International Journal of Pure and Applied Mathematics
Volume 117 No. 21 2017, 391-406
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue
ijpam.eu

Web Personalization through Probabilistic Graphical Model Incorporated Passive Feedback Approach

[1]B.J. Doddegowda, [2]Sunil Kumar S. Manvi and [3]G.T. Raju

[1]School of Computing & IT,
Reva University,
Bengaluru, Karnataka.
bjdgowda10@gmail.com

[2]School of Computing & IT,
Reva University,
Bengaluru, Karnataka.
sunil.manvi@revainstitution.org

[3]Departmen of CSE,
RNSIT, Bengaluru,
VTU, Karnataka.
gtraju1990@yahoo.com

Abstract

Currently, due to the explosive growth of the Web, millions of new Web documents are being added in a short span of time. These documents provide information in a number of semantic categories such as sports, politics, consumer goods etc. Due to the presence of large number of semantic categories, identifying the users intended information requirement becomes challenging, because quite often, the information querying is performed through keyword queries which may create ambiguity in interpreting the user query.

Web Personalization is a technique which orients the Web search engine to cater for user's information requirements. The passive feedback framework for Web Personalization collects the statistics related to user information requirements without voluntarily involving the user. This framework is advocated for those users who are not comfortable in providing active feedback about their information requirements.

Veritably, learning models seem attractive to implement passive feedback framework. But, most of the learning models require training set

to perform effective functioning. Since, the users are un-comfortable in providing any direct feedback for their information requirements, creating the required training set for a specific user becomes a challenge. Some of the contemporary techniques advocate using small degree of active feedback, which defeats the very purpose of building a complete passive feedback framework for Web personalization.

In this work, a new passive feedback framework for Web Personalization, which utilizes probability graphical model is presented. The training set for the proposed framework is generated without the need of direct feedback from the user. This proposed framework empirically out-performs the contemporary technique in-terms of substantial increase in search effectiveness.

1. Introduction

Overview on Web Personalization

Web is facing the challenge of Big Data, wherein, the challenge not only lies in the query execution performance, but, also in providing user relevant results. Quite often, the Web search engines provide keyword query interface, which may result in ambiguity regarding the exact context interpretation of the query. For example, the query John Adams might easily refer to the former U.S president, but, the user might be interested about a scientist who also shares the same name. Without the aid of Web Personalization, and due to the Big Data effect, extremely large number of Web documents related to the query, which contain information about all the individuals, events and monuments sharing the same name might be returned. This humongous result set makes it extremely difficult for the user to search for the relevant documents.

One of the straightforward methods to overcome ambiguity is to force the user to submit keyword queries with context description. But, some of the users might not be comfortable in revealing their exact context, and even with context description, ambiguity might still persist. To overcome this issue, a technique called Web Personalization was proposed in the literature [1-3]. This technique aligns the Web search engine to cater specific users information needs.

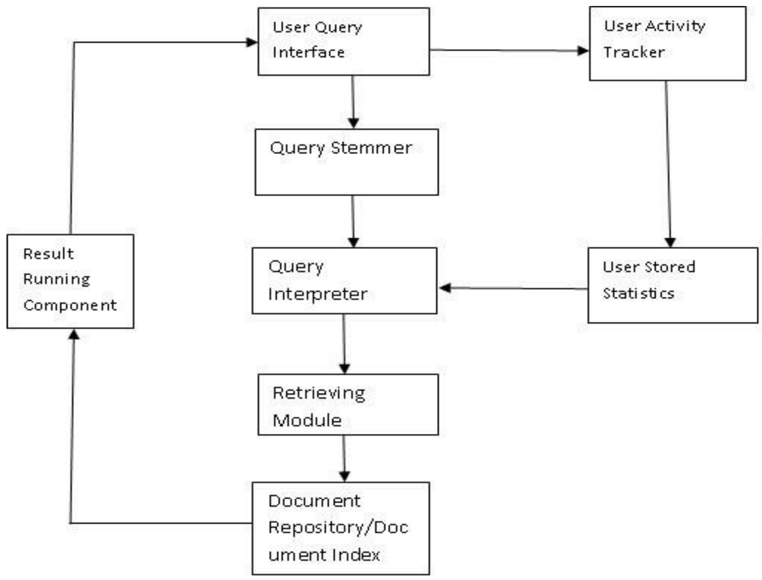

Figure 1: Passive Feedback Framework Flow Diagram

The two major frameworks for Web Personalization are: active feedback and passive feedback. The active feedback framework expects the user to provide feedback on the result set of each query, indicating the relevance level of each document in the result set. This framework might not appeal to every user, because some of the users might not be keen in providing the required feedback, and such users, might be deprived from the benefits of Web Personalization.

Since, each user to a large degree has uniqueness in information requirements, utilizing the feedback knowledge of other users is largely ineffective.

The passive feedback framework does not require user's feedback, but, collects the user specific search statistics indirectly by using automated trackers, query logs etc. This framework, quite often, provides lesser effectiveness compared to active feedback framework, but, it is an inevitable choice for those users who are uncomfortable in providing the required result feedback.

The generic flow diagram of the Web Personalization technique, which employs passive feedback framework, is illustrated in Figure 1. The query stemmer module performs stemming of the user query submitted in the user query interface to remove unwanted terms. The query interpreter module interprets the context of the query, and maps the user query to the related semantic category/s by utilizing the user’s stored statistics from previous search sessions (user activity tracker). The retrieving module retrieves the related documents or document addresses from the document repository or document index confined to the selected semantic category/s, and presents the result set to the user. Optionally, ranking of the result set is performed to sort the documents on user relevance (result ranking component).

Motivation

Since, deciding the semantic category of a user query is based upon learning from user's stored statistics, application of learning models to solve this problem appears attractive. But, most of the effective learning models require training set to train the model parameters. In a purely passive feedback framework, the user cannot be forced to give the required feedback to generate the training set. Currently, in the literature, the problem of generating the training set for learning models utilized in purely passive feedback systems has not been effectively addressed.

Contributions

In this work, the following contributions are made:

1. A novel passive feedback framework for Web Personalization based on probability graphical model is proposed. This model provides better search effectiveness with respect to user relevance than contemporary technique when demonstrated empirically.
2. The required training set is generated without the need of any active feedback from the user. The training set generation mechanism justifies its design by providing excellent user relevant results.

Organization of the Paper

This paper is organized as follows: Section 2 describes the related work. The proposed passive feedback oriented Web Personalization technique is presented in Section 3. The empirical results and their corresponding discussions are presented in Section 4. Finally, the work is concluded with future directions in Section 5.

2. Related Work

Multiple contributions have been made to perform Web Personalization by using users activity data [1-3]. Some of the considered parameters were user goals, personal interest, context, prior knowledge, language, country etc. It was highlighted in [4-7] that, inaccurate Web Personalization can have great negative impact on the user's satisfaction. It was also highlighted that, passive feedback framework are especially vulnerable to go astray from the intended goal, because required knowledge(including training set) might not be available.

Community based Web Personalization techniques [8-12] cluster a group of individuals, and provide a common Personalization for all the members in the cluster. The grouping of individuals is automated based on the common pattern of searching. By using learning models, the users are clustered, and a common Personalization is provided to all the users in the cluster. It was argued in [8, 10] that, extreme individual Personalization might reduce the overall effectiveness of the system. In [9], the individual Personalization model was augmented through the community Personalization model. However, community based Web Personalization is effective in only some cases, and the remaining cases demand individual Web Personalization.

Aggregate level Web Personalization systems [13-16] do not perform Personalization over specific users, but, the documents which are most likely to be accessed in future will be identified, and given preference in the final ranking of the result set. In a stricter sense, aggregate level Web Personalization cannot be recognized as genuine Personalization scheme, but, only as a result ranking module.

It was argued in [17] that, the user keyword queries require transformation to understand the intended context. There are multiple techniques proposed in the literature to perform query transformation. In [18-22] the terms in the user stored statistics are selected to expand the query. In [23-26] the terms present in the top ranked documents in the result set of the previous queries submitted by the user are used for expanding the query. However, query transformation can only aid in the context understanding of the user query, but, usually cannot on its own decide the intended context of the query.

Recently, a purely passive feedback framework oriented Web Personalization system was proposed [27]. This system achieves Personalization based on user's click data. However, the system is too naive to provide robust effectiveness.

3. Proposed Passive Feedback Framework for Web Personalization

Passive Feedback Framework Parameters

The user Web activity information will be stored for each session. Each session can be considered as the time interval in which, the user submits a query, and goes through the corresponding result set. The Web documents are divided into m semantic categories, and each document is assumed to have a single category affiliation. The session Web activity information will be stored in the form of tuples. Here, each tuple represents a semantic category, so, there will be m tuples in each session. Let, t_{ij} indicate the tuple belonging to the i^{th} session and j^{th} semantic category. Let, 's' be the number of sessions for which the user activity data is stored. Here, $1 \le i \le s$ and $1 \le j \le m$. The following attributes are considered for each tuple:

1. *Activity Time*: The first attribute considers the amount of time spent by the user in going through the documents belonging to a particular semantic category, and for a particular session. Here, $at(t_{ij})$ indicates the attribute value stored in tuple $\mathbf{t}_{ij}$ regarding activity time for the i^{th} session, and corresponding to j^{th} semantic category.
2. *Number of Document Access:* The second attribute considers the number of documents corresponding to a semantic category visited by the user during a particular session. Here, $nd(\mathbf{t}_{ij})$ indicates the attribute value stored in tuple $\mathbf{t}_{ij}$ regarding number of visited documents for the i^{th} session, and corresponding to j^{th} semantic category.
3. *Number of Consecutive Document Access:* The third attribute considers the number of times the user accesses two documents belonging to the same semantic category consecutively in a particular session. Here, $cd(\mathbf{t}_{ij})$ indicates the attribute value stored in tuple $\mathbf{t}_{ij}$ regarding number of times the user consecutively accesses two documents belonging to the j^{th} semantic category, and for the i^{th} session.
4. *User Location:* The fourth and the final attribute considers the location of the user in a particular session. The location refers to some country, and each country has a unique identification which $\in$ N. Here, $co(\mathbf{t}_{ij})$ indicates the attribute value stored in tuple $\mathbf{t}_{ij}$ regarding the location of the user corresponding to the i^{th} session.

Generation of Training Set

The user activity data is split into two exclusive subsets. The rst subset called the training set contains session data for $s_1(s_1 < s)$ sessions, and the second subset called the test set contains session data for $s_2(s_2 < s)$ sessions. Here, $s_1 + s_2 = s$ and $s_1 >> s_2$. Let, *tr_set* indicate the training set, and *te_set* indicate the test set. Also, $|tr\_set| = s_1$ and $|te\_set| = s_2$.

Each tuple $t_{ij} \in tr\_set$ is labeled with the label l_{ij} = 0 or 1. Here, $l_{ij} = 1$ indicates that, the tuple $\mathbf{t}_{ij}$ is relevant for the query corresponding to the i^{th} session. Only a single tuple is considered relevant for a particular session, and $l_{ij} = 0$ indicates that, the tuple $\mathbf{t}_{ij}$ is irrelevant for the query corresponding to the i^{th} session.

The labeling of i^{th} session tuples is performed by performing the optimization task represented in Equation 1. Here, $\widehat{t_{ij}}$ is that tuple, which satisfies the optimization condition. The labeling of the tuples ∈ i^{th} session is performed as represented in Equation 2.

$$\hat{t}_{ij} = \max(nd(\mathbf{t}_{ij}) + at(\mathbf{t}_{ij}) + cd(\mathbf{t}_{ij})), \text{ Here session } \mathbf{t}_{ij} \in tr\_set \quad (1)$$

$$l_{ij} = \begin{cases} 1, & for \ \widehat{t_{ij}} \\ 0, & \text{other tuples} \end{cases} \quad (2)$$

Problem Statement

Let, $l_j = 1$ indicate that the j^{th} semantic category is relevant to the user, and $l_j = 0$ indicate that it is irrelevant. The task is to design the probability density function, which accomplishes the selection of user required semantic categories by calculating the relevance probability as represented in Equation 3.

$$\boxed{\text{Relevance probability for } j^{th} \text{ semantic category} = P\left(l_j = 1 \middle| tr\_set, te\_set \right)} \quad (3)$$

Graphical Model

Each semantic category is associated with a random vector, which is represented in Equation 4. Here, $\mathbf{T}_j$ is the random vector associated with the j^{th} semantic category, and l_j;, $at(j)$, $nd(j)$, $cd(j)$ and $co(j)$ are the random variables which ∈ $\mathbf{T}_j$.

$$\mathbf{T}_j = \begin{bmatrix} l_j \\ nd(j) \\ cd(j) \\ at(j) \\ c0(j) \end{bmatrix} \quad (4)$$

For any $\mathbf{t}_{ij} \in tr\_set$ $(1 \leq i \leq s_1)$ is considered as an observation or sample of the random vector $\mathbf{T}_j$. So, l_{ij}, $nd(\mathbf{t}_{ij})$, $at(\mathbf{t}_{ij})$, $cd(\mathbf{t}_{ij})$ and $co(\mathbf{t}_{ij})$ are considered as the observations or samples of the random variables l_j, $nd(j)$, $at(j)$, $cd(j)$ and $co(j)$ respectively. Based on the empirical analysis, the graphical model illustrated in Figure 2 is defined. Here, $cd(j)$ and $nd(j)$ are correlated random variables, because there is a tendency that, as the number of consecutive document access corresponding to a semantic category increases, it also influences the increase in the number of document access in the same semantic category.

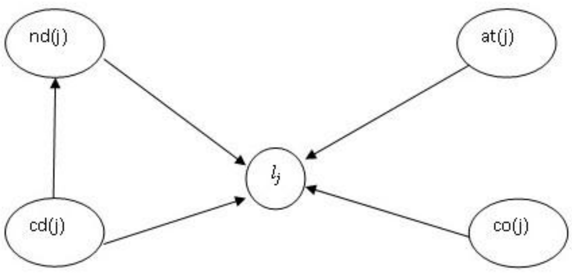

Figure 2: Proposed Graphical Model

Based on the graphical model and chain rule of probability, the conditional probability of l_j with respect to *nd(j), at(j), cd(j)* and *co(j)* is represented in Equations 5 and 6.

$$P(l_j | nd(j), cd(j), at(j), co(j)) = \frac{P(l_j, nd(j), cd(j), at(j), co(j))}{P(nd(j), cd(j), at(j), co(j))} \quad (5)$$

$$\boxed{P(l_j | nd(j), cd(j), at(j), co(j)) = \frac{P(l_j, nd(j), cd(j), at(j), co(j))}{P(nd(j), cd(j)), P(at(j))P(co(j))}} \quad (6)$$

Based on the empirical analysis, the distribution functions are represented in Equations 7, 8, 9 and 10. Here, Φ refers to normal distribution, $\mu 1 \in R^5$, $\Sigma_1 \in R^{5x5}$, $\mu 2 \in R^2$, $\Sigma_2 \in R^{2x2}$, $\mu 3$, $\mu 4$, σ_3^2 and $\sigma_4^2 \in R$. The calculation techniques of these distribution parameters are outlined in the Appendix section.

$$P(lj, nd(j), cd(j), at(j), co(j)) \sim \Phi(\mu_1, \Sigma_1) \quad (7)$$

$$P(nd(j), cd(j)) \sim \Phi(\mu_2, \Sigma_2) \quad (8)$$

$$P(at(j)) \sim \Phi(\mu_3, \sigma_3^2) \quad (9)$$

$$P(co(j)) \sim \Phi(\mu_4, \sigma_4^2) \quad (10)$$

Generation of Master Tuple from Test Set

For each semantic category, a master tuple is created by using the tuples $\in$ *te_set*. This scenario is illustrated in Equation 11. Here, $\vec{t_j}$ is the master tuple for the j^{th} semantic category.

$$\boxed{\vec{t_j} = \sum_{i=1}^{s_2} t_{ij} \quad where \quad t_{ij} \in te\_set} \quad (11)$$

The relevance probability for $\vec{t_j}$ is represented in Equation 12.

Here, $nd(j) = nd(\vec{t_j})$, $cd(j) = cd(\vec{t_j})$,

$at(j) = at(\vec{t_j})$, *and* $co(j) = co(\vec{t_j})$.

$$\boxed{P(lj|\vec{t_j}) = P(lj | nd(j), cd(j), at(j), co(j))} \quad (12)$$

Algorithm

The proposed Web Personalization technique is outlined in Algorithm 1. The user activity data will be used in generating the required training set as described in Section 3.2. The distribution parameters, which are required in calculating the relevance probability of a semantic category represented in Equation 6 are calculated as described in the Appendix.

The master tuple for each semantic category is generated by using Equation 11, and the corresponding relevance probability is calculated by using Equation 12. The semantic categories are sorted in non-increasing order, based on their relevance probability values. This ordered semantic categories for user U is used to resolve ambiguity if it arises in the future queries submitted by U. Higher the ordered list entry, more relevant will be the semantic category to the user U.

Algorithm 1 Proposed Web Personalization Technique (User U)
Generate the training set as outlined in Section 3.2.
Calculate the distribution parameters represented in Equation 6 for each semantic category $j(1 \leq j \leq m)$ by using the procedure outlined in Appendix.
for j = 1 to m **do**
Calculate $\vec{t_j}$ by using Equation 11.
Compute the relevance probability P ($l_j = 1 | \vec{t_j}$) by using Equation 12.
end for
Sort the semantic categories based on their relevance probability in non-increasing order.
Higher the position of a semantic category in the ordered list, higher will be its priority.
For any query launched by user U, if ambiguity arises due to the semantic interpretation of the query, then, the qualified semantic categories are given preference based on their priority values.

4. Results and Discussions

System Setup

For the goal of empirical performance analysis, the proposed Web Personalization technique is implemented inside a customized prototype Web search engine. Web documents belonging to three different Web datasets namely: IMDB, Wikipedia and Mondial are used for performing Web query search. Each user submits the query to the Web search engine interface, and the relevant documents extracted by the searching component of the Web search engine is ranked by using state of the art IR ranking functions.

If the user submits an ambiguous query, then, the qualifying Web documents will be ordered in two phases.

In the first phase, the Web documents are grouped based on their semantic categories. The group priority will be decided based on their corresponding semantic category priority values. Higher priority groups will be presented first to the user. In the second phase, inside each group, the documents will be ranked based on the IR ranking functions.

The empirical setup parameter values are outlined in Table 1. Totally, 20 users were used in the empirical analysis. The required user activity information was recorded through the user activity recording module incorporated inside the Web search engine.

Table 1: Empirical Setup Parameter Values

Empirical Setup Parameters	Used Values
Total number of Web documents	Around 10^5
Number of semantic categories (m)	100
Number of users	20
Number of Queries submitted by each user (s)	100
Number of queries in the training set (s_1)	80
Number of queries in the test set (s_2)	20

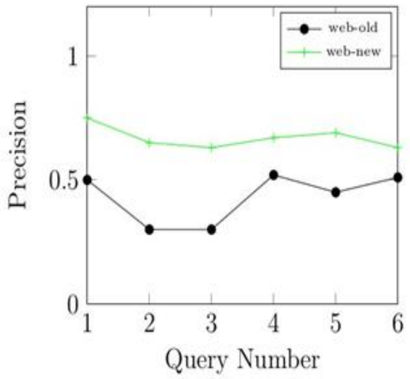

Figure 3: Queries vs Precision

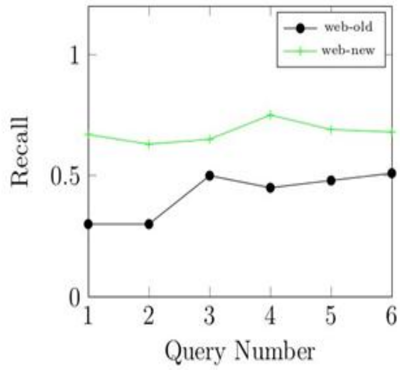

Figure 3: Queries vs Recall

Each user had submitted 100 queries, and the user activity in the provided non-personalized result set was recorded to create the required test set and training set.

To perform the empirical analysis, randomly six users are selected from the available 20. Each user will submit a new single query to the Web search engine, and the provided personalized results will be subjected to user feedback. The user will assign a relevance value between [0 10] to each Web document of the result set, wherein, 0 indicates that the document is irrelevant, and 10 indicates it is the most relevant. Top-100 results for each query will be presented to the user.

The proposed Web Personalization technique, for the ease of reference will be referred as web-new, and it will be compared against contemporary passive feedback oriented Web Personalization technique [27], which will be referred as web-old.

Empirical Results Discussions

The designed experiment analyzes the performance of *web-old* and *web-new* with respect to the six IR metrics: *Precision, Recall, Average Precision, Reciprocal Rank, Fall-Out* and F_β. The result of this analysis is illustrated in Figures 3, 4, 5, 6, 7 and 8. It is evident that, *web-new* out performs web-old in all the six IR performance metrics, mainly due to the effectiveness of the proposed model, which considers the most effective parameters to decide user affinity towards the available semantic categories.

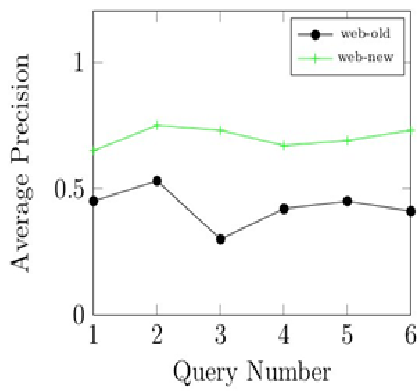

Figure 5: Queries vs Avg Precision

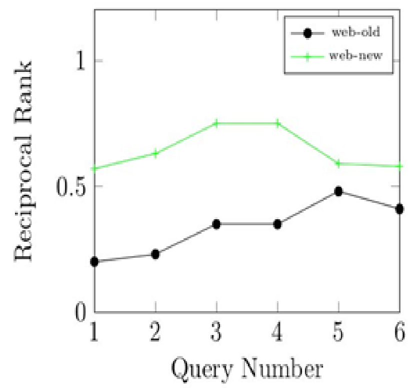

Figure 6: Queries vs Reciprocal Rank

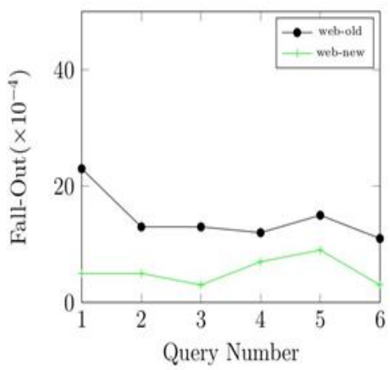

Figure 7: Queries vs Fall-Out

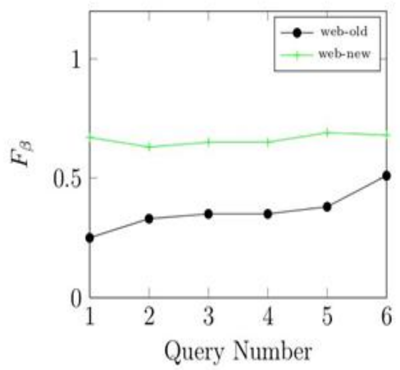

Figure 8: Queries vs F_3

5. Conclusion

In this work, the importance of Web Personalization was highlighted. The requirement of passive feedback framework for Web Personalization and its associated challenges were highlighted. The related work in the literature, and corresponding shortcomings were presented. For the proposed Web Personalization technique, the training set was generated without any active intervention of the user. A graphical model oriented Web Personalization technique was presented, which substantially outperformed the contemporary technique [28] in the empirical evaluations.

In future, it will be exciting to extend the proposed Web Personalization technique from a purely passive feedback framework to an active feedback framework, and analyze the gain in effectiveness, when compared with the passive feedback framework and other contemporary active framework Web Personalization techniques.

References

[1] Steichen B., Oconnor A., Wade V., Personalization in the wild: providing personalization across se-mantic, social and open-Web resources, In 22nd ACM Conference on Hypertext and Hypermedia (Hypertext) (2011).

[2] Stamou S., Ntoulas A., Search personalization through query and page topical analysis, User Model, User-Adapt. Interact., (2009).

[3] Speretta M., Gauch S., Personalized search based on user search histories. In IEEE/WIC/ACM International Conference on Web Intelligence, Compiegne University of Technology, Compiegne (2005), 622-628.

[4] Shen X., Tan B., Zhai C., Implicit user modeling for personalized search, 14th ACM International Conference on Information and Knowledge Management (CIKM (2005), 824831.

[5] Vassiliou C., Stamoulis D., Spiliotopoulos A., Martakos D., Creating adaptive Web sites using personalization techniques: a unied, integrated approach and the role of evaluation, In Patel, N.V. (ed.) Adaptive Evolutionary Information Systems, IGI Publishing, Hershey (2003), 261-285.

[6] Espinoza F., HK K., An interactive interface to an adaptive information system, In User Modeling for Information Filtering on the World Wide Web Workshop, Hawaii, USA, (1995).

[7] De La Passardiere B., Dufresne A., Adaptive navigational tools for educational hypermedia, In Tomek, I. (ed.) Computer Assisted Learning, Lecture Notes in Computer Science, Springer, Berlin (1992), 555-567.

[8] Teevan J., Morris M.R., Bush S., Discovering and using groups to improve personalized search, 2nd ACM International Conference on Web Search and Data Mining (WSDM), (2009).

[9] Sugiyama K., Hatano K., Yoshikawa M., Adaptive Web search based on user pro le constructed without any sort from users, 13th International Conference on World Wide Web (2004), 675-684.

[10] Mei Q., Church K., Entropy of search logs: how hard is search? With personalization? With backo In International conference on Web search and Web data mining (WSDM) (2008).

[11] Brajnik G., Guida G., Tasso C., User modeling in intelligent information retrieval. Inf. Process. Manag., 23 (1987), 305-320.

[12] Pazzani M., Billsus D., Content-based recommendation systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web, Springer, Berlin (2007), 325-341.

[13] Agichtein E., Brill E., Dumais S., Ragno R., Learning user interaction models for predicting Web search result preferences, 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR) (2006).

[14] Gao W., Niu C., Nie J.Y., Zhou D., Hu J., Wong K.F., Hon H.W., Cross-lingual query suggestion using query logs of die rent languages, In 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR) (2007), 463-470.

[15] Sun J.T., Zeng H.J., Liu H., Lu Y., Chen Z., Cube SVD: a novel approach to personalized Web search, In 14th International Conference on World Wide Web (WWW) (2005), 382-390.

[16] Smyth B., Balfe E., Anonymous personalization in collaborative Web search, Inf. Retr., (2006), 165-190.

[17] Furnas G.W., Landauer T.K., Gomez L.M., Dumais S.T., The vocabulary problem in human-system communication, Commun. ACM 30 (1987), 964-971.

[18] Zhou D., Lawless S., Wade V., Improving search via personalized query expansion using social media, Inf. Retr., (2012).

[19] Chirita P.A., Firan C.S., Nejdl W., Personalized query expansion for the Web, 30th Annual International ACMSIGIR Conference on Research and Development in Information Retrieval (SIGIR) (2007).

[20] Psarras I., Jose J., A system for adaptive information retrieval, In: Lecture Notes in Computer Science, 4th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH) Springer, Heidelberg (2006), 313-317.

[21] Yin Z., Shokouhi M., Craswell N., Query expansion using external evidence, Lecture notes in computer science, 31st European Conference on Information Retrieval (ECIR), Springer, Toulouse (2009), 362-374.

[22] Cui H., Wen J.R., Nie J.Y., Ma W.Y., Query expansion by mining user logs, IEEE Trans. Knowledge Data Engg., 15 (2003), 829-839.

[23] Leveling J., Jones G.J.F., Classifying and ltering blind feedback terms to improve information retrieval electiveness, Adaptivity, Personalization and Fusion of Heterogeneous Information

(RIAO), Le Centre De Hautes Etudes Internationales DInformatique Documen-taire, Paris (2010), 156-163.

[24] Ogilvie P., Voorhees E., Callan J., On the number of terms used in automatic query expansion. Inf. Retr., (2009), 666-679.

[25] Cao, G., Nie, J.Y., Gao, J., Robertson, S.: Selecting good expansion terms for pseudo-relevance feedback, 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR) (2008), 243-250.

[26] De Luca E.W., Nrnberger A., Adaptive support for cross-language text retrieval, Lecture Notes in Computer Science. 4th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH) (2006), 425-429.

[27] Nesrine Zemirli, Web Cap: Inferring the users Interests Based on a Real-Time Implicit Feed-back, Seventh International Conference on Digital Information Management (ICDIM) (2012).

Appendix A
Parameters Calculation Techniques

If X is a random variable, and ($x_{,}$ $x_2, \ldots x_n$) are the observations or samples of X, then, the expectation μ and variance σ^2 of X are calculated as represented in Equations 13 and 14 respectively. Here, $\bar{x}_n$ is the sample mean of X.

$$\mu \approx \bar{x}_n = \frac{\sum_{i=1}^{n} x_i}{n} \tag{13}$$

$$\sigma^2 \approx \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 \tag{14}$$

If Y is a random variable, and ($y_1, y_2, \ldots y_n$) are the observations or samples of Y, then, the sample covariance of X and Y indicated by $S(X,Y)$ is calculated as represented in Equation 15. Here, $\bar{y}_n$ is the sample mean of Y. If $X = Y$, then, $S(X, X)$ becomes the sample variance of X.

$$S(X,Y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)(y_i - \bar{y}_n) \tag{15}$$

If $Z \in R^d$ is a random vector as represented in Equation 16, then, the covariance matrix Σ of Z is calculated as represented in Equation 17.

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \\ . \\ . \\ Z_d \end{bmatrix} \tag{16}$$

$$\Sigma \approx \begin{bmatrix} S(Z_1, Z_1) & S(Z_1, Z_2) & \ldots. & S(Z_1, Z_d) \\ S(Z_1, Z_2) & S(Z_1, Z_2) & \ldots. & S(Z_1, Z_d) \\ . & . & & . \\ . & . & & . \\ S(Z_1, Z_2) & S(Z_1, Z_2) & \ldots. & S(Z_1, Z_2) \end{bmatrix} \tag{17}$$

The expectation of **Z** indicated by $\mu \in R^d$ is calculated as represented in Equation 18. Here, (z_1, $z_{2,\ldots}z_n$) are the observations or samples of **Z**.

$$\mu \approx \frac{\sum_{i=1}^{n} Z_i}{n} \tag{18}$$