

# A Tutorial on Information Filtering Concepts and Methods for Bio-medical Searching

Vinaitheerthan Renganathan<sup>1\*</sup>, Ajit N Babu<sup>2</sup> and Suptendra Nath Sarbadhikari<sup>3</sup>

<sup>1</sup>Skyline University College, University City, Sharjah, United Arab Emirates

<sup>2</sup>Center for Advancement of Global Health, Kochi, Kerala, India

<sup>3</sup>National Institute of Health and Family Welfare, Center for Health Informatics New Delhi, New Delhi, India

## Abstract

**Objective:** This tutorial presents an overview of information filtering and its application in bio-medical searching. It is intended for bio-medical investigators and students who have a basic understanding of the information retrieval concept, and would like to understand the workings of information filtering systems in various contexts to improve the efficiency and productivity of searches. This paper throws light on different information filtering models, types of information filtering, and the difference between information retrieval and information filtering.

**Keywords:** Information filtering system; Filtering agent; User profiles; Methods of gathering user information; Filtering models; Information filtering types

## Introduction

Vast amounts of information are now widely accessible on the web. Customarily, when a user wants to find interesting documents or data sources, the user has to actively search the World Wide Web. Searchers required effective means to efficiently find the information that they really need, and avoid the irrelevant information that does not match their interests. Information retrieval [1,2], and information filtering are two major information access techniques:

Information retrieval is concerned with retrieving the relevant documents from a large collection of material efficiently. It is concerned with the collection, representation, storage, organization, access, manipulation and display of information.

The immense volume of source information, however, often leads to query results which are too long and unwieldy for human users to manage effectively. The need, therefore, arises for more “intelligent” aids for information access tasks. Information filtering is an example of such an information access process.

Unlike information retrieval, information filtering generally focuses on users’ long-term information needs, often being stable preferences. It operates on dynamically changing information streams (e.g. email and news). Based on a user’s profile, which is initially derived from his or her interests, a filtering system processes a new item and takes appropriate actions that either ignore it or bring it to the user’s notice.

## Information Filtering and Information Retrieval

Information filtering and information retrieval both have similar aims, [3]. Each wants to retrieve information according to the user request; they try to minimize as much as possible the amount of irrelevant information. But, there are key differences between information retrieval and information filtering, [4], as noted in Table 1.

## Need for information filtering techniques in the biomedical field

Information filtering is a critical resource for biomedical researchers as the Internet contains vast volumes of heterogeneous data and analytical tools.

- Information filtering tools can reduce the time spent by a

biomedical researcher in getting relevant information, which can in turn, enhance the timeliness and quality of research.

- It will help the researcher in transforming the information to knowledge in a shorter time period.

Need for information filtering felt in different biomedical sub fields

- Medicine
- Public health
- Bioinformatics
- Medical informatics

A well-known example is PubMed. There are different sources of information available for medical professionals, such as Pubmed, which necessitates the information filtering role in getting the relevant information which suites to the profile of interest of the medical professionals.

	Information retrieval	Information filtering
Information need	Dynamic	Static
Information source	Static	Dynamic
User profile	Not necessary	Essential
Scope	Generalized	Specific
Information seeking Behavior	Short term	Long term
User Query	Brief	Description or explanation of the information
User interaction with the system	Single information seeking episodes	Series of information seeking episodes

Table 1: Information retrieval vs filtering system.

**\*Corresponding author:** Vinaitheerthan Renganathan, Head of Institutional Research, Skyline University College, P.O. Box 1797, University City, Sharjah, United Arab Emirates, Tel: 971-6-5441155; Fax: 971-6-5441166; E-mail: [vinairsearch@yahoo.com](mailto:vinairsearch@yahoo.com)

**Received** May 13, 2013; **Accepted** September 03, 2013; **Published** September 09, 2013

**Citation:** Renganathan V, Babu AN, Sarbadhikari SN (2013) A Tutorial on Information Filtering Concepts and Methods for Bio-medical Searching. J Health Med Informat 4: 131. doi:10.4172/2157-7420.1000131

**Copyright:** © 2013 Renganathan V, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## Information Filtering System Architecture

Information filtering system consists of a filtering agent and user profile (Figure 1).

### Filtering Agent

Filtering agent [5,6] acts an interface between the user and the document system, and helps the user in finding the relevant topics of a given topic through the user agent. It reduces the user's time and effort in locating the relevant document through the specialized domain knowledge it possesses.

The filtering agent filters out irrelevant incoming documents and presents to the user only those documents which match the user's interest. During the course of time, the filtering system becomes more effective as it learns the user's preferences and develops accuracy in performing the filtering tasks. Its roles include interfacing with the source document subsystem, managing the user-profile (which is discussed in the Section 7), calculating the relevance of a document-vector to the user-profiles and communicating with the user.

The process flow associated with the filtering agent:

1. The topic for the current document filtering session is obtained from the user
2. User profile vector is initialized using the current topic's title and text
3. Document Vector  $d$  is obtained from source document subsystem
4. Current user-profile vector  $t$  is retrieved
5. The following probabilities are calculated  
 Probability of Relevance (PR)=Probability (Relevant/ $d$ ;  $t$ )  
 Probability of Non-Relevance (PN)=Probability (Non-Relevant/ $d$ ;  $t$ )
6. If  $PR > PN$ , then for the given relevant document  $d$ :
  - (a) Text of  $d$  is obtained from source document subsystem
  - (b) User is informed text of  $d$
  - (c) Actual relevance judgment is obtained from the user for the document  $d$
  - (d) User-profile vector  $t$  is updated with the actual relevance judgment of  $d$
7. Repeat from Step 3.

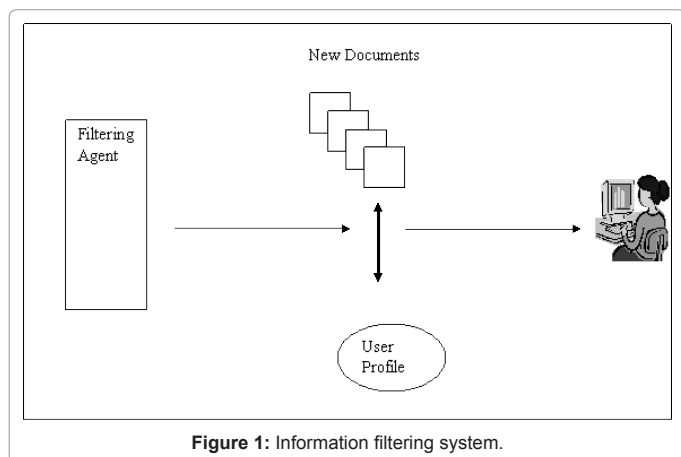


Figure 1: Information filtering system.

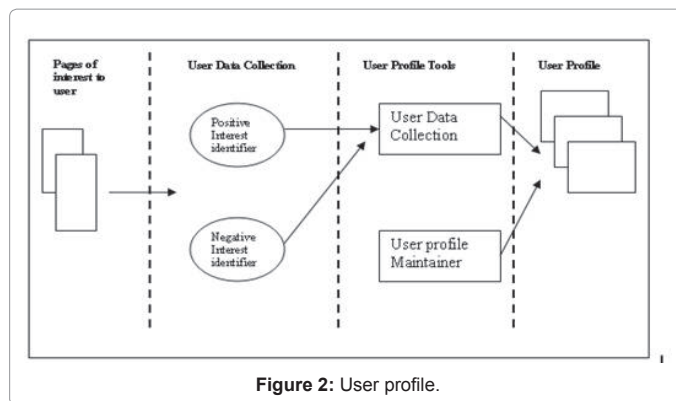


Figure 2: User profile.

### User Profiling

User profile [7-10] (Figure 2) represents users' details (User interest, needs, goals, and behavior), and is constantly updated in response to user feedback. The quality of user profiles has a major impact on the performance of information retrieval and filtering systems.

In order to provide personalized information to a user, the system creates and maintains a description of the type of information that the user interested to access. Personalized content is retrieved based on information matching the user profile.

The user profile is divided into two categories static or dynamic [11]. Static profiling is the process of acquiring a user's characteristics, such as age, gender, profession, etc., through direct input from the user.

Dynamic profile is created based on the future or current actions of the user, such as current location, position and occupation and browser used, etc.

### Example of user profile and topic of interest

User profile of a cardiologist contains the topics of his interest (for example myocardial infarction), and it will help the filtering system to deliver articles on myocardial infarction, whenever it is published through the user profile created in the system. Apart from the article, the user profile can be used to recommend a movie or music based on his hobbies mentioned in the user profile.

### User profile

User profile of medical physician who wishes to use an information filtering system

- Age: 38
- Sex: Male
- Education: DM (sub-specialization in cardiology)
- Field of work: Cardiology
- Topic of interest: Myocardial infarction
- Hobbies: Soccer, music

### User profile for personalized cancer information system for patients

**Types of interest:** Stability of interests

Direct to short-term interests

Main interest groups Explanation of medical issues concerning the

patient, especially medical problems and treatments

### User preferences

Important preferences-

Privacy sacrifice anonymous use by the server

### Task and goal characteristics

Available Time-short to medium

Frequency of use-irregular

Types of task to gain knowledge about the personal medical situation and treatments

### Importance

Urgency-high

Delivery patterns-almost synchronously

### User characteristics

**Current medical problems:** Past medical problems

Treatments

Test results

Personals details like age and sex

### User profile

Initialization-implicit via medical dossier

Maintenance-implicit via medical dossier, which is updated by the medical staff

### Methods for gathering user information

User profile helps the Filtering system in the personalization process and the user profile can be created using three methods, namely explicit, implicit and mixed method.

### Explicit methods

The explicit method requires the user to furnish the information, such as preferences explicitly in the form of questionnaire or template. This type of personalization is known as customization. The explicit methods are adapted for the acquisition of the characteristics of the users, such as name, age, preferences and priorities. Explicit method also involves relevance feedback [12-14] process, where the user evaluates (in the form of rating or ranking) the relevancy of the document the filtering system returned. Relevance feedback helps the filtering system to explicitly update the user profile.

### Implicit methods

The implicit method employs procedures, which records the information about the changing behavior or characteristics of the user. In this method, the user will not know the process of collecting information. Implicit methods are appropriate to record the actions of the user, such as time spent; items purchased, history of movement, key pressed, mouse movement. This method helps the filtering system to maintain the updated user profile and helps the system to detect change in the user's interest [15].

### Mixed methods

The mixed method uses a combination of explicit and implicit

methods. The system is constantly looking for patterns in the behavior of a user combined with explicit questions to the users to verify the assumptions about user characteristics, such as the rating of the users on what extent the output of the system matches his preferences. Also, a mixed system begins its task by requesting the users to provide information about the preferences which solves the cold start problem.

### The Filtering Model

Information filtering models may be interpreted as decision functions whose domain is the set of all possible document features, and whose range is the set (relevant, non relevant). Even though it has its roots in the techniques of information retrieval, the filtering algorithm differs from personalized search [16]. The following are the important models used in the information filtering process.

### String matching model

In the string matching traditional model [17], the user specifies his/her information needs by a string of words. A document would match the information need of a user if the user-specified string exists in the document. This method is one of the earliest and simplest approaches. The method is less able to match the documents that require contextual and experiential knowledge, and also it suffers from the following problems homonymy (words are spelled same, but have different meanings), synonymy (different words having the same meaning), polysemy (words with multiple meaning) and bad response time.

### Boolean model

The Boolean model [18,19] is a modification of the above method where the user can combine words using primitive Boolean operators such as AND, OR and NOT. This method gives the user a tool to better express his/her information need, but at the same time, requires some skills on behalf of the user. It assigns equal weights to all terms in the query. This model also has the same problems of the string matching models.

### Vector space model

The Vector Space [18-21] Model represents queries (profiles) or documents in the user profile as vectors in a vector space, with component terms weighted.

A document  $K$  is represented as a vector of dimension  $n$ , where  $n$  is the total number of terms. Each term is given a weight, which represents its importance in the document and in the whole document collection. The vector  $K$  is represented as  $K=(w_1 \dots w_n)$ , where  $w_i$  is the weight assigned to the  $i$ -th term. Similarly, the query (profile) vector is represented as  $U=(x_1 \dots x_n)$ .

Vector is created from the list of words the document contains and words with high frequency and low content discriminating powers (such as 'and', 'the') are excluded. The weight of the terms is calculated by multiplying the term frequency (tf) and inverse document frequency (idf). tf is the how frequently the term appears in the document and idf is related to how frequently it appears in the whole document collection.

The relevance of an unseen document to a given query (profile) is judged by calculating the distance between the query (profile) vector and the document vector, and comparing this distance to a threshold value  $P$  (vector distance is established using some pre specified distance-metric). Normally, the distance is calculated using the cosine similarity measure [21].

### Latent semantic index model

Latent Semantic Index (LSI) model [22,23] works on the assumption that there is a latent structure in the pattern of words usage across the documents that can be exploited to overcome one drawback of vector space model, which assumes that the words are orthogonal or independent. It creates multi dimensional semantic structure of information using implicit higher order structure of association of terms with the documents. Latent Semantic index works on singular value decomposition technique (SVD) [22,23], which is also used to reduce the dimensionality of the large volume of data. LSI creates a semantic space for a set of documents which are previously judged by the user as relevant or not. New document is categorized as relevant if it is close to the interesting documents in the semantic space. On the other hand, if it is close to the non interesting documents in the semantic space, it is categorized as irrelevant.

### Bayesian network model

Bayesian network principles can be used in the collaborative filtering system. The rating data can be used to learn a Bayesian network, [24] where each item is represented by a node, and directed arrows between items signify user's interest on items influence interest of other items. The Bayesian network is used to create probabilistic decision trees for each item, where leaf nodes are likelihoods of the target user's interest on the target item, and intermediary decisions are based on the target user's view on the parent items of the target item from the network. Bayesian networks is useful when the user preferences changes slowly with respect to the time needed to build the model, but are not suitable for environments in which user preference models needs to be updated frequently.

### Artificial neural network models

Artificial neural networks models in the information filtering systems [25,26], are used to automatically create the terms in user profiles by the training of neural network models through examples. The ability of the ANN to model non-linear relationships can be applied to the matching of the documents to the user profile. Neural networks can be used to represents the user's preferences where words from documents are represented as nodes and strength of association between words in the same document. There are two types of learning algorithms are available in the neural networks: one is supervised learning and another unsupervised learning.

### Markov models

Markov models based information system works on the Markovian Decision Process [MDP] [27], which involves the stochastic model of sequential decisions. Given a number of observed events (i.e. past history of the users actions, such as last visited web page or characteristic of the document), the next event is predicted from the probability distribution of the events which have followed these observed events in the past [28]. Mitios online book store [27] is an example of such systems which uses the MDP Model.

### Filtering Types

#### Content-based

Content Based filtering system [29-31] recommends a document by matching the document profile with the user profile, using traditional information retrieval techniques such Term Frequency and Inverse Document frequency (TF-IDF). User characteristics are gathered over time and profiled automatically based upon a user's prior feedback and

choices. The system uses item to item correlation in recommending the document to the user. The system starts with the process of collecting the content details about the item, such as treatments, symptoms etc. for disease related item and author, publisher etc. for the book items. In the next step, the system asks the user to rate the items. Finally, system matches unrated item with the user profile item and assign score to the unrated item and user is presented with items ranked according to the scores assigned.

News Dude [32], is one of the examples of content based filtering system which uses short term TF-IDF technique and long term Bayesian classifier for learning on an initial set of documents provided by the user.

Content based information filtering systems are not affected by the cold start problem and new user problem, as the system focuses on the individual user needs Content based information filtering systems are not suitable for multimedia items, such as images, audio, video. Multimedia documents must be tagged with a semantic description of the resource which will be a time consuming process. Content-based filtering methods cannot filter documents based on quality and relevance.

#### Collaborative filtering

Collaborative filtering systems [33-36] filters information based on the interests of the user (past history), and the ratings of other users with similar interests. It is widely used in many filtering systems or recommender systems, especially in ecommerce applications. One of the examples of such system are Amazon.com and e-Bay, where a user's past shopping history is used to make recommendations for new products.

Collaborative filtering system involves the computation of similarity between user interests. Similarity between the users interest are calculated using different methods such as Pearson correlation coefficient. The system collects the ratings of each item from different users explicitly or through their behavior, and then calculates the similarity between the ratings of the users. The ratings can be explicit on a numeric scale [18,37], or implicit such as purchases, clicks and mouse movement. Then, the users are grouped based on the calculated similarity measures and future items are recommended to the user based on the recommendation of other users in the group.

Consider a group of users  $U_1, U_2, \dots, U_n$  and items  $I_1, I_2, \dots, I_m$ . The Table 2 shows the rating given by the users on different items.

For example: If similarity rating between the user  $U_1$  and  $U_5$  is high, then user  $u_1$  and  $u_5$  can be grouped and new items will be recommended to each user based on the other user's interest. Here, item  $I_3$  will be recommended to the user  $U_5$ , as a new item based on the high rating given by the other user in the group  $U_1$ . Similarly, item  $I_m$  will be recommended to User  $U_4$  based on the rating of other user  $U_3$ .

There are two types of collaborative filtering systems are available:

	I1	I2	I3	.....	Im
U1	1	4	4		4
U2	1	3	4		3
U3	2	4	3		5
U4	2	4	3		
U5	1	4			4
....					
Un	3	4	1		4

Table 2: Rating given by the users on different items.



one is memory based and another model based. Memory based collaborative filtering systems works on the neighborhood principles to recommend items and model based collaborative systems uses models in recommending items by estimating models based on the ratings. The models are built using different machine learning algorithm, such as Bayesian network, Clustering and Markov models.

The collaborative systems can be used to filter all types of items, including the multimedia items. It suffers from the cold-start problem and early rater problem. It involves the issues of filtering a new item, if it is not rated by any one of users yet. This system also suffers when data are sparse, which makes the recommendation difficult, as there are few common items present in calculating the similarity measures.

### Application of collaborative filtering in bio-Medical field

The following section provides examples of application of collaborative filtering in the Health Care field.

1. Collaborative filtering is used in the form of recommender systems in the Health Care field. In nursing clinical recommender system [38], the user (say nurse) selects all the required items for a particular treatment plan, and not selecting a particular item of interest based on his likes, as in the commercial recommender systems. The user behavior in the clinical recommender system is binary, either accept or reject the item, and the here, the requirement is objective not subjective.

2. Collaborative filtering is used to overcome the omission of medication [39], by detecting the omission from the Patient's observed medication list and reconciling the medication lists.

3. Collaborative filtering is applied to make personalized medical predictions using the CARE system [40], a collaborative recommendation engine for prospective and proactive health care, using the assumption that patients with similar history will continue to develop similar conditions.

4. CHORUS [41,42], a Collaborative filtering system, is used to organize online radiology resource effectively and retrieve information from the resources in a useful manner.

5. Diabetic recommender system [43,44], a collaborative filtering system which suggests a diabetic friendly food option when a person shops for groceries or eats at restaurant based on algorithm developed by Yehuda Koren from Yahoo Research.

6. Walk et al. [45] used collaborative filtering tool to perform high level mapping of recommender techniques to collaborative ontology engineering platforms in the biomedical domain, and as a proof-of-concept in the form of implementations in the context of the ICD-11 project

7. Smart H tweet engine [46], provides users with personalized health related recommendations and alters based on stored clinician knowledge. It extracts user interest, health condition and emotions from social media to create rich user profile.

8. Elsten [47] built life style coach a recommender system, which facilitated the decision process of its users by filtering large quantities of data and recommended only a few personally relevant item using collaborative filtering techniques.

9. Kotevska [48] proposed a collaborative patient-centered health care system model which provides tool for personal health care by generating different recommendation, notification and suggestion to the users, using collaborative filtering technique.

10. Song et al. [49] proposed a Health Social Network Recommender System which works on the principles of Support Vector Machine (SVM), to provide a social networking frame work for patient care, in particular, parents of children with Autism Spectrum Disease (ASD).

### Hybrid filtering systems

The hybrid filtering systems combines features of both the content and collaborative filtering systems. The hybrid system overcomes the problem of cold start and early rater problem by using the content based approach in the initial stage. In the subsequent stages, it uses collaborative filtering systems features, which helps the system to recommend all types of items, including multimedia items and overcomes the problem related to content based filtering techniques.

### Social filtering

The most successful social filtering [50], system is Yahoo. Yahoo employs humans to evaluate documents and puts documents, which are interesting, into its structured information database. The simplest and most common filtering is by organizing discussions into groups (newsgroups, mailing lists, forums, etc.). Each group has a topic, and wants only contributions within that topic. Sometimes, the right to submit contributions is restricted.

### Filtering against spamming

Filtering against spamming [51], involves both content based filtering techniques and collaborative filtering techniques [52]. Content based spam filtering techniques which use the entire content of the emails to find out the words and phrases, which are classified as spam by the users. The content based spam filtering uses the rule based techniques, nearest neighborhood method and Bayesian networks to filter the Spam. Spam Assassin [53], is one of the spams filtering, which uses Bayesian network principles to filter the emails.

In collaborative spam filtering, when the users classifies the email as spam, it will be added to the central database in the email servers through signatures added to the email. A signature is computed for every new email and is compared with the database, and if it matches, it will be classified as spam. It works on the principles of near duplicate similarity matching techniques [54]. Yahoo! uses the collaborative spam filtering system to classify the email as spam.

### Effectiveness of the Filtering System

The effectiveness of the filtering system is determined by the precision and recall concept [55,56]. Precision is number of relevant document retrieved divided by the total number of documents.

Recall is the number of relevant document retrieved divided by the total number of relevant document.

### Information filtering in different forms

Information filtering system is used to filtering the different forms of information items, such as

- Articles
- Emails
- Music and movies
- News
- Images

## Information Filtering Software and Systems

The perceived benefits of information filtering have motivated several researchers to develop information filtering software for biomedical researcher, some of which are:

- BioSifter [57,22], which is a powerful tool, which can retrieve relevant documents from biological literature based on the user's interest profile.
- PubCrawler [58], is a free altering service that keeps users informed of the current contents of Medline and Genbank.
- A personalized cancer information system [12], is available for cancer patients to receive information about treatment, disease measurement related to their medical condition.
- Oncosifter [59], is available for retrieving the latest news, diagnosis and treatment information related to cancer.
- MARVIN [60], filters relevant documents from a set of Web pages and follows links to retrieve new documents.
- PURE [61], a PubMed article recommender system, which automatically finds articles, which are relevant to user interest using content based filtering technique

Apart from these programs, there is other generic filtering software, which are also called recommender systems are available:

- SIFT [62], is an information filtering system, and helps the users to receive new documents based on the profiles submitted to the system through the profile index.
- Beehive [63], a distributed system for social sharing and filtering of information, provides a simple and intuitive interface for distributing relevant information.
- Tapestry system [64], support both content-based filtering and collaborative filtering. It involves people in collaborative filtering by recording their reactions to documents they read, and which helps to filter the documents effectively.
- InRoute [46], an Information filtering system works on the inference network model
- GroupLens, [65,66]-GroupLens is a system for collaborative filtering of netnews, to help people find articles they will like in the huge stream of available articles.
- RAMA [67], helps to retrieve useful information from various Internet sources  
including USENIX news and anonymous FTP servers.
- PI-Agent [6], implements an information filtering system using importance based classification realized by implicit pre-classification profiles.
- NewsSIEVE [51], is an information filtering agent which automatically generates simple user profiles to make the review process easier and improves the performance of the filtering system.
- SysKill and Webert [68], a software agent that learns to rate pages on the Worm Wide Web (WWW) capable recommending pages to the users which might of interest to the user.
- INFOS [69]-System reduces the user's search burden by automatically eliminating Usenet news articles predicted to be irrelevant.

- iAgent [70] is an Information filtering system to filter the newspapers.
- NewT [71], a system which helps the user filter Usenet Netnews.
- Ringo [72]- Personalized music recommender system accessible to users via email.

## Challenges of Information Filtering

- Explication of long-dated information needs in the shape of a user profile is a difficult job, as the user needs and interests depend on many parameters, including the personal, behavioral characteristics. Some of the parameters are not easily translated into system components (user's mood, work load, financial position, etc.)

- User profile has to specify current and future documents correctly. The quality of user profiles is a key to making a filtering system work. From the user's point of view, there are two potential problems. One is if a large proportion of the items that the system sends to a user are irrelevant, and then the system becomes more of an annoyance than help. Conversely, if the system fails to provide the user with enough relevant information, then the benefit of information filtering delivery is largely lost, because the user will still have to actively hunt for information.

- Total utility of an Information Filtering system is maximal, if effort in programming the user profile is low, and the quota of correct classified documents is high. If the use profile is of high volume, make filtering system over burden.

- Humans are often not able to explain in general how they evaluate the relevance (class affiliation) of a document, i.e. the user may not be able to explain the basis on which he has given the relevance feed back to filtering system.

- Synonymy, Homography, while in user profile may pose a problem, etc.

- Privacy is a important issue in maintaining the user profiles.

## Conclusion

Efficient information retrieval requires information filtering and search adaptation to the user's current needs, interests, knowledge level, etc. From early days of SDI (selective Dissemination of information) to current modern information retrieval, information filtering has undergone a tremendous change. In this paper I have tried to focus on methods of Information Filtering. I have not covered all aspect of the topic like the technical details behind the methods because it is behind the scope of this article. My main aim of this paper is to pinpoint the highlights of Information Filtering, and to draw a comprehensive picture of methodology behind Information Filtering.

## Author's Contribution

Vinaitheerthan wrote the first draft. Dr. Ajit N Babu and Dr. SN Sarbadhikari have meticulously rewritten and enhanced the quality of the manuscript for making it more acceptable to biomedical researchers.

## References

1. Baeza-Yates R, Ribeiro-Neto B (1999) Modern information retrieval. ACM press, New York, USA.
2. Si L, Jin R (2003) Flexible mixture model for collaborative filtering. Pittsburgh, PA 15232 USA.
3. O'Riordan A, Sorensen H (1995) An intelligent agent for high-precision text

- filtering. Proceedings of the fourth international conference on Information and knowledge management, Baltimore, Maryland, USA.
4. Callan J (2000) System and method for filtering a document stream. Google Patents US.
  5. Yan Wang, Jie Liang, Jianguo Lu (2013) Discover hidden web properties by random walk on bipartite graph. *Information Retrieval*.
  6. Kuroпка D, Serries T (2001) Personal information agent.
  7. Foltz PW (1990) Using latent semantic indexing for information filtering. Proceedings of the Conference on Office Information Systems Cambridge, MA, UK 40-47.
  8. Golemati M, Katifori A, Vassilakis C, Lepouras G, Halatsis C (2007) Creating an ontology for the user profile: Method and applications. Proceedings of the First RCIS Conference.
  9. Osman A, Mantoro T (2011) A user profile for information filtering using RFID-SIM card in pervasive network. *Multimedia Computing and Systems (ICMCS), International Conference, IEEE*.
  10. Sorensen H, O'Riordan A, O'Riordan C (1997) Profiling with the informer text filtering agent. *Journal of Universal Computer Science* 3: 988-1006.
  11. Liaquat H, Iftikhar N, Ali S, Iqbal ZZ (2007) A frame work for query results refinement in multimedia databases. *World Academy of Science, Engineering and Technology*.
  12. Cawsey AJ, Jones RB, Pearson J (2000) The evaluation of a personalised health information system for patients with cancer. *User Modeling and User-Adapted Interaction* 10: 47-72.
  13. Van Setten M (2001) Personalized Information Systems. Giga CE Project Part of Gigaport Project, Telematica Instituut.
  14. O'Riordan C, Sorensen H (1997) Information filtering and retrieval: An overview. *Citeseer*.
  15. Lam W, Mukhopadhyay S, Mostafa J, Palakal M (1996) Detection of shifts in user interests for personalized information filtering. Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval New York, NY, USA 317-325.
  16. Liu F, Yu C, Meng W (2004) Personalized web search for improving retrieval effectiveness. *Knowledge and Data Engineering IEEE Xplore* 16: 28-40.
  17. Croft WB, Das R (1989) Experiments with query acquisition and use in document retrieval systems. Proceedings of the 13th annual international ACM SIGIR conference on Research and development in information retrieval 349-368.
  18. Salton G, McGill MJ (1986) Introduction to modern information retrieval. New York, NY, USA.
  19. Zhou W, Smalheiser NR, Yu C (2006) A tutorial on information retrieval: Basic terms and concepts. *Journal of Biomedical Discovery and Collaboration Chicago, USA* 1: 2.
  20. Becker J, Kuroпка D (2003) Topic-based vector space model. *Business Information Systems, Proceedings of BIS 2003, Colorado Springs, USA*.
  21. Yan TW, Garcia-Molina H (1994) Index structures for selective dissemination of information under the boolean model. *ACM Transactions on Database Systems New York, NY, USA (TODS)* 19: 332-364.
  22. Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R (1990) Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41: 391-407.
  23. Foltz PW (1990) Using latent semantic indexing for information filtering. *ACM SIGOIS Bulletin ACM*.
  24. Pennock DM, Horvitz E, Lawrence S, Giles CL (2000) Collaborative filtering by personality diagnosis: A hybrid memory-and model-based approach. Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence (UAI-2000), Morgan Kaufmann Publishers Inc 473-480.
  25. Boger Z, Kuflik T, Shapira B, Shoval P (2000) Information filtering and automatic keyword identification by artificial neural networks. Proceedings of the 8th European Conference on Information Systems.
  26. Jennings A, Higuchi H (1993) A user model neural network for a personal news service. *User Modeling and User-Adapted Interaction* 3: 1-25.
  27. Brafman R, Heckerman D, Shani G (2003) Recommendation as a stochastic sequential decision problem. Proceedings of ICAPS, USA.
  28. Zukerman I, Albrecht DW (2001) Predictive statistical models for user modeling. *User Modeling and User-Adapted Interaction* 11: 5-18.
  29. Melville P, Mooney RJ, Nagarajan R (1999) Content-boosted collaborative filtering for improved recommendations. Proceedings of the National Conference on Artificial Intelligence (AAAI-02) Edmonton, Alberta, Canada 187-192.
  30. Yu K, Schwaighofer A, Tresp V (2002) Collaborative ensemble learning: Combining collaborative and content-based information filtering via hierarchical bayes. Proceedings of the 19th conference Uncertainty in Artificial Intelligence: Proceedings of the 19th Conference (UAI-2003) 616-623.
  31. Rao KN (2010) Application domain and functional classification of recommender systems-A survey. *DESIDOC Journal of Library & Information Technology* 28: 17-35.
  32. Billsus D, Pazzani MJ (2000) User modeling for adaptive news access. *User modeling and User-adapted Interaction* 10: 147-180.
  33. Goldberg D, Nichols D, Oki BM, Terry D (1992) Using collaborative filtering to weave an information tapestry. *Commun ACM* 35: 61-70.
  34. Breese JS, Heckerman D, Kadie C (1998) Empirical analysis of predictive algorithms for collaborative filtering. Morgan Kaufmann Publishers, USA.
  35. Hill W, Stead L, Rosenstein M, Furnas G (1995) Recommending and evaluating choices in a virtual community of use. Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY, USA 194-201.
  36. Su X, Khoshgoftaar TM (2009) A survey of collaborative filtering techniques. *Advances in Artificial Intelligence* 2009: 19.
  37. Sheth B, Maes P (1993) Evolving agents for personalized information filtering. *Artificial Intelligence for Applications. Proceedings Ninth Conference, IEEE*.
  38. Duan L, Street W, Xu E (2011) Healthcare information systems: data mining methods in the creation of a clinical recommender system. *Enterprise Information Systems* 5: 169-181.
  39. Hasan S, Duncan GT, Neill DB, Padman R (2011) Automatic detection of omissions in medication lists. *J Am Med Inform Assoc* 18: 449-458.
  40. Davis DA, Chawla NV, Blumm N, Christakis N, Barabasi AL (2008) Predicting individual disease risk based on medical history. Proceedings of the 17th ACM conference on Information and knowledge management 769-778.
  41. Kahn Jr CE (2005) Collaborative filtering to improve navigation of large radiology knowledge resources. *Journal of Digital Imaging* 18: 131-137.
  42. CHOURS-Collaborative Hypertext of Radiology. National Library of Medicine, USA.
  43. Chin CM (2012) Mobile Health Monitoring: The Glucose Intelligence Solution. University of California, Berkeley, USA.
  44. <http://www.diabeats.co>.
  45. Walk S, Strohmaier M, Tudorache T, Noy NF, Nyulas C, et al. (2011) Recommending Concepts to Experts: An Exploration of Recommender Techniques for Collaborative Ontology Engineering Platforms in the Biomedical Domain. Stanford, CA 94305-5479, USA.
  46. Batool R, Khan WA, Hussain M, Maqbool J, Afzal M, et al. (2012) Towards personalized health profiling in social network. *Information Science and Service Science and Data Mining (ISSDM), 6th International Conference on New Trends Yongin, South Korea* 760-765.
  47. Elsten T (2012) Supporting the selection of health improvement measures by means of a recommender system NIVEL Research Institute, Utrecht.
  48. Kotevska O, Vlahu-Gjorgievska E, Trajkovic V, Koceski S (2012) Towards a patient-centered collaborative health care system model. *IJCTE* 4: 1025-1029.
  49. Song I, Dillon D, Goh TJ, Sung M (2011) A health social network recommender system. *Agents in Principle, Agents in Practice, Springer* 7047: 361-372.
  50. Shardanand U, Maes P (1995) Social information filtering: algorithms for automating "word of mouth". Proceedings of the SIGCHI conference on Human factors in computing systems: ACM Press/Addison-Wesley Publishing Co, USA.

51. Cormack GV (2007) Email spam filtering: A systematic review. *Foundations and Trends in Information Retrieval* 1: 335-455.
52. Kong JS, Rezaei BA, Sarshar N, Roychowdhury VP, Boykin PO (2006) Collaborative spam filtering using e-mail networks. *Computer* 39: 67-73.
53. [www.spamassassin.apache.org](http://www.spamassassin.apache.org).
54. Mehta B, Nangia S, Gupta M, Nejd W (2008) Detecting image spam using visual features and near duplicate detection. *Proceedings of the 17th international conference on World Wide Web*, New York, NY, USA 497-506.
55. Belkin NJ, Croft WB (1992) Information filtering and information retrieval: two sides of the same coin? *Communications of the ACM New York, NY, USA* 35: 29-38.
56. Klinkenberg R, Renz I (1998) Adaptive information filtering: Learning drifting concepts Dortmund, Germany.
57. Palakal M, Mukhopadhyay S, Mostafa J, Raje R, N'Cho M, et al. (2002) An intelligent biological information management system. *Bioinformatics* 18: 1283-1288.
58. An update alerting service for pubmed and GenBank. Pubcrawler.
59. Mane K, Sidharth Thakur, Oncosifter: A Customized Approach to Cancer Information.
60. Baujard O, Baujard V, Aurel S, Boyer C, Appel RD (1998) A multi-agent softbot to retrieve multilingual medical information on the Internet. *Stud Health Technol Inform* 23: 187-191.
61. Yoneya T, Mamitsuka H (2007) Pure: A pubmed article recommendation system based on content-based filtering. *Genome Inform* 18: 267-276.
62. Garcia-Molina TWYH (1995) Sift- A tool for wide-area information dissemination. *USENIX New Orleans, Louisiana, USA* 16-20.
63. Huberman BA, Kaminsky M (1996) Beehive: A system for cooperative filtering and sharing of information.
64. Goldberg D, Nichols D, Oki BM, Terry D (1992) Using collaborative filtering to weave an information tapestry. *Communications of the ACM* 35: 61-70.
65. Resnick P, Iacovou N, Suchak M, Bergstrom P, Riedl J (1994) GroupLens: an open architecture for collaborative filtering of netnews. *Proceedings of the ACM conference on Computer supported cooperative work*, Chapel Hill, NC, USA 175-186.
66. Group Lens Research. About GroupLens.
67. Binkley J, Young (1995) Rama: An architecture for internet information filtering. *Journal of Intelligent Information Systems* 5: 81-99.
68. Pazzani MJ, Muramatsu J, Billsus D (1996) Syskill and Webert: Identifying interesting web sites. *Proceedings of the 13th national conference on artificial intelligence*.
69. Mock KJ (1996) Intelligent information filtering via hybrid techniques: Hill climbing, case-based reasoning, index patterns and genetic algorithms.
70. Nakashima T, Nakamura R (1997) Information filtering for the newspaper. *Communications, Computers and Signal Processing, IEEE Pacific Rim Conference Victoria, BC* 1: 142-145.
71. Maes P (1994) Agents that reduce work and information overload. *Communications of the ACM New York, NY, USA* 37: 30-40.
72. Shardanand U (1994) Social information filtering for music recommendation. *Massachusetts Institute of Technology, USA*.