served and held several Academic and Administrative positions including Director of Evaluation, Principal, Head of the Department, Chairman and Member of Boards of Studies and Students' Advisor.