Pageranked Conceptual Search for Android domain

Ms Deepika Bhatia*
Department of Computer Sc. & Engg.,
G.H.Raisoni College of Engineering
Nagpur, India
deepika.1885@rediffmail.com

Ms Smita Nirkhi
Department of Computer Sc. & Engg.,
G.H.Raisoni College of Engineering
Nagpur, India
smita811@gmail.com

Dr Preeti Bajaj
Department of Electronic Sc.& Engg.,
G.H.Raisoni College of Engineering
Nagpur, India
principal@ghrce.edu.in

Abstract*— As the web is rapidly growing, there is a need to cater the user's interests and needs from their behavior to extract the relevant information from the contents of the web. Web mining is used to categorize users and pages by analyzing the user's behavior, the contents of the pages and order of the URLs accessed. It has become an increasingly important task. There is also a sharp increase in the requirements for location-based services using location information, due to the development of wireless internet and mobile systems. Web structure mining approach has been user in our study. Concept mining and user profile updation approaches have been used. The concept of weighted page ranked algorithm is used to display the final results.

Keywords—Web search, Personalised Search, page ranking, GPS, URL

I. INTRODUCTION

Mobile web search represents a different paradigm in case of desktop search. In case of mobile environment, user needs a time and cost effective web search and also interested in the significant results also taking into account the context that surrounds him all in a limited-interface mobile device with a higher bandwidth cost. To improve information retrieval effectiveness, most search engines provide query suggestions semantically related to the typed context in order to help the user identify the terms he/she is interested in. Yahoo and Google provide such kind of suggestions for the queries which are more specific and general to user. But even then these systems provide same kind of suggestions to each and every user typing the same query. Our approach is based on method that provides personalized query suggestions depending on the user personalized context based techniques. Our approach uses clickthrough data to estimate the user's conceptual preferences and also provides results according to every individual's conceptual need. For example, if a user queries the search engine about the term "jaguar", it may refer to animal or car, as it is ambiguous term. Our method provides results according to the user's need, if interested in car, he will get results accordingly or if he/she interested in animal cheetah, he will get different results according to his needs. Our proposed approach is based upon proactive search application [1]. In this approach, the resources retrieved by the search engine are automatically filtered against the user's preferences and requirements and several other parameters to automatically select the most appropriate content, thus reducing the cognitive load imposed on the user.

Also our approach uses the location based services. LBS provide users with local information, i.e. the information that belongs to a particular domain of interest to the user and can be of used while he/she remains in a particular geographical area. In addition, LBS need to take into account the interactions and dependencies between user and context for filtering information and deliver it in order to fulfill the needs and constraints of mobile users. We argue that consequently it brings up a series of technical challenges in terms of context-awareness and personalization, as well as query formulation and answering etc. Instead, they need a better solution.

Various modules have been listed in our study. These modules are- first, when a user submits a query and concepts, their relations are mined online from the web-snippets that are clicked by the user. Second, clickthroughs are collected to predict user's conceptual preferences. Third, pageranking is applied to the returned conceptual results. Fourth, with these modules, location is being concatenated so that the user will get results independent of his location. He/she will get the desired information regarding the concepts and location preferences. Google has been used as a middleware to get the web results.

The rest of this paper is organized as follows: in section 2, we compare our method with other similar approaches. Related work to the given approach has been discussed. In section 3, our previous paper module is described in brief. In section 4 the module implementations and snapshots of the project is presented. Experimental results for personalized conceptual approach and weighted page ranking is discussed. A brief comparison of a general page ranking and used weighted page ranking approach is being given. Also in this section, the results are evaluated based upon the backend search engines such as Google etc. Further sections, give the conclusion and future work. Bibliographic references have been given at the end of paper.

II. RELATED WORK

There are many research and developments related to context-aware web search and studies from various points of view. But the vast majority of the efforts have been based on desktop environments. Therefore, the environment of mobile devices is less explored related to its desktop environment. However, there are some previous researches related to web search

improvement and personalization based on mobile devices. Search engines such as Google and Yahoo offer a mobile version that includes an autocompletion and word recommendation mechanisms, mostly based on their gathered knowledge of general previous searches.

Daniel Aréchiga et. al. [11] suggested an approach to improve the web search for a mobile user related to the mobile device interface and input method. Because most of the actual devices have limited numbers of keys on the keypad, and only a minority of them includes a full physical or virtual keypad in reality. An intelligent interface is presented in the paper which will help the users to write fewer letters with a recommender or autocompletion system for such kind of keyboard devices [2]. If these recommendations also consider context information, the recommended contexts will be more precise and effective. To solve this issue, a Context Model has been proposed which includes context information by means of the use of ontology and thesaurus concepts. This model creates and manages the user and device profiles which consist of environmental data like time, place, user preferences, device capabilities etc. The use of ontologies and a thesaurus help to offer better results. The ontologies helped to model the spatial, temporal, user dependent and device capabilities concepts easily.

There are some proposals which include an ontological user profile [3] and presented a context model to determine user's interests [4] or re-rank search results according to the profile created while user is searching the web [5] [6], but basically they consider desktop users only. Some earlier proposals used ontologies to represent user profile [7], others include some context [8] elements but usually limited to the geographic location. Some approaches used clustering of the web results [9]. Some researchers dealt with word recommendation and auto completion [2] mechanisms on mobile devices, but they hardly include the ontology concept knowledge. Also, there is a newer propose for a standardized mobile ontology [10], but mainly focus was on mobile services instead of the web search, so it does not include elements of context-aware approach.

Page L. et.al. presented a random surfing model underlying the PageRank [13] algorithm used for navigating the Web. In this model, the surfer stars searching by jumping to a random URL. After visiting a page, the surfer jumps to a new random page or he follows one link from the current page. That is he either searches the URLs with probability'd' and randomly chooses a link from the page to follow to a new page or probability (1-d) that he jumps to another random URL. This model is a simple random walk model given by a Markov chain. Based upon this surfing model, the basic PageRank algorithm calculates the rank (indicating popularity rather than relevance) for each web page by iteratively propagating the rank until convergence.

Topic-Sensitive PageRank [14] has been given earlier in which documents are accessed non-uniformly according to their topics. In weighted PageRank extensions [15, 16, 17] approaches, links are followed non-uniformly according to their popularity. Several link-semantics-aware extensions [18, 19] recognize links with different meanings and compute a PageRank weighted by the link semantics.

To alleviate the above problems, we introduce the notion of concept based data by considering the concepts extracted from web snippets and adapt the concept of weighted page ranked algorithm in our approach. It will provide effective personalization effects by using the concept preference profiles built upon the extracted concepts and clickthrough links.

III. PROPOSED APPROACHES

Our approach uses the Web usage mining concept that can be defined as the automatic discovery and analysis of patterns in the click-streams and associated data collected or generated as a result of user interactions with Web resources on different web sites. The goal of Web usage mining is to capture, model, and analyze the behavioral patterns and profiles of users interacting with a Web site. The discovered patterns are usually represented as collections of pages or resources that are frequently accessed by different users with common needs or interests.

Also when the user submits his query to the web, unfortunately a category hierarchy list shown to a user is usually very large, and as a result, an ordinary user may have difficulty in finding the proper links leading to the required categories. Furthermore, users are often too impatient to identify the proper categories before submitting their queries to the search engine. An alternative to browsing is to obtain a set of categories for a user query and then filter them according to the user's clickthrough pattern. Implementation of our modules is done on the following platforms:-

- Software used : Android operating system (on SDK)
- IDE: Eclipse
- SQLite as backend database
- For the layout design of display on emulator; XML file format is used.
- Jsoup parser
- Android device

As an important first step, we focus on two major types of concepts, **content concepts** and **location concepts**. A content concept, in a Web page, is like a keyword or phrase and defines the content of the page, whereas a location concept refers to a physical location related to the page returned by the web search. Table I shows an example query apple, with the content and location concepts extracted.

TABLE I

Example results for the query word APPLE

CONCEPT	LOCATION
FRUIT	NAGPUR
COMPUTER	DELHI
SOFTWARE	BOMBAY
IPOD	CALCUTTA

We argue that the concepts extracted from the user's search results can show the user's interests in long run. The extracted concepts indicate. The concepts extracted from the user's search pages can represent a possible concept space arising

from a user's queries, which can be maintained along with the clickthrough data for future preference adaptation. To find the user's intentions behind the query, we propose a two-step strategy to improve retrieval effectiveness.

- In the first step, the system automatically deduces, for each user, a small set of categories for each query and list of categories submitted by the user, based on his/her search history.
- In the second step, the system uses the set of categories and also the location information to augment the query to conduct the web search.

Specifically, we provide a strategy in order to

1. Model and gather the user's search history about categories given by the user earlier,
2. Construct a user profile based on the search history
3. Deduce appropriate categories for each user query based on the user's profile and
4. Improve web search effectiveness by using these categories as a context for each query.

In the proposed system architecture shown in picture (Fig.1) the personalized search model is given. In this the user requests for the required service by typing a query to the middleware. Here in our approach, middleware used is Google search engine. With the help of various data mining techniques the user's context and location information is extracted from the captured data and this information is combined with the clickthrough information. Our proposed methodolody of frequent pattern mining extracts the required information and gives weighted page rank to the clickthrough data according to user's interestingness. The ontology kept helps to update the user's profile and rerank the searched results. The output is displayed according to user's preferences combined with location of the user. The search results according to location and user's context are displayed hierarchically in list box.

Various snapshots are shown to demonstrate that our strategy of personalized web search is both effective and efficient. The main problem is to personalize web search for improving retrieval effectiveness. Some ranking algorithms depend only on the link structure of the documents i.e. their popularity scores (called web structure mining), whereas others look for the actual content in the documents (called web content mining), while some use a combination of both. PageRank can be thought of as a model of user behavior. Assume there is a "random surfer" who is given a web page at random and he/she keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page. The probability that the random surfer visits a page is its PageRank. And the *d* damping factor is the probability at each page the "random surfer" will get bored and request another random page.

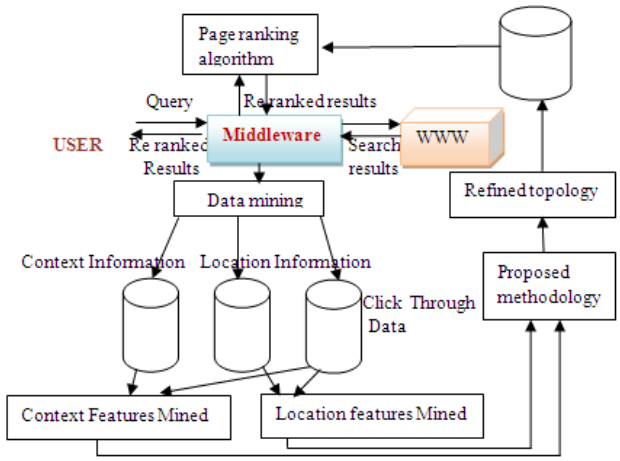

Figure1. Personalized search model

3.1 Weighted page ranking approach

Weighted PageRank Algorithm assigns larger rank values to more important (popular) pages instead of dividing the rank value of a page evenly among its outlink pages. Each outlink page gets a value proportional to its popularity (that is number of inlinks and outlinks). Our approach uses the concept of weighted page ranking algorithm. Weighted page ranking is much efficient than normal page rank algorithm. Both page rank and weighted page rank algorithms are compared [12] in Table II below:-

TABLE II

Comparison of Page Ranking algorithms

Criteria	**Page ranking algorithm**	**Weighted page rank algorithm**
Mining technique used	Web structure mining(WSM)	Web structure mining (WSM)
Methodology	Computes scores of pages at indexing time.	Computes scores based upon incoming and outgoing links of pages.
Input parameters	Back links	Both back and forward links
Complexity	O(log n)	<O(log N)
Search engine	Google	Our research model
Limitation	Query dependent	Query independent
Quality of results	Medium	Higher than pagerank algorithm

3.2 Content based filtering approach

In this approach a user's profile is created apriori for user's preferences. When an unseen query is typed by the user and requested from the service provider, its data is compared with the already existing context records collected according to user's past clickthrough collection. Depending upon this comparison the items with minimum distance i.e. maximum related items are recommended to the user according to his context interests.

3.3 Geographical information extraction

In this approach the information about the location of the user's mobile device is extracted from the GPS (global positioning system). Here the longitude and latitude information is extracted with the help of yahoo search engine. Three types of geographical information are taken into account-

- Global, Local implicit and Explicit locations.

In case of global information, for example- a user requesting for ringtones, the user is interesting in getting global results from the web service provider. Local implicit queries contain the search words with current location implicit in them, for example search for nearby south Indian restaurant. Explicit queries need to specify the exact location of the user. For example, the user may be interested in getting the desired information about restaurants at other cities or places. Our location based services approach uses all these types of aspects in the current implementation.

IV. EXPERIMENTAL EVALUATION

In the following, we describe our experimental data, our evaluation protocol and then present and discuss the obtained results. In the evaluation of the quality of the context aware personalization method, we use the ranked results returned by the backend search engines (i.e., Google and Yahoo etc.) as the baseline.

Our experimental methodology is designed to achieve two main objectives:

- To evaluate the effectiveness of our location based model approach for different types of search queries like local explicit, local implicit or global categories, and also,
- To evaluate its efficiency when actually integrating it into a search engine with our personalization process.

Further work is being carried out in improving the location based search results. The following snapshots demonstrate our results performance. Figure 1 requests the user to enter his categories of interest. After that the context words are linked with user's current location. Figure 2 demonstrates the search words resulted after linking user's preferences with the main web search. When a user clicks on any of the returned URLs, the search word with interested categories gets stored as clickthrough data also with the geographical information. When a user requests for the same web search, he/she will get the interested information after the implementation of frequent pattern mining approach. It will increase the effectiveness of the search results as compared to normal search with Google or MSN search. Android based domain has been used for the implementation of our approaches. Now user will get desired information at any place from his limited screen size in an efficient and personalized manner.

Figure 1.User's context entered

Figure 2. Results with location

With normal Google search the location has to be manually provides to the search engine. Also the results will not be personalized with Google, when user is going to search again for the same information. User should get the same results according to his/her interests even after 4 to 5 different searches, if he is more interested in the same earlier result. Figure 3 shows the normal Google search results for the hospitals query search.

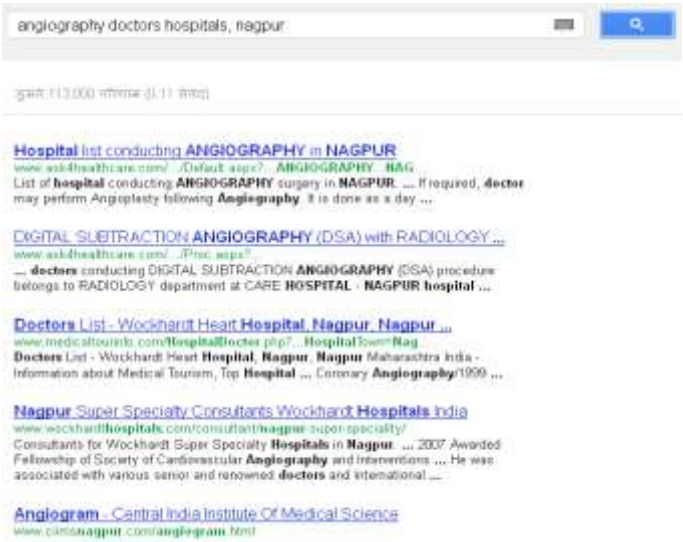

Figure3. Google search results for query hospitals

V. CONCLUSION

In this paper, we proposed a multi-facet personalization framework for automatically extracting a user's content and location preferences based upon the user's clickthrough data. The personalized information combined with user's location is given by the middleware after applying weighted page ranking with our refined approach. The experimental results using a sample of queries from the top documents returned by Google search, show that the proposed method achieves higher

accuracy in identifying location sensitive queries and shows significant improvement on search relevance when integrated to a search engine.

VI. FUTURE ADVANCEMENTS

In this paper we have presented a comprehensive discussion of the Web personalization process viewed as an application of data mining techniques which must therefore be supported during the various phases of a typical data mining cycle. A large-scale web search is a complex system. Stemming could be done to improve the retrieval effectiveness. In our proposed work, the source data could be cleaned and preprocessed so that user can get error free and more efficient results. As for the future work, we plan to study the effectiveness of other kinds of concepts such as people names, time and environment for personalization. Users access different items for different reasons and under different contexts. The modeling of context and its use within recommendation generation needs to be explored further. Also, user interests and needs change with time. Identifying these changes and adapting to them is the main goal of personalization.

REFERENCES

[1] Vincenzo Della Mea, Luca Di Gaspero, Davide Menegon, Danny Mischis, Stefano Mizzaro, Ivan Scagnetto, and Luca Vassena, "The Context-Aware Browser"University of Udine, Italy 1541-1672 2010 ieee intelligent systems Published by the IEEE Computer Society

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

[3] Gauch, S., Chaffee, J., and Pretschner, A.: Ontology-based personalized search and browsing. Web Intelligence and Agent Systems Vol. 1, Issue 3-4 pp. 219-234 (2003)

[4] Sieg, A., Mobasher, B., Burke, R.: Representing Context in Web Search with Ontological User Profiles. In: Proceedings of the Sixth International and Interdisciplinary Conference on Modeling and Using Context, (2007)

[5] Pretschner, A., Gauch, S.: Ontology-based personalized search. In: Proceedings of the11th IEEE International Conference on Tools with Artificial Intelligence pp. 391--398(1999)

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

[7] Weiβenberg, N., Voisard, A., and Gartmann, R.: Using ontologies in personalized mobile applications. In: Proceedings of the 12th Annual ACM international Workshop on Geographic information Systems, pp 2--11 (2004)

[8] Hattori, S., Tezuka, T., and Tanaka, K.: Query Modification Based on Real-World Contexts for Mobile and Ubiquitous Computing Environments. In: Proceedings of the 7th international Conference on Mobile Data Management, pp. 10--12, (2006)

[9] 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)

[10] Villalonga, C., Strohbach, M., Snoeck, N., Sutterer, M., Belaunde, M., Kovacs, E.,Zhdanova, A., Goix, L.W., Droegehorn, O.: Mobile Ontology: Towards a Standardized Semantic Model for the Mobile Domain. In: Service-Oriented Computing - ICSOC 2007Workshops, Springer, 248--257 (2009)

[11] Devices Daniel Aréchiga, Jesús Vegas and Pablo de la Fuente Redondo Ontology Supported Personalized Search for Mobile Departamento de Informática, Universidad de Valladolid, 47011 Valladolid, Spain) Proceedings of ONTOSE 2009)

[12] Ashutosh Kumar Singh ,A Comparative Study of Page Ranking Algorithms for Information Retrieval International Journal of Electrical and Computer Engineering 4:7 2009

[13] Page, L., Brin, S., Motwani, R.,Winograd, T.: The pagerank citation ranking: Bringing order to the web. Technical report, Stanford University (1998)

[14] Haveliwala, T.H.: Topic-sensitive pagerank. In: WWW'02. (2002) 517–526

[15] Jeh, G., Widom, J.: Scaling personalized web search. In: WWW '03. (2003) 271–279

[16] Xing, W., Ghorbani, A."Weighted pagerank algorithm."In: Proc. of the 2nd AnnualConference on Communication Networks and Services Research. (2004) 305–314

[17] Nie, Z., Zhang, Y.,Wen, J.R., Ma,W.Y.: Object-level ranking: Bringing order to web objects.In: WWW'05. (2005) 567–574

[18] Zhuge, H., Zheng, L.: Ranking semantic-linked network. In: WWW'03 Posters. (2003)

[19] Baeza-Yates, R., Davis, E.: Web page ranking using link attributes. In: WWW'04 Posters.(2004) 328–329

[20] Baeza-Yates, R., Davis, E.: Web page ranking using link attributes. In: WWW'04 Posters.(2004) 328–329