Volume 2, Issue 5, May 2017 ISSN (online): 2456-0006
International Journal of Science Technology Management and Research
Available online at: www.ijstmr.com

EFFICIENT WEB SEARCH BASED ON USER INTEREST AND BROWSING HISTORY

Sagar Bapu Gojare
BE Computer
Shatabdi Institute of Engineering & Research
Nashik, India

Hemant Ramesh Raut
BE Computer
Shatabdi Institute of Engineering & Research
Nashik, India

Akbar Hanif Shiakh
BE Computer
Shatabdi Institute of Engineering & Research
Nashik, India

Vivek Eknath Shinde
BE Computer
Shatabdi Institute of Engineering & Research
Nashik, India

Suhas B. Gote
Asst. Professor
Shatabdi Institute of Engineering & Research
Nashik, India

Abstract: ***At the point when distinctive users give same question, same outcome will be returned by a normal web search tool, regardless of which user presented the inquiry. This won't not be proper for users which require diverse data. While hunting down the data from the web, users require information in view of their advantage. For a similar watchword two users may require diverse bit of data. This reality can be clarified as takes after: a scientist and a developer may require data on "infection" however their fields are is completely unique. Researcher is scanning for the "infection" that is a microorganism and developer is hunting down the pernicious programming. For this sort of inquiry, various archives on particular themes are returned by bland web indexes. Henceforth it gets to be distinctly troublesome for the user to get the important substance. In addition it is likewise tedious. Customized web inquiry is considered as a promising answer for handle these issues, since various query items can be given relying on the decision and data needs of users. It misuses user data and inquiry setting to learning in which sense a question allude.***

Keywords: Profile, Domain knowledge, DMOZ directory.

I. INTRODUCTION

Today's web crawlers more often than not can't recognize diverse users' needs well for instance, a pc researcher may utilize the pursuit question "panther" to find data on apple operating system x panther and a scientist may utilize a similar inquiry for the creature panther; in any case, a web index for the most part treats the two inquiries a similar way. On the other hand, customized seek gives modified outcomes. In our past work we presented a scoring capacity for customizing list items. The capacity utilizes four to score a term that matches the user profile, which is found out from the user's bookmarks. we utilize the page scores to re-rank recovered site pages in light of accuracy and review, we demonstrated that our customized re-positioning methodology outflanked a web crawler at lower positions, yet not at the main 5 positions in this paper we enhance the scoring capacity by changing the components/attributes utilized as a part of the capacity and including report length standardization.

II. LITERATURE SURVEY

Implemented a wrapper around the search site that collects information about user's search activity and builds user profile by classifying collected information (queries).They have used these profiles to re-rank the search results and the rank-order of the user-examined results before 493 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT) and after re-ranking were compared [1].

Identified that current web search engines do not consider the special needs of user or interests of user and proposes a novel technique which uses search history of user to learn user profiles. This work uses user's search history for learning

of user profile and category hierarchy for learning of a general profile and then combines both profiles to categorize user's query to represent user's search intention and to disambiguate the words used in query [2].

Identifies that different users may have need of different special information, when they use search engines and techniques of personalized web search can be used to solve the problem effectively. Three approaches Rocchio method, k-Nearest method and Support Vector Machines have been used in to build user profile to present an individual user's preference and found that k-Nearest method is better than others in terms of its efficiency and robustness[3].

In this paper the authors have given a personalized web search outcome which is in accordance with the need of user in various situations. The analysis of model has resulted in three concepts to implement the model, which is semantic indexing for web resources, modeling and acquiring user context and semantic similarity matching between web resources and user context. The author has defined it as context based adaptive personalized web search [4].

In this paper the location and content concept has been separated and is organized into different ontology to make an ontology-based, multi-facet (OMF) profile which is captured by web history and location interest. This model actually gives results by outlining the concepts in accordance with the preference of user. By keeping the diverse interest of the users in mind, location entropy is introduced for finding the degree of interest and information related to location and query. The personalized entropies actually establish the relevant output content and location content. At last, an SVM based on the ontology is derived which can be used for future purpose for ranking or re-ranking. The experiments show that the results produced by OMF profiles are more accurate in comparison with the ones which use baseline method [5].

Proposed a personalized web search model that combines community based and content based evidences based on novel ranking technique. Nowadays, uploading data on internet has become a daily activity. A massive amount of data is uploaded in the form of web pages, news, and blogs etc. on a regular basis. So, it becomes very difficult for the user to search for relevant content. Not only for users but also for search engines like Google and Yahoo it becomes difficult. Information overload is the only reason behind this difficult situation. Other than this user's preference is the second problem, which is not taken into consideration while producing the results. The author tried to solve this problem through this model which produces results on the basis of preference and interest of the user. In this paper, authors proposed a unique approach to find out the interest and preference of the user. It's a two way approach, first it will find out the activities of user through his/her profile in social networking sites. Secondly, it will find out information from what the social networking sites provide to the user through friends and community. Based on the results, user's interest and preference will be prioritized by the web search or it is personalized [6].

III. EXISTING SYSTEM

With the advancement of Internet, web search tools have contributed a considerable measure in looking data from the web. They help in finding data on the web fast and simple. Be that as it may, there is still opportunity to get better. Momentum web indexes don't consider specific requirements of client and serve every client similarly. It is difficult to tell the web crawler what we the client really need. Nonspecific web search tools are taking after the "one size fits all" model which is not versatile to individual clients.

IV. PROPOSED SYSTEM

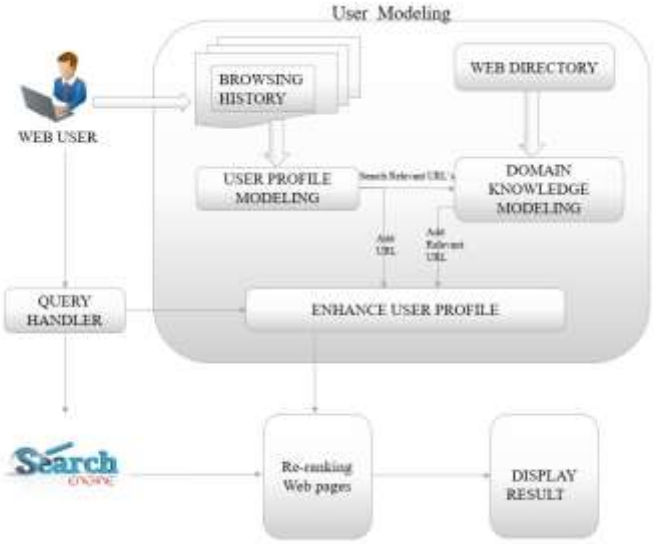


Fig. 1 Proposed System

We have proposed a system for productive web seek in view of Client intrigue and perusing history. In view of the Client Profile and the Space Learning, the framework continues overhauling the client profile and in this way assembles an upgraded client profile. This Improved client profile is then utilized for proposing important pages to the client. The proposed system has been actualized by playing out a few analyses. These analyses demonstrate that the execution of the framework utilizing upgraded client profile is superior to anything those which are gotten through the basic client profile. Our work is huge as it enhances the general hunt effectiveness, taking into account the individual enthusiasm of the user's. Accordingly, it might be a little stride in the field of customized web look. This examination is about customized web seek on sight and sound substance. Our approach included building an intrigue profile in view of his collaboration with web indexed lists and his perusing conduct as indicated by area insightful pursuit. Area based Personalization of list items of mixed media substance is accomplished by positioning indexed lists in light of nearness to the client premium profile. The quality to perceive client interests in a totally non-obtrusive way and the exactness of the customized results are a portion of the real favorable circumstances of our approach.

V. SYSTEM ARCHITECTURE

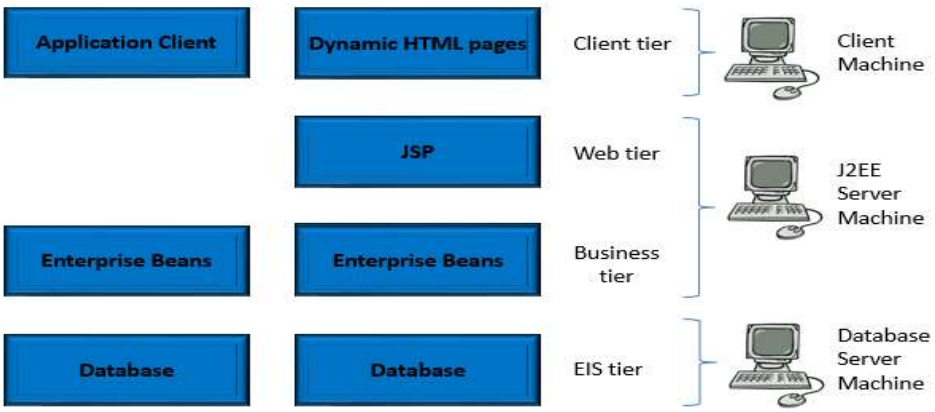


Fig. 2 System Architecture

VI. MODULE DESCRIPTION:

After careful analysis the system has been identified to have the following modules:

A. Personalized Web Search Module.
B. User Modeling Module.
C. Domain Knowledge Modeling Module.
D. Enhanced User Profile Module.

A. Personalized Web Search Module:

Personalized web search which considers individual's interest into mind and enhances the traditional web search by suggesting the relevant pages of his/her interest. We have proposed a simple and efficient model which ensures good suggestions as well as promises for effective and relevant information retrieval. In addition to this, we have implemented the proposed framework for suggesting relevant web pages to the user.

B. User Modeling Module:

Our system considers user's profile (based on user's weblog/navigation browsing history) and Domain Knowledge in order to perform personalized web search. Using a Domain Knowledge, the system stores information about different domain/categories. Information obtained from User Profile is classified into these specified categories. The learning agent learns user's choice automatically through the analysis of user navigation/browsing history, and creates/updates enhanced User Profile conditioning to the user's most recent choice. Once the user inputs query, the system provides good suggestions for personalized web search based on enhanced user profile. Further our model makes good use of the advantages of popular search engines, as it can re-rank the results obtained by the search engine based on the enhanced user profile.

C. Domain Knowledge Modeling Module:

Domain knowledge is the background knowledge that we used to enhance the user profile. The source which we have used for preparing Domain Knowledge is DMOZ directory. For preparing Domain Knowledge, first we have crawled the Web pages from DMOZ directory for some specified categories, where each category is represented by collection of URL's present in that category.

D. Enhanced User Profile Module:

Using the information of user browsing history and domain knowledge, we create an Enhanced User Profile. Once the Enhanced User Profile is created, we take the user query and suggest the relevant web pages with respect the query. In our Experiment, we have used User Profile as a base case for suggesting the relevant pages and compared the results with the pages suggested from Enhanced User Profile. For each query, we suggest top 20 relevant documents from User Profile and for the same query we also suggest top 20 relevant documents from Enhanced User Profile. In order to compare the efficiency of the result, we compared the similarity of suggested documents with the user query.

Following steps explain the process of preparing the Enhanced User Profile. Perform the following steps for each document (URL) in user profile:

a. Select the URL from the User Profile.
b. Add the URL to the Enhanced User Profile.
c. Find the cosine similarity of this URL with the URLs present in user specific categories from the Domain Knowledgebase.
d. Rank the URLs on descending order of cosine similarity.
e. Retrieve top 20 URLs.
f. Calculate the average of the cosine similarity of these top 20URLs.
g. From the top 20 URLs add only those URLs to the enhanced user profile whose similarity value is above the average value.

VII. TECHNICAL SPECIFICATION

i. Hardware Specification

- Memory - 1GB
- System - AMD/Dual core
- Hard Disk - 30GB

- Monitor – SVGA

ii. Software Specification

- Eclipse Mars.1 Release (4.5.1)
- Java 1.8.0_65
- Mysql-installer-community-5.7.9.0
- Mysql-query-browser-1.1.20-win
- Mongodb-win32-x86-64-2008+
- Apache-tomcat-7.0.67-windows-x64
- Apache Jersey1.19
- Hibernate-distribution-3.6.4

VIII. ALGORITHM

Serialization Algorithm steps:

1) Start.
2) User login to maintain a user's history profile
3) If users is not already register then firstly register, then proceed to next step.
4) After login user enter the keyword query.
5) If keyword matches with existing query those stored in profile history according to their domain Wise.
6) Semantic Clustering of domains are performed using weighs of keyword, when similar domain keyword occur.
7) Than it displays recommended link according to domain.
8) According to domains displayed to user, user will search data or pages in easy and faster way.

Flowchart:

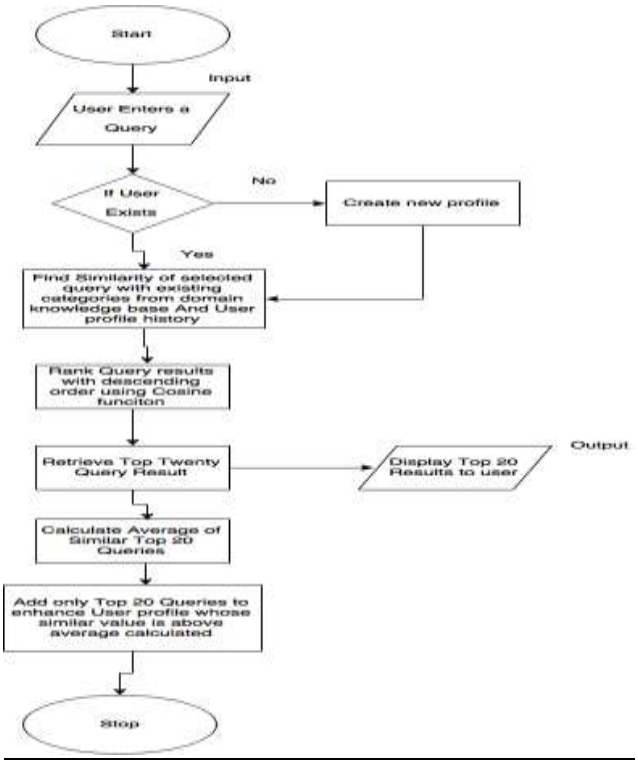


Fig.3 Flowchart

- Above flowchart show the detailed working of algorithm and process to step by step operation of system.

IX. APPLICATIONS

- Generic search engines are important for retrieving relevant information from web.
- The first study was conducted by the researchers to discover that reducing search cost led to lower quality choices.
- The reason behind this discovery was that consumers make worse choices because lower search costs cause them to consider inferior options.
- It also showed that if consumers have a specific goal in mind, they would further their search, resulting in an even worse decision

X. EXPERIMENTAL RESULTS AND ANALYSIS

We have designed our own dataset. In our Experiment, we have used the browsing history of 8 different users from our class. Each user have different area of interest. Our Experiment is conducted for 20 queries in which different queries from different area of interest of user such as java, computer, movies, etc. In order to collect the domain knowledge, we have crawled the datasets from DMOZ for the selected topics using Mongodb database, while MySQL query browser is used to view the different schemas. By setting crawling parameters of Mongodb we have restricted the crawling to specific DMOZ topic.

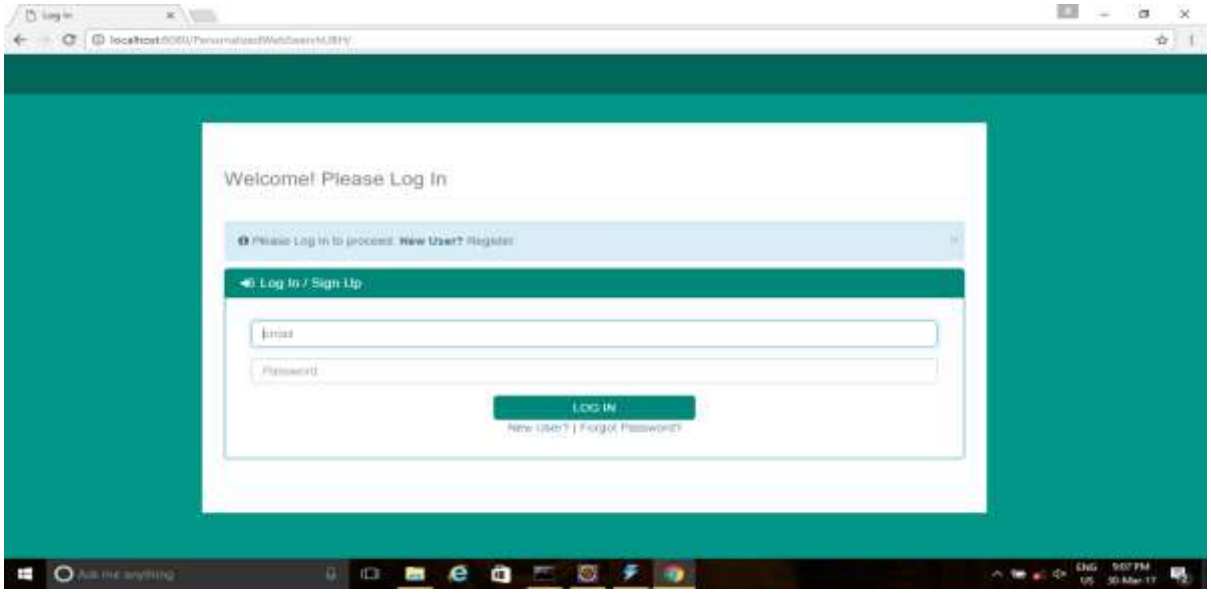


Fig.4 Login Page

- In login page new user can register and login to the system.

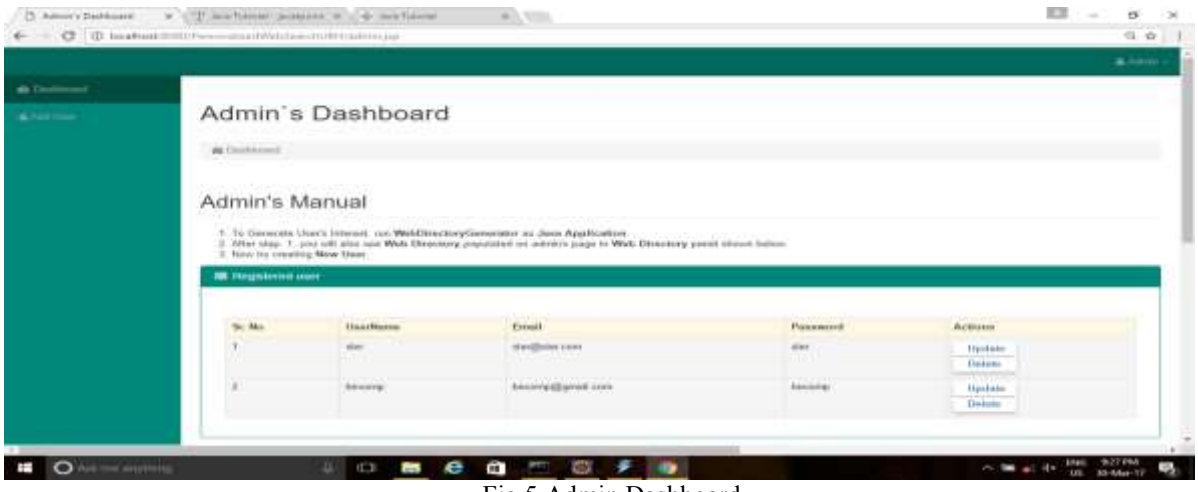


Fig.5 Admin Dashboard

- Admin can login to the system and view the history of users and manage all the activity.

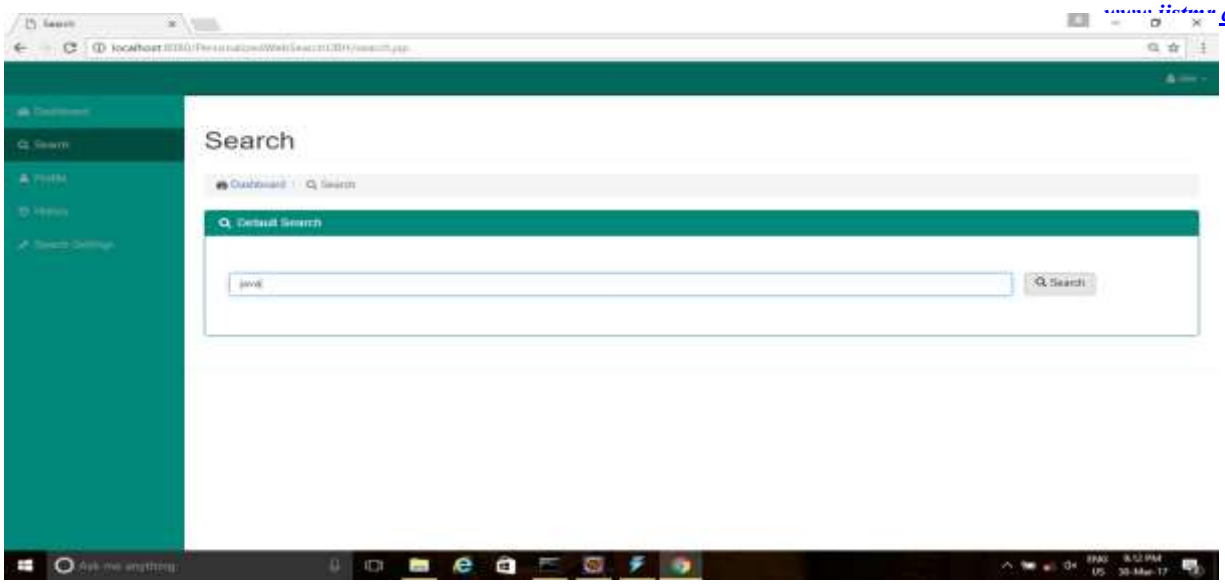
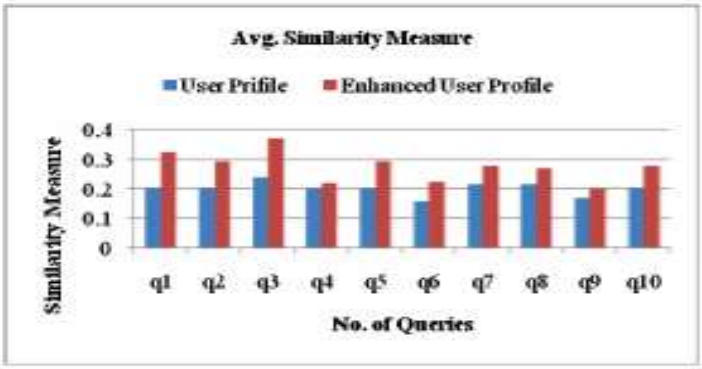


Fig.6 Search Page

- Users can search from search page and get the relevant result as per the weight of keyword..

Fig.7 Web Directory

- Web Directory used to populate the HTML web pages, Admin can add or remove the pages or documents.
- (Figure) represent the bar graph of average cosine similarity obtained for top 10 queries. For all 10 queries, the average cosine similarity measure of User Profile and Enhanced User Profile has been calculated.
- The average improvement in Enhanced User Profile for all 10 queries as compared to User Profile. Experimental results show that our proposed model for personalize web search is effective for focused information retrieval and suggests good Web pages.



CONCLUSION AND FUTURE WORK

The proposed framework has been implemented by performing some experiments. These experiments shows that the performance of the system using enhanced user profile is better than those which are obtained through the simple user profile. Our work is significant as it improves the overall search efficiency, catering to the personal interest of the user's. Thus, it may be a small step in the field of personalized web search. In future this framework may be applied for re-ranking the web pages retrieved by search engines on the basis of user priorities. We may also apply collaborative filtering for personalized web search in our framework

REFERENCE

1. M Speretta and S Gauch, "Personalized Search Based on User Search Histories", Proceeding Of International Conference on Web Intelligence, pp. 622-628, 2005.
2. F Liu, C Yu and W Meng, "Personalized Web Search for Improving Retrieval Effectiveness", IEEE Transactions On Knowledge And Data Engineering, pp. 28-40,Volume 16, 2004.
3. C Liang, "User Pro_le for Personalized Web Search", International Conference on Fuzzy Systems and Knowledge Discovery, pp. 1847- 0, 2011.
4. X Pan, Z Wang and X Gu, "Context-Based Adaptive Personalized Web Search for Improving Information Retrieval Effectiveness", International Conference on Wire-less Communications, Networking and Mobile Computing, pp. 5427 - 5430, 2007.
5. K.W.T. Leung, D.L. Lee and Wang-Chien Lee, "Personalized Web search with location preferences", IEEE 26th International Conference on Data Engineering, pp.70 I - 712, 2010.
6. O. Sha_q, R. Alhajj and 1. G. Rokne, "Community Aware Personalized Web search", International Conference on Advances in Social Networks Analysis and Mining, pp.3351 - 355,2010.
7. Li, L. Yang, Z. and Kitsuregawa, M. (2009) Rank Optimization of Personalized Search. Journal of the Database Society of Japan, 8, pp.35-40.
8. Page, L., et al., The PageRank Citation Ranking: Bringing Order to the Web, Technical Report, (1999).
9. Brin, S., R. Motwani, and T. Winograd, What can you do with a web in your pocket, Data Engineering Bulet in, 21(2), (1998), pp.37-47.
10. Bharat, K. and G.A. Mihaila, when experts agree: using non-a_liated experts to rank popular topics . ACM Trans. Inf.Syst., 20(1), (2002), pp.47-58.
11. Haveliwala, T.H., et al., Evaluating strategies for similarity search on the web, in Proceedings of the 11th nternationalconference on World Wide Web, ACM: Hon- olulu, Hawaii,USA, (2002), pp. 432 - 442.