ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

Personalized Mobile Search Engine Using Ontologies

Dr.C.R Rene Robin[1], Dr. C. Nalini[*2]

[1]Professor & Head, Dept. of CSE, Jerusalem College of Engineering, Chennai, Tamil Nadu, India

[2]Associate Professor, Dept. of CSE, Bharath University, Chennai, Tamil Nadu, India

[*]Corresponding Author

ABSTRACT: Web search is a frequent activity on Internet connected devices, yet web search is still a nuisance when user is using a mobile device, due to their default keypad and small screen, and that search results could be mostly irrelevant for the user needs given its mobile context. User needs an efficient way to introduce query terms and receive more precise information. In this paper, we propose a new web search personalization approach that captures the user's interests and preferences in the form of concepts by mining search results and their click through. Onto-Search is based on a client server model. Server performs heavy tasks such as training, re ranking, To preserve privacy only the feature vectors is passed to the server. Location SVM is used for re-ranking of future search results.

KEYWORDS: Personalization, content ontology, location ontology, re ranking search results, mobile search engine, click through data

I. INTRODUCTION

In today's era internet is widely used. Using internet through mobile has become common nowadays. But there arises some issues like small keypad, limited display screen. To be efficient mobile search engine must be able to profile users interests and return highly relevant results by submitting shorter queries. Leung et al developed a system that personalizes search results based on concept preferences and showed that it is more effective than page preferences[2].Several personalization techniques have been proposed to model users' content preferences via analysis of users' clicking and browsing behaviors [3],[4],[5],[6].In this paper we personalize search using content and location ontology, user preferences. For example if the user provides a query "hotel", a set of results will be returned to the client. Based on the click through of the user Onto-Search understands that the user's content preference is "reservation" and location preference is "India". Further GPS locations of the user are also tracked. If the user is at the location "Chennai" then it will be effectively used in personalization. For Personalization of future queries the user preference "reservation","Chennai" will be taken into account. Thus for future queries Onto-Search will favour results concerning hotels in Chennai.GPS locations can be obtained by mobile devices. To provide effective personalization Onto-Search is capable of combining users location preferences and GPS locations.

The Onto-Search client is responsible for receiving the user's requests,submitting the requests to the Onto-Search server, displaying the returned results, and collecting his/her click through in order to derive his/her personal preferences. The Onto-Search server, on the other hand, is responsible for handling tasks such as training and re ranking of search results before they are returned to the client. The profiles of users are stored on the Onto-Search clients, thus preserving privacy to the users.Onto-Search client performs tasks such as submitting the query to the Onto-Search server, providing the results obtained from the server, collecting the click through of the user that can be used in personalization. The Onto-Search performs computationally heavy tasks such as training and re ranking. Onto-Search client is a Google Android platform and Onto-Search server is a PC server. Only the feature vectors obtained from the users click through are passed to the server thus preserving the privacy of the user. These vectors are used for personalization of future queries.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

II. RELATED WORK

Most search engines return roughly the same results to all users which may be relevant or irrelevant. However, different users may have different information needs even for the same query. For example, a user who is looking for a laptop may issue a query "apple" to find products from Apple Computer, while a housewife may use the same query "apple" to find apple recipes. The personalization of search results of the search engines were based on users clickthrough [7]. There are some proposals which include an ontological user profile and even define a context model to determine user's interests or re-rank search results according to the profile created while the user is surfing the web [8] [9], but essentially they consider desktop users only. Focusing on the mobile environment, there is previous research related to web search. Some proposals already uses ontology to represent user profile, others include some context elements but usually limited to the geographic location. Some others approaches use ontology to create clusters of results.

There are other works which deal with word recommendation and autocompletion[10] on mobile devices,but they rarely include the ontology knowledge. Also, there is a newer propose for a standardized mobile ontology [11] but it is focused on mobile services instead the web search, so it does not include elements relevant to obtain better search results. RankNet, scalable implementation of neural networks, is then employed to learn the user behavior model from the cleaned clickthrough data. Earlier personalization techniques were based solely on the computational behavior of the user to model his interests regardless of his surrounding environment[12]. The main limitation of such approaches is that they do not take into account the dynamicity of the user interests.

Gan et. al suggested that search queries can be classified into two types, content (i.e.non-geo) and location (i.e.geo).A classifer was built to classify geo and non-geo queries, and the properties of geo queries were studied in detail.It was found that a significant number of queries were location queries focusing on location information[13]. Hence, a number of location-based search systems designed for geo queries have been proposed.These include Yokoji et al search engine who proposed a location-based search system for web documents[14]. A parser was employed to extract location information from web documents, which was converted into latitude longitude pairs or polygons .When a user enters a query together with the location information specified in a latitude longitude pair, the system creates a search circle within the specified latitude-longitude pair and retrieves documents containing location information within the search circle.

III. SYSTEM ARCHITECTURE

An ontology-based, multi-facet (OMF) user profling strategy to capture both of the users' content and location preferences (i.e., .multi-facets.) is used for building a personalized search engine for mobile users[15]. Figure 1 shows the general process of our approach, which consists of two major activities: 1) Reranking and 2) Profile Updating.

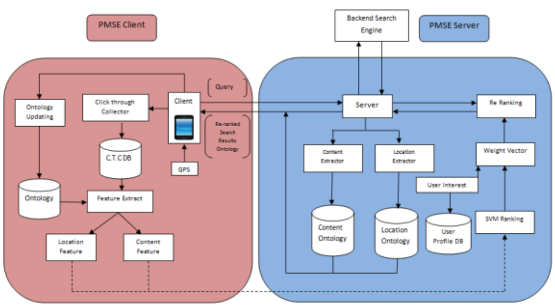

Fig.1 System Architecture

When a user enters a query, the search results are obtained from the backend search engines (e.g.,Google, MSNSearch, and Yahoo). The results obtained from the search are combined and reranked according to the user's profile trained from the user's previous search activities. After the search results are obtained from the backend search

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

engines, the content and location concepts (i.e. important terms and phrases) and their relationships from the search results and stored, respectively, as content ontology and location ontology.

After the search results are obtained from the backend search engines, the content and location concepts (i.e. important terms and phrases) and their relationships from the search results are stored, under content ontology and location ontology[16]. When the user clicks on a search result, the results,content and location concepts are stored in a database. RSVM training is done to using the feature vectors to obtain a content weight vector and a location weight vector for reranking.The search results for the user GPS locations are used along with location preferences for effective personalization. Weight vector is incremented for the particular GPS locations and is updated in user profile.When the user issues the same query again the personalized results according to the user content and location preferences are displayed to the user[17]. Thus the user can get effective results which helps him in making quick decision making without any confusion.

IV.CONCEPT EXTRACTION

Our personalization approach is based on concepts to profile the interests and preferences of a user. Therefore, an issue we have to address is how to extract and represent concepts from search results of the user. We propose in this paper an ontology-based, multi-facet (OMF) profling method in which concepts can be further classifed into different types, such as content concepts, location concepts.Here we define two concepts namely, content concepts and location concepts. A content concept, like a keyword or key-phrase in a Web page, is related to the content of the page, whereas a location concept provides information about the places related to the page.

4.1 Content Ontology

If a keyword/phrase exists frequently in the web-snippets arising from the query q, it can be considered as an important concept related to the query.Thus, our content concept extraction method first extracts all the keywords and phrases from the web-snippets arising from q.The possible concept space determined for the query “hotel,” while the clickthrough data determine the user preferences on the concept space.In general, the ontology covers more than what the user actually wants.

The concept space for the query “hotel” consists of “map,” “reservation,” “room rate,”..., etc. If the user is indeed interested in information about hotel rates and clicks on pages containing “room rate” and “special discount rate” concepts, the captured clickthrough favors the two clicked concepts. Feature vectors containing the concepts “room rate” and “special discount rate” as positive preferences will be created corresponding to the query “hotel.” When the query is issued again later, these feature vectors will be transmitted to the Onto-Search server and transformed into a content weight vector to rank the search results according to the user’s content preferences.

4.2 Location Ontology

In the location ontology, we organize all the cities as children under their provinces, all the provinces as children under their regions, and all the regions as children under their countries[1].The location ontology extraction method first extracts all of the keywords and key-phrases from the documents returned for q. If a keyword or key-phrase in a retrieved document d matches a location name in our predefined location ontology,it will be treated as a location concept of d. For example, given the concept “Tamil Nadu” from document d, we would match it against our location ontology. If a concept matches several nodes in the location ontology, all matched locations will be associated with the document.

V.CLICKTHROUGH COLLECTION

When the user issues a query a set of results are returned to the client. When the user clicks on a search result the clickthrough is captured and the clicked concept is favored. The content ontology together with the clickthrough serve as the user profile in the personalization process.They will then be transformed into a linear feature vector to rank the search results according to the user's preferences.These search preferences,along with a set of feature vectors are submitted along with future queries to the server for search results reranking.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

SpyNB learns user behavior models from preferences extracted from clickthrough data. SpyNB assumes that users would only click on documents that are of interest to them. Thus, it is reasonable to treat the clicked documents as positive samples. However, unclicked documents are treated as unlabeled samples because they could be either relevant or irrelevant to the user.Based on this interpretation of clickthroughs, the problem becomes how to predict from the unlabeled set reliable negative documents which are irrelevant to the user.

VI. RSVM

Ranking SVM is an application of SVM to solve certain ranking problems. Its purpose is to improve the performance of the internet search engine.Using clickthrough data as the input, RSVM aims at finding a linear ranking function, which holds for as many document preference pairs as possible. It maps the similarities between queries and the clicked pages onto certain space. It calculates the weights between any two of the vectors obtained. Reranks the search results based on the weights. If a user has visited the GPS location l_r, the weight of the location concept is incremented. Hence, we assume that the location that the user has visited a long time ago is less important than the location that the user has recently visited.Thus GPS plays a important role in location information.

VII. EXPERIMENTAL EVALUATION

Finding the required information quickly and easily on the web remains a major challenge. Most search engines compute a numeric score on how well each search results matches the query, and rank the results according to this value. The top ranking results are then shown to the user.

The users are asked to perform relevance judgment on the top 100 results for each query by filing in a score for each search result to reflect the relevance of the search results to the query. The score indicates three levels of relevancy (Good, Fair"and"Poor).Documents rated as Good" are considered relevant (positive samples), while those rated as Poor" are considered irrelevant(negative samples) based on the user's needs. The documents grouped as Fair are treated as unlabeled. Documents rated as "Good"(relevant documents) are used to compute the average relevant rank improvements (i.e., the difference between the average ranks of the relevant documents in the search results before and after personalization) and top N precisions, the two primary metrics for our evaluation[18].

Considering our proposed system, we use two straightforward evaluation measures, namely, precision and recall. Different queries may induce from the search results different concept spaces which are different in both sizes and diversities. Additionally, different users may have different interests and preferences on the search results. The diversity of content and location information of a query can be obtained from the extracted content and location concepts, while the interest/preference of a user can be found by his clickthrough behavior. In information retrieval contexts, precision is the fraction of retrieved documents that are relevant to the search. Recall is the fraction of the documents that are relevant to the query that are successfully retrieved.

Precision and recall are defined in terms of a set of retrieved documents (e.g. the list of documents produced by a web search engine for a query) and a set of relevant documents (e.g. the list of all documents on the internet that are relevant for a certain topic).

VIII. CONCLUSION

In this paper, we proposed an Ontology-Based Personalization framework for automatically extracting and learning a user's content and location preferences based on the user's clickthrough. Here we develop different methods for extracting content and location concepts, which are maintained along with their relationships in the content and location ontologies. As for the future work, we plan to study the effectiveness of other kinds of concepts such as people names and time for personalization.

REFERENCES

[1] K.W.-T. Leung, W. Ng, and D.L. Lee, "Personalized Mobile Search Engine," *IEEE Trans. Knowledge and Data Eng*, VOL. 25, NO. 4, APRIL 2013.

[2] *Lydia Caroline M., Vasudevan S.,"Growth and characterization of l-phenylalanine nitric acid, a new organic nonlinear optical material", Materials Letters, ISSN : 0167-577X, 63(1) (2009) pp. 41-44.*

[3] K.W.-T. Leung, W. Ng, and D.L. Lee, "Personalized Concept-Based Clustering of Search Engine Queries," *IEEE Trans. Knowledge and Data Eng*, vol. 20, no. 11, pp. 1505-1518, Nov. 2008.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 5, May 2015

[4] *Langeswaran K., Gowthamkumar S., Vijayaprakash S., Revathy R., Balasubramanian M.P., "Influence of limonin on Wnt signalling molecule in HepG2 cell lines", Journal of Natural Science, Biology and Medicine, ISSN : 0976-9668, 4(1) (2013) PP. 126-133.*

[5] W. Ng, L. Deng, and D. L. Lee, "Mining user preference using spy voting for search engine personalization," *ACM TOIT*, vol. 7, no. 4, 2007.

[6] T. Joachims, "Optimizing search engines using click through data," In *Proc. of ACM SIGKDD Conference*, 2002.

[7] *Jayalakshmi T., Krishnamoorthy P., Ramesh Kumar G., Sivamani P., "Optimization of culture conditions for keratinase production in Streptomyces sp. JRS19 for chick feather wastes degradation", Journal of Chemical and Pharmaceutical Research, ISSN : 0975 – 7384, 3(4) (2011) PP.498-503.*

[8] W. Ng, L. Deng, and D. L. Lee, "Mining user preference using spy voting for search engine personalization," *ACM TOIT*, vol. 7, no. 4, 2007.

[9] Q. Tan, X. Chai, W. Ng, and D. Lee, "Applying co-training to clickthrough data for search engine adaptation," in *Proc. of DASFAA Conference*, 2004.

[10] *Jebaraj S., Iniyan S., "Renewable energy programmes in India", International Journal of Global Energy Issues, ISSN : 0954-7118, 26(4Mar) (2006) PP.232-257.*

[11] E. Agichtein, E. Brill, and S. Dumais, "Improving web search ranking by incorporating user behavior information," in *Proc. of ACM SIGIR Conference*, 2006.

[12] Pretschner, A., Gauch, S "Ontology-based personalized search" *Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence pp. 391—398* (1999).

[13] Sieg, A., Mobasher, B., Burke, R "Ontological User Profiles for Representing Context in Web Search", *Proceedings of the 2007 IEEE/WIC/ACM international Conferences on Web intelligence and intelligent Agent Technology*, pp. 91--94 IEEE Computer Society (2007)

[14] *Gopalakrishnan K., Prem Jeya Kumar M., Sundeep Aanand J., Udayakumar R., "Thermal properties of doped azopolyester and its application", Indian Journal of Science and Technology, ISSN : 0974-6846, 6(S6) (2013) PP. 4722-4725.*

[15] De Luca, E. W. and Nürnberger, A "Using clustering methods to improve ontology-based query term disambiguation", *International Journal of Intelligent Systems*, vol. 21, pp. 693--709 (2006).

[16] Kamvar, M. and Baluja, S "Query suggestions for mobile search understanding usage patterns", *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing System*, pp. 1013-1016 (2008).

[17] Dr. T.Nalini, Dr.K.Manivannan, Vaishnavi Moorthy, Efficient Remote Data Possession Checking In Critical Information Infrastructures Ensuring Data Storage Security In Cloud Computing, International Journal of Innovative Research in Computer and Communication Engineering, ISSN (Online): 2320 – 9801,pp 12-18, Vol. 1, Issue 1, March 2013

[18] Dr.K.P.Kaliyamurthie, Cervical Cancer Screening and Classification Using Acoustic Shadowing, International Journal of Innovative Research in Computer and Communication Engineering, ISSN(Online): 2320-9801,pp 1659-1662, Volume 1, Issue 8, October 2013

[19] Dr.K.P.Kaliyamurthie, Automated Information Retrieval System Using Correlation Based Multi- Document Summarization Method, International Journal of Innovative Research in Computer and Communication Engineering, ISSN(Online): 2320-9801,pp 4328-4335, Vol. 1, Issue 7, September 2014

[20] Dr.K.P.Kaliyamurthie, D.Parameswari, Modelling Cloud Storage, International Journal of Innovative Research in Computer and Communication Engineering, ISSN: 2249-2651,pp 1-5, Volume1 Issue3 Number2–Dec2011

[21] Dr.K.P.Kaliyamurthie, D.Parameswari, Implementation of Customized Greencall Algorithm for Energy-Efficient Of Wireless LANs, International Journal of Innovative Research in Computer and Communication Engineering, ISSN: 2249-2615,pp 15-21, Volume1 Issue 1 Number2-Aug 2011

[22] http://www.dur.ac.uk/p.h.shaw/teaching/msc/ protege_tutorial.pdf

[23]https://mailman.stanford.edu/mailman/listinfo/protege-owl