Incorporating Contextual Cues into Electronic Repositories

Jie-Mein Goh
National University of Singapore
gohjm@comp.nus.edu.sg

Danny Chiang-Choon Poo
National University of Singapore
dpoo@comp.nus.edu.sg

Klarissa Ting-Ting Chang
Carnegie Mellon University
changtt@cmu.edu

Abstract

Large document archives such as electronic knowledge repositories are important sources of information whereby knowledge can be derived. Despite efficient information retrieval technologies, users are still not satisfied with the search process and the results presented. Capturing the context of the search can help to enhance retrieval and alleviate the problem of information overload. Contextual information of what and where the user task is, what the user knows, and what the system capabilities are, can greatly enhance an information system's ability to retrieve information from electronic repositories thereby facilitating users to discover knowledge. In this paper, we present a framework that helps to incorporate contextual cues in information systems. Our experiment suggests that there is potential in adopting such a framework in future information systems to ameliorate the retrieval process.

Keywords: contextual cues, electronic repositories

1. Introduction

Studies have shown that the perceived output quality is an essential factor for successful implementation of knowledge management technologies (Kankanhalli et al. 2001). Based on information retrieval studies, the perceived output quality can be measured by the relevancy of the documents returned, it follows that highly relevant and accurate results will encourage users to utilize electronic knowledge repositories. However, despite the very efficient ranking techniques available, users are still not satisfied with the retrieval process.

A very important issue that has arisen in information retrieval literature is the problem of information overload. Information overload refers to the situation whereby users are so overwhelmed by the amount of information that they have to digest such that they are unable to process the information effectively (Wurman 1989). Research has suggested that users feel bored or frustrated when they receive too much information (Roussinov 1999) which can lead to the state where the individual is no longer able to effectively process the amount of information he is exposed, giving rise to a lower decision quality in a given time set.

This problem is exacerbated due to the ever increasing amount of information available in electronic repositories in organizations and the World Wide Web (Farhoomand and Drury 2002). Therefore, it is important to understand how we can achieve relevant and accurate results to achieve greater success in knowledge management initiatives. Denning (1982) stated that we should shift our attention from generating information to controlling and filtering information. With the increasing access to various information sources, presenting the "right information to the right person at the right time and place" becomes a critical challenge.

Two main methods have been identified in past literature are used to mitigate this problem by capturing contextual cues through user profiling and information retrieval. Improving the retrieval process requires an understanding of individual needs and goals in a given context. However, a thorough analysis of the conditions under which context can be incorporated into information systems have yet to be clearly established.

In this paper, we will use a two-pronged approach to develop a model that can help to establish and acquire contextual cues for the system by applying the lens of personalization and information filtering. We present a review of current personalization strategies being used in information systems and information retrieval studies to examine how to bridge the gap between perceived utility of electronic repositories and the quality of information retrieved by utilizing both profiling and information filtering strategies. By extracting information personalized to the user's context, we can ensure better retrieval and use of the information. Drawing on existing studies in user profiling and information filtering, we propose a model that incorporates both user profiling and information filtering concepts. We will also present strategies for capturing contextual cues identified by applying the framework that can be employed in other information systems. Experiments were conducted on an electronic repository implemented with a search facility that incorporated contextual cues derived from our proposed framework and the results are presented to corroborate the usefulness of this work.

2. Related Literature

Two main streams of literature was identified and applied in our work, i.e. user profiling and information filtering. We review techniques that have been employed and give an analysis of each technique and describe our proposed framework.

2.1 User Profiling

User profiling is the ability to represent and reason about the interests or preferences of a user. Several approaches in information retrieval have been developed to produce better search results or to guide users towards more relevant results. These systems request users to provide explicit feedback on documents in terms of ratings or preferences. Employing user's feedback to improve systems had shown to be effective. However, in the real world, it is difficult to ensure that all users will voluntarily offer their feedback to such systems. The approach taken here will not only focus on the filtering techniques, but the means to get a user's feedback implicitly.

Previously, there has been several literature discussing various approaches to alleviate the problem of information overload namely through user profiling and information filtering techniques. Reviewing previous studies on user profiling have led us to identify two main types of profiling approaches namely, static profiling and dynamic profiling:

Static profiling is the process of analyzing a user's static and predictable characteristics and might change only occasionally. It includes demographic information such as name, gender, designation, date of birth and place of residence and long-term interests, that can be captured once and change very rarely. Such information usually comes from users themselves e.g. electronic registration or survey forms. Pazzani, Muramatsu and Billsus (1996), Asnicar and Tasso (1997) created intelligent agents that will analyze user feedback based on ratings defined by the user on the visited page as a measure of user interest. They performed an extended navigation of related pages and graphically show the set of the pages found, classified according to the user's interest.

Through static profiling we usually know what kind of information the user is generally interested as soon as the user have supplied the information. There are several problems when we rely solely on static profiling. Firstly, the profile is static, and is only valid for a certain period of time until the user changes their interest. Hence, a static profile degrades in quality over time. In addition, the input is based on the individual's interest, prone to users' subjectivity and may not accurately reflect an objective view that can infer the interests of other users with similar interests.

Dynamic profiling on the other hand is the process of analyzing a user's activities or actions to determine what the user is interested in over a period of time. In this aspect of profiling, the user's behavior is of interest to us and it is sometimes referred to as behavioral profiling. Dynamic profile contains information that change more frequently than the information of the static profile. Examples of dynamic profiling include the access logs, search queries, history, bookmarks and tracking web browsing characteristics. Although the users' information needs are captured at real time, the general interests of the user cannot be traced.

2.2 Information Filtering

Similarly, there has been a lot of literature that discussed on information filtering techniques. The goal of information filtering is to remove irrelevant data and present only the adequate and relevant information to the user that will satisfy his or her information requirements (Belkin and Croft, 1992). Two kinds of approaches for information filtering have been presented in previous literature:

Content based filtering compares the contents of items associated with a user profile and selects those documents whose contents best match the contents of another user profile using some similarity measures. In Avery's work (1997), a system that receives explicit user feedback through ratings of relevant pages uses filtering strategies to suggest pages of interest to users was developed. INVAID (Kelly and Dunnion 1999), is another system developed to receive explicit user feedback through ratings of relevant pages and suggests pages of interest to users based on the feedback of the user coupled with filtering strategies. Stewart and Davies (1997) created intelligent agents that analyze the user feedback based on well defined ratings of visited pages as a measure of user interest. All the above systems request users to provide explicit feedback on documents in terms of ratings or preferences. The content of the profile dominates in this approach and depends on how well the profiles match that of other users.

The main limitation of this approach is that some users are reluctant to furnish details about themselves or offer their viewpoints. In the real world, it is difficult to ensure that all users will voluntarily offer their feedback to such systems due to the cost of examining and rating an item (Ramscar et al. 1997). Unless the user perceives that there is additional value in participating in such evaluation, the system with all the best filtering strategies may still result in the lack of any ratings at all (Hirashima et al. 1998). Thus, implicit rating is needed such that it removes the cost of examination of an item from any evaluator. In addition, the computational cost of such implicit ratings must be at best hidden away from the user. Content based filtering systems can uniquely identify and capture users' characteristics but still have several limitations that collaborative filtering systems have some main advantages over them (Herlocker et al. 1999).

Collaborative filtering organizes users with similar interest into peer groups, thus enabling the recommendation of documents considered interesting by peers to other members of that group. Several studies have attempted to cluster profiles of their users. Examples are BIRCH (Zhang et al. 1996) and DBSCAN (Ester et al. 1998). We feel that collaborative filtering can help to capture contextual cues of a group of like minded individuals which will be beneficial to users of electronic repositories because new knowledge can promote creativity and stimulate innovativeness.

As this approach relies heavily on user clusters, its effectiveness highly depends on how well the clustering of profiles correlates the users. Collaborative filtering systems have the ability to generate recommendations in domains where there is no or little content associated with the items. Hence, collaborative filtering systems have been successful in certain domains (Goldberg et al. 1992). Resnick, Iacovou, Sushak, Bergstrom and Reidl (1994) has also implemented the idea of collaborative filtering in GroupLens.

3. The Framework for Capturing Contextual Cues

Clearly, each of the approaches described in the previous section has its own shortfalls. Recognizing the deficiencies of each approach, there have been attempts to show the value of combining concepts of both content and collaborative techniques in personalization systems (Balabanovic and Shoham 1997; Good et al. 1999; Melville et al. 2001). However, as we have pointed out, building a good personalized system requires both the part of the user and the system which can be bridged closer by capturing the context from the user and content of the repository.

Thus, to provide a complete analysis of a personalized system, we propose a framework that unifies the concepts from user profiling and information filtering. There are two aspects captured in our proposed framework: First, we consider the disparate sources of contextual cues that can be obtained from the domain. This can be accomplished with the use of four new concepts that have been defined. Second, we consider the means to incorporate contextual cues which have been identified as feasible and important to the domain. This refers to the extent of user involvement which can be *explicit* or *implicit*. If the user manually supplies the information and requires an active involvement, this refers to an explicit means to incorporate contextual cues. On the other hand, if the system automatically captures some information of the user without the user's knowledge, we deem this as an implicit means to capture context.

In our framework, we attempt to retain the superior qualities of user profiling and information filtering by proposing an integrative approach. We will adopt the concepts of user profiling and information filtering by providing explicit and implicit ratings, as well as both content and collaborative filtering to capture contextual cues in knowledge management systems. We have defined four new concepts (Figure 1) that will help us identify, categorize and generate new sources of contextual cues, they are:

- Static Content Sources
- Dynamic Content Sources
- Static Collaborative Sources
- Dynamic Collaborative Sources

CONTEXT		CONTENT	COLLABORATIVE
	DYNAMIC	***Dynamic Content Sources*** refers to the contextual cues derived from the dynamic changes in the behavior of the users.	***Dynamic Collaborative Sources*** refers to the contextual cues derived from organizing users with similar actions and behavior into peer groups and filtering information pertaining to group's interest.
	STATIC	***Static Content Sources*** refers to the contextual cues derived from the information that changes rarely such as the demographic information of the user.	***Static Collaborative Sources*** refers to contextual cues derived from the information that changes rarely after organizing users with similar profiles into peer groups.

INFORMATION FILTERING

Figure 1: Framework to identify contextual cues

Each individual category should be considered when designing and implementing information systems such as electronic knowledge repositories. Deriving contextual cues from each category should be carefully considered since strategies used in each category are unique and will contribute contextual cues from different sources. Search facilities or search engines, which filter search results generated from users' queries should apply these concepts in their system design in order to derive information gathered from these strategies to personalize search results. In the next section, we will describe the strategies to implement each of these concepts.

3.1 Static Content Sources

Static Content Sources refer to contextual cues derived from the gathering of static information regarding the user that changes rarely. This information is usually captured upon a registration process either online or offline.

User's Demographic Information & Interests. Typically, systems allow users to enter a simple profile when they first register with the system. For instance, in many e-commerce websites, users enter information via a registration interface that allows the system to capture and store their personal attributes such as gender, occupation and interests. The static content sources are captured explicitly as users are required to register with the system manually by entering vital information that will help the system to learn more about the user. It is static as the registration is usually done once. In organizations, an implicit form of static content source would be the information derived from repositories where large amounts of employee data are already stored.

However, the limitation of static content sources is that feedback may not be received accurately and consistently. For instance, users may leave optional fields in the registration form blank. Static content sources captures information that encompasses what the user already knows. In addition, the static content sources are not able to capture the user's task applicable at a particular point in time. Therefore, further concepts will be described in the following sections, which are required to address these issues.

3.2 Dynamic Content Sources

The system gathers contextual cues based on the dynamic changes in the behavior of the user, leaving only those that represent the user's profile. This means that the system captures the longitudinal temporal dimension by keeping track of the user's behavior during his interactions with the system over a period of time. There are three main ways to capture this type of contextual cues:

User's Actions. Browsing patterns and click streams provides a source of information about users. Such activities are analyzed to determine topics and concepts of interest through off-line data mining. For instance, some systems collect data that detect the user's behavior such as the documents the user has saved and printed. Other actions that can be captured and studied are documents that are read or ignored, saved or deleted, and items that have been replied or not replied (Stevens 1992). Other usage data such as whether a user evaluates or recommends an item, deletes an item, cites or refers to item, marks item as interesting are more actions that can allow system to record contextual cues (Nichols et al. 1997). This information can then be used to recommend items that would be of interest to the user. The user can then explicitly indicate if he is interested. System can also implement relevance feedback techniques to refine the content of future articles. This offers an implicit way of gathering feedback about a user's preference without having the user to offer explicit views about his or her interest.

User's History. Terveen et al. (2002) found that there was effective use in keeping track of users' history especially in reuse tasks. Users made use of their history effectively which allowed them to shorten the time taken for certain tasks. Owing to the limited information processing capability of humans, it is helpful to keep track of his history as he may not be able to remember past actions or events. Some users may not remember the whole process of how they did a search and how they arrived at the results they wanted but rely on keywords to help them recollect their search routines. For example, a user in his last search may have used additional keywords recommended by the system to refine the search. Contextual cues derived from historical data are helpful as it helps users recollect keywords that they have used previously. An example is Bharat's (2000) work who has implemented a system called to keep track of the "search context" by following the different search session and collecting useful queries and promising results. Amalthaea (Moukas 1997), is another system that makes use of dynamic content personalization by examining hotlinks and browsing history of the users.

User's Preferences. User preferences can be captured explicitly through the use of ratings. The use of ratings is common in our daily life especially on the web. Forms of ratings range from free text form to ratings on a discrete scale. Autodesk (Baclace 1992), lets a user selects a discrete rating value for each document read. The evaluator or the user has to examine the item and assign some value on the rating scale. The central limitation of rating is the cost of rating which requires effort and time. As Oard and Marchionini (1996) points out, expert annotations contribute to the economic value and thus the marketplace will assign them a price. In line with this idea, when implementing this approach, system designers should consider various measures to increase the positive value of rating by the users where appropriate. For instance, providing greater recognition to the experts or awarding monetary rewards can help to encourage participation.

It is often very difficult to gather feedback without having the user to explicitly indicate his preferences. This is especially so if a user is searching over the Internet. However, as user's preference is a very important source of contextual cues, implicit feedback can come in the form of items that have been read. Another source of user's preferences is the length of time that they spend on certain articles. Liang and Lai (2002) have shown empirically using browsing content and time to determine user preferences can be effective in discovering user interests. In their work, the results show that the longer time the user spends on an article, the greater the preference of the user in the subject. By analyzing the documents that have been read during the search, we can determine what a user preferred as well as determine what a user the user will not find appealing. Based on this information, we suggest additional words to augment the query so that we can filter away irrelevant information and return only relevant information.

3.3 Static collaborative sources

Community profiles are used to provide a shared community level of feedback that can be used by members of that community. An example of such a system is Footprints (Wixelbalt and Maes 1997). Visitors can see common paths through a website as an aid to navigation at that site. Terveen et al. (2002) observed that users wanted novel recommendations and closely related to what they were interested thus supporting the proposition that keeping track of users' history must be combined with collaborative filtering such that users receive support in finding like-minded users.

Static collaborative sources refers to the contextual cues obtained by clustering users with similar profiles based on information that changes rarely, i.e., static content sources, either automatically or via a user's explicit request. Every time a new user is added into the system, the system will take a period of time to collect information about the user and to construct the user's profile with information that will aid the system in serving the user's needs.

Clustering Static Sources. In this technique users are grouped according to the static content sources such as information captured during registration. This can be performed by the system through some supervised machine learning or clustering algorithm. The system will make use of the learned algorithm to recommend groups that the user may be interested to join.

For instance, by learning the information provided by different users during registration, it allows individual user's with similar behavior to share information with one another via their recommendations and preferences on items of interest. The user's explicit feedback allows user to have control over the items that he or she would be informed. Otherwise, the system can also automatically adjust the relevance rating of a user as an implicit form of static collaborative source. For example, the categories or terms listed in the user's profile captured via registration are matched across other users' profiles. If the term or keyword in the user's profile is found in another user's profile, the similarity measure for these two users is increased accordingly.

3.4 Dynamic Collaborative Sources

Dynamic collaborative sources refers to the contextual cues obtained by clustering users with similar behavior into peer groups based on the user's profile and filtering information pertinent to group's interest. This will be an important source of contextual cues as Terveen's et al. (2002) observation supports the use of dynamic collaborative personalization in information systems.

Clustering Dynamic Sources. This technique is similar to that in "Clustering Static Sources", but the difference is that the system will cluster users based on dynamic sources i.e. via the users' behavior or actions. For instance, the system can automatically cluster a user's click stream data, recommend items of interest to the user and allow the user to indicate his interest. This serves as an explicit means to introduce dynamic collaborative contextual cues in the system. Otherwise, the system can also implicitly introduce dynamic collaborative sources by automatically adjusting the relevance of results presented to the user when the user issues a search query. The relevance scores are derived from other like-minded users' actions and behaviors.

4. Putting the Framework in Action

We applied the framework on an electronic repository of a library system and incorporated the contextual cues that were identified. Figures 2 and 3 show the results after application of the framework.

Firstly, we identified the sources of contextual cues. Figure 2 shows the sources of contextual cues identified. Following this step, we considered the means to incorporate these sources. Figure 3 shows the strategies derived from the application of framework. We will not go into the details of the implementation as this is not the focus of this paper.

CONTEXT		**CONTENT**	**COLLABORATIVE**
	DYNAMIC	***Dynamic Content Sources*** • User's loans • User's reservations • User's search history • User's preferences	***Dynamic Collaborative Sources*** • Clustering loans and reservations information
	STATIC	***Static Content Sources*** • User's demographic information • User's role • User's interests	***Static Collaborative Sources*** • Clustering demographic information

INFORMATION FILTERING

Figure 2: Sources of contextual cues

CONTEXT	*Dynamic Content Sources*		*Dynamic Collaborative Sources*	
	Explicit	***Implicit***	***Explicit***	***Implicit***
DYNAMIC	• User indicates if item is relevant via loans and reservations • User select suggested terms based on user’s preferences • User views search history	• System tracks loans and reservations and adds new terms to profile derived from selected items • System filters off irrelevant terms based on user’s selected items • System tracks search queries issued	• User selects search query suggested by system based on relevant items of other like-minded users	• System clusters users’ profiles based on loans and reservations • System assigns higher relevance score to items found in same cluster derived from like-minded individuals’ loans and reservations
	Static Content Sources		***Static Collaborative Sources***	
	Explicit	***Implicit***	***Explicit***	***Implicit***
STATIC	• User registers online		• User selects cluster based on work interests	• System clusters users based on work interests
	CONTENT		**COLLABORATIVE**	

INFORMATION FILTERING

Figure 3: Strategies to derive contextual cues

5. Performance Evaluation

Preliminary results of the application of the framework were assessed by inviting twenty participants to assess the pages returned by an electronic repository with a search engine with and without applying the framework. In the experiment, the twenty participants were divided into four different user groups where each group has its common objective. Each user group consists of five individuals where each individual assumed a role. Each user was asked to use the search facility based on their roles assigned to them during the experiment. The users first searched for relevant documents using the original search engine and to note down the query terms issued. Following this, the users are asked to issue the same query terms by using the search engine incorporated with contextual cues using the framework. As all the search queries for each session were recorded, it was ensured that there is no order effect as the search queries were repeated for every user.

Two measures, relative precision and total number of records, were employed to assess the performance of our approach. As the total number of relevant records in the system is not known, we use relative precision, P, as a means to define our metric.

Relative precision is defined as

$$P = (N_r/N)*100$$

where N_r is the number of relevant records retrieved and N is the total number of records that are retrieved.

The experimental results showed an improvement for a majority of users in relative precision and an average reduction of total relevant records by incorporating contextual cues. In general, the results suggest that there is potential in applying the four concepts that we have identified to alleviate the problem of information overload as the majority of the users have benefited from the system, by an improved percentage of relevant records or a reduction in total records, or both. The details of the results are shown in Tables 1 and 2. The better search query terms due to contextual cues that have been captured through the framework could have led to an increase in relative precision yet a drop in the number of records that the user needs to sieve through. However, there is an exceptional case that failed to improve the relative precision and caused an increase in the total records. This was because the user entered very specific queries that were different from other users belonging to the same group. Because of this disparity, the system recommended terms that did not belong to users of the same group. This led to a drop in relative precision and increased the total number of records.

User	Total Records (without contextual cues)	Total Records (with contextual cues)
1	263	209
2	306	95
3	420	77
4	432	215
5	514	340
6	184	108
7	165	88
8	299	262
9	419	118
10	506	186
11	236	243
12	248	254
13	167	194
14	113	9
15	561	258
16	523	540
17	35	56
18	516	575
19	34	34
20	249	119

Table 1: Detailed results showing total number of records retrieved

User	Relative Precision (without contextual cues)	Relative Precision (with contextual cues)
1	28.5	33.9
2	4.9	12.6
3	4.0	9.1
4	31.5	60.9
5	28.9	35.9
6	33.2	50.0
7	10.9	15.9
8	32.1	36.6
9	19.3	43.2
10	16.8	32.3
11	16.5	26.3
12	45.2	48.0
13	65.3	92.8
14	1.8	55.6
15	39.4	60.9
16	54.1	37.8
17	14.3	44.6
18	15.9	41.2
19	38.2	55.9
20	17.3	65.5

Table 2: Detailed Results Showing Relative Precision

6. Conclusion

The main contribution of this paper is that it presents a framework for deriving and implementing strategies in any information system to maintain or even enhance the perceived utility of the system. This paper identified the shortfalls of user profiling and information filtering research and proposed a novel recommendation to incorporate contextual cues through our proposed framework. While this paper elaborated on four different sources of contextual cues, static content sources, dynamic content sources, static collaborative sources, dynamic collaborative sources, we are not advocating the use and implementation of all the different types of concepts that have been mentioned in the paper in all information systems.

The purpose of this framework allows system designers to identify the different sources of contextual cues and the means to capture such cues methodically when implementing retrieval systems for electronic repositories. Our experiment suggests that employing such a framework in electronic repositories to be useful. However, the success of the technology depends on the domain of the system and most importantly the users. This suggests that further research could explore which quadrant of the framework could be important in different situations and even extend the current framework.

Electronic repository is an important component in knowledge management systems. This paper provides support for ongoing development of electronic repositories that are contextually aware and useful to users. Designers of knowledge management systems might find it beneficial to incorporate this framework. The results of our research suggest that incorporating contextual cues ameliorate the problem of information overload. It is hoped that through such a framework, not only can we improve the quality of the retrieved content, the incorporation of contextual cues such as suggesting novel relevant items via dynamic collaborative sources could also help to enhance the perceived utility of the information, promote knowledge sharing and even create new knowledge.

7. References

Asnicar, F. A., and Tasso, C. "ifWeb: a Prototype of user model-based intelligent agent for document filtering and navigation in the World Wide Web," *Proceedings of the Workshop on Adaptive Systems and User Modelling on the World Wide Web, 6th International Conference on User Modeling*, Chia Laguna, Sardinia, 1997, pp. 3-12.

Avery, C., and Zeckhauser, R. "Recommender systems for evaluating computer messages," *Communications of the ACM*, (40:3), 1997, pp. 88-89.

Balabanovic, M., and Shoham, Y. "Content-Based, Collaborative Recommendation," *Communications of the ACM*, (40:3), 1997, pp. 66-72.

Baclace, P.E. "Competitive Agents for Information Filtering," *Communications of the ACM*, (35:12), 1992, pp. 50.

Belkin, N.J., and Croft, B.W. "Information Filtering and Information Retrieval: Two Sides of the Same Coin?," *Communications of the ACM*, (35:12), 1992, pp. 29-36.

Bradford, C., and Marshall, I. "Analysing Users WWW Search Behaviour," *Proceedings of the IEE colloquium*, 1999, pp. 149.

Brezillon, P. "Context in problem solving: A survey," *The Knowledge Engineering Review*, (14:1), 1999, pp. 1-34.

Denning, P. "Electronic Junk," *Communications of the ACM*, (23:3), 1982, pp. 163-165.

Ester, M., Kriegel, H., Sander, J., Wimmer, M., and Xu, X. "Incremental Clustering for Mining in a data warehousing environment," *Proceedings of Very Large Databases Conference*, 1998, pp. 323-333.

Farhoomand, A.F., and Drury, D.H. "Managerial Information Overload," *Communications of the ACM*, (45:10), 2002, pp. 127-131.

Goldberg, D., Nichols, D., Oki, B., Terry, D., "Using Collaborative Filtering to Weave and Information Tapestry," *Communications of the Association of Computing Machinery*, (35:12), 1992, pp. 61-70.

Good, Nathaniel, Schafer, J.B., Konstan, K.A., Borchers, A., Sarwar, B., Herlocker, J. and Riedl, J. "Combining Collaborative Filtering with Personal Agents for Better Recommendations," *Proceedings of the 16th national conference on Artificial Intelligence and 11th Innovation Applications of AI Conference on Artificial Intelligence and Innovative Applications of Artificial Intelligence*, 1999, pp. 439-446.

Herlocker, J., Konstan, J., Borches, A., and Riedl, J. "An Algorithmic Framework For Performing Collaborative Filtering," *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR99)*, 1999, pp. 230-237.

Hirashima, T., Matsuda, N. and Toyoda, J. "Context Sensitive Filtering for browsing in hypertext," *Proceedings of Intelligent User Interfaces (IUI98)*, 1998, pp. 119-126.

Kankanhalli, A., Tan, B.C.Y., and Wei, K.K. "Seeking Knowledge in Electronic Knowledge Repositories: An Exploratory Study," *Proceedings of the Twenty-Second International Conference on Information Systems*, December 2001, New Orleans.

Kelly, L., and Dunnion, L. "INVAID: an intelligent navigational aid for the World Wide Web," *IEEE Two-day Seminar. Searching for Information: Artificial Intelligence and Information*, 1999, pp. 14-17.

Konstan, J.A., Miller, B.N., Maltz, D. "Grouplens: Applying Collaborative Filtering to Usenet News," *Communications of the ACM*, (40:3), 1997, pp. 77-87.

Liang, T.-P., Lai, H.-J. "Discovering User Interests from Web Browsing Behavior: An Application to Internet News Services," *Proceedings of 35th Hawaii International Conference on System Sciences (HICSS-35)*, Big Island, HI, USA, 2002.

Lierberman, H. "Letizia: An Agent That Assists Web Browsing," *Proceedings of International Joint Conferences on Artificial Intelligence (IJCAI95)*, 1995, pp. 924-949.

Melville, P., Mooney, R.J., Nagarajan, R. "Content-Boosted Collaborative Filtering," *Proceedings of the SIGIR-2001 Workshop on Recommender Systems*, New Orleans, LA, 2001.

Moukas, A. "Amalthaea: Information Discovery and Filtering using a Multiagent Evolving Ecosystem," *Applied Artificial Intelligence: An International Journal*, (11:5), 1997, pp. 437-457.

Newell, S.C. "User Models and Filtering Agents for Improved Internet Information Retrieval," *User Modelling and User-Adapted Interaction* 7, 1997, pp. 223-237.

Nichols, D.M., Twidale, M.B. and Paice, C.D. "*Recommendation and Usage in the Digital Library*," Technical Report CSEG/2/97, Computing Department, Lancaster University, 1997.

Oard, D. (1995) University of Maryland Department of Electrical and Computer Engineering. "*Information Filtering Defined*," (available online at: http://www.enee.umd.edu/medlab/filter/filter_definition.html)

Oard, D.W. and Marchionini, G. "*A Conceptual Framework for Text Filtering*," Technical Report CAR-TR-830, Human Computer Interaction Laboratory, University of Maryland at College Park, 1996.

Pazzani, M., Muramatsu, J., Billsus, D. "Syskill and Webert: Identifying interesting web sites," *13th National Conference on Artificial Intelligence*, Portland, Oregon, 1996, pp. 54-61.

Resnick, P., Iacovou, N., Sushak, M., Bergstrom, P., Riedl, J. "An Open Architecture for Collaborative Filtering of Netnews," *In Proceedings of the 1994 Computer Supported Cooperative Work Conference, New York, ACM*, 1994, pp. 175-186.

Ramscar, H., Pain, H. and Lee, J. "Do we know what the user knows, and does it matter? The epistemics of user modeling," *Proceedings of the 6th International Conference on User Modeling UM97*, 1997, pp. 249-243.

Roussinov, D.G. and Chen, H.-C. "Document clustering for electronic meetings: an experimental comparison of two techniques," *Decision Support Systems*, (27:1-2), 1999, pp. 67-79.

Shardanand, U. and Maes, P. "Social Information Filtering: Algorithms for Automating "Word of Mouth"," *Proceedings of CHI'95 Conference on Human Factors in Computing Systems*, 1995, pp. 210-217.

Stevens, C. "*Knowledge-Based Assistance for Accessing Large, Poorly Structured Information Spaces*," PhD Thesis, Department of Computer Science, University of Colorado, 1992.

Stewart, S. and Davies, J., "User profiling Techniques: A Critical Review." *Proceedings of the 19th Annual BCS-IRSG Colloquium on IR Research*, 1997, pp. 8-9.

Terveen, L., McMackin, J., Amento, B., Hill, W. "Specifying Preferences Based On User History," *Proceedings of CHI 2002 Conference on Human Factors in Computing Systems, Minneapolis, Minnesota*, 2002.

Wixelbalt, A., and Maes, P. (1997) "*Footprints: visualizing histories for the Web*," (available online at: http://www.media.mit.edu/people/wex/Footprints/footprints1.html)

Wurman, Richard Saul. *Information Anxiety*. New York: Doubleday. 1989.

Wu, Y.H., Chen, Y.C., and Chen, Arbee L.P., "Index Structures of User Profiles for Efficient Web Page Filtering Services," *Proceedings of IEEE Conference on Distributed Computing Systems*, 2000, pp. 644-651.

Zhang, T., Ramakrishnan, R., and Linvy, M. "BIRCH: An Efficient Data Clustering Method for Very Large Databases," *Proceedings of ACM SIGMOD Conference*, 1996, pp. 103-114.