



## Sustaining Confidentiality Protection in Personalized E-Learning for Ontology

<sup>1</sup>A. Devi, <sup>2</sup>P. Hariharan

<sup>1</sup>Research Scholar, <sup>2</sup>Assistant Professor

<sup>1,2</sup> Dept. of Computer Science, Adhiparasakthi College of Arts and Science (Autonomous),  
Kalavai, Vellore, Tamil Nadu, India

---

**Abstract---** *Personalized E-learning (PEL) has demonstrated its effectiveness in improving the quality of various searches on the Internet. However, evidences show that users' reluctance to disclose their private information during search has become a major barrier for the wide proliferation of PEL. We propose a personalized E-learning engine, PEL that captures the users' preferences in 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 concept 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 o offline tasks that includes data preparation ontology creation and usage mining and on of online tasks that concerns the production of recommendations. We also provide an online predication mechanism for deciding whether personalizing a query is beneficial. Extensive experiments demonstrate the effectiveness of our framework.*

**Keywords:** *PEL, Ontology, Clustering, Content Concept, Location Concept, Rival Search Engine*

---

### I. INTRODUCTION

The E-learning engine has long 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 users' contexts and backgrounds, as well as the ambiguity of texts. Personalized E-learning (PEL) is a general category of search techniques aiming at providing better search results, which are tailored for individual user needs. As the expense, user information has to be collected and analyzed to figure out the user intention behind the issued query. The solutions to PEL can generally be 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 user's query history. Although this strategy has been demonstrated to perform consistently and considerably well, it can only work on repeated queries from the same user, which is a strong limitation confining its applicability. In contrast, profile-based methods improve the search experience with complicated user-interest models generated from user profiling techniques. Profile-based methods can be potentially effective for almost all sorts of queries, but are reported to be unstable under some circumstances. Although there are pros and cons for both types of PEL 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 click-through data bookmarks user documents, and so forth. Unfortunately, such implicitly collected personal data can easily reveal an amount of user's 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 also dampen the data-publisher's enthusiasm in offering personalized service. In fact, privacy concerns have become the major barrier for wide proliferation of PEL services.

### 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 traditional way of teaching, practicing, and assessing the teacher designs or chooses assignments for a weekly exercise sheet according to the course, the exercise sheet may be distributed as a printed document or made available online. So that, the students can work through the exercise sheet at home and present their solutions at the blackboard. So 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 learning modules into several distances learning course to attract more Learning providers. Online courses are mostly instructional content which is delivered through online and the Hybrid courses-content in the class rooms and Web facilities courses content are delivered partially in the class room setting.

## II. RELATED WORKS

To protect user privacy in profile-based PEL, researchers have to consider two contradicting effects during the search process. On the one hand, they attempt to improve the search quality with the personalization utility of the user profile. On the other hand, they need to hide the privacy contents existing in the user profile to place the privacy risk under control.

A 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, significant gain can be obtained by personalization at the expense of only a small (and less-sensitive) portion of the user profile, namely a generalized profile. Thus, user privacy can be protected without compromising the personalized search quality. In general, there is a tradeoff between the search quality and the level of privacy protection achieved from generalization.

### Profile Based Personalization

Previous works on profile-based PEL mainly focus on improving the search utility. The basic idea of these works is to tailor the search results by referring to, often implicitly, a user profile that reveals an individual information goal. In the remainder of this section, we review the previous solutions to PEL on two aspects, namely the representation of profiles, and the measure of the effectiveness of personalization.

Many profile representations are available in the literature to facilitate different personalization strategies. Earlier techniques utilize term lists/vectors or bag of words to represent their profile. However, most recent works build profiles in hierarchical structures due to their stronger descriptive ability, better scalability, and higher access efficiency.

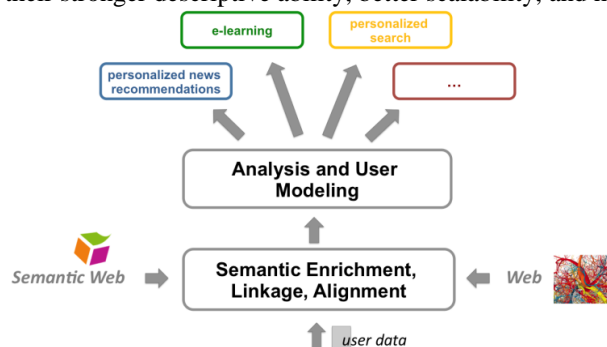


Fig 1: the Search Engine based Profile Search

The majority of the hierarchical representations are constructed with existing weighted topic hierarchy/graph, such as and so on. Another work in builds the hierarchical profile automatically via term-frequency analysis on the user data. In our proposed UPS framework, we do not focus on the implementation of the user profiles. Actually, our framework can potentially adopt any hierarchical representation based on taxonomy of knowledge.

### User Interest Profiling

E- LEARNING uses “concepts” to model the interests and preferences of a user. Since location information is important in mobile search, the concepts are further classified into two different types, namely, content concepts and location concepts. The concepts are modeled as ontology’s, 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 ontology’s indicate a possible concept space arising from a user’s queries, which are maintained along with the Ontologies data for future preference adaptation. In E- LEARNING, we adopt ontology’s to model the concept space because they not only can represent concepts but also capture the relationships between concepts due to the different characteristics of the content concepts and location concepts.

### Personalized Ranking Functions

Upon reception of the user’s preferences, Ranking SVM (RSVM) is employed to learn a personalized ranking function for rank adaptation of the search results according to the user content and location preferences. 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. Using the preference pairs as the input, RSVM aims at finding a linear ranking function, which holds for as many document preference pairs as possible. An adaptive implementation, SVM light available at, is used 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 combine the content and location weight vectors into one integrated weight vector.

## III. PREVIOUS IMPLEMENTATION

The E-learning engine has long 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 intention. A major problem in E-learning is that the interactions between the users and search engines are limited by the

small form factors of the web devices. As a result, web users tend to submit shorter, hence, more ambiguous queries compared to their E-learning counterparts. In order to return highly relevant results to the users, E-learning engines must be able to profile the users' interests and personalize the search results according to the users' profiles. A practical approach to capturing a user's interests for personalization is to analyze the user's Ontologies data. However, most of the previous work assumed that all concepts are of the same type. Observing the need for different types of concepts.

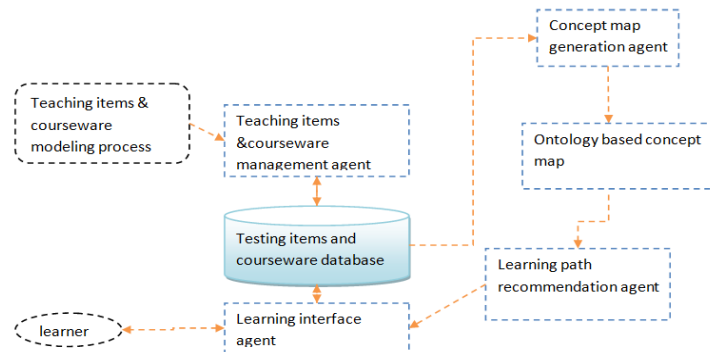


Fig 2: Previous System Architecture based on Semantic Search Technique

### User Preferences Extraction and Privacy Preservation

Given that the concepts and clickthrough data are collected from past search activities, user's preference can be learned. These search preferences, inform of a 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 section, we first review a preference mining algorithms, namely SpyNB Method that we adopt in E-LEARNING, and then discuss how E-LEARNING preserves user privacy. SpyNB learns user behavior models from preferences extracted from Ontologies data. Assuming that users only click 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 IMPLEMENTATION

It proposes a privacy-preserving personalized E-learning framework UPS, which can generalize profiles for each query according to user-specified privacy requirements. Relying on the definition of two conflicting metrics, namely personalization utility and privacy risk, for hierarchical user profile, we formulate the problem of privacy-preserving personalized search as 5-Risk Profile Generalization. In order to handle the queries that focus on location information, a number of location-based search systems designed for location queries have been proposed. Proposed a location-based search system for web documents. Location information was extracted from the web documents, which was converted into latitude-longitude pairs.

### Advantages:

1. To generate the quick reports
2. To make accuracy and efficient calculations
3. To provide proper information briefly
4. To provide data security
5. To provide huge maintenance of records
6. The Flexibility of transactions can be completed in time

### Profile-Based Personalization

It introduces an approach to personalize digital multimedia content based on user profile information. For this, two main mechanisms were developed: 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 her profile to metadata descriptions of the content. Both features are integrated into a personalization system.

### 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 simple but effective generalization algorithms for user profiles allowing for query-level customization using our proposed metrics. We also provide an online prediction mechanism based on query utility for deciding whether to personalize a query in UPS. Extensive experiments demonstrate the efficiency and effectiveness of our framework.

### Generalizing User Profile

The generalization process has to meet specific prerequisites to handle the user profile. This is achieved by preprocessing the user profile. At first, the process initializes the user profile by taking the indicated parent user profile into account.

The process adds the inherited properties to the properties of the local user profile. Thereafter the process loads the data for the foreground and the background of the map according to the described selection in the user profile.

**ALGORITHMIC 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:**

- Select randomly k objects (patterns) to be the seeds for the centroids of k clusters.
- Assign each pattern to the centroid closest to the example, forming in this way k exclusive clusters of examples.
- Calculate new centroids of the clusters. For that purpose average all attribute values of the examples belonging to the same cluster (centroid).
- 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 entire first page of results provides a more informative comparison. I find that Google and at least one other engine return Google content on the first page of results in 7% of the queries. Google refers to its own content on the first page of results without agreement from either rival search engine in only 7.9% of the queries. 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:

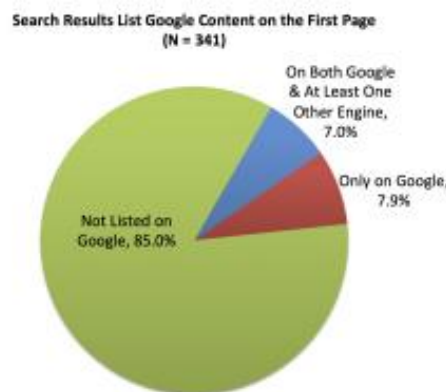


Fig 3: Search Results list Google Content on the First page

Table 1: Percentage of Google Search results with Google Content based Search 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
<b>Top 1</b> N = 14	78.6% 11	57.1% 8	50.0% 7
<b>Top 3</b> N = 24	37.5% 9	58.3% 14	29.2% 7
<b>Top 5</b> N = 31	38.7% 12	64.5% 20	35.5% 11
<b>First Page</b> N = 45	51.1% 23	68.9% 31	48.9% 22

So 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 v. 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 display the same data presented above in order to facilitate direct comparisons between Google and Bing.

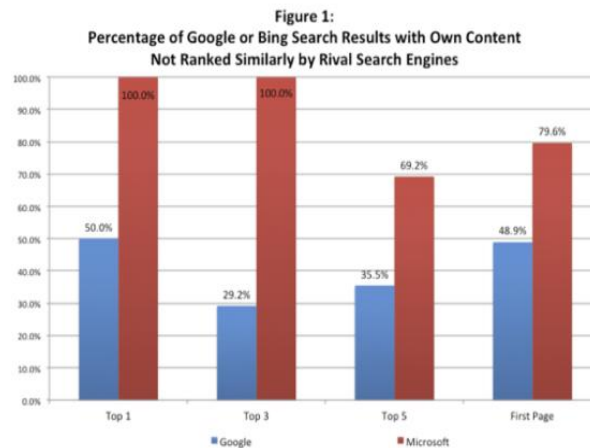


Fig 4: Percentage of Google or Bing Search Result with Own Content Not Ranked Similarly by Rival Search Engine

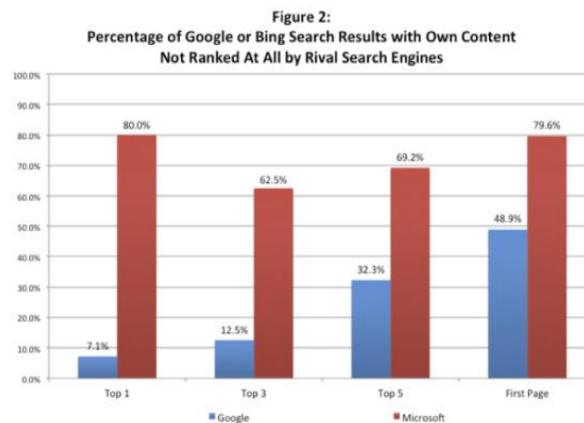


Fig 5: Percentage of Google or Bing Search Result with Own Content Not Ranked at all by Rival Search Engine

As Figures 4 and 5 illustrate, 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.

My 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.

### V. CONCLUSION

In this paper, we presented firstly basic Semantic Web and Web Usage Mining notions. Then, we discussed about 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 Ontologies. We expounded and argued about our proposed approach for producing recommendations to users in a given e-Learning corpus. Finally, we 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 with 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 item sets 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
- [5] John S. Breese, David Heckerman, and Carl Kadie, "Empirical Analysis of Predictive Algorithms For Collaborative Filtering".2010.
- [6] T. Urdan, & C. Weggen, "Corporate e-learning: exploring a new frontier", WR Hambrecht and Co, 2000.
- [7] T. Wentling, C. Waight, J. Gallaher, J. La Fleur, C. Wang, & A. Kanfer, "E-learning: a review literature", Knowledge and Learning Systems Group, National Center for Supercomputing Applications, University of Illinois, 2000.
- [8] K. Fry, "E-learning markets and providers: some issues and prospects", Education and Training, Emerland, 43(4), 2001, 233-239.
- [9] K. Markellos, P. Markellou, M. Rigou, S. Sirmakessis, & A. Tsakalidis, "Web personalizationfor enhancing e-learning experience", Proc. 5<sup>th</sup> International Conf. on Information Communication Technologies in Education, ICICTE, Samos Island, Greece, 2004.
- [10] P. De Bra, P. Brusilovsky, & G.J. Houben, "Adaptive hypermedia: from systems to framework", ACM Computing Surveys, 31(4es), 1999.