SUSTAINING CONFIDENTIALITY PROTECTION IN PERSONALIZED WEB SERACH FOR ONTOLOGY

*1Ms. Devi A , *2 Mr. Hariharan P.

*1 M.Phil Research Scholar, Department of Computer Science Adhiparasakthi College of Arts and Science (Autonomous), Kalavai, Vellore, Tamil Nadu, India

*2 Assistant Professor, Department of Computer Science Adhiparasakthi College of Arts and Science (Autonomous), Kalavai, Vellore, Tamil Nadu, India,

---***---

Abstract - ***Personalized E-Learning (PEL) improves the quality of various search services on the Internet. However, the reluctance to disclose the user's private information during search has become a major barrier for the wide proliferation of PEL. In this paper we propose a PEL engine that captures the user's preferences into the form of concepts by mining their Ontologies data. Due to the importance of location information in web search, PEL classifies these concepts into content concepts and location concepts. The user preferences are organized in an ontology-based, multi-facet user profile, which are used to adapt a personalized ranking function for rank adaptation of future search results. We propose a framework for personalized e-learning based on aggregate usage profiles and domain ontology. We have distinguished two stages in the whole process, one of offline tasks that includes data preparation, ontology creation and usage mining and one of online tasks that concerns the production of recommendations. We also provide an online prediction mechanism for deciding whether personalizing a query is beneficial. Extensive experiments demonstrate the effectiveness of our framework.***

Key Words: *Web Usage Mining, Adaptive hypermedia PEL, Ontology, Clustering, Privacy Preserving, Greed Algorithm*

I. INTRODUCTION

The E-learning engine has become the most important portal for ordinary people looking for useful information on the web. However, users might experience failure when search engines return irrelevant results that do not meet their real intentions. Such irrelevance is largely due to the enormous variety of user's contexts and backgrounds, as well as the ambiguity of texts. Personalized E-Learning (PEL) is a general category of search technique that aims at providing better search results, which are tailored for individual users needs. User's information has to be collected and analyzed to figure out the user intention behind the query issued. The solution to PEL can be generally categorized into two types, namely click-log-based methods and profile-based ones. The click-log based methods are straightforward. They simply impose bias to clicked pages in the users query history. Although this strategy has been demonstrated to perform consistently and it can only work on repeated queries from the same user, which is a strong limitation confining its applicability. The profile-based methods improve the search experience with complicated user-interest models generated from user profiling techniques. This method can be effective for almost all sorts of queries, but are reported to be unstable under some circumstances. Although there are pros and cons for both techniques, the profile-based PEL has demonstrated more effectiveness in improving the quality of E-learning recently, with increasing usage of personal and behavior information to profile its users, which is usually gathered implicitly from query history browsing history browsing click-through data bookmarks user documents and so forth. Such implicitly collected personal data can easily reveal a amount of users private life. Privacy issues rising from the lack of protection for such data, for instance the AOL query logs scandal, not only raise panic among individual users, but the data-publishers enthusiasm in offering personalized service. In fact, privacy concerns have become the major barrier for wide proliferation of PEL services.

1.1 PROBLEM STATEMENT

The two web based systems consisting of Generic E-Learning system and Semantic Search Engine system with objective of e-learning is to develop the Generic E-learning frame work and semantic search engine. In the traditional way of teaching, practicing, and assessing, the teachers design or choose assignments for weekly exercise sheet according to the course, the exercise sheet may be distributed as a printed document or made available online. The students can work through the exercise sheet at home and present their solution at the blackboard. The teacher gives feedback and the tutor may take notes about student's performance. For large groups of students, manual correction is labor and time-intensive but the problems are especially grave for programming assignments, with the rise in online education, the CDL wishes to integrate their modules into several distances learning course to attract more learning providers.

Online courses are instructional content which are delivered through online. The Hybrid courses content are in the class room settings and Web facilities courses content are partially in the classroom settings.

II. Related work

To protect the user privacy in profile based PEL, researchers have to consider the two contradicting effects during the search process. This attempts to improve the search quality with personalization utility of the user profile. Also they need to hide the privacy contents that are present in the user profile to place the privacy risk under control.

Few previous studies suggest that people are willing to compromise privacy, if the personalization by supplying user profile to the search engine yields better search quality. In an ideal case, gain can be obtained by personalization at the expense of only a small (less sensitive) portion of the user profile called generalized profile. Thus the privacy of the user can be protected without compromising the personalized search quality. There is a tradeoff between the search quality and at the level of the privacy protection achieved from the generalization.

2.1. Profile Based Personalization

The main focus of profile-based PEL in the previous works was on improving the search utility. The basic idea is to tailor the search results by referring to, often implicitly, a user profile that reveals an individual information goal. The previous solutions to PEL can be reviewed on two aspects, namely representation of profiles, and the measure of the effectiveness of personalization.

Many profile representation are available in literature to facilitate different personalization strategies. Previous techniques utilize term lists or bag of words to represent their profile. The most recent works of profiles are built in hierarchical structure due to their stronger descriptive ability, better scalability and higher access efficiency. Mapping from one ontology to another one is expressing of the way how to translate statements from ontology to the other one. Often it means translation between concepts and relations. In the simplest case it is mapping from one concept of the first ontology to one concept of the second ontology.

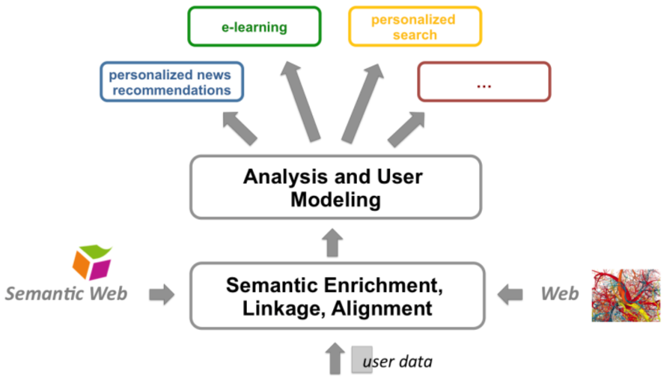

e-learning
personalized search
personalized news recommendations
...
Analysis and User Modeling
Semantic Web
Semantic Enrichment, Linkage, Alignment
Web
user data

Fig 1: the Search Engine based Profile Search

The hierarchical representations are constructed with the existing weighted topic hierarchy/graph and so on. Another work is built automatically via term-frequency analysis on the user data. in our proposed UPS framework, our focus is not on the implementation of the user profiles. Our framework can adopt any hierarchical representation based on taxonomy of knowledge.

2.2 User Interest Profiling

E-learning uses "concepts" to model the interests and preferences of a user. In mobile searches the location information is important so the concepts are further classified into two different types as content concepts and location concepts. The concepts are modeled as ontology's to capture the relationships between the concepts. The characteristics of the content concept sand location concepts are dissimilar. So, we propose two different techniques to build the content ontology and location ontology. This ontology's indicate a possible concept space from the user's queries which are maintained with Ontologies data for further preference adaptation. Ontologies are adopted to model the concept space in E-learning since they not only represent concepts but also capture the relationships between the concepts.

2.3. Personalized Ranking Functions

Ranking SVM (RSVM) is employed to learn a personalized ranking function for rank adaptation of the results according to the user content and location preferences while receiving the user's preferences. From the search results of the document features, a set of content concepts and location concepts can be extracted for a given query. Since each document can be represented by a feature vector, it can be treated as a point in the feature space. RSVM aims at finding a linear ranking function which holds many document preference pairs as possible, when preference pairs are used as the input. An adaptive implementation, SVM light available at, is used in our experiments. The two main issues in the RSVM training process are discussed below:

1. How to extract the feature vectors for a document;
2. How to combine the content and location weight vectors into one integrated weight vector.

III. PREVIOUS IMPLEMENTATIONS

The E-learning engine has become the most important portal for ordinary people who are looking for useful information on the web. However, when the search engine returns irrelevant results that do not meet their real intentions the users experience failures in this case. A major problem in E-learning is that the interactions between the users and search engines are limited by small form factors of the web devices. This result in submission of shorter more ambiguous queries by the web users compared to their E-learning counterparts. In order to return highly relevant results to the users, the E-learning engines must be able to profile the users interests and personalize the search according to the users profiles. To capture a users interests fro personalization, analyze the users Ontologies data. Most of the previous work assumed that all concepts are of the same type.

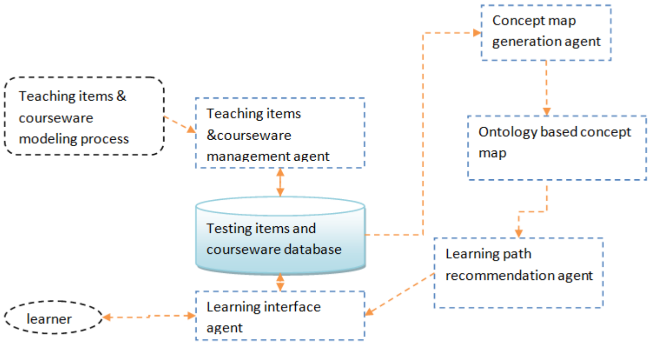

Concept map generation agent
Teaching items & courseware modeling process
Teaching items &courseware management agent
Ontology based concept map
Testing items and courseware database
Learning path recommendation agent
learner
Learning interface agent

Fig 2: Previous System Architecture based on Semitic Search Technical

3.1 USER PREFERENCES EXTRACTION AND PRIVACY PRESERVATION

Users preferences can be learned by collecting the concepts and click through from past search activities. These set of feature vectors are to be submitted along with future queries to the E-Learning server for search result re-ranking. Instead of transmitting all the detailed personal preference information to the server, E-learning allows the users to control the amount of personal information exposed. In this part, we first review a preference mining algorithm namely SpyNB Method, that is adopted in E-learning and also discuss how E-learning preserves user privacy. User behavior models are learned by SpyNB from preferences extracted from Ontologies data. Assuming that users click only on documents that are of interest to them, SpyNB treats the clicked documents as positive samples, and predict reliable negative documents from the unlabeled (i.e. unclicked) documents.

IV. SYSTEM IMPLEMETNATION

The system proposes a privacy-preserving personalized E-Learning framework UPS, which can be used to generalize profiles for each query according to user-specified privacy requirements. We formulate the problem of privacy-preserving personalized search as 5-Risk Profile Generalization based on the definition of two conflicting metrices: personalization utility and privacy risk for hierarchical user profile. A number of location-based search systems are designed to handle the queries that focus on location information. A location-based search system is designed for web documents. Location information's are extracted from the web documents which are converted into latitude-longitude pairs.

Advantages:

1. Generate quick reports

2. Make accurate and efficient calculations

3. Provide proper information briefly

4. Provide data security

5. Provide huge maintenance of records

6. The Flexibility of transactions can be completed in time

4.1 PROFILE-BASED PERSONALIZATION

It is an approach to personalize digital multimedia content based on the user profile information. Two main mechanisms were developed for this purpose: a profile generator that automatically creates user profiles representing the user preferences and a content-based recommendation algorithm that estimates the user's interest in unknown content by matching their profile to metadata descriptions of the content. These both features are integrated into a personalization system.

4.2 PRIVACY PROTECTION IN PEL SYSTEM

We propose a PEL framework called UPS that can generalize profiles in for each query according to user-specified privacy requirements. Two predictive metrics are proposed to evaluate the privacy breach risk and the query utility for hierarchical user profile. We develop two effective generalization algorithms for user profiles that allow query-level customization using the metrices that are proposed. An online prediction mechanism is provided based on the query utility for deciding whether to personalize a query in UPS or not. The efficiency and effectiveness of this framework is demonstrated using extensive experiments.

4.3 GENERALIZING USER PROFILE

The generalization process has to meet some of the specific prerequisites to handle the user profile. This is achieved by preprocessing the user profile. At first, the process initializes the user profile by considering the indicated parent user profile into account. This process adds the inherited properties to the properties of the local user profile. Then the process loads the data for the foreground and the background of the map according to the described selection in the user profile.

4.4 ONLINE DECISION

The profile-based personalization contributes little, even reduces the search quality while exposing the profile to a server would for sure risk the user's privacy. To address this problem, we develop an online mechanism to decide whether to personalize a query. The basic idea is if a distinct query is identified during generalization, the entire runtime profiling will be aborted and the query will be sent to the server without a user profile.

4.5 ALGORITHM IMPLEMENTATION ALGORITHMIC ANALYSIS:

1. Genetic Algorithm scheme:

- Generate the initial population of individuals
- Calculate the fitness value for each individual in that population
- Repeat on this generation until stop condition is met: (time limit, sufficient fitness achieved, etc.)
- Select the best-fit individuals for reproduction
- Create new individuals by applying crossover and mutation operations
- Evaluate the individual fitness of new individuals

2. K-means clustering algorithm:

1. Select k objects (patterns) randomly to be the seeds for the centroids of k clusters.
2. Assign each pattern to the centroid closest to the example, in this way k exclusive clusters are formed.
3. Calculate new centroids of the clusters. To do so average all attribute values of the examples belonging to the same cluster (centroid).
4. Check if the cluster centroids have changed If yes, start again the step 2. If not, cluster detection is finished and all patterns have their cluster memberships defined.

EVALUATION RESULT:

The first page provides more informative comparison. I found that Google and at least one other search engine returns 7% of results of queries in the first page. Google refers 7.9% queries to its own content on the first page of results without agreement from either rival search engine. Meanwhile, Bing and at least one other engine refer to Microsoft content in 3.2% of the queries. Bing references Microsoft content without agreement from either Google or Blekko in 13.2% of the queries:

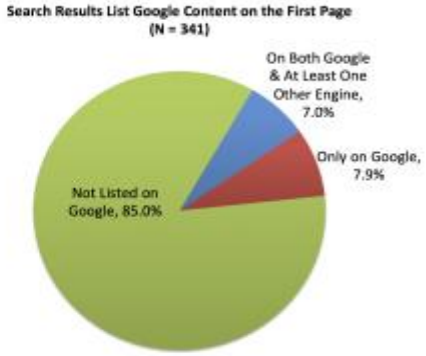
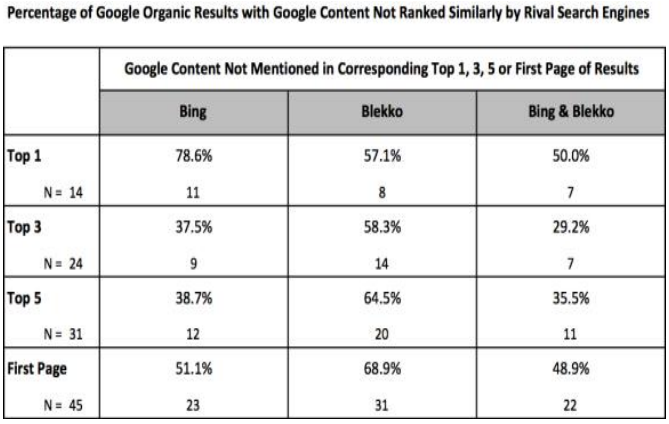


Fig 3: Search Results list Google Content on the First page



Percentage of Google Organic Results with Google Content Not Ranked Similarly by Rival Search Engines

	Google Content Not Mentioned in Corresponding Top 1, 3, 5 or First Page of Results		
	Bing	Blekko	Bing & Blekko
Top 1	78.6%	57.1%	50.0%
N = 14	11	8	7
Top 3	37.5%	58.3%	29.2%
N = 24	9	14	7
Top 5	38.7%	64.5%	35.5%
N = 31	12	20	11
First Page	51.1%	68.9%	48.9%
N = 45	23	31	22

Table: Percentage of Google Search results with Google Content based Search

When Google ranks its own content highly, at least one rival engine typically agrees with this ranking. For example, when Google places its own content in its Top 3 results, at least one rival agrees with this ranking in over 70% of queries. Bing especially agrees with Google's rankings of Google content within its Top 3 and 5 results, failing to include Google content that Google ranks similarly in only a little more than a third of queries.

A Closer Look at Google vs Bing

On E&L's own terms, Bing results are more biased than Google results; rivals are more likely to agree with Google's algorithmic assessment (than with Bing's) that its own content is relevant to user queries. Bing refers to Microsoft content other engines do not rank at all more often than Google refers its own content without any agreement from rivals. Figures 4 and 5 shows the same data presented above in order to facilitate direct comparisons between Google and Bing.

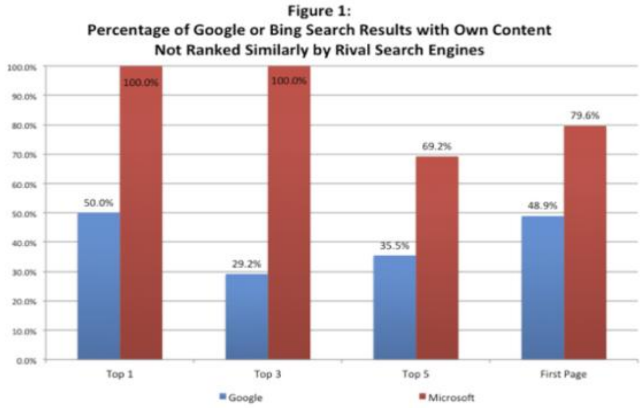


Fig 4: Percentage of Google or Bing Search Result

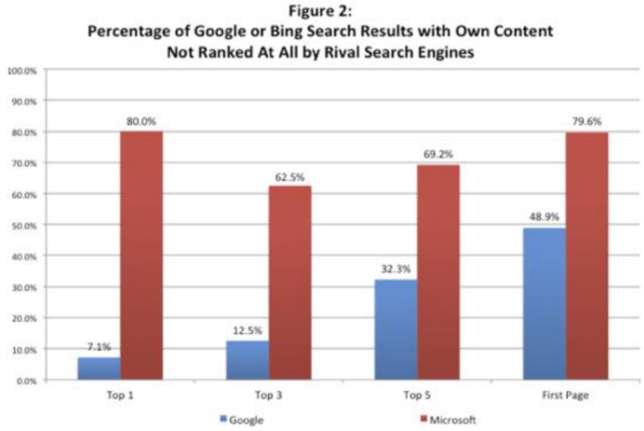


Fig 5: Percentage of Google or Bing Search Result

The Bing search results for these 32 queries are more frequently "biased" in favor of its own content than are Google's. The bias is greatest for the Top 1 and Top 3 search results.This study finds that Bing exhibits far more "bias" than E&L identify in their earlier analysis. For example, in E&L's study, Bing does not refer to Microsoft content at all in its Top 1 or Top 3 results; moreover, Bing refers to Microsoft content within its entire first page 11 times, while Google and Yahoo refer to Microsoft content 8 and 9 times, respectively. Most likely, the significant increase in Bing's "bias" differential is largely a function of Bing's introduction of localized and personalized search results and represents serious competitive efforts on Bing's behalf.

CONCLUSION

In this paper, firstly basic Semantic Web and Web Usage Mining notions are presented. Then the application of techniques coming from the new emerging area of Semantic Web Mining in the domain of E-Learning systems and analyzed the significant role of ontology's are discussed. We expounded and argued about our proposed approach for producing recommendations to users in a given e-Learning corpus. Finally, we have concluded with the description of the recommendation engine's operation and presented an algorithm for making effective recommendations.As shown in the paper, the proposed personalization scenario tries to integrate the Semantic Web vision by using Ontologies Using Mining techniques in order to better service the needs and the requirements of learners. We strongly believe that the combination of domain's ontology and frequent item sets, which include all the information about users' navigational attitude, enhances the whole process and produces better recommendations. The system first finds an initial recommendation set and then uses the frequent itemsets to enrich it, taking into consideration other users' navigational activity. In this way, we reduce the time we spend on parsing all frequent item sets and association rules. We focus only on those sets that come out from the combination of the active user session and the ontology's recommendations. The time reduction arises because of the fact that frequent item sets are filtered through the ontology's recommendation set resulting in a smaller searching space

References:

[1] Z. Dou, R. Song, and J.-R. Wen, "A Large-Scale Evaluation and Analysis of Personalized Search Strategies," Proc. Int'l Conf. World Wide Web (WWW), pp. 581-590, 2007.

[2] Andreas Krause, Eric Horvitz" A Utility-Theoretic Approach To Privacy in Online Services", 2010

[3] Yabo Xu, Benyu Zhang, Zheng Chen, Ke Wang, "Privacy-Enhancing Personalized Web Search", 2009

[4] Zhicheng Dou, Ruihua Song, JiRong Wen, "A Largescale Evaluation and Analysis of Personalized Search Strategies", 2001