ISSN:0975-9646

Context-aware Personalized Mobile Web Search Techniques-A Review

Deepika Bhatia[1*], Smita Nirkhi[1], Preeti Bajaj[2]

[1]*Department of Computer Science & Engineering,*

[2]*Department of Electronics Science & Engineering, G.H.R.C.E. Nagpur*

***Abstract*-The advent of the Web has raised different searching and querying problems. Keyword matching[1] based querying techniques used by various search engines, return thousands of Web documents for a single query, and most of these documents are generally are unrelated to the users' required information. Queries submitted to a search engine may have ambiguous meanings. For example, depending on the users [2], the query "apple" may refer to a fruit, the company Apple Computer and so forth. Thus, providing personalized query suggestion (e.g., users interested in "apple" as a fruit get suggestions about fruit, while users interested in "apple" as a company) will help the user in getting the desired information. This makes the current retrieval systems far from optimal. Also because of geography's important role in search requests, methods are required which aim at giving improved answers to geographic search requests. Geo search applications can use a standard keyword interface and extract geographic information (implicit or explicit) from queries, employ graphic interfaces such as interactive maps, or use the current location of a mobile user. This paper focuses on the survey of many efficient Context aware Personalized Web search approaches with geographical information which were proposed by many authors.**

***Keywords*- Web search, geographic search, local search, ontology, Personalized Search, LBS**

I. INTRODUCTION

"A system is a context–aware system if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's tasks".

Given different backgrounds of users, different interests of users and ambiguities in natural language, it is very likely that query words of two different users may appear exactly same even though their information needs are different. A particular word could mean many things in different contexts and the actual context can be determined by the user alone. For example; a search keyword "Java", can mean the Java island in Indonesia or the Java programming language. A typical search engine provides similar set of results without considering of who submitted the query. Therefore, there is the requirement to have personalized web search system which gives outputs appropriate to the user in the form of highly ranked pages. The current retrieval systems perform a 'word to word' match of the query. It makes the current retrieval systems far from optimal. The problem of Personalized Search aims to customize search results according to each individual user according to his/her context. This would possibly satisfy them in fulfilling their information needs. Web search queries that have a location intent, are also studied. While some queries have explicit location information in the query like "pizza hut Kansas City", many others do not, like "airport shuttle", but still expect search engines to return localized search results. Methods, for automatic identification of location sensitive queries incorporated in a personalization process in order to return local search results ordered higher in the search results are studied. Some Web search systems use relevance feedback to refine user needs or ask users to register their demographic information beforehand in order to provide better service. Since these systems require users to engage in additional activities beyond search to specify their preferences manually, approaches that are able to implicitly capture users' information needs should be developed.

Also in mobile search, the interaction between users and mobile devices are constrained by the small form factors of the mobile devices. To reduce the amount of user's interactions with the search interface, it is required to understand the users' needs, and provide him highly relevant information. By capturing the users' interests in user profiles, a personalized search middleware is able to adapt the search results obtained from general search engines to the users' preferences through the method of personalized reranking of the search results. In the personalization process, user profiles is an important part in reranking search results and thus need to be trained regularly depending upon the user's search activities. Several personalization techniques have been proposed to model users' content preferences. In this paper, we recognize the importance of location information in mobile search and user's location preferences in addition to context preferences in user profiles.

The rest of the paper is organized as follows:- In section II literature survey from previous papers is presented. Also various context aware approaches are studied. By various techniques and measures, geographical information is searched for implicit or explicit queries of users. Section III gives the conclusion drawn from the review paper. Section IV gives the problems and future directions that can help to explore the related issues.

II. LITERATURE SURVEY

D. E. Rose et al. [5] and U. Lee, Z. Liu et al. [6], studied users' click-through behavior, to understand the user's intentions. Clickthrough data plays an important role for tracking user actions on a search engine. They find that over 60% of queries were informational, and nearly 40% seemed to give unrelated information as per user's request. The classification by U. Lee, Z. Liu et al. [6], uses click-through data to identify the information need reflected by a query.

T.Joachims [4] presented an approach to learn retrieval functions by analyzing which links the users click on in the presented ranking. This led to a problem of learning with preference examples like "for query q, document da should

be ranked higher than document db". It studied the problem of learning a ranking function over a finite domain in terms of empirical risk minimization. A Support Vector Machine (SVM) algorithm was given that led to a convex program. It can be extended to non-linear ranking functions. Experiments showed that the method can successfully learn a highly effective retrieval function for a meta-search engine.

Zhengyu Zhu et al., [7] proposed query expansion approach based on a personalized web search model. The novel system, as a middleware connecting a user and a Web search engine, is fixed on the client machine. It can study the user's favorites implicitly. After studying user favorites, it can produce the user profile automatically. When the user enters query keywords, related personalized expansion words are produced by the suggested approach, and the words which are common with the query keywords are forwarded to a famous search engine such as Google. These expansion words can facilitate search engine retrieval information for a user based on implicit search criteria.

Hyun-suk Hwang et al., [8] suggested that different kinds of contextual information is collected via various sensors such as a GPS device, temperature, noise, and air pollution and other information providing services. The data transmitted from other sensors, with the exception of GPS sensors, cannot be generally used in context aware systems without the installation of sensors and construction of wireless sensor networks. Thus, most approaches of context-aware systems have service-oriented architecture based on Web services with context information such as user's location, profile, and activities [9][10].

Capila[10] proposed a context aware architecture based on Web services for building service-oriented system. The approaches of context-aware systems are integrated in LBS(location based services) with geographical location and user's context [9][10]. It also provides a software architecture which can be used for different types of context systems. This system is operated on the Web services

Damião [10] presented a context-aware LBS architecture based on map services, routing service, advertisement service through a standard interface. It helps users to find products, Points of Interest (PoI), and helps in the management of his or her daily tasks according to context. However, the systems need more specific map services with search functions adaptive to users.

FLAME2008 [12] is a context-aware tool which has been designed for the Beijing Olympics Games in 2008. It is a situation-aware system, which uses user profiles and context to deliver adaptive information during certain time intervals considering user location and user's environment to all mobile visitors. The system classifies context information into a three-level approach, which are sensor values, meaning the user's location, temperature, and context data such as surrounding information of a user, and situation level according to the user's demand at a certain time in bottom order. Also, it has an ontology browser with integrated semantic query functionality using a structured ontology including location, time, content, and situation, and domain sub-ontologies.

Buriano et al., [11] describe roles of ontologies in context-aware systems. They mention that ontologies help in interpreting data coming from sensors or observations using defined context modeling. Also, it is used to exploit reasoning on context ontologies to solve incomplete context information. Also the ontologies are used to share and exchange context information among different recommenders to integrate related data for the user.

B.Faltings et al., [3] presented contextual and personalized location based services called LBS. It provides mobile users with context-aware and personalized services in terms of information delivery and exchange. LBS knowledge should describe the characteristics of the user, the content available in the data sources, and the environment in which LBS/user interactions are embedded. Hence, context information, user profiles and data profiles play an important role. Assumption is asserted that everything is context-dependent. For example, the name of a person is in fact context-dependent as the same person may have an official name but also a nickname in an email address book, a familiar name used at home by family members, an alias used to chat on Internet, etc. The assumption that everything is context dependent materializes in the fact that every user query is checked against context data to see if the query should be reformulated differently because of the additional knowledge extracted from the context repository. For example, the query for a shopping area with a Chinese restaurant needs knowledge about the local shopping areas, knowledge that would be stored in the current spatial context.

Kenneth Wai-Ting Leung [13] presented methods to improve user's search experience. Most of search engines provide query suggestions to help users formulating more effective queries. When a user submits a query, a list of terms that are semantically related to the submitted query is provided to help the user. Query clustering techniques have been developed. Similarity between queries was measured based on overlapping keywords or similar phrases in the queries. Each query is represented as a keyword vector. Similarity functions such as cosine similarity or Jaccard similarity [14] were used to measure the distance between two queries. One major limitation of the approach is that common keywords also exist in unrelated queries. A Google middleware for clickthrough data collection has been designed to study the performance of the system. This middleware is tested with test queries selected from a spectrum of topical categories. Using the standard recall-precision measures, the performance of the middleware is evaluated. Beeferman and Berger's agglomerative clustering algorithm [15] is used as the baseline to compare with proposed concept-based approach. Experimental results showed that the average precision at any recall level is better than the baseline method. He introduced the notion of concept-based graphs by considering concepts extracted from web-snippets and adapt BB's method to this new context. It provided personalization effect by using the concept preference profiles; built upon the extracted concepts and clickthrough data.

Qingqing Gan et al., [15] suggested that due to geography's important role in search requests, (e.g., for hotels etc.), geographic search technologies (also called *local search*), based methods aimed at giving improved answers to geographic search requests. Geo search applications can use a standard keyword interface and extract geographic terms from queries of the users. What sites users visited as a result of a geo query, how different

geographic terms were used by the same user, and what non-geographic terms are associated with geographic terms are studied in the paper.

Mario Arias et al. [17] presented challenges for mobile data. Mobile Web Search deals with the user's querying on mobile phones in mobile environment, also doing some other work. An autocompletion engine is given that can be helpful both by saving typing time and by sending related terms. The autocompletion mechanism will be helpful if the user intends to use the same word that the system is expecting, but if a synonym is in user's mind, the system will be unable to recognize and autocomplete it. It led to the introduction of a concept-driven semantic and context-aware autocompletion engine. A semantic autocompletion engine is extended with context-aware suggestions, which filters out the non-contextually-relevant concepts. Semantic and context-aware autocompletion systems are integrated into a working prototype. It also has a user interface designed for the mobile environment. Feasibility of the system is tested by making a qualitative analysis of the improvement obtained in the user experience.

Feng Gui et al., [21] suggested that Mobile search is quite different from standard PC-based web search in a number of ways:

- the user interfaces and I/O are limited by screen real estate,
- key pads are tiny and inconvenient for use,
- limited bandwidth and costly connection fees.

Earlier, many personalized search algorithms are studied in the context of PC-based web search. An architecture is proposed which collects user information (at mobile device and carrier network) and derives user context in given situations. It is shown that personalized mobile search perform well for ambiguous queries and localized searches.

Lyon [18] gave fusion of geo-location and documents which enable queries that take into account both location proximity and text relevancy. For example, Google Maps supports location-aware text retrieval queries of users. A new kind of top-k query system is designed providing a location-aware top-k text retrieval (LkT) query. Output of the systems is a list of k objects ranked according to a ranking function that combines their distances to the query location and the relevance of their textual descriptions to the query phrase. The L*k*T query is different from the query that retrieves relevant documents within a geographic range. The text relevancy of a query result is computed by means of language models and a probabilistic ranking function. Framework is designed which leverages the inverted file for text retrieval and the R-tree for spatial proximity querying. Several indexing approaches are explored within the framework. The framework encompasses algorithms that utilize the proposed indexes for computing the top-k query. R-tree is used to find the nearest neighbors and then for each neighbor, an inverted file is used to rank the objects according to text relevancy.

Kenneth Wai-Ting Leung et al., [22] suggested to separate concepts into content concepts and location concepts, and organize them into ontologies, to create an *ontology-based, multi-facet (OMF)* profile to capture the user's content and location interests for improving the search accuracy. The differences between their work and existing works are: Earlier works required the users' to manually enter their location preferences explicitly (with latitude-longitude pairs or text form). They suggested method that does not require users to explicitly enter their location interests manually. In it both of user's content and location preferences, are automatically learnt from the user's clickthrough data from the user's profile. The method studies entropies derived from a query's search results and a user's clickthroughs to estimate the query's content and location ambiguities. The approach consists of two major activities :-

- **Reranking**: The search results are obtained from the backend search engines (e.g., Google), when a user submits a search keyword. These search results are combined and reranked according to the user's profile trained from the user's previous search activities.
- **Profile Updating**: After the search results are obtained, the content and location concepts and their relationships are mined online from the search results and stored, as content ontology and location ontology. When the user clicks on a search result, the user's clickthrough data is updated. The content and location ontologies, along with the clickthrough data, are then employed in RSVM(ranked support vector machine) training to obtain a content weight vector and a location weight vector for reranking the search results for the user.

F. Akhlaghian et al., [23] proposed a personalized search engine using ontology-based fuzzy concept networks. In this paper the authors personalize the search engine outputs with the help of automatic fuzzy concept networks. The concept of ontology is used to improve the common fuzzy concept networks built according to user's profile. Experimental output shows the enhancement in personalized search engine outputs using enriched fuzzy concept networks in contrast to common fuzzy concept networks.

Ourdia Bouidghaghen et al., [24] introduced the importance of the geographic needs of mobile users with lesser interaction with a device in mobile environment. The proposed work aims at detecting user's implicit local intent from global one and classifying local sensitive queries on the basis of location name in the query(present or not). Method is proposed to estimate location language models as location profiles for mobile search queries, by exploiting the top N results returned by a general Web search engine. Two measures are exploited namely the Kurtosis and Kullback-Leibler divergence measures. Depending on the contextual factor exploited to personalize mobile Web search, three different approaches are differentiated: user's preferences based personalization approaches [25],[26],[27], [28], location-based personalization approaches [29], and social-based personalization approaches [30]. In personalized search systems, the personalization component can affect the search in three distinct phases:

(1) as a part of the retrieval process such as topic-sensitive PageRank [31],
(2) in a distinct re-ranking activity by combining initial document score and a personalized score expressing user's thematic interests [27]

(3) a combination of interests and location preferences in [28] or by query refinement by adding terms to the query from the user context [32].

III. CONCLUSION

In the literature survey many Context Aware Personalized Web Search approaches have been discussed in various environments. Personalized Web search is to carry out retrieval for each user incorporating his/her own information need. As the competition in search market increases, some search engines have offered the personalized search service. For example, Google's Personalized Search allows users to specify the Web page categories of interest. An autocompletion engine has been built which helps the user to automatically complete his search query. LBS are studied that helps to know the geographical information related to the users. Top-k query system is designed providing a location-aware top-k text retrieval system. User's clickthrough data and pageranking algorithms play an important role in user's concept extraction. Fuzzy based networks also help to improve the performance of context aware personalized systems.

IV. PROBLEMS AND DIRECTIONS

The literature survey will help the researchers for extending the work in future. We can plan to improve various techniques and methods according to user's needs. Also in future researchers can :-

- Study the effectiveness of other kinds of concepts such as people names and time for personalization.
- Investigate methods to exploit a user's content and location preference history to determine regular user patterns or behavior.
- Estimate the computational costs induced by various studied technique when integrated in the information retrieval process.
- Earlier focus was only on the user's location to personalize the search results. Besides this, mobile user's queries seems to be also sensitive to others factors such as user's interests, time, etc. Personalization can be extended in relevance to user's interests and time together with the initial score of the search engine.

REFERENCES

[1] Selma Ay¸se Ozel M*etadata-based and Personalized Web Querying*, January, 2004

[2] Wai-Ting Leung *Personalized Concept-Based Clustering of Search Engine Queries* ,1041-4347/08 IEEE .2008.

[3] Février Lausanne,*Contextualized and Personalized Location-based Services* EPFL .pdf ,2008

[4] Thorsten Joachims T. Joachims, "*Optimizing Search Engines Using Clickthrough Data*," IEEE Proc. ACM SIGKDD, 2002.

[5] D. E. Rose and D. Levinson. Understanding user goals in Web search. In *Proc. Of the 13th Int. Conf. on the World Wide Web*, pages 13–19, 2004.

[6] U. Lee, Z. Liu, and J. Cho. *Automatic identification of user goals in Web search.* In *Proc. of the 14th Int. Conf. on the World Wide Web*, pages 391–400, 2005.

[7] Zhengyu Zhu, Jingqiu Xu, Xiang Ren, Yunyan Tian and Lipei Li, " *Query Expansion Based on a Personalized Web Search Model,"* Third International Conference on Semantics, Knowledge and Grid, Pp. 128 –133, 2007.

[8] Hyun-suk Hwang et al. *Context-aware System Architecture using Personal Information based on Ontology* IEEE 2006

[9] Damião R. de Almeida, Cláudio de Souza Baptista, Elvis R. da Silva, Cláudio E. C. Campelo, Hugo F. de Figueirêdo, Yuri A. Lacerda: *A Context-Aware System Based on Service-Oriented Architecture.* Advanced Information Networking and Application, pp18-20, 2006

[10] Capilla, R.: *Context-aware Architectures for Building Service-Oriented Systems*. Proceeding of the Conference on Software Maintenance and Reengineering, IEEE, pp300-303, 2006

[11] Buriano, L., Marchetti, M., Carmagnola, F., Cena, F., Gena, C., Torre, I.: *The Role of Ontologies in Contextaware Recommender Systems.* 7th International Conference on Mobile Data Management, 2006

[12] WeiBenberg, N., Gartmann, R.*: Ontology Architecture for Semantic Geo Services for Olympia* 2008. Conference of GI-, pp267-283, 2004

[13] Kenneth Wai-Ting Leung *Personalized Concept-Based Clustering of Search Engine Queries* 2008 IEEE

[14] G. Salton and M.J. Mcgill, *Introduction to Modern Information Retrieval* McGraw-Hill, 1983 .

[15] D. Beeferman and A. Berger, *"Agglomerative Clustering of a Search Engine Query Log,"* Proc. ACM SIGKDD, 2000.

[16] Qingqing Gan et al. *Analysis of Geographic Queries in a Search Engine Log LocWeb 2008*, April 22, 2008, Beijing, China.

[17] Mario Arias et al *ContextBased Personalization for Mobile Web Search* VLDB '08, August 2430,2008, Auckland, New Zealand .

[18] G. Cong and C.S. Jensen and D. Wu, *Efficient Retrieval of the Top- k Most Relevant Spatial Web Objects*, in proceedings of the VLDBEndowment, 2(1): 337-348, ACM, 2009.

[19] J. M. Ponte and W. B. Croft. *A language modeling approach to information retrieval.* In *SIGIR*, pp. 275–281, 1998.

[20] C. Zhai and J. Lafferty. *A study of smoothing methods for language models applied to information retrieval*. *ACM TOIS*, 22(2):179–214,2004.

[21] Feng Gui, Malek Adjouadi et al.*Personalized Approach for Mobile Search* 8-0-7695-3507-2008 IEEE

[22] Kenneth Wai-Ting Leung et al. *Personalized Web Search with Location Preferences* -4244-5446-2010 IEEE

[23] F. Akhlaghian, B. Arzanian and P. Moradi, *"A Personalized Search Engine Using Ontology-Based Fuzzy Concept Networks,"* International Conference on Data Storage and Data Engineering (DSDE), Pp.137 –141, 2010.

[24] Ourdia Bouidghaghen et al. *Personalizing Mobile Web Search for Location Sensitive Queries*,IEEE,2011

[25] M. Daoud and L. Tamine and M. Boughanem and B. Chebaro, *Learning Implicit User Interests Using Ontology and Search History for Per- sonalization*, in International Web Information Systems Engineering- International Workshop on Personalized Access to Web Information (WISE-PAWI), pp. 325-336, Springer, 2007.

[26] L. Tamine and M. Boughanem and N. Zemirli, *Exploiting Multi-Evidence from Multiple User's Interests to Personalizing Information Retrieval*, in IEEE Int'l Conference on Digital Information Management, YouakimBadr, Richard Chbeir, Pit Pichappan (Eds.), IEEE Engineering Manage-ment Society, pp. 7-12, 2007.

[27] V. Varma and N. Sriharsha and P. Pingali, *Personalized web search engine for mobile devices*, In International Workshop on Intelligent Information Access (IIIA'06), 2006.

[28] K.W.T. Leung and D.L Lee and W.C. Lee, *Personalized Web Search with Location Preferences*, in proceedings of the 26th International Conference on Data Engineering (ICDE), pp. 701-712, IEEE, 2010.

[29] D. Mountain and A. MacFarlane, *Geographic information retrieval in a mobile environment: evaluating the needs of mobile individual*, Journal of Information Science, 33(5):515530, 2007.

[30] P. A. Salz, *Making search social Unleashing search for the mobile gen- eration*, Taptu White Paper, Publisher and Chief Analyst Msearch Groove, 2008.

[31] T. Haveliwala, *Topic-sensitive page rank*, In Proceedings of the 11th International world wide web conference, 2002.

[32] S. Hattori and T. Tezuka and K. Tanaka, *Context-aware query refinement for mobile web search*, In 2007 International Symposium on Applications and the Internet Workshops, 2007.