DESIDOC Bulletin of Information Technology, Vol. 26, No. 4, July 2006, pp. 3-15

Generating User Profiles by Translating Content Queries to Concepts Using Thesaurus

K. Nageswara Rao* and Dr V.G. Talwar**

* Defence Research & Development Laboratory, Hyderabad-500 058
e-mail: knrao_hyd@rediffmail.com

** Professor, DOS in Library & Information Science, University of Mysore, Mysore

ABSTRACT

As more information becomes available electronically, information retrieval or filtering tools for finding information of interest to users become increasingly important. Building tools for assisting users in finding relevant information is often complicated by the difficulty in articulating user interest in a form that can be used for searching. The aim of the approach described here is to build instantaneous user profiles in a single episode information-seeking environment that instantaneously and accurately captures user interests with minimum user interaction. The research work described here focuses on the importance of accuracy and preciseness of user profiles that could be represented in terms of concepts, i.e., keywords or subject descriptors assigned to technical documents. Studies conducted here are in the context of understanding and building content-based profiling system for technical documents, specially in the field of aerospace science and technology. Sample data of aerospace grey literature in the form of technical papers with title and associated descriptors have been considered for the design and development of database and web-based profile builder for generating user profiles. The descriptors assigned to technical papers with the help of technical thesaurus were considered. Equal importance or weightage was given to broader and narrower terms, because there is no pre-defined mechanism to assign different weightage to different subject descriptors. Generated user profiles in terms of weighted vector of descriptors could directly be used for recommending technical documents with relevance ranking for latest addition of documents.

Keywords: Information retrieval, information filtering, recommender system, user profiles, grey literature, profile builder

1. INTRODUCTION

As more and more information becomes available electronically, the need for effective personalised information filtering has become indispensable. Recently, personalisation has become an important marketing tool for e-commerce applications, specially on internet. Personalisation is the ability to provide customised content and services tailored to

Received on 23 June 2006

individuals on the basis of knowledge about user preferences and their behaviour. Personalisation applications range from customised web content presentations to books, CDs and stock purchase recommendations. The major issues that must be taken care while implementing personalisation are: how to provide personal recommendations based on knowledge of individual user, how users behave while searching for documents or items, how similar a particular user is to other users of the system, and how to extract this knowledge from the available data and store it in user profiles for providing better services?

This paper discusses the various factors to be considered such as input/feedback and profile representation techniques to design, develop, and realise a technical document recommender system (RS) that has inbuilt capabilities to recommend documents to a user based on his interests.

There are many difficulties in developing good model of a user's interests. A variety of factors could be used to describe a user's interests. For example, a user provides a set of keywords or terms to describe his interests. The documents the user has read in the past, documents the user has purchased, subjective community the user belongs to, etc., could be the other sources for representing user's interests. Even though there is a clear-cut idea of which factors are important for predicting user interests, there is no guarantee that those factors alone can decide the information requirements of users.

A simple method of determining whether particular piece of information satisfies a user's interests is through keyword matching. If user's interests are described by certain keywords, then a set of documents containing those words should be treated as relevant. But in reality, it is not the case because inappropriate matches arise when people do not exactly reflect the topic or content of interest. It is due to the fact that a single word can have more than one meaning, and conversely, single concept can be described by many different words. Research survey shows that two people use the same word to describe an object 10 to 20 per cent of the time only[1].

Various information filtering and recommender system have addressed developing user profiles for better recommendations. Most of the RS use either content-based filtering or collaborative filtering approach for building user/customer profiles for recommendation of items. Some systems integrate both methods to best match the user requirements.

To overcome the problems in understating and building user profiles in content-based filtering, we have developed a novel approach. The Profile Builder constructs instantaneous single-episode user profiles for better recommendation of technical documents, i.e., books, articles, reports, etc., for which a number of keywords/subject descriptors were assigned using standard technical thesaurus. It considered the factors like descriptors, narrower terms (NTs), broader terms (BRs), related terms (RTs), and number of appearances of descriptors for building user profiles based on subject interest. For brevity, this paper discusses only the profile construction part of the RS. The importance of user queries in information retrieval and user profiles in information filtering for retrieving most relevant documents required by the user have also been discussed along with various techniques, algorithms, and approaches of building user profiles, both in content-based and collaborative filtering systems.

2. QUERIES IN INFORMATION RETRIEVAL

In a traditional information retrieval (IR) system, documents in the collection remain relatively static while dynamic queries are submitted to the retrieval system. The most common approach of the user for seeking information has been termed as 'ad hoc information retrieval'. In a similar but distinct approach, the queries remain relatively static while documents delivered are dynamic. This operational mode has been termed as 'information filtering'. Most of the users face difficulty in choosing correct descriptors due to lack of knowledge about thesaurus used by indexers. They also face difficulties in formulating a query and in using Boolean operators. The general criterion adopted to retrieve most relevant documents in the information retrieval process is through the relevance feedback mechanism. The main idea in relevance feedback cycle is the selection of important terms assigned to documents that have been identified as relevant by the user and enhance the importance of these terms in the formation of new query. The expected result of this process is that the new query will be moved towards retrieving more relevant documents instead of retrieving non-relevant documents.

The main advantage of relevance feedback over other query reformation strategies is that the mechanism shields the user from the details of query reformation process. When the user provides relevance judgment on retrieved documents, the system provides a systematic process to emphasise some terms that are relevant and others that are non-relevant.

Relevance feedback approaches are a form of supervised learning where a user indicates which retrieved documents are relevant or irrelevant. Experiments and studies using the smart system[2] and experiments using probabilistic weighting model[3-8] have shown good improvements in precision when relevance feedback is used. The improvements in precision are due to addition of new terms from relevant documents and modification of term weights based on the user-relevance judgments.

In addition, several statistical and artificial techniques have been used to capture better term associations and semantics, as information could be lost in the vector-based model. One such method to tackle the deficiencies in normal vector-based methods and handle synonymy and polysemy is the latent semantic indexing (LSI). In this system the latent structure in the pattern of word usage across documents is estimated. The description of terms, documents and user queries based on the underlying latent semantic structure is used for representing and retrieving information[9].

3. INFORMATION FILTERING—USER PROFILES

The earliest mechanism of electronic information filtering originated from the concept of selective dissemination of information (SDI). SDI was designed as an automatic way of keeping scientists informed of new documents published in their areas of specialisation. It helped scientists to create and modify user profile of keywords that described his/her interests. SDI system used the profile to match the keywords against new articles to predict which new articles would be most relevant to his/her interests. Recently, more or less the similar concept has been incorporated in personalisation, information filtering and RS by building user profiles interactively and intelligently.

Conventional information retrieval is closely related to information filtering: It has the goal of retrieving information relevant to the user while minimising the retrieval of irrelevant information[10,11]. The crux of the matter in information filtering is that the user profile describing user's preferences is constructed to compare incoming documents to determine which documents might be of interest to a particular user. There are three primary differences between information retrieval and information filtering[12]. First, user preferences (profiles) in information filtering typically represent long-term interests, while queries in information retrieval tend to represent short-term interests. Second, information filtering is typically applied to streams of incoming data, while in information retrieval, changes to the database do not occur often and retrieval is not limited to only the new items in the database.

Finally, information filtering involves the process of removing irrelevant information from a dynamic stream of data, while information retrieval involves the process of finding relevant information from static database. For example, information filtering is used for selection of new articles of interest from thousands of articles broadcast daily, selection of preferred judicial decisions or selection of articles from daily newspaper, etc.

In information filtering, the important aspect is not only ranking relevant documents but also the construction of a user profile, which truly reflects the user's preferences. A simplistic approach for constructing a user profile is to describe the profile through a set of keywords. If the user is not familiar with the type of incoming documents, he might find it difficult to provide keywords, which appropriately describe his preferences. In addition, an attempt by the user to familiarise with the vocabulary of the incoming documents might turn into a tedious and time-consuming task. In such an environment, the user may not precisely describe his profile. An alternative approach is to collect information from the user about his preferences implicitly and use such information to build user profiles.

In the very beginning stage of information filtering, the user provides a set of keywords which describe an initial profile of his preferences. As and when new documents arrive the system uses the initial profile to select documents, which are potentially of interest and shows them to the user. The user goes through relevance feedback cycle in which he indicates not only the relevant but also the non-relevant documents.

The system then uses this information to adjust or fine tune the user profile description in a way that it reflects the new preferences declared. The process of refining the user profile continues until the system stabilises and filters documents actually required by the user. Once the user profile is stabilised and there are no more changes in the profile, unless the user's interest shifts suddenly from one subject to another, the information filtering task can be viewed as a traditional information retrieval task in which the documents keep arriving into the system.

4. EXPLICIT *VS* IMPLICIT FEEDBACK FOR USER PROFILING

Relevance feedback has its history in information retrieval that dates back well over thirty years[13]. Relevance feedback is typically used for query expansion during short-term modelling of a user's immediate information needs and for user profiling during long-term modelling of a user's persistent interests and preferences. The feedback is nothing but the explicit or implicit input data given by the user while interacting with the retrieval or filtering system in order to build user's tastes or profiles.

4.1 Explicit Feedback

Traditional relevance feedback methods expect users to give the feedback explicitly. For example, specifying keywords, selecting and marking documents, or answering questions about their interests. Such relevance feedback methods force users to engage in additional activities beyond their normal searching behaviour. More often, it is difficult to collect the necessary data from users limiting the effectiveness of explicit techniques. It is observed from the developed information filtering systems that three ways of obtaining explicit relevance feedback are like/dislike, ratings, and provision of text comments. In like/dislike, users explicitly judge items on a binary scale, i.e., relevant/not relevant, interesting/not interesting, or like/hate. In rating, users are required to provide a relevance judgment on a discrete scale. For example, five-point numeric scale, hot/lukewarm/cold, or a graphical bar mapped to a numeric scale. In the third type feedback, users provide text comments about the usefulness of a single item for further processing and extraction of relevant items from the filtering system. But, explicit relevance feedback has the following disadvantages[14]:

- The type of numeric scales implemented in the filtering systems are not sufficient for the user to represent the relevance of items.

- Users are generally reluctant to provide relevance feedback when it is meeting their immediate goals.

- The relevance feedback must always be relative to the changing information need of a user and the relevance judgments of individual items are assumed to be independent.

4.2 Implicit Feedback

Implicit feedback techniques obtain information about users by intelligently observing their natural interactions with the filtering system. Behaviour of some users, that have been observed as source of implicit feedback, includes documents reading time/ time spent on a particular web page, links followed by the user, history of purchases, book-marking a web page, saving a document, printing a document, etc. The properties of documents that relate to the content a user is looking, i.e., language, document structure, i.e., format (text, image, audio, video) and document source, i.e., URL, publisher, author, etc., may also be considered for capturing the interests of users.

Other sources of implicit feedback include replying or forwarding e-mail, scrolling, maximising, minimising or resizing the widow containing the document or web page. Research in the area of learning to adapt user behaviour has shown good results by relying on techniques based on user-relevance feedback[15-22].

Implicit feedback techniques have been used to retrieve, filter, and recommend a variety of items; hyperlinks, web documents, academic and professional journal articles, e-mail messages, internet news articles, movies, books, television programmes, jobs, stocks, etc. The primary advantage that can be achieved by using implicit techniques is that such techniques require no direct feedback form the users. Implicit measures are generally considered less accurate than explicit measures, and large quantities of implicit data can be gathered at no extra cost to the user[23]. Further, an implicit feedback method seems to be promising for data in smaller and laboratory contexts[14]. However, implicit feedback measures can be combined with explicit ratings to obtain precise and more accurate representation of user interests thereby reducing the user's efforts in rating the items.

Some systems use both implicit feedback and explicit feedback techniques for better recommendation of items to their users. *Appendix 1* shows filtering/recommender systems, which utilises explicit feedback, implicit feedback, and sometimes both feedback mechanisms.

5. REPRESENTATION OF USER PROFILES

Several techniques have been developed to represent user profiles. Most profiles are constructed either directly by user-supplied items of interest or by automatic methods in which an agent is able to learn user preferences and build user profiles. Automatic profiling mechanisms can be classified in three main paradigms, i.e., statistical keyword analysis, social filtering algorithms, and machine-learning techniques[24].

Statistical keyword analysis is common and relies on standard information retrieval techniques. In this method, as keywords are analysed in isolation, there are some losses of contextual information that affects the accuracy of profiles. Balabanovic proposed an adaptive agent for web browsing wherein the user profile was represented by a single-feature-weighted vector using the term frequency-inverse document frequency (TF-IDF) technique[25]. The vector weights are increased or decreased based on the explicit positive or negative feedback from the users. Social filtering algorithms, instead of learning profiles, compare different users' profiles for constructing and representing user profiles. It generally needs a large community of users data to operate effectively.

Several machine-learning approaches can be used to learn a user profile, such as Bayesian classifier, nearest neighbour, PEBLS, decision trees, TF-IDF, neural networks, genetic algorithms, and memory-based reasoning[26-30]. Neural network techniques have been used to learn user's profile in research papers of many authors[31-33]. Others explore, genetic algorithms to learn user interests by incremental relevance feedback in NewT[34], Amalthea[35], and Widyantoro[36]. Most of the learning approaches also include relevance feedback analysis. Machine-learning techniques require a large set of training examples to train the system, which may be a serious problem in practical implementation of machine learning for user profile generation. A partial solution to this problem can be incremental buildup and improvement of user profiles based on relevance feedback.

6. PURPOSES OF GENERATING USER PROFILES

The main purposes of building user profiles are:

Automatic Notification: Profiles are used to provide automatic notification and SDI supported by electronic mail system. Through this users receive new events or know the arrival of items of their interests as and when these are added to the database.

Searching and Retrieval: Profiles are used to search the documents of interest by matching the attributes of profiles and documents. Profiles are also used to rank search results and presented in decreasing order of relevance. Precisely and accurately generated user profiles enable highlighted documents that better match user's interests. Previous research has shown that users fail to define their information needs accurately. For example, the query terms provided by users are poor predictors for relevancy of e-mail messages in comparison to terms identified automatically by an artificial neural net[37].

7. ROLE OF THESAURUS IN INDEXING & RETRIEVAL

According to Foskett, the main purpose of a thesaurus is basically to provide standard vocabulary for indexing and searching to assist users in locating terms for proper query formulation[38]. This mechanism provides classified hierarchies that allows broadening and narrowing of current query request according to the needs of the user.

The fundamental idea behind building a thesaurus is to use controlled vocabulary for helping in the indexing and retrieval of documents. A controlled vocabulary provides important advantages such as normalisation of indexing concepts, reduction of noise, identification of indexing terms with a clear semantic meaning, and retrieval based on concepts rather than words. Such advantages are practically important in specific domains such as science and technology disciplines for which there exist a large amount of compiled knowledge.

The terms are the indexing components of the thesaurus. Usually, a term in a thesaurus is used to denote a concept, which is a basic semantic unit for conveying ideas. Terms can be individual words, group of words, or phrases but most are single words. Further, terms are basically nouns and nouns are also verbs in gerund form when these are used as nouns. Whenever a concept cannot be expressed by a single word, a group of words is used instead.

A set of terms related to a given thesaurus term is mostly composed of synonyms and near-synonyms. In addition to these, relationships can be induced by patterns of co-occurrences within documents. Such relationships are usually of hierarchical nature and most often indicate broader terms (BR) or narrower terms (NR). However, the relation must also be a lateral or non-hierarchical in nature. In this case, the terms are called related terms (RT).

The BT and NT relationships define a classification hierarchy where the BT is associated with a class and its instances of that class. Further, it might be that a NT is associated with two or more BTs. While BT and NT relationships can be identified in a fully automatic manner, dealing with RT relationships is much difficult. RT relationships are dependent on the specific context and particular needs of the group of users and are difficult to identify without the knowledge provided by specialists.

8. PROFILE BUILDER—TRANSLATING CONTENT QUERIES TO CONCEPTS

The general approach for finding any document in a library setup by the user is with the help of card catalogue or online public access catalogue (OPAC). According to Hildreth, online catalogs have been classified into three generations[39]. In the first generation, OPACs were largely known as item-finding tools, typically searchable by author, title, and accession number. They contained relatively short, non-standard bibliographic records. In the second generation, OPACs comprised increased search functionality, i.e., access by subject heading, keywords, some Boolean search capability and ability to browse subject headings. These also offered a choice of display formats (short, medium, long) and improved usability (error messages). Third generation systems included strategy, assistance, free text, controlled vocabulary, and individualised displays.

Further, features incorporated in new systems are: Improved graphical user interface, support for Z39.50, hyperlinks, Dublin core metadata standard, and incorporation of Java programming. In general, OPACs only provide documents where the word occurrences match the user query without any relevance ranking. When the user wants to retrieve the most relevant documents of his interest with relevance ranking, OPACs would fail to provide such information. So, we have adopted an approach for building user profile in a single-episode information-seeking environment by capturing the documents retrieved by normal OPAC search mechanism. OPAC results are displayed to the user to select the most relevant documents. Based on the relevant documents selected by the user, a user profile is generated with the help of thesaurus terms and represented as weighted vector of descriptors for easy and accurate representation of his interests. We have studied and selected aerospace grey literature in the form of conference papers as a prototype database for experimental setup.

8.1 Source of Literature

For the design and development of experimental setup, aerospace grey literature i.e., the database of titles of technical papers/ articles, and subject descriptors/thesaurus terms assigned to these from the product called STAR (scientific and technical aerospace reports) was selected. The product has been selected because it is a comprehensive engineering and technology information resource providing world wide bibliographic coverage of published and unpublished scientific and technical literature.

8.2 Design of Database

Aerospace grey literature in the form of conference papers/articles database has been developed with corresponding subject descriptors for each paper as assigned by the technical team of NASA. Three table were created in Oracle with the following field specifications and established relationships between the tables as shown below:

Document Table	Doc_Key Table	Keyworld Table
Acc_No. Title	Acc_No. Key_No.	Keyworld Key_No.

The document table (DT) contains accession number and title data. Keyword table (KT) contains unique keywords with corresponding key numbers. Both DT and KT are linked through Doc_Key table with accession number from DT and key number from KT. The descriptors/keywords assigned to the titles of conference papers are considered equally important and equal weightage has been given to BT and NT because there is no existing mechanism for evaluating the relevancy of these terms as most or least relevant.

8.3 Development of Profile Builder

The profile builder (PB), a system for building user profile in terms of weighted vector of descriptors has been developed in Java programming language. It provides web-based interface for the user to search the document collection, followed by the selection of most relevant documents of his interest for building the user profile. All the descriptors with multiple occurrences are extracted from the database corresponding to the articles exclusively selected by the user. Based on the number of descriptors, weightage for each descriptor is calculated. When a particular descriptor is assigned to more than one title, its weightage is calculated accordingly. It is a common understanding that similar articles share more or less the same descriptors. For every user interacting with the system, user profile was created in terms of weighted vector of subject descriptors, which were tagged as most relevant documents by the individual user of the system. We felt the importance of descriptors as allocated by the subject experts to the articles during indexing process. Based on this concept, user profiles were developed because the descriptors are the authoritative terms for the representation of thought content of the subject or the articles. The PB system screen shots show the searching and selection of most relevant articles (Fig.1), and generating weighted vector of user profiles in terms of descriptors (Fig. 2). The system also has the provision for the modification and updation of already constructed user profiles.

9. FURTHER RESEARCH & CONCLUSIONS

In many of the information filtering and recommender system, building a user profile in terms of word occurrences has been used for retrieving and recommending documents. However, for recommending technical documents, such a study of building user profile in terms of weighted descriptors/concepts helps in retrieving precise and most relevant documents required by the user. The concept could be visualised as like-minded people or peers share similar characteristics of interests and read similar characteristic documents. Here, in our study, it is a general consensus that technical documents dealing with similar subjects share similar concepts or descriptors in common. Further research on designing a system to retrieve and recommend documents from a large database using weighted vector of user profile, which was built by the user normally in a single-episode information-seeking environment for the latest papers added to the database, is in progress. The process will enable the user not to search the database repeatedly thereby reducing the search efforts and valuable time. It is also proposed to develop a recommender system by initially displaying the most relevant documents, followed by the least relevant documents so that the user can decide whether or not to consult documents based on the relevance ranking.

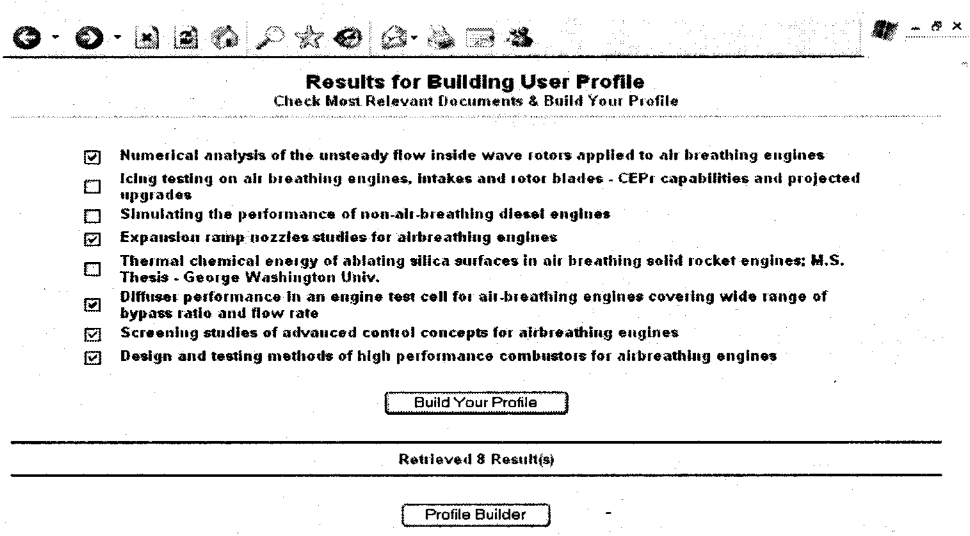
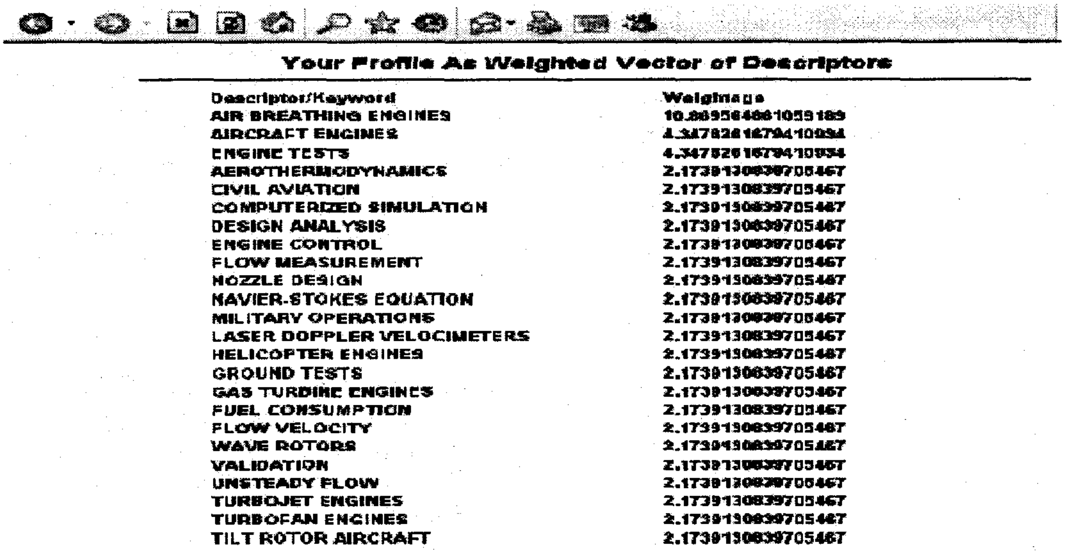


Figure 1. Searching and selection of most relevant articles using air-breathing engines as search worlds.



Your Profile As Weighted Vector of Descriptors

Descriptor/Keyword	Weightage
AIR BREATHING ENGINES	10.8695648661055189
AIRCRAFT ENGINES	4.3478261679410934
ENGINE TESTS	4.3478261679410934
AEROTHERMODYNAMICS	2.1739130839705467
CIVIL AVIATION	2.1739130839705467
COMPUTERIZED SIMULATION	2.1739130839705467
DESIGN ANALYSIS	2.1739130839705467
ENGINE CONTROL	2.1739130839705467
FLOW MEASUREMENT	2.1739130839705467
NOZZLE DESIGN	2.1739130839705467
NAVIER-STOKES EQUATION	2.1739130839705467
MILITARY OPERATIONS	2.1739130839705467
LASER DOPPLER VELOCIMETERS	2.1739130839705467
HELICOPTER ENGINES	2.1739130839705467
GROUND TESTS	2.1739130839705467
GAS TURBINE ENGINES	2.1739130839705467
FUEL CONSUMPTION	2.1739130839705467
FLOW VELOCITY	2.1739130839705467
WAVE ROTORS	2.1739130839705467
VALIDATION	2.1739130839705467
UNSTEADY FLOW	2.1739130839705467
TURBOJET ENGINES	2.1739130839705467
TURBOFAN ENGINES	2.1739130839705467
TILT ROTOR AIRCRAFT	2.1739130839705467

Figure 2. Generated weighted vector of user profile in terms of descriptors.

REFERENCES

1. Furnas, G.W.; Landauer T.K.; Gomez, L.M. & Dumais, S.T. Statistical semantics: Analysis of the potential performance of keyword information systems. *Bell Sys. Tech. J.*, 1983, **62**(6), 1753-806.

2. Salton, G. The SMART retrieval system–experiments in automatic document processing. 1971. Englwood Cliffs, NJ, USA, Prentice Hall Inc.

3. Robertson, S.E. & Sparck Jones, K. Relevance weighting of search terms. *J. of Ame. Soc. for Inf. Sci.*, 1976, **27**(3), 129-46.

4. Belew, R. Adaptive information retrieval: using a connectionist representation to retrieve and learn about documents. *In* Proceedings of the 12th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Cambridge, Massachusetts, USA, 1989. pp.11-20.

5. Salton, G. & Buckley, C. Improving retrieval performance by relevance feedback. *J. Am. Soc. Infor. Sci.* 1990, **41**, 288-97.

6. Harman, D. Relevance feedback revisited, edited by Nicholas J. Belkin; Peter Ingwersen, and Annelise Mark Pejtersen. *In* Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Copenhagen, Denmark, 1992. pp.1-10.

7. Haines, D. & Croft, B. Relevance feedback and inference networks. *In* Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Pittsburgh, Pennsylvania, USA, 1993. pp. 2-11.

8. Buckley, C.; Salton, G. & Allan, J. The effect of adding relevance information in a relevance feedback environment. *In* Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Ireland, 1994. pp. 292-300.

9. Deerwester, S.; Dumais, S.T.; Furnas, G.W.; Landauer, T.K. & Harshman, R. Indexing by latent semantic analysis. *J. Am. Soc. Infor. Sci.* 1990. **41**(6), 391-407.

10. Salton, G. & McGill, M.J. Introduction to modern information retrieval. McGraw Hill, New York, USA, 1983.

11. van Rijsbergen, C.J. Information retrieval. London, Butterworths, 1979.

12. Belkin, N. & Croft, B. Information filtering and information retrieval: Two sides of the same coin? *Communications ACM*, 1992, **35**(12), 29-38.

13. Spink, A. & Losee, R.M. Feedback in information retrieval. *Ann. Rev. Infor .Sci. Technol.*, 1996, **31**, 33-78.

14. Montaner, M.; Lopez, B. & De La Rosa, J.L. A taxonomy of recommender agents on the internet. *Artific. Intell. Rev.*, 2003, **19**, 285-330.

15. Amati, G.; Crestani, F. & Ubaldini, F. A learning system for selective dissemination of information. *In* Proceedings of the 15th International Conference on Artificial Intelligence, Nagoya, Japan, 1997. pp. 764-69.

16. Bharat, K.; Kamba, T. & Albers, M. Personalised, interactive news on the web. *Multimedia Systems*, 1998, **6**(5), 349-58.

17. Fitzpatrick, L. & Dent, M. Automatic feedback using past queries: Social searching? *In* Proceedings of the 20th International Conference on Research and Development in Information Retrieval, Philadelphia, USA, 1997. pp. 306-13.

18. Goker. A. & McCluskey, T.L. Towards an adaptive information retrieval system, edited by Zbigniew W. Ras and Maria Zemenkova. *In* Proceedings of the 6th Intenational Symposium on Methodologies for Intelligent Systems, Charlotte, N.C, USA, 1991. pp. 348-57.

19. Kindo, T.; Yoshida, H.; Morimoto, T. & Watanabe, T. Adaptive personal information filtering system that organises personal profiles automatically, edited by M.E. Pollack. *In* Proceedings of the 15th International Joint Conference on Artificial Intelligence, San Francisco, CA, USA, 1997. pp. 716-21.

20. Salton, G. & McGill, J.M. Introduction to modern information retrieval. Addison Wesley Publishing Co, Reading, Massachusetts, USA, 1989.

21. Velez, B.; Weiss, R.; Sheldon, M. & Gifford, D.K. Fast and effective query refinement. *In* Proceedings of the 20th International Conference on Research and Development in Information Retrieval (SIGIR-97), Philadelphia, USA, 1997. pp. 6-15.

22. Goldberg, D.J.; Nichols, D.; Oki, B.M. & Terry, D. Using collaborative filtering to weave an information tapestry. *Communications ACM*,1992, **35**(12), 61-70.

23. Nichols, D.M. Implicit ratings and filtering. *In* Proceedings of the fifth DELOS Workshop on Filtering and Collaborative Filtering Budapest, Hungary, 1997. pp. 31-36.

24. Soltysiak, S. & Crabtree, B. Automatic learning of user profiles towards the personalisation of agent services. *BT Technol. J.*, 1998, **16**(3), 110-17.

25. Balabanovic, M. An adaptive web page recommendation service. *In* Proceedings of the First International Conference on Autonomous Agents, Marina del Rey, California, USA. ACM Press, New York. 1997. pp. 378-85.

26. Joachims, T.; Freitag, D. & Mitchell, T. WebWatcher: A tour guide for the world wide web. *In* Proceedings of the 15th International Conference on Artificial Intelligence, Nagoya, Japan, 1997. pp. 770-75.

27. Lieberman, H. Letizia: An agent that assists web browsing. *In* Proceedings of the International Joint Conference on Artificial Intelligence, Montreal, Canada, 1995. pp. 924-29.

28. Pazzani, M.; Muramatsu, J. & Billsus, D. Syskill and Webert: Identifying interesting websites. *In* Proceedings of the 13th

National Conference on Artificial Intelligence and Eighth Innovative Applications of Artificial Intelligence Conference, Oregon, Portland, 1996. pp. 54-61.

29. Pazzani, M. & Billsus, D. Learning and revising user profiles: The identification of interesting web sites. *Machine Learning*, 1997, **27**(3), 313-31.

30. Sheth, B.D. A learning approach to personalised information filtering. Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 1994. Master's Thesis.

31. Wiener, E.; Pederson, J. & Weigend, A. A neural network approach to topic spotting. *In* Proceedings of the Fourth Annual Symposium on Document Analysis and Information Retrieval, Las Vegas, NV, USA, 1995. pp. 317-32.

32. McElligot, M. & Sorensen, H. An evolutionary connectionist approach to personal information filtering. *In* Proceedings of the Fourth Irish Neural Network Conference, University College, Dublin, Ireland, 1994. pp.141-46.

33. Tan, A. & Teo, C. Learning user profile for personalised information dissemination. *In* Proceedings of 1998 International Joint Conference on Neural Networks, Anchorage, AK, 1998. pp. 183-88.

34. Sheth, B. & Maes, P. Evolving agents for personalised information filtering. *In* Proceedings of the Ninith Conference on Artificial Intelligence for Applications, Orlando, FL, USA, 1993. pp. 345-52.

35. Moukas, A. Amalthaea: Information filtering and discovery using a multiagent evolving system. *Appl. Artific. Intell.: Inter J.*, 1997, **11**(5), 437-57.

36. Widyantoro, D.H.; Yin J.; El Nasr, M.S.; Yang, L.; Zacchi, A. & Yen, J. Alipes: A swift messenger in cyberspace. *In* Proceedings of the Spring Symposium on Intelligent Agents in Cyberspace, Palo Alto, CA, 1999. pp. 62-67.

37. Boger, Z.; Kuflik, T.; Shapira, B. & Shoval, P. Automatic keyword identification by artificial neural networks compared to manual identification by users of filtering systems. *Infor. Process. Manag.*, 2001, **37**(2), 187-98.

38. Foskett, D.J. Thesaurus: Readings in information retrieval, edited by K. Sparck Jones & P. Willet. Morgan Kaufmann Publishers Inc, San Francisco, USA, 1997. pp.111-34.

39. Hildreth, C.R. Online catalogue design models: Are we moving in the right directions? Report to the Council on Library Resources, Washington, D.C., Council on Library Resources, 1995.

Filtering/recommender systems with adopted feedback mechanism

Filtering/recommender system	Explicit feedback	Implicit feedback	Explicit + implicit feedback
The Adaptive Place Advisor (Goker & Thompson, 2000)	✓		
ACR News (Mobasher, Cooley & Srivastava, 2000)		✓	
Amalthaea (Moukas, 1997)	✓		
Anatagonomy (Sakagami, Kamba & Koseki, 1997)			✓
Beehive (Huberman & Kaminsky, 1996)		✓	
Bellcore Video Recommender (Hill, Stead, Rosenstein & Furnas, 1995)	✓		
Casmir (Berney & Ferneley, 1999)	✓		
CDNow (Hardie & Fader, 2001)			✓
CoFind (Dron, Mitchell, Siviter & Boyne, 2000)	✓		
Community Search Assistant (Glance, 2001)	✓		
Dietorecs (Arslan & Ricci, 2002)	✓		
Entrée (Burke, 2000)	✓		
Expertise Recommender (McDonald & Ackerman, 2000)	✓		
ExplaNet (Masters, 2004)			✓
Fab (Balabanovic, 1997b)	✓		
Fairwis (Buono, Costabile, Hemmje, Jaschke & Muscogiuri, 2001)			✓
Foxtrot (Middleton, De Roure & Shadbolt, 2002)	✓		
GroupLens (Konstan, Miller, Maltz, Herlocker, Gordon & Riedl, 1997)			✓
GroupMark (Pemberton, Rodden & Procter, 2000)		✓	
IfWeb (Asnicar & Tasso, 1997)	✓		
InfoFinder (Krulwich & Burkey, 1996)	✓		
INFOrmer (Sorensen, Riordan & Riordan, 1997)	✓		
InterestMap (Liu & Maes, 2005)	✓		
Jester (Goldberg, Roeder, Gupta & Perkins, 2001)	✓		
Krakatoa Chronicle (Kamba, Bharat & Albers, 1995), (Bharat, Kamba & Albers, 1998)			✓
LaboUr (Schwab & Pohl, 1999), (Schwab, Pohl & Koychev, 2000)		✓	
Let's Browse (Lieberman,Van Dyke & Vivacqua, 1999)		✓	
Letizia (Lieberman, 1995)		✓	
LIBRA (Mooney & Roy, 2000)	✓		
LifeStyle Finder (Krulwich, 1997)			✓
METIOREW (Bueno, Conejo & David, 2001),	✓		
MIAU (Baldes, et.al., 2003)	✓		
MovieLens (Good, et.al., 1999)	✓		
MyVU (Geyer-Schulz, Hahsler & Jahn, 2000),	✓		
News Dude (Billsus & Pazzani, 1999)	✓		
NewsWeeder (Lang, 1995)	✓		
NewT (Sheth & Maes, 1993)	✓		
Personal WebWatcher (Mladeni, 1998)		✓	
ProfBuilder (Wasfi, 1999)		✓	
PSUN (Sorensen & McElligot, 1995)	✓		
P-Tango (Claypool, Gokhale, Miranda, Murnikov, Netes & Sartin, 1999)	✓		
RACOFI (Anderson, et.al., 2003)	✓		
RASCAL (McCarey, O`Cinneide & Kushmerick, 2004)		✓	
Recommender (Basu, Hirsh & Cohen, 1998)	✓		
Ringo (Shardanand, 1994) (Shardanand & Maes, 1995)	✓		
SELECT (Alton-Schiedl, et.al., 1999)			✓
SIFT Netnews (Yan & Garcia-Molina, 1995)	✓		
SiteIF (Stefani & Strappavara, 1998)		✓	
Smart Radio (Hayes & Cunningham, 2000) (Hayes,			✓

Filtering/recommender system	Explicit feedback	Implicit feedback	Explicit + implicit feedback
SIFT Netnews (Yan & Garcia-Molina, 1995)	✓		
SiteIF (Stefani & Strappavara, 1998)		✓	
Smart Radio (Hayes & Cunningham, 2000) (Hayes, Cunningham & Smyth, 2001)			✓
Syskill & Webert (Pazzani, Muramatsu & Billsus, 1996) (Pazzani & Billsus, 1997)	✓		
Tapestry (Goldberg, Nichols, Oki & Terry, 1992)			✓
TiVo (Ali & van Stam, 2004)			✓
WebSail (Chen, Meng, Zhu & Fowler, 2002)	✓		
WebSell (Cunningham, Bergmann, Schmitt, Traphöner, Breen & Smyth, 2001)	✓		
Websift (Cooley, Tan & Srivastava, 1999).		✓	
WebWatcher (Armstrong, Freitag, Joachims & Mitchell, 1995) (Joachims, Freitag & Mitchell, 1997)			✓

Contributor

Shri K. Nageswara Rao received Masters degree in Physics in 1988 and Masters in Library and Information Science in 1992 from Sri Venkateswara University, Tirupati, and Annamalai University, respectively. From 1993 to 1995, he worked at the National Informatics Centre, Hyderabad. During 1995 to 1999, he worked at the Naval Physical and Oceanographic Laboratory, Kochi, as Scientist. Presently, he is Head, Technical Information Centre, Defence Research and Development Laboratory, Hyderabad. His research interests encompass internet technologies, web mining, design and development of digital libraries, and recommender systems. He has published/ presented six technical papers in national/international journals/ conferences.