

A Novel Survey On Personalized Mobile Search Engine

CHETAN MAHAMUNI¹, SAURAB YEOLE², SHASHANK MISHRA³

S. M. NALAWDE⁴

^{1,2,3}BE Department of Computer Engineering Sinhgad Institute of Technology, Lonavala, SPPU, Pune. Maharashtra, India.

⁴Professor Department of Computer Engineering Sinhgad Institute of Technology, Lonavala, SPPU, Pune. Maharashtra, India.

ABSTRACT

We propose a personalized mobile search engine, PMSE that captures the users' predilections in the form of concepts by mining their click through data. Due to the consequentiality of location information in mobile search, PMSE relegates these concepts into content concepts and location concepts. In integration, users' locations (situated by GPS) are acclimated to supplement the location concepts in PMSE. The utilizer predilections are organized in an ontology-predicated, multi-facet utilizer profile, which are acclimated to habituate a personalized ranking function for rank adaptation of future search results. To characterize the diversity of the concepts associated with a query and their relevance's to the users need, four entropies are introduced to balance the weights between the content and location facets. Predicated on the client-server model, we withal present a detailed architecture and design for implementation of PMSE. In our design, the client accumulates and stores locally the click through data to bulwark privacy, whereas cumbersome hefty tasks such as concept extraction, training and re ranking are performed at the PMSE server. Moreover, we address the privacy issue by restricting the information in the utilizer profile exposed to the PMSE server with two privacy parameters. We prototype PMSE on the Google Android platform. Experimental results show that PMSE significantly amends the precision comparing to the baseline.

Key words: - Search Engine, User Preferences, PMSE, and multiple preferences.

1. INTRODUCTION

A major quandary in mobile search is that the interactions between the users and search engines are inhibited by the minuscule form factors of the mobile contrivances. As a result, mobile users incline to submit shorter, hence, more equivocal queries

compared to their web search counterparts. In order to return highly pertinent results to the users, mobile search engines must be able to profile the users' intrigues and personalize the search results according to the users' profiles. A practical approach to capturing a user's fascinates for

personalization is to analyze the user's clickthrough data. Leung, et. al., developed a search engine personalization method predicated on users' concept predilections and showed that it is more efficacious than methods that are predicated on page predilections. However, most of the precedent work postulated that all concepts are of the same type. Observing the desideratum for different types of concepts, we present in this paper a personalized mobile search engine, PMSE, which represents variants of concepts in different ontologies. In particular, apperceiving the paramountcy of location information in mobile search, we dissever concepts into location concepts and content concepts.

2. RELEGATED WORK

2.1 Existing System

A major quandary in mobile search is that the interactions between the users and search engines are inhibited by the minute form factors of the mobile contrivances. As a result, mobile users incline to submit shorter, hence, more equivocal queries compared to their web search counterparts. In order to return highly germane results to the users, mobile search engines must be able to profile the users' intrigues and personalize the search results according to the users' profiles. A practical approach to

capturing a user's fascinations for personalization is to analyze the user's clickthrough data. Leung, et. al., developed a search engine personalization method predicated on users' concept predilections and showed that it is more efficacious than methods that are predicated on page predilections. However, most of the antecedent work postulated that all concepts are of the same type. Observing the desideratum for variants of concepts.

2.2 Proposed System

Many subsisting personalized web search systems are predicated clickthrough data to determine users' predilections. Joachims proposed to mine document predilections from clickthrough data. Later, Ng, et. al. proposed to coalesce a spying technique together with a novel voting procedure to determine utilizer predilections. More recently, Leung, et. al. introduced an efficacious approach to prognosticate users' conceptual predilections from clickthrough data for personalized query suggestions. Search queries can be relegated as content (i.e., non-geo) or location (i.e., geo) queries. Examples of location queries are "hong kong hotels", "museums in london" and "virginia historical sites". In, Gan, et. al., developed a classifier to relegate geo and non-geo queries. It was found that a consequential

number of queries were location queries fixating on location information. In order to handle the queries that fixate on location information, a number of location-predicated search systems designed for location queries have been proposed. Yokoji, et. al. proposed a location-predicated search system for web documents. Location information were extracted from the web documents, which was converted into latitude-longitude pairs.

3. IMPLEMENTATION

3.1 User Interest Profiling :

PMSE uses “concepts” to model the fascinates and predilections of a utilizer. Since location information is paramount in mobile search, the concepts are further relegated into two variants, namely, content concepts and location concepts. The concepts are modeled as ontologies, in order to capture the relationships between the concepts. We observe that the characteristics of the content concepts and location concepts are different. Thus, we propose two different techniques for building the content ontology and location ontology. The ontologies designate a possible concept space arising from a user’s queries, which are maintained along with the clickthrough data for future predilection adaptation. In PMSE, we adopt ontologies to model the

concept space because they not only can represent concepts but additionally capture the relationships between concepts. Due to the different characteristics of the content concepts and location concepts.

3.2 Diversity and Concept Entropy

PMSE consists of a content facet and a location facet. In order to seamlessly integrate the predilections in these two facets into one coherent personalization framework, a consequential issue we have to address is how to weigh the content predilection and location predilection in the integration step. To address this issue, we propose to adjust the weights of content predilection and location predilection predicated on their efficacy in the personalization process. For a given query issued by a particular utilizer, if the personalization predicated on predilections from the content facet is more efficacious than predicated on the predilections from the location facets, more weight should be put on the content-predicated predilections; and vice versa.

3.3 User Preferences Extraction and Privacy Preservation

Given that the concepts and clickthrough data are accumulated from past search activities, user’s predilection can be learned. These search predilections, apprise of a set

of feature vectors, are to be submitted along with future queries to the PMSE server for search result re-ranking. In lieu of transmitting all the detailed personal predilection information to the server, PMSE sanctions the users to control the amount of personal information exposed. In this section, we first review a predilection mining algorithms, namely SpyNB Method, that we adopt in PMSE, and then discuss how PMSE preserves utilizer privacy. SpyNB learns utilizer comportment models from predilections extracted from clickthrough data. Postulating that users only click on documents that are of interest to them, SpyNB treats the clicked documents as positive samples, and presage reliable negative documents from the unlabeled (i.e. unclicked) documents. To do the presage, the “spy” technique incorporates a novel voting procedure into Naïve Bayes classifier to soothsay a negative set of documents from the unlabeled document set. The details of the SpyNB method can be found in. Let P be the positive set, U the unlabeled set and PN the presaged negative set ($PN \subset U$) obtained from the SpyNB method. SpyNB postulates that the utilizer would always prefer the positive set over the soothsaid negative set.

3.4 Personalized Ranking Functions:

Upon reception of the user’s predilections, Ranking SVM (RSVM) is employed to learn a personalized ranking function for rank adaptation of the search results according to the utilizer content and location predilections. For a given query, a set of content concepts and a set of location concepts are extracted from the search results as the document features. Since each document can be represented by a feature vector, it can be treated as a point in the feature space. Utilizing the predilection pairs as the input, RSVM aims at finding a linear ranking function, which holds for as many document predilection pairs as possible. An adaptive implementation, SVM light available at, is utilized in our experiments. In the following, we discuss two issues in the RSVM training process: 1) how to extract the feature vectors for a document; 2) how to coalesce the content and location weight vectors into one integrated weight vector.

4. EXPERIMENTAL RESULTS

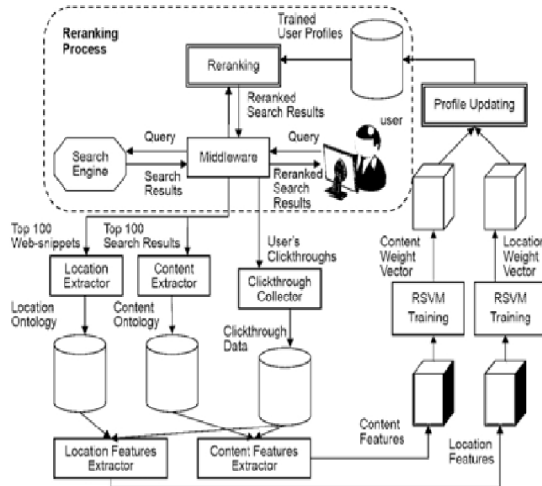


Fig 1 Architecture Diagram

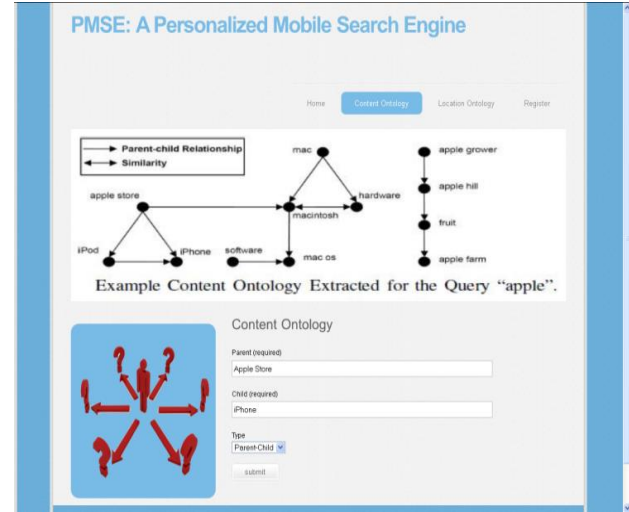


Fig 4 ontology (content):

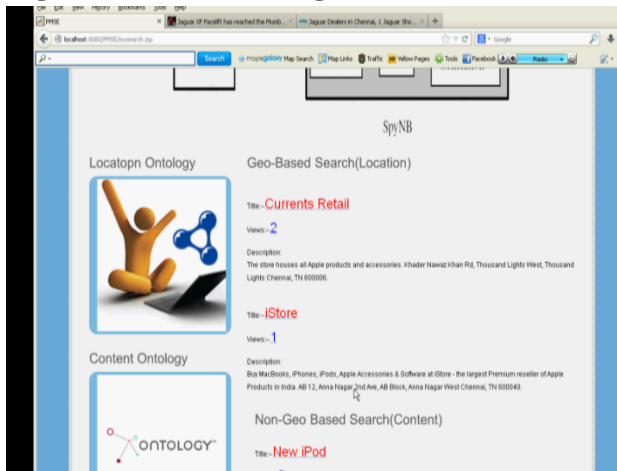


Fig2ScreenShot (Before algorithm applying)

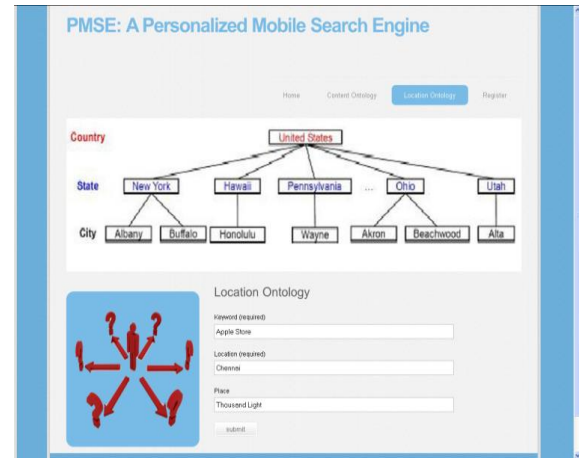


Fig 5 Ontology (Location):

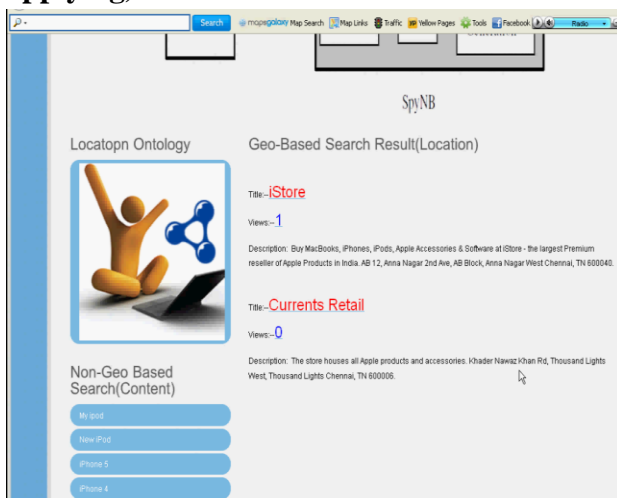


Fig 3 After NB algorithm applied:

5. CONCLUSION

To habituate to the utilizer mobility, we incorporated the user's GPS locations in the personalization process. We observed that GPS locations avail to ameliorate retrieval efficacy, especially for location queries. We additionally proposed two privacy parameters, minDistance and expRatio, to address privacy issues in PMSE by sanctioning users to control the amount of personal information exposed to the PMSE

server. The privacy parameters facilitate smooth control of privacy exposure while maintaining good ranking quality. For future work, we will investigate methods to exploit conventional peregrinate patterns and query patterns from the GPS and click through data to further enhance the personalization efficacy of PMSE.

6. REFERENCE

- [1] Pingdom, “Internet 2009 in Numbers,” <http://royal.pingdom.com/2010/01/22/internet-2009-in-numbers/>, 2010.
- [2] J. Grau, “US Retail e-Commerce: Slower But Still Steady Growth,” http://www.emarketer.com/Report.aspx?code=emarketer_2000492, 2008.
- [3] Internetretailer, “Web Tech Spending Static-But High-for the Busiest E-Commerce Sites,” <http://www.internetretailer.com/dailyNews.asp?id=23440>, 2007.
- [4] D. Dhyani, W.K. Ng, and S.S. Bhowmick, “A Survey of Web Metrics,” *ACM Computing Surveys*, vol. 34, no. 4, pp. 469-503, 2002.
- [5] X. Fang and C. Holsapple, “An Empirical Study of Web Site Navigation Structures’ Impacts on Web Site Usability,” *Decision Support Systems*, vol. 43, no. 2, pp. 476-491, 2007.
- [6] J. Lazar, *Web Usability: A User-Centered Design Approach*. Addison Wesley, 2006.
- [7] D.F. Galletta, R. Henry, S. McCoy, and P. Polak, “When the Wait Isn’t So Bad: The Interacting Effects of Website Delay, Familiarity, and Breadth,” *Information Systems Research*, vol. 17, no. 1, pp. 20-37, 2006.
- [8] J. Palmer, “Web Site Usability, Design, and Performance Metrics,” *Information Systems Research*, vol. 13, no. 2, pp. 151-167, 2002.
- [9] V. McKinney, K. Yoon, and F. Zahedi, “The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach,” *Information Systems Research*, vol. 13, no. 3, pp. 296-315, 2002.
- [10] T. Nakayama, H. Kato, and Y. Yamane, “Discovering the Gap between Web Site Designers’ Expectations and Users’ Behavior,” *Computer Networks*, vol. 33, pp. 811-822, 2000.