Collaborative Framework with User Personalization for Efficient web Search : A D^3 Mining approach

V.Vijayadeepa , Dr.D.K.Ghosh

Abstract - User personalization becomes more important task for web search engines. We develop a unified model to provide user personalization for efficient web search. We collect implicit feedback from the users by tracking their behavior on the web page based on their actions on the web page. We track actions like save, copy, bookmark, time spent and logging into data base, which will be used to build unified model. Our model is used as a collaborative framework using which related users can mine the information collaboratively with little amount of time. Based on the feed back from the users we categorize the users and search query. We build the unified model based on the categorized information, using which we provide personalized results to the user during web search. Our methodology minimizes the search time and provides more amount of relevant information.

Index terms - Feature extraction ,similarity measure ,SQ categorization and page classification

Motivation

World Wide Web contains huge data-both structured and unstructured and it is a dynamic environment as the data and the user change frequently. In such a dynamic environment, the task of finding desired information quickly and exactly becomes crucial and tracking user search behavior is also difficult. This motivated to personalization of web search where customizing the user's search environment according to their interests and thus aids the user to identify their information need without much difficulty.

Background

Ioannis Anagnostopulos and Ilias Maglogiannis in their work proposed, Adapting user's browsing and web evolution features for effective search in medical portals [2] is an intelligent meta-search algorithm exclusively for health-related web search, capable of self-adapting over the continuous changes that occur on medical sites, using a web evolution adaptation mechanism. The algorithm also supports personalization features based on user's browsing behavior. When user's preferences are used jointly with dynamic survey mechanism that adapts to web evolution events and changes, a more efficient medical related search is provided. Transparency is achieved for both personalization and web evolution adaptation mechanisms, requiring virtually none effort from the user's part. The construction of the personalized preferences is performed in a totally transparent way, without interferences in the users' browsing behavior, while the merged meta-results are presented without labeling their source, ensuring that the user is completely unbiased to his preferences. The only feedback that the user receives is a text paragraph regarding the URL as most of web search engines do. The personalized preferences are recorded on client –side and they are updated continuously according to the meta- results acquired by the user, the time spent for their exploration as well as their download volume. Thus, the user's profile is also adjusted to any possible changes in respect to his information needs. The paper provides the implementation details for the proposed meta-search algorithm along with its initial assessment.

Practical algorithms and lower massive bounds for similarity search in massive graphs, proposed by Daniel Fogares and Balazs Racz [1] is scalable to graphs with billions of vertices on a distributed architecture. The similarities of multi step neighborhoods of vertices are numerically evaluated by similarity functions using SimRank and PSimRank similarity functions for the application of fingerprint verification. The SimRank, includes a recursive refinement of co- citation, and PSimRank, includes a novel variant with better theoretical characteristics. These methods are presented in a general framework of Monte Carlo similarity algorithms that pre-compute an index database of random fingerprints, and at query time, the similarities are estimated from the fingerprints. It is justified in the paper that in approximation the method is asymptotic worst–case lower bounds. The above-specified algorithm is utilized in the proposed system to exploit the similarity information between various set of nodes. Here the set of nodes refers to a set of web pages the user visited for a search query in each session.

An Evaluation of Personalized Web Search for Individual User, by S.Sendhilkumar and T.V. Geetha [4] supports user searches by learning about user preferences and by observing responses to prior search experiences aided by User Conceptual Index (UCI) to improve user search in the World Wide Web. The paper also directly focused on the evaluation of the results of individual user's UCI based search and introduces three measures for the purpose. The paper addressed the problem of irrelevant web search results for a novice user and suggested two mechanisms to solve the problem: Context oriented and Individual oriented personalized search. Context includes factors like the nature of information available, the information currently being examined, when and what applications in use and so on. The Individual oriented search encompasses elements like the user's goals, prior and tacit knowledge, past information seeking behaviors, among others. With the aim of developing a personalized search system for an efficient web search a new search – aiding index called the UCI that incorporates both the contents of search as well as the individual's search activity has been proposed in the paper. The main aim of the UCI is to assess the relationship between the ranks of a search result as returned by a web search engine and the individuals' perceived relevancy of the result. The common evaluation method applied in Information Retrieval systems is precision and recall and it usually requires relevance feedback from users. However, obtaining relevance feedback explicitly from users for personalized web search systems is extremely challenging and time consuming due to the large size of WWW. For reducing the time needed for evaluation the paper suggests the user was asked to provide relevance judgments only for the top 30 pages among the result pages. The paper also presents a personalized web search using UCI, recommending pages for individual user. Also it focuses on three measures namely the Page-Query relevancy, the Page-Interest relevancy and the Query-Interest relevancy that are used along with the UCI to evaluate the relevancy of a page with respect to the users' context of search. These measurements are utilized in the proposed system to evaluate the final results.

Orland Hoeber and Xue Dong Yang [5] in their work developed two systems to support the visual exploration of web search results rather than list based representation. They are HotMap and Concept Highlighter. In both of these systems, the search results are provided at two levels of detail: an overview map that provides a compact and abstract representation of the top 100 documents returned by the underlying search engine and a detail window that shows 20-25 documents at a time. The paper addresses one of the reason for users do not delve deeply into the set of search results is due to the factor that the list based representation does not readily support the user's relevance decision making tasks. Document surrogates must be considered one–by–one in blocks of ten at a time, resulting in an inability to explore the search results easily. To address this shortcoming of list based representations, they have developed two interfaces to support the visual exploration of web search results:HotMap and ConceptHighlighter. In HotMap, the frequencies of each of the query terms from the user's queries are depicted visually using color-coding. This allows the users to easily identify "hot" documents based on the frequent appearance of the query terms within the document surrogates. In addition to this visual representation, the search results can be dynamically resorted based on the query term frequencies, supporting an interactive exploration of the search results. In Concept Highlighter, a set of relevant concepts is generated from a concept knowledge base using the query terms; interactive concept-based fuzzy clustering is used to cluster the search results with respect to these concepts. The same kind of dual window projection but in web graphical format is made in the proposed system.

Wenxue Tao and Wanli Zuo in their work of Query Sensitive Self Adaptable Web Page Ranking Algorithm [7] analyses HITS and PageRank, ranking algorithms and points out their limitations in capturing both global and local importance scopes. A new query-sensitive algorithm termed as Page-Rank satisfying both global and local authority is introduced and several strategies for combining the algorithm with traditional PageRank are also proposed. The global importance scope incorporates a broader area covering the area just represented by the query. And it could be the whole Internet, total Web pages indexed by a search engine, or a predefined category of a search engine according to the specific topic represented by the query. Global importance scope involves Web page content, Web page out-links, Web page outside environment, the authority of the web site to which the Web page belongs, etc. Different from global importance scope, the local importance scope implies the exact area represented by the query. As the user's query is unpredictable, the topic or the area represented by the query is often quite different from the predefined topics in the search engine. Therefore, the global importance scope and the local importance scope are often diverse, and cannot substitute each other. The ranking of result pages should reflect not only the authority in the broader scope but also the authority in the exact scope represented by the query, i.e. the query sensitiveness. In HITS algorithm, the link structure of a graph is analyzed to find the set of hub pages and a set of authority pages in an iterative fashion. The computation is carried out at query time and efficiency becomes a practical concern. Hence the algorithm is not fit for the proposed system. In PageRank algorithm, though computation is offline, it involves whole portion of the Internet. Thus only global scope is considered and not the local scope. As [7] addresses the importance of both the scope, the PageRank algorithm is not taken into account for the proposed system. Topic sensitive algorithm is not considered since it contains enormous topics to categorize and clustering them is a big cumbersome. Moreover, too much of clusters lead to sacrifice global importance. In Query-Sensitive PageRank algorithm, a voting mechanism is involved. For each list of inverse table, the top 'n' Web pages is selected to construct a voting set. The algorithm is also not taken into account since it is uncertain that all existing web pages remain unchanged throughout the time.

Methodology

The user interface is a tabbed web browser, which is a part of the system. Through this browser the user can provide short-term query simultaneously in multiple tabs for his information need. The user interacts with the system to give search query, to view the ranked results and to view the re ranked results. The re ranking is done based on the past search behavior of the user with the system. The browser also supports for providing actions like *SAVE, COPY, PRINT AND E-MAIL*, which depicts the importance of the web page for his need. The browser also projects the re-ranked results in an interactive graph like structure rather than list based representations.

Each user-visited web page is represented by a set of index words that comes under top list. The usage time of each search query and usage-time of each visited page is calculated transparently without disturbing the user.

Based on the search query, index word and usage-times, the User Conceptual Index (UCI) is calculated. The UCI can be represented mathematically as a function of weights of above parameters. The usage time directly indicates whether previous search results were relevant or irrelevant to the user's information need.

The search queries that have similar or related meaning are categorized to a group using word-dictionary in-order to avoid inconsistencies that arise in above strategies.

The visited-pages that have similar or related index words are also categorized to recommend the pages for a novice user. The users with similar search behavior are categorized to a group to improve the efficiency of personalization mechanisms. The pages are re-ranked by analyzing individual's behavior and are projected to them in dual window. Some of the strategy for personalization of web search is described as follows:

1. A user's search history can be collected without direct involvement.
2. The user's profile can be constructed automatically from the user's search history and is dynamically updated.
3. The categories that are likely to be of interest to the user are deduced based on his search behavior.

Feature Extraction

The first step in the project is to extract the feature of the user-visiting page. The feature of a page, P is defined as a set of top 'n' frequently occurring terms. In order to extract the features, the source content of each page is extracted and it is de-tagged. From the de-tagged page, the stop words are removed and the terms in the page are extracted. From the set of terms, the top 'n' frequently occurring terms is extracted. These 'n' terms form as index words of the page.

Algorithm : Feature Extraction

{

Given: User visiting page P_i

Procedure:

step1: The de- tagged and stop words eliminated page P_i can be represented as

$IW= \{IW_1, IW2, IW3, IW_n\}$ and

$F= \{F_1, F2, F3\ldots F_n\}$ where IW is the index word set and F is the Frequency set corresponding to IW and n is the number of index words in the page.

step2: Select top 'k' frequency words

$F_{topk}= \{F_1, F2 \ldots F_t\}$ which corresponds to $IW_{topk}= \{IW_1, IW2\ldots IW_t\}$

Where $k<=t<=n$.

step3: Compute the mean for the above set

$\mu (F_{topk}) =[F_1+F_2+\ldots+F_t]/t$

step4: The keywords in IW_{top} that have frequency above $\mu (F_{topk})$ form the feature of the page.

step5: Now represent the feature of the page as

$Feature (P_i) = \{F_1, F2 \ldots Fm\}$

Where $1<=m<=k$.

step6: End

}

User Association Analysis

The user association is analyzed to find the similarity of search among different users. From the set of visited-pages, the actions performed by the user are monitored. From the action it is concluded whether the page is useful to the user. The order in which the page is visited is also tracked and a directed graph is constructed. The usage-time for each page forms the weight of the page.

Algorithm : Similarity measure

```
{
 Given: User behavior graph
 Procedure:
            step1:  Indexing
            for i:=1 to N do
                        for every vertex j of the web graph do
            Behavior [i] [j] []: =reversed path of    length l starting from j.
end
end
            step2:  User Sim(i,j)
            Sim:=0
            for i:=1 to N do
                        for j:=1 to N do
             let k be the smallest offset with
      Behavior[i][u][k]=Behavior[j][u][k]
if such k exists then
            Sim=Sim+c^k
end
            end
            return Sim/N
step 3: End
}
```

The above algorithm is used to identify the similarity behavior between two users. Whenever the search behavior is common, then it is certain that the users might come from same source point. Thus, higher the length l in the above algorithm, greater is the similarity.

Search Query Categorization

Greater the number of times a user uses a particular SQ, greater is the interest of the user on the particular topic associated with the search query terms. If the previous search result is not relevant to the user's information need, then the user might modify the SQ to fit into their context of search. Even though the search keyword gets changed, the information need on a topic doesn't get changed in that session. Hence the alternate keyword supplied by the user may also be intended to search exactly for same topic. So it is necessary to identify the alternate meanings of user's search query, which leads to categorization of the search query. The query categorization is necessary to reduce the limitations in key word based search.

Algorithm : SQ Catergorization

{

Given: Search Queries given by the user in a due course of time

Procedure :

Step1: Collect all the search queries given by the user in a due course of time.

Step2: Find the alternate meaning of the search query using a word dictionary.

Step3: Find whether the result of step2 exists in the search query set provided by the user.

Step 4: If such commonalities exist, update the TF matrix and ST matrix.

}

Visited Page Categorization

Higher the similarity ranks between two users, greater the commonalities of search between them. Two pages can be said to similar even they spoke of exactly the same topic with different keywords. Hence it is necessary to identify the alternate meanings of the Index words, which leads to categorization of the visited page. The categorization of the page is used to expand the similarity rank calculation, which aids to identify common search behavior.

Algorithm: Page Classification

{

Given: The index terms of all the visited pages by the user in a due course of time

Procedure:

Step1: Collect all the index terms of all the visited pages by the user in a due course of time.
Step2: Find the alternate meaning of the index terms using a word dictionary.
Step3: Find whether the result from step2 exists in the index terms set collected from user history.
Step 4: If such commonalities exist, update the SFmatrix

}

Results and Analysis

Fig1. shows the name of user

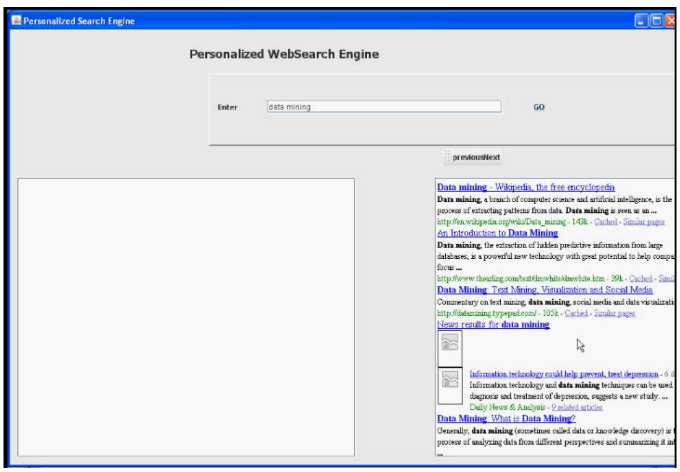

Fig 2. shows the result for query Data mining

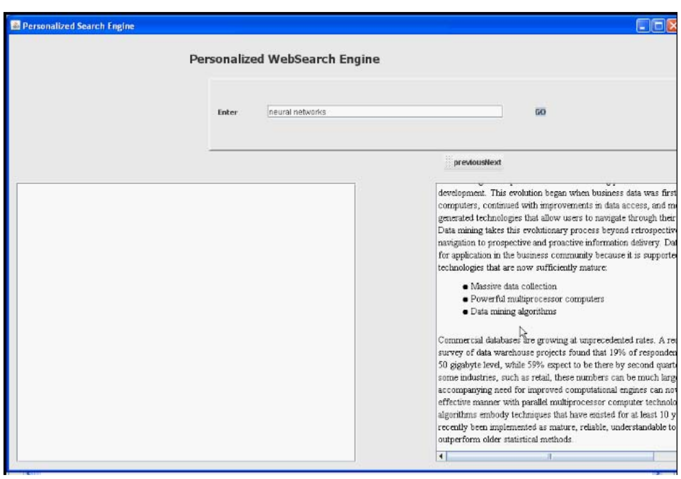

fig 3. shows the page source of url clicked

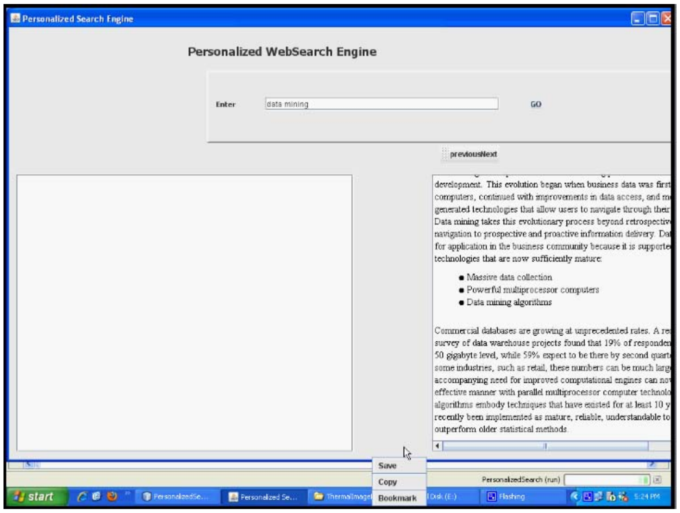

Fig 4 displays the result of action performed.

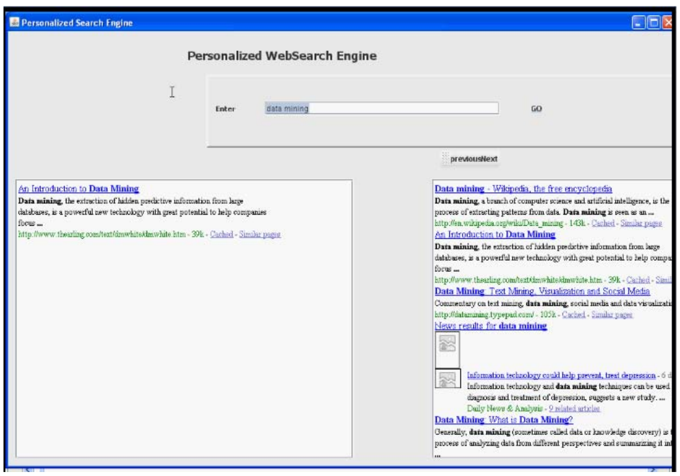

Fig 5 Displays the proposed results.

The results displayed above shows , the recommended results from our framework according to the interest of the users.

Conclusion

With the semi-structure of information on the Internet and the arbitrariness of releasing the enormous amount of web pages, turns finding desired information quickly and exactly to be a crucial task. Search engine is playing an increasing important role in information retrieval on the Internet. The search results given by search engines are generally sorted on descendent importance of its usage. Humans think in terms of concepts but the concept may be differing from one another. Hence the importance of a page is gained from users with different concepts. Thus this contradictory importance does not be feasible in future. Hence user centric personalization is the only solution to solve the problem. We further investigate this framework to increase the relevancy of the web links to the search query.

References

[1] Daniel Fogares and Balazs Racz, Practical Algorithms and Lower Massive Bounds for Similarity Search in Massive Graphs, *IEEE transactions on knowledge and Data Engineering,vol 19,No5*,pages 585-598, *May 2007*.

[2] Ioannis Anagnostopulos,Ilias Maglogiannis, Adapting user's Browsing and Web Evolution features for Search in medical portals, *First IEEE International Workshop on semantic media adoption and personalization*, pages 37-42, *2006*. effective

[3] Liang Deng, Martin D. F. Wong, An Exact Algorithm for the Statistical Shortest Path Problem, *ACM conference on Asia South Pacific design automation*, pages 965-970, *2006*.

[4] S.Sendhilkumar and T.V. Geetha, An Evaluation of Personalized Web Search for Individual User, *International Conference on Artificial Intelligence and Pattern Recognition (AIPR07)*, FL, USA, pages 484 -490, 2007.

[5] Orland Hoeber and Xue Dong Yang, ExploringWeb Search Results Using Coordinated Views, *Fourth IEEE International Conference on Coordinated & Multiple Views in Exploratory Visualization,* pages 3-13, *2006.*

[6] O. Hoeber and X.D. Yang,The visual exploration of web search results using HotMap, *International Conference on Information Visualization,* pages157-165, *2006.*

[7] Wenxue Tao, Wanli Zuo, QuerySensitive SelfAdaptable Web Page Ranking Algorithm, *Second International Conference on Machine Learning and Cybernetics, Xi,* pages 413-418, *2003.*

[8] Rahul R. Joshi and Y. Alp Aslandogan, Conceptbased Web Search using Domain Prediction and Parallel Query Expansion, *IEEE International Conference on Information Reuse and Integration*, Waikoloa, Hawaii, pages16-18, 2006.

[9] A. Broder, "A Taxonomy of Web Search," ACM SIGIR Forum, vol. 36, no. 2, pp. 3-10, 2002.

[10] U. Lee, Z. Liu, and J. Cho, "*Automatic Identification of User Goals in Web Search,*" Proc. 14th Int'l World Wide Web Conf. (WWW '05), pp. 391-400, 2005.

[11] B.J. Jansen, A. Spink, J. Bateman, and T. Saracevic, "*Real Life Information Retrieval: A Study of User Queries on the Web,*" ACM SIGIR Forum, vol. 32, no. 1, pp. 5-17, 1998

[12] K. Sugiyama, K. Hatano, and M. Yoshikawa, "AdaptiveWebSearch Based on User Profile Constructed without Any Effort from Users," Proc. 13th Int'l World Wide Web Conf. (WWW '04), pp. 675-684, 2004.

[13] F. Liu, C. Yu, and W. Meng, "*Personalized Web Search for Improving Retrieval Effectiveness*," IEEE Trans. Knowledge and Data Eng., vol. 16, no. 1, pp. 28-40, Jan. 2004.

[14] M. Speretta and S. Gauch, "*Personalized Search Based on User Search Histories*," Proc.IEEE/WIC/ ACM Int'l Conf. Web Intelligence (WI '05), pp. 622-628, 2005.

[15] L. Page, S. Brin, R. Motwani, and T. Winograd, "*The PageRank Citation Ranking: Bringing Order to the Web,"* technical report, Computer Science Dept., Stanford Univ., 1998.

[16] T.H. Haveliwala, *"Topic-Sensitive Pagerank*," Proc. 11th Int'l World Wide Web Conf. (WWW), 2002.

[17] Y. Zhou and W.B. Croft, "*Query Performance Prediction in Web Search Environmen*ts," Proc. 30th Ann. Int'l ACM SIGIR Conf.Research and Development in Information Retrieval (SIGIR '07), pp. 543-550, 2007.

[18] P.A. Chirita, C. Firan, and W. Nejdl, "*Summarizing Local Context to Personalize Global Web Search*," Proc. ACM Int'l Conf. Information and Knowledge Management (CIKM), 2006.

[19] J. Chaffee and S. Gauch, "*Personal Ontologies for Web Navigation*," Proc. ACM Int'l Conf. Information and Knowledge Management (CIKM '00), pp. 227-234, 2000.

[20] S. Gauch, J. Chaffee, and A. Pretschner, "*Ontology-Based Personalized Search and Browsing," Web Intelligence and Agent Systems*, vol. 1, no. 3/4, pp. 219-234, 2003.

[21] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar, "*Rank Aggregation Methods for the Web*," Proc. 10th Int'l World Wide Web Conf., pp. 613-622, 2001.

[22] P.W. Foltz and S.T. Dumais, "*Personalized Information Delivery: An Analysis of Information Filtering Methods*," Comm. ACM, vol. 35, no. 12, pp. 51-60, 1992.

[23] W.B. Frakes and R. Baeza-Yates, " *Information Retrieval: Data Structures and Algorithms*". Prentice Hall, 1992.

[24] N. Fuhr, "*A Decision-Theoretic Approach to Database Selection in Networked IR*," ACM Trans. Information Systems (TOIS), vol. 17, no. 3, pp. 229-249, 1999.

[25] S. Gauch, G. Wang, and M. Gomez, "*ProFusion: Intelligent Fusion from Multiple*, *Distributed Search Engines*," J. Universal Computer Science, vol. 2, no. 9, pp. 637-649, 1996.

[26] E.J. Glover, G.W. Flake, S. Lawrence, W.P. Birmingham, A. Kruger, C.L. Giles, and D.M. Pennock, "*Improving Category Specific Web Search by Learning Query Modifications*," SAINT,pp. 23-34, 2001.

[27] G.H. Golub and C.F. Van Loan, *"Matrix Computations*", third ed. 1996.

[28] L. Gravano and H. Garcia-Molina, "*Generalizing GlOSS to Vector- Space Databases and Broker Hierarchies*," Proc. 21st Int'l Conf. Very Large Databases (VLDB), pp. 78-89, 1995.

[29] D.A. Grossman and O. Frieder, "*Information Retrieval: Algorithms and Heuristics*". 1998.

[30] A.E. Howe and D. Dreilinger, "*SavvySearch: A Meta-Search Engine that Learns which Search Engines to Query*," AI Magazine, vol. 18, no. 2, pp. 19-25, 1997.

[31] P. Ipeirotis, L. Gravano, and M. Sahami, "*Probe, Count, and Classify: Categorizing Hidden Web Databases*," ACM SIGMOD, pp. 67-78, 2001.

[32] T. Joachims, D. Freitag, and T. Mitchell, "*Webwatcher: A Tour Guide for the World Wide Web*," Proc. 15th Int'l Joint Conf. Artificial Intelligence (IJCAI), pp. 770-777, 1997.

[33] D. Koller and M. Sahami, "*Hierarchically Classifying Documents Using Very Few Words*," Proc. 14th Int'l Conf. Machine Learning (ICML), pp. 170-178, 1997.

[34] Y. Labrou and T. Finin, "*Yahoo! as an Ontology: Using Yahoo! Categories to Describe Documents*," Proc. Eighth ACM Int'l Conf. Information and Knowledge Management (CIKM), pp. 180-187, 1999.

[35] H. Lieberman, "*Letizia: An Agent that Assists Web Browsing*," Proc. 14th Int'l Joint Conf. Artificial Intelligence (IJCAI), pp. 924-929, 1995.

[36] W. Meng, W. Wang, H. Sun, and C. Yu, "*Concept Hierarchy Based Text Database Categorization*," Int'l J. Knowledge and Information Systems, pp. 132-150, Mar. 2002.

[37] T. Mitchell, "*Machine Learning*". McGraw Hill, 1997.

[38] M. Montague and J.A. Aslam, "*Condorcet Fusion for Improved Retrieval*," Proc. 11th ACM Int'l Conf. Information and Knowledge Management (CIKM), pp. 538-548, 2002.

[39] M. Pazzani and D. Billsus, *"Learning and Revising User Profiles: The Identification of Interesting Web Sites*," Machine Learning, vol. 27, pp. 313-331, 1997.

[40] A.L. Powell, J.C. French, J.P. Callan, M.E. Connell, and C.L. Viles, "*The Impact of Database Selection on Distributed Searching*," Proc. 23rd Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 232-239, 2000.

[41] A. Pretschner and S. Gauch, "*Ontology Based Personalized Search*," Proc. Eighth IEEE Int'l Conf. Tools with Artificial Intelligence (ICTAI), pp. 391-198, 1999.

[42] J. Rocchio, "*Relevance Feedback in Information Retrieval*," The Smart Retrieval System: Experiments in Automatic Document Processing, pp. 313-323, 1971.

[43] S.E. Robertson and I. Soboroff, "*The TREC-10 Filtering Track Report*," Proc. Text REtrieval Conf. (TREC-10), 2001, http://trec.nist.gov/pubs/trec10/papers/filtering_track.pdf.

[44] G. Salton and M. McGill, "Introduction *to Modern Information Retrieval* ". New York: McGraw-Hill, 1983.

[45] E.M. Voorhees, "*Overview of TREC 2001*," Proc. Text Retrieval Conf. (TREC-10), 2001, http://trec.nist.gov/pubs/trec10/papers/overview_10.pdf.

[46] "*Common Evaluation Measures*," Proc. Text REtrieval Conf. (TREC-10), E.M. Voorhees and D. Harman, eds., p. A-14, 2001.
[47] D.H. Widyantoro, T.R. Ioerger, and J. Yen, "*An Adaptive Algorithm for Learning Changes in User Interests*," Proc. Eighth ACM Int'l Conf. Information and Knowledge Management (CIKM), pp. 405-412, 1999.
[48] J. Xu and W.B. Croft, "*Cluster-Based Language Models for Distributed Retrieval*," Proc. 22nd Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 254-261, 1999.
[49] T.W. Yan and H. Garcia-Molina, "*SIFT—A Tool for Wide-Area Information Dissemination*," Proc. 1995 USENIX Technical Conf., pp. 177-186, 1995.
[50] Y. Yang and C.G. Chute, "*An Example-Based Mapping Method for Text Categorization and Retrieval*," ACM Trans. Information Systems (TOIS), vol. 12, no. 3, pp. 252-277, 1994.
[51] Y. Yang, "*Noise Reduction in a Statistical Approach to Text Categorization*," Proc. 18th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 256-263, 1995.
[52] Y. Yang and X. Liu, "*A Re-Examination of Text Categorization Methods*," Proc. 22nd Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 42-49, 1999.
[53] C.T. Yu, W. Meng, W. Wu, and K.-L. Liu, "*Efficient and Effective Metasearch for Text Databases Incorporating Linkages among Documents*," ACM SIGMOD, pp. 187-198, 2001.

V.Vijayadeepa received her B.Sc degree from university of Madras and M.Sc degree from Periyar University. She has completed her M.Phil at Bharathidasan University.She is having 10 years of experience in collegiate teaching and She is a Head of the department of computer applications in Muthayammal college of Arts and Science affiliated by Periyar University. Her main research interests include personalized Web search, Web information retrieval, data mining, and information systems.