PERGAMON

Information Processing and Management 38 (2002) 671–694

INFORMATION PROCESSING & MANAGEMENT

www.elsevier.com/locate/infoproman

An experiment in building profiles in information filtering: the role of context of user relevance feedback

Luz M. Quiroga [a,*], Javed Mostafa [b]

[a] *Information and Computer Sciences/Library and Information Science, University of Hawaii at Manoa, 1680 East-West Road, POST 314B, Honolulu, HI 96822, USA*

[b] *Information Science and Informatics, Indiana University, 1320 E 10th St. Bloomington, Indiana, IN 47405-3907, USA*

Abstract

An experiment was conducted to see how relevance feedback could be used to build and adjust profiles to improve the performance of filtering systems. Data was collected during the system interaction of 18 graduate students with SIFTER (Smart Information Filtering Technology for Electronic Resources), a filtering system that ranks incoming information based on users' profiles. The data set came from a collection of 6000 records concerning consumer health. In the first phase of the study, three different modes of profile acquisition were compared. The explicit mode allowed users to directly specify the profile; the implicit mode utilized relevance feedback to create and refine the profile; and the combined mode allowed users to initialize the profile and to continuously refine it using relevance feedback. Filtering performance, measured in terms of Normalized Precision, showed that the three approaches were significantly different ($\alpha = 0.05$ and $p = 0.012$). The explicit mode of profile acquisition consistently produced superior results. Exclusive reliance on relevance feedback in the implicit mode resulted in inferior performance. The low performance obtained by the implicit acquisition mode motivated the second phase of the study, which aimed to clarify the role of context in relevance feedback judgments. An inductive content analysis of thinking aloud protocols showed dimensions that were highly situational, establishing the importance context plays in feedback relevance assessments. Results suggest the need for better representation of documents, profiles, and relevance feedback mechanisms that incorporate dimensions identified in this research.

Keywords: Information filtering; Relevance feedback judgments; User models; User profiles

* Corresponding author.

E-mail addresses: lquiroga@hawaii.edu (L.M. Quiroga), jm@indiana.edu (J. Mostafa).

0306-4573/02/$ - see front matter
PII: S0306-4573(01)00058-9

1. Introduction

Literature frequently refers to the problem of information overload and to difficulties in finding accurate and pertinent information. The amount of information we create and exchange is far more than a person can understand and manage and the task of retrieving a piece of relevant information takes a considerable amount of time and effort (Morita & Shinoda, 1994).

Proposed solutions to information overload emphasize the need for specialization in information retrieval services. For example, personalization of information retrieval can help people to find information with potential value for their needs, "at least 99% of the available data is of no interest to at least 99% of the users" (Bowman, Danzig, Manber, & Schwartz, 1994, p. 106). Personalization of information delivery relies on systems that selectively weed out the irrelevant information based on user preferences (Foltz & Dumais, 1992). Profiles, i.e., representations of users' information preferences, are applied as filters to streams of documents. Information filtering systems deliver personalized information or they have the "exclusive objective of identifying the relevance of documents according to interest profiles" (Mostafa & Lam, 2000, p. 415).

Most of the work on information filtering (IF) is being conducted within the framework of user modeling research that considers profiles as user models. Profiles fit these definitions of a user model: sets of beliefs about the users (Kay & Kummerfeld, 1994), or attempts to model salient aspects of the user (Self, 1990). Shapira and colleagues directly defines IF in terms of user models and profiles: "All information filtering models and systems are based on modeling the user and presenting his information needs in the form of a profile" (p. 273). For Belkin and Croft (1992) filtering is "the process of determining which profiles have a high probability of being satisfied by a particular object from the incoming stream" (p. 34).

Profiles are the basis for the performance of IF systems. Many researchers share the belief that "the construction of accurate profiles is a key task – the system's success will depend to a large extent on the ability of the learned profiles to represent the user's actual interest" (Balabanovic & Shoham, 1997, p. 68).

Fidel and Crandall (1997) performed one of the few known evaluations of the effectiveness of a filtering system, based on user perceptions. They concluded that, "building a 'good' profile is still the central obstacle to achieving reasonable performances" (p. 204). Their recommendation was "in addition to developing filtering algorithms and agents, research in this area should focus on methods to create and improve filtering profiles" (p. 204). Additional research could lead to developing a general typology of attributes that could be included in filtering profiles. Oard (1997) mentioned that very little is known about the effectiveness of filtering systems, because most studies focus on the efficiency of algorithms for filtering techniques instead.

User relevance feedback is used to create and refine profiles in IF and also for query reformulation in information retrieval (IR). Relevance is conceptualized as the "user decision to accept or reject information retrieved from an information system" (Schamber, 1994, p. 3). Relevance feedback is a cyclic process whereby the user feeds back into the system decisions on the relevance of retrieved documents and the system then uses these evaluations to automatically modify the retrieval process (Korfhage, 1997).

Our study follows Fidel and Crandall's line of research and includes a comparison of profile acquisition modes. In some of these modes, the profile is adapted by means of relevance feedback assessments provided by the user. Our intention in this experiment was to see how relevance

feedback could be used to build and adjust profiles in a way that improves the performance of filtering systems. Eighteen graduate students were studied while using SIFTER (Smart Information Filtering Technology for Electronic Resources), a filtering system developed by the Indiana University Filtering Research Group that ranks incoming information based on profiles (http://sifter.indiana.edu). For this study, profiles were based on topical classes in the consumer health domain.

In the *first phase* of the study, three different modes of profile acquisition were compared. The explicit mode allowed users to directly specify the profile; the implicit mode utilized relevance feedback; and the combined mode allowed users to initialize the profile and to continuously refine it using relevance feedback. Filtering performance, measured in terms of Normalized Precision, showed that the three approaches were significantly different ($\alpha = 0.05$ and $p = 0.012$). The explicit mode of profile acquisition consistently produced superior results. Introducing feedback for the acquisition of user profile might have a benefit in the long term; however, acquiring the profile based exclusively on feedback resulted in the lowest effectiveness. The low performance obtained by the implicit acquisition mode motivated the *second phase* of the study, which aimed to clarify the role of context in relevance feedback judgments. An inductive content analysis of thinking aloud protocols, showed dimensions that were highly situational, establishing the importance context plays in feedback relevance assessments.

Numerous factors may account for the lower than expected effectiveness of feedback. This study identified some characteristics of the user's background, information needs, and the document collection that influenced relevance feedback judgments. These results have important implications for filtering design. They suggest attributes that could be included in the representation of documents and profiles as well as in the feedback mechanism to improve filtering performance. These results may contribute to the general typology of profile attributes that Fidel and Crandall (1997) stated is needed.

This paper is organized into these sections: theory and background, the problem, research questions and methods; data and results; and, finally, implications and conclusions.

2. Background

In this section research about the concept of relevance feedback is reviewed; SIFTER, the filtering system used in our study is described, focusing on the modules that deal with profile acquisition based on relevance feedback.

2.1. The role of context in the process of retrieval, filtering and organization of information

Since the mid-1960s researchers in IR have experimented with the use of relevance feedback to reformulate a query incorporating terms from documents that the user has found relevant and deemphasizing terms from non-relevant documents, a simple process that "can prove unusually effective results" (Salton & Buckley, 1990, p. 288). On the other hand, a growing literature has suggested that relevance assessments involve more than an agreement between terms in documents and queries. Relevance assessments depend on user context, and they are subjective, dynamic, cognitive, situational, and multidimensional (Schamber, 1994). Cognitive, interactive

feedback models are being proposed that extend the original model incorporating cognitive and situational user states (Spink & Losse, 1996). Following, selected studies are reviewed, showing how context plays a role in the way people organize information and judge the outcomes of IR and IF systems.

2.1.1. Context in information retrieval: user relevance feedback

The limitations of relevance feedback have been widely recognized by the IR research community. In a call for papers for a workshop on relevance feedback (RF) in IR hosted by the Department of Computer Science of the University of Glasgow, some of the problems were highlighted. "As a rule, RF is a successful practical solution to the uncertainty inherent in information seeking. However, the performance of individual techniques can vary over queries, collections and users. RF has also been criticized for not being accessible to users: the basic operation is simple (marking relevant documents) but how users should make relevance decisions to get the best performance from a RF system is not always obvious" (Dunlop, 1999).

Among the topics of interest for researchers in IR are RF and the user (e.g., human relevance), types of feedback (e.g., positive and negative), and RF and user profiling. One goal has been to build models of interactive systems that incorporate different types of relevance feedback.

Studies on human relevance have found that users evaluate information based on criteria beyond topical appropriateness of documents. Schamber (1994) mentioned 80 relevance factors suggested in the literature through the 1970s. Factors include characteristics of users (e.g., education, experience), requests (e.g., specificity, subject), the documents (e.g., aboutness, style), information system (e.g., access, cost), assessment condition (e.g., time for judging, order of presentation), and choice of scale (e.g., ease of use, response required). Other categories related to the user situation (e.g., emotions, beliefs and preferences, knowledge, need).

Barry (1994) conducted open-ended interviews with 18 faculty and students, who were asked to discuss the relevance of their search results, and all printed textual information. She identified 23 relevance criterion categories and grouped them into seven broad categories. Four pertained to the document: content, sources of documents, the document as a physical entity, and other information or sources within the environment. The other three categories pertained to the user: situation, beliefs/preferences, and experience/background.

Thirty respondents from Schamber (1991) were users of weather information, asked to evaluate the source and presentation formats. Using structured time-line interviewing and questionnaires, she identified 22 detail categories that were grouped in 10 broad categories of relevance, all of them document-related: accuracy, currency, specificity, geographic proximity, reliability, accessibility, verifiability, clarity, dynamism, and presentation quality.

In a more recent joint study Barry and Schamber (1998) compared and contrasted their own previous individual studies. They agreed on 10 broad categories common to both studies, most of them document related: depth/scope/specificity, accuracy/validity, clarity, currency, tangibility, quality of sources, accessibility, availability of information/sources of information, verification and affectiveness. They concluded that the high degree of overlap and similarities among criteria mentioned by such diverse sets of users and situations "provide evidence for the existence of a finite range of criteria that are applied across type of information problem situations and information sources" (Barry & Schamber, 1998).

Several researchers have proposed models of IR that emphasize human–computer interaction (HCI) and have the user context as an important component. Saracevic (1996) discussed the kind of questions that needed to be addressed, types of interaction and variables involved, and their effect on performance. He considered that answers to these questions should be sought through observation of the human side of participants, their intervention, and purposes of interaction.

Saracevic (1996) proposed a "stratified interaction model" that he proposed includes levels, or strata, for information resources, computer resources, and the interface; as for the user, there are strata for query characteristics, user characteristics (knowledge, intent, task, belief), situation, and environment. A simplified view of the model depicts three main levels: surface, cognitive, and situational. At the cognitive level, users interact with text representation and interpret and make cognitive relevance judgments; in the situational level, they judge their utility on the solution of their information need. In a general conclusion Saracevic (1996) pointed out that interaction had been ignored in IR research, which had focused more on text representation. "IR interaction is a complex process that is very much situation or context dependent: it starts from and relates to the user, their tasks or problems, competencies, knowledge states and intents..." (p. 4). The process of user modeling, that is, of capturing user dimensions, needs to be better understood to support interaction and adequate interface design.

Wang (1997) identified user relevance criteria on document selection and use that include topicality, orientation/level (e.g., theoretical, empirical), discipline, recency, novelty, quality, availability, authority (reputation), relation/origin (relationship author–user: e.g., colleague, advisor), personal knowledge, and epistemic, functional or emotional value. She compared her own studies with those by Barry (1994), Cool, Belkin, and Kantor (1993), Park (1992) and Schamber (1991) and found that the "non-topical criteria identified by these studies, although named and categorized slightly different by individual researchers, are compatible across the studies" (Wang, 1997, p. 163). The overlap among these studies and observations of data elements used as cues to criteria were then translated into a "cognitive model of document selection" and nine principles for the design of Document Retrieval Systems that support user modeling. Wang's model includes components for (a) processing information based on document attributes, (b) applying criteria, (e.g., topicality, orientation, quality), (c) assessing values (e.g., functional, social, emotional), and (d) making decisions (rejection, acceptance, uncertain). The model depicts relationships between components, showing how document attributes offer cues for criteria (e.g., author affiliation might be cue to discipline and quality), and how each criteria is linked to assessment values (e.g., how authority and relation/origin associate to social values) that affect acceptance. The nine principles range from references to document elements that should be made available, ways to group and present retrieved documents, how searches may be automatically expanded to include, for example, recognized authors, and similar or cited documents. One of the nine principles refers to the feasibility of building a knowledge base of the user that can be shared. This notion is also of interest in current research on building individual or collaborative profiles in information filtering.

In 1997, Spink prepared an overview of feedback concepts within the framework and models of cybernetics, the social sciences, HCI, and IR. She sought to enlarge the information seeking and retrieving (ISR) feedback concept beyond relevance feedback to what she calls interactive feedback. An interactive feedback loop is the "human interactive loop process of generating meanings, and presents a dynamic and responsive view with greater applicability and explanatory power for information seeking and retrieving" (Spink, 1997, p. 737). Concerns in interactive

relevance feedback are nature, elements, types of feedback, "interestingness", magnitude and frequency of feedback, positive versus negative strategies, content versus term relevance feedback, and situational and cognitive relevance.

Much research has been done. Still, there is no consensus as to the factors that contribute to human relevance assessments. What researchers agree on is the situational, cognitive, and dynamic nature of relevance (e.g., Park, 1992; Schamber, 1994; Barreau, 1995; Saracevic, 1996). However, still problematic is how these human dimensions can be described in enough detail to allow computer modeling.

2.1.2. Context in information filtering: modeling users' interest

A user profile generally contains keywords and topics of interest to the user; however, some researchers look for the inclusion of elements related to the user as a person. This "extended" profile might be used in the selection of documents, or at least could be used to organize the retrieved set in a way most compatible with user preferences (Korfhage, 1997). Fidel and Crandall (1997) performed an empirical study on filtering performance dentifying attributes for filtering profiles, based on users' perceptions. They examined criteria users employed to determine whether a document was relevant. Profiles were based on topics, but they found that users employed many criteria beyond topics or subjects. Thirteen criteria for relevance and fourteen for non-relevance assessments were identified. Some of the criteria for relevance were directly related to the user's specific working status, or to a product or service of the company, or to previous work. Other criteria were attributes of documents: a case study, hard data, trends, background, or technicality. Relevance criteria depended on the situation of the individual, confirming beliefs, and updating. When judging documents as non-relevant, the following type of reasons were given: too basic, general, detailed or technical; not technical enough; not familiar with content, nothing new, or technology not available at company. Based on these results Fidel and Crandall recommended developing methods to create and maintain useful profiles, incorporating relevance and non-relevance criteria into the profiles to improve filtering.

Other researchers have identified profile elements useful for a particular domain. Ardissono and Goy (1999) wrote that user profiles of potential customers for electronic shops should include domain expertise, lifestyle, and destination use of items. These elements can direct systems to recommend and describe products and regulate how much technical information to include and the linguistic form to employ. Petrelli, De Angeli, and Convertino (1999), in a study of visitors to museums, categorized them according to "classical" dimensions such as age, profession, education, and specific knowledge or background. They also considered situational dimensions such as time available and motivation for the visit. They then incorporated in user profiles settings for language style (expert versus naive) and verbosity (depending on time available).

2.1.3. Context in organizing information behavior

Donovan (1991) stated that individuals organize their information spaces according to their interests, thereby facilitating future reference. Organizing is "the human-guided process of deciding how to interrelate information, usually by placing it into some sort of a hierarchy" (Bowman et al., 1994, p. 105). Theories of categorization have evolved from the classical to the prototype to the contextual view in search of a cognitive approach where building categories

includes not only attributes of the object to be classified but also characteristics of the human being who is performing the classification task (Lakoff, 1987).

Malone (1983) investigated how people organize their personal information environments. He found two types of units of office organization used by his subjects: piles and files. Piles seem more oriented to goals and situations while files (bookshelves and folders in cabinets) reflect more permanent topical categories. Piles were organized according to dimensions such as events, priorities, deadlines, current projects, etc.; they were also used as finding and memory aids (67% of piles were reminders).

Case's research focused on the information needs and uses of scholars (cf. Case, 1991). He compared storage habits of researchers in social sciences and in humanities. His interest was to determine how scholars acquired, sorted, filed, and gave priority to a heavy intake of information. He found filing and indexing methods were influenced by situational factors. A scholar's priority could be current research, teaching, projects and publications, methodologies, and sources and circumstances of interest in a topic.

Kwasnik's doctoral dissertation explored the "influence of context on the process by which people organize and classify their own documents in their own personal information spaces" (Kwasnik, 1989, p. 1). By context she meant the overall situation of a person, from physical environment to goals, history, expectations, predisposition, time and space constraints, understandings, levels of expertise, and "taken for granted" (TFG) knowledge. Her underlying assumption was there is a relationship between cognitive organization and organization of physical things, but there is no one set of organizing principles; they vary not only from person to person but from one situation to another. Observing how faculty members organized documents and sorted incoming e-mail, Kwasnik found context was a determinant in the clustering process, with context taking precedence over topical attributes of the document. Therefore, representation of TFG knowledge, common sense, and conditions is critical for the design of systems emulating human behavior.

From these studies it may be concluded that context influences human information behavior, a notion also supported by findings in the second phase of our study.

2.2. Multilevel filtering and the SIFTER system

The SIFTER system, designed by Indiana University Filtering Research Group, was the filtering system used in this study. The filter's main function is to establish relevance for each incoming document according to the interest profile. An ideal filter would take into account all the different words, their relationships, and ultimately the semantic essence represented in each document before determining relevance. However, in a dynamic and heterogeneous environment, directly determining relevance based on complete content of documents is a highly computationally demanding task.

To cope with the computational demands, SIFTER performs filtering in multiple levels. Specifically, filtering is conducted in two steps: (1) incoming documents are classified into a relatively small and stable set of topics for which the user's interest is known, and (2) the documents are subsequently sorted according to the interest values established for individual topics. The profile in SIFTER, therefore, constitutes a set of interest values whereby each value represents the user's interest in a specific topic. After documents are classified into the internally maintained set of

topics, the documents can be pruned, sorted or organized (a specific filtering mode chosen by the user) based on the interest value associated with the topics.

SIFTER decomposes the overall filtering operation into three subtasks: representation, classification, and profile acquisition. In reality, the domain may contain numerous topics and the topics may change over time, and so to deal with topic diversity and change, the representation subtask incorporates the usage of thesauri and is supplemented with automated term discovery. The default algorithm for document representation is based on the vector space model. Each document is transformed to a vector of concepts with associated weights. SIFTER supports various means for classification, ranging from unsupervised clustering to supervised neural-network based approaches. The profile acquisition process and the associated components critical to this research are briefly described below. Other publications provide more details on the representation, classification and profile acquisition techniques and algorithms implemented by the SIFTER research group (Mostafa, Quiroga, & Palakal, 1998; Mostafa & Lam, 2000).

2.2.1. Profile acquisition

SIFTER supports three major modes of profile acquisition: (1) explicit, (2) implicit and (3) a combination of explicit and implicit. In SIFTER, the explicit profile is an array or vector of real values: r_i $(i = 1, \ldots, k)$. Each real value in r represents the user's interest in a specific topic, and k is the total number of topics covered by SIFTER. If the explicit mode is selected, in the first session when SIFTER is executed the user is given the option to directly select each value in r.

In the first session SIFTER displays a form window in which the user can identify their interest in each topic (a scale ranging from VI = very interested to NI = not interested). The selected option for each topic is translated to a numerical relevance value (0–10) and is stored in the profile.

In each session SIFTER presents the user a fixed number of newly accumulated documents and ranks them according to the relevance values in the profile. In subsequent sessions in the explicit mode, users can view their explicit profile but they cannot modify it. In contrast, the implicit mode captures the profile indirectly, based on the user's rating of documents. In this mode, after the user views a document, he/she is given the option to rate the document using a slide scale from 0 to10 (0 = not interested, 10 = highly interested). The rating information is treated as relevance feedback for the topic corresponding to the document. SIFTER employs a reinforcement learning algorithm to incrementally capture and update the profile based on the relevance feedback data.

Finally, in the combined explicit and implicit mode, the user is given the option of directly setting the profile in the first session. In subsequent sessions, the user is given the option of providing relevance feedback. The relevance feedback data is used to continuously update the profile.

2.2.2. Relevance judgment

SIFTER was designed with built-in instrumentation and transaction logging capability to conduct evaluation. For every session executed, SIFTER logs the individual documents presented, their topic classification, the corresponding relevance feedback (if any), and the ranking. In addition to this data, SIFTER can measure the quality of ranking based on each user's own relevance judgment, and store this information as part of the session log. To establish the quality of ranking, at the end of each session SIFTER presents the document list again and requests that the user identify the documents he/she considers as relevant.

The document numbers and their user rankings are ultimately transformed to normalized precision and recall scores. In this paper, we employed only normalized precision (NP) for evaluation. NP is frequently applied to systems that rank retrieved documents. The performance of the system, the ranking quality score, is measured by its ability to show the most relevant documents at the top, and can be measured using the formula below

$$\text{Normalized Precision (NP)} = 1 - \frac{\sum_{i=1}^{\text{REL}} \log \text{Rank}_i - \sum_{i=1}^{\text{REL}} \log i}{\log(N!/(N-\text{REL})!\text{REL}!)},$$

where N is the total number of documents displayed, REL is the total number of documents the user found relevant, and Rank_i is the ranking of the relevant document i in the final output (Salton & McGill, 1983). This formula normalized results according to the number of relevant hits and number of items to be ranked.

3. Research problems and methodology

Experiments were conducted in order to investigate several research questions related to profile acquisition for IF systems (Quiroga, 1999; Quiroga & Mostafa, 2000). This paper describes two of the phases of the study.

3.1. Research problems

3.1.1. First phase: profile acquisition

In the first phase, different profile acquisition modes were analyzed to see how much the automated processes could be improved by increasing human involvement. Data were collected based on interaction of 18 users with the filtering system, SIFTER. Data included system-generated transaction logs, written statements, as well as verbalization provided by the subjects during thinking aloud processes. Three modes of profile acquisition were compared: explicit, implicit, and combined explicit and implicit. The explicit mode allowed users to directly specify the profile; the implicit mode utilized relevance feedback; the combined mode allowed users to initialize the profile using explicit means and to continuously refine it using relevance feedback. The posed research question was: What is the influence of different modes of profile acquisition on filtering performance?

In our experimental design, the independent variable was the three modes of profile acquisition and the dependent variable was effectiveness of filtering measured with the NP metric.

3.1.2. Second phase: the role of context in user relevance feedback

The second phase included an analysis of factors that may influence relevance feedback judgments. It was conducted in part to further clarify the results of the first phase experiments. An inductive content analysis of the user thinking aloud protocol, collected during interaction with the filtering system, was conducted. Data included in this analysis came from the 12 users who provided feedback, those assigned to either the implicit or combined mode. The posed research question was: What are the dimensions that influence user's feedback judgments and what is the role of context in these judgments?

3.2. Methodology

Subjects. Eighteen graduate students, eight of them women, were recruited from three departments at Indiana University (six per department): School of Health, Physical Education and Recreation, School of Library and Information Science, and Cognitive Science Program. In a short demographic survey, 13 of them graded their computer and Internet skills as "intermediate", four considered themselves as "experts," and only one graded himself as "novice".

Document data set and profiles. The data set for this study came from a collection of 6000 records (title, author, abstract, and keywords) concerning consumer health retrieved from the Ebsco Health Sources Plus database, whose target audience is the general public. Records were classified by the IF system, SIFTER, according to 15 classes pre-established by the database producers: anxiety, allergy, heart, cholesterol, depression, diet, environment, exercise, eyes, headache, lungs, medicine, teeth, men's health, and women's health. User's profiles were based on these classes.

Laboratory setting. The experiments were run in a usability lab with facilities for running interactive software, recording user verbal protocols, and videotaping the computer screen.

Procedures and activities. Before running the experiments users completed a short demographic survey and wrote statements about their information needs. The subjects were randomly assigned to one of the three experimental modes. In the explicit mode, users could specify their preference for classes only in the first session. In the implicit mode, the preference for classes was inferred from the feedback provided to documents in any of the 15 sessions. In the combined mode, the preference for classes provided in the first session was continuously modified based on feedback in subsequent sessions.

Data were collected during user interaction with SIFTER and included system-generated transaction logs, written statements, and users' verbalizations. Users were asked to think aloud to express motivations for providing feedback to a document. Each user performed 15 filtering sessions. A session with SIFTER consisted of: execution of the system, presentation of a list of documents ranked according to the user's profile, review of documents by the user and feedback to documents (in the modes that allowed feedback). Through feedback SIFTER tuned up profiles.

When a user ended a session, SIFTER presented the ranked list of documents and requested the user to identify relevant documents. The list of relevant documents was used to calculate effectiveness of filtering, measured in terms of NP. After the experiments users were asked to express their satisfaction with the profile built by the machine and to provide a revised profile if they thought one was needed.

4. Data analysis, results and discussion

4.1. First phase

Analysis of the different profile acquisition modes showed the three approaches to be significantly different ($\alpha = 0.05$ and $p = 0.012$). The explicit means of profile acquisition consistently produced superior results. Exclusive reliance on relevance feedback in the implicit mode resulted in inferior performance (Fig. 1).

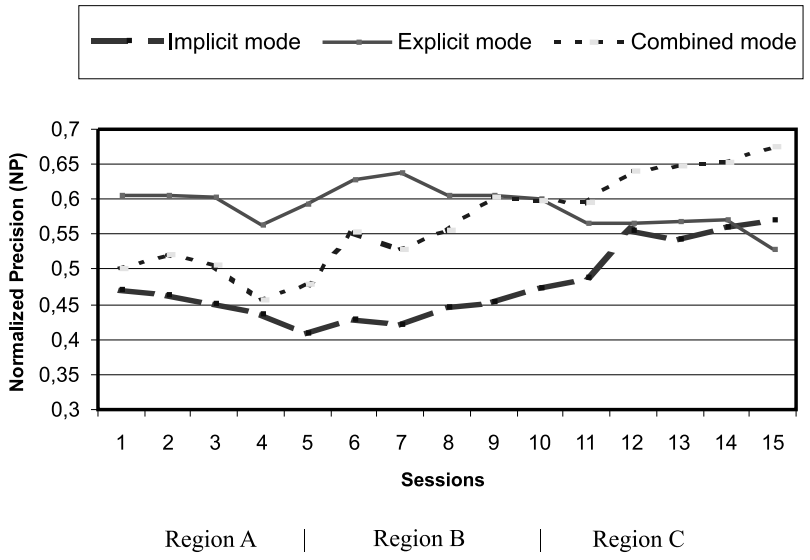

Fig. 1. Filtering performance (NP) values throughout the 15 sessions, for the three experimental profile acquisition modes.

A more detailed analysis was performed to test whether, within the same experimental mode, there was a significant change in the mean NP across regions – Region A (sessions 1–5), Region B (sessions 6–10), Region C (sessions 11–15). There were several motivations for the across-regions study. First, users behavior might change as they become familiar with the system in general and with the effect of their feedback in the profile tuning which would have a positive effect in the performance of the system. On the other hand, users could change preferences with exposure to document contents, which would have a negative effect on the performance. From the point of view of the system, the learning algorithm, based on feedback, will have more data to better tune the profile as the sessions proceed.

For each of the experimental modes, an ANOVA ($\alpha = 0.05$) was performed across the regions. Only the combined mode showed a significant difference in NP across regions (p-value $= 0.0299$). This result suggests that introducing feedback for the acquisition of user profile might have a benefit in the long term.

We expected that as users increased their involvement in the acquisition of their profiles, the system performance would improve. In that way, the combined mode, on average, would obtain the highest NP and the explicit mode would obtain the lowest. In the long term, we expected the implicit mode to converge with the combined mode. Because the user was not required to provide initial preferences in the implicit mode, we expected the reduction in cognitive load to be most advantageous to profile acquisition. However, results of the first phase showed that in general, across experimental sessions the performance of the implicit mode was significantly lower than the other two modes. The performances of the explicit and combined modes were not significantly

different from each other, but were better than the implicit mode. Findings suggested an advantage in providing initial explicit preferences. Feedback might have a benefit in the long term as the gradual improvement in NP, across regions, for the combined mode suggests. Surprisingly, acquiring the profile based only on feedback resulted in the lowest filtering performance, which lead us to the second phase of the study.

4.2. Second phase

Understanding the users' rationales in providing feedback on a document was the goal of the second phase of our study. Results of the content analysis show characteristics of the documents but, more important, characteristics of the users that influenced feedback. These elements were not part of the representation of documents and user profile, which partially explains the low performance of the implicit mode of profile acquisition.

4.2.1. Characteristics of the user that influence feedback

Five categories emerged from the analysis: demographic data (age, gender, population/community, marital status), domain expertise (profession/occupation, education, projects), lifestyle (hobbies, habits, health disabilities), health status (illness, intolerance, propensities), and health concerns of friends and relatives. The categories in Table 1 chart user statements, indicating number of subjects mentioning each. Frequently cited attributes were gender and age, almost all subjects had preferences for materials related to their gender and age. Population/community, occupation/profession, field of study, and projects were important factors in accepting a document. Lifestyles, habits, and environment influenced feedback. Some labeled themselves as vegetarian, smoker, or sport or outdoor person when explaining why they considered a document relevant or not. The individuals' health statuses also influenced their information needs. If subjects had a diagnosed illness, they preferred documents on treatment rather than prevention but those "prone to" an illness preferred information on prevention or symptoms.

Some individuals wanted only to be alerted to illness and detection methods. Immediacy of illness was reason for giving preference to related documents. Chronic illness and scheduled surgery caused one user to postpone reading about that illness. Concern about the health problems of friends and relatives was interestingly a frequent category.

4.2.2. Characteristics of documents that influence feedback

Two main groups of document attributes were found to influence relevance judgments: topical factor and non-topical factors. Three influencing topical factors were identified (Table 2). Orientation or facet (e.g., surgical procedures, genetic aspects, or drug side effects), specificity (e.g., kind of allergy, exercise, or anxiety) and the combination of topics and facets (e.g., exercise and heart attack; homeopathy and headache).

Non-topical factors (Table 3) included credibility of the information source, comprehensibility and approach (e.g., research, technical paper), novelty, format (e.g., reference book, handbook, descriptive, statistics), and target audience (technical, medical, patients, general public). A variety of factors, ranging from trust in the source and knowing or ignoring the information to preferring certain types of documents and considering the availability of them, attracted user attention. The

Table 1
Characteristics of the users that influence feedback assessments. Number of subjects and examples

		Frequency	Examples
Demographic	Age	9	"looks interesting, that would be me, because I feel like I am middle aged and I need to work on that"; "I am too young for that"; "no, not in the treatment of obese adolescent"; "headache in children, headache in older people, I don't really fit either one of those"; "this may be interesting to me later on in life"
	Gender	8	"Alcohol, well maybe not on men, if it was alcohol on women"; "no, it turned out to be about women"; "this one about men's health, pass, the computer seems to think I am a male"
	Marital, family status	5	"married men are healthier. That would definitely be interesting, I am about to get married"; "this looks interesting, I don't have children yet but, I would like someday"
	Population	7	"Hispanic women: neither Hispanic or a woman"; "I am not interested in depressed senior citizens"; "if I were African American it would pertain to me even more"; "for the Italian community, not interested"; "about groups of people I am not part of"
Domain expertise	Education, occupation, projects	6	"yeah, this is my area, nutrition, important, very important"; "very similar to what I just had done, doing a research project on salaries in professional sports..."; "actually some of my research deals with working children and anxiety... that would give me some baseline on the anxiety in kids"; "interesting for me being a psychology graduate student"; "protection for a human research subject, being a doctoral student this could be of interest to me"
Lifestyle	Habits, hobbies	8	"I am definitely interested in this, I am an exercise science person"; "I am an environmental person"; "no, this sounds way too new-agey for me; I am not a runner but I do play basketball"; "it might be related to my smoking"; "it's for a vegetarian"; "preoperative anxiety music therapy, no, I want them to give you drugs"; "it sounds kind of interesting, music therapy, I'd really like to get more on environmental and women's health"; "recipes are always good, I can't cook but I like to read them";
	Work, home environment, conditions	7	"living in Hawaii, that's sort of interesting"; "it's about tuberculosis, I am not worried"; "tropical, infectious diseases, doesn't look at all like anything I need"; "working in cold weather, here's Indiana"

(*continued on next page*)

Table 1 (*continued*)

		Frequency	Examples
Own health status	Status	7	"interesting to me because I have bleached my teeth"; "I use contact lenses, obviously it pertains to me"; "here is one on women, but is on breast feeding which doesn't really relate to me very much"; "a lot of things dealing with pregnancy, somewhat interesting"; "on women after menopause, so doesn't do anything for me either"; "I do weigh 270 pounds"; "alcohol, stress, depression, I have no problems with any of those things"; "hearing aids, not interested"; "I am going to read it because I take that medicine"
	Illness, intolerance, (diagnosis)	10	"I don't need to worry about that, according to my doctor my blood pressure is wonderful"; "I have an injury, I'm curious, let's see"; "I do have an allergy to ANSAID so let's see what it says"; "I have liver problem myself"; "cataracts, not interested"; "until I am diagnosed I don't feel like reading this"
	Being prone to (prevention)	6	"it is not strictly about prevention but about healing, when you are already ill"; "Since I am looking for ways to prevent it, I would read this"
	Alert, detect, symptoms	5	"doesn't say if it is a factor, symptom..."; "ah something that is not a good test for detecting depression, sure that's interesting"
	Immediateness, urgency, priority	4	"I had to rate teeth because I just went to the dentist today"; "my dental problems are here and now while the heart is in the future"
Relatives' and friends' concerns	Demographic	2	"As I age and my parents age, this would be very interesting to me"
	Lifestyle	2	"my boyfriend doesn't know about nutrition, so I think this would be interesting"; "pet separation anxiety, my friend has a dog that has that problem" "my mother as well as one of my older sisters is a smoker, I'll give an 8"
	Health status	8	"this could be interesting to me and my brothers because my dad had heart defects"; "my dad has been diagnosed with cancer so I am interested in a lot of cancer stuff"; "anything having to do with diabetes interests me because my father had diabetes"; "because my family has a history of lung cancer, it pertains to me personally" "because my sister is going to get breast implant... it doesn't pertain to me but I can point out to her about the potential of breast cancer"

Table 2
Characteristics of the documents that influence feedback assessments: Topical factors. Number of subjects and examples

		Frequency	Examples
Orientation, facets	e.g. side effects, procedures, aspects	8	"it is more genetic aspects and this is not really what I am looking for"; "ok, I definitely want to know about clinical side effects"; "Oh man, why is it that the only stuff on eyes is stuff about surgery"; "I am more interested in the preventive side of it rather than after the fact"; "not too interested in the drug aspect of it"; "no, it is about ophthalmologist not eye care"; "alternative medicine, that sounds like it could be handy"
Specificity level	e.g. subtopic, kind of allergy, type of anxiety, kind of exercise	10	"it's just not sifting that information because that's sort of a narrow topic"; "ok, deals with teeth but a little too specific"; "too general for me to be interested"; "not that type of allergic reaction"; "nut allergy"; "social anxiety,... it's the kind of anxiety I might be interested in"; "weight lifting type of exercise ok"; "that one would be interested: that type of cancer"
Combining topics	Intersecting topics, facets, population, approach	10	"almost exactly what I am interested in, sports and nutrition"; "not relevant to me cause I want to know about eyes and weight, not eyes and smoke"; "ok, it is about reducing anxiety in athletes"; "homeopathy and headaches, well homeopathy might be interesting but headache really aren't"; "no, it's experimental research that is interesting but the subject wasn't really"; "proper diet and preventing cancer, so that makes it more interesting to me"; "Interesting as I have lung problems from allergies"

approach of the document was the most frequently cited, with users appreciating information that targeted the average person rather than the specialist. The wording of titles was a reason for reading or not reading a document. Some users stated that they consistently read weird or mysterious titles while others avoided them. A document dealing with current affairs motivated some users. General interest documents, something "everybody should know", were of general acceptance regardless of profiles. The last non-topical factor pointed out by users was the "interestingness" of materials. Users made distinctions between relevance, "interestingness", and topicality: "interesting but not relevant" or "right on the topic but not promising". Numerous judgments were based on how interesting the article was although it was not considered relevant, or the opposite, being relevant but not interesting.

The same criteria can result in article selection by one person and rejection by another. Opposite reactions to the same document are illustrated in the following users' statements: "preoperative anxiety music therapy, no, I want them to give you drugs" versus "it sounds kind of interesting, music therapy, I'd really like to get more on environmental and women's health". Users often expressed degrees of relevance. Subjects used words like "partially" relevant, "mildly" relevant, and "some kind" of relevance in their statements. One user, reflecting on his interest in lungs, said: "Talks about bronchial muscle response, about asthma, I don't think this is interesting to me. Although maybe it might have something to do with lungs but only in a roundabout way".

Table 3
Characteristics of the documents that influence feedback assessments: Non-topical factors. Number of subjects and examples

	Frequency	Examples
Credibility	5	"this is more political stuff than science based"; "questionnaires... I am interested in the validity of them"; "a good article to look at, there is a lot of misinformation out there"; "I would read the article just to see the justification, research they did to support the article"; "it also gives observations by a doctor"
Novelty (not known; known)	5	"this is definitely interesting, I did not know that drinking water helps prevent cavities"; "that's very interesting, I've heard about that before"; "a visual problem I did not know about, sure that's always interesting"; "I've never heard of chew sticks so that's kind of interesting"
Format (e.g. descriptive, reference book, statistics, facts, sheets, equipment, press release)	7	"very relevant, descriptive, explain what an allergy is"; "I wish the synopsis here would give a little bit more information, that's probably the weakest of identifying relevance"; "it is only announcing the availability of fact sheets"; "just a press release"
Availability	3	"ok, it's a book on... actually I can run off and get that"; "this would be relevant, we have a WalMart in Spencer too, so I could go right down the street for that"; "handy because it tells you where to go to get the book... and cost"
Approach, comprehensibility (e.g. basic or applied research, technical paper)	10	"no, too clinical"; "Acronyms I don't know anything about so I am not interested in"; "the abstract just has more jargon"; "if I can't pronounce it I probably can't read it, nope"; "not as applied as I thought it would be, it's more biochemical"; "it seems a more philosophical article"; "I need something that will sort out the technical from the non-technical for me"; "yes, it looks like a mathematical model, that's pretty good"; "Scientific methodology, questionnaires, good stuff"; "my purpose of using science health information would be I answer my own questions and this one I'd have to use a dictionary to look up every other word so that's not very helpful"
Target audience (e.g. managers, specialist, researchers, technicians, doctors, patients, layman)	4	"I don't know drugs names at all,... for medical experts"; "I think a lot of these would only be of interest to specialists in those fields"; "too scientific, too complex, no, it's not for me"; "that sounds like a layman's article"; "looks like something more for a dentist than for its patients"; "this is more of interest to chemists, biologists and possible doctors, but not at all interesting to me"; "highly technical, they don't really seem to be aimed at just the average person"
Title appeal	5	"that's too weird, I must have a look at that"; "good, ok, I haven't been clicking on things like mysterious titles, this one I almost passed by anyway"; "sometimes you've to read them just because the title is cool"
Current affairs	2	"cholesterol controversy, all right"; "interesting because it has been in the news"

Table 3 (*continued*)

	Frequency	Examples
Of general interest	5	"Diet and cancer is always important"; "that looks very interesting, it's always important to have enough energy"; "cancer is everywhere, so that's important to me"; "imaging... in general regardless of the topic area"; "news updates are always good, cancer is always scaring"; "vision benefits are always worth looking at"
Interesting but not relevant/ relevant but not interesting	8	"they are amusing, but not relevant"; "well it is not a promising title but it's right on the topic"; "I am going to look at because it's intriguing, what eyes got to do with football? Oh, not relevant to my concerns but it's a real interesting thing"; "it doesn't pertain but I selected it because is kind of funny and interesting"

4.3. Comparison with other studies on relevance

How do these categories compare with those found in previous research on relevance? In order to put the results of this study in the framework of research on relevance, a rough comparison of our findings with previous findings was attempted. We chose two studies for this comparison: one by Barry and Schamber (1998), because it is the most recent, rigorous, and comprehensive study on human relevance in information retrieval, and one by Fidel and Crandall (1997), because it is one of the few conducted in relation to information filtering.

It is important to note the difference in the broadest level of categories used in each study. Of the 10 categories identified by Barry and Schamber, nine related to characteristics of the document (depth/scope/specificity, accuracy/validity, clarity, currency, tangibility, quality of sources, accessibility, availability of information/sources of information, verification) and only one refers explicitly to the user (affectiveness). Fidel and Crandall divided their categories in two broad groups: criteria for judging a report relevant (13) and criteria for judging a report non-relevant (14). Our study grouped criteria in three broad categories: user related (5), topicality of documents (3) and non-topical attributes of a document (10).

Table 4 shows how each one of the criteria found in our study is contrasted with those found in Barry and Schamber (1998) and Fidel and Crandall (1997). In the Fidel and Crandall column, R indicates a judgment of relevant, NR a judgment of non-relevant. In the Barry and Schamber column, categories common to their own individual studies are printed in bold; categories identified in only one of their studies are prefixed with the individual author name.

Overlapping categories: data in Table 4 show a considerable overlap among the three studies. Nine of the 10 categories common to Barry and Schamber are present in our study, "currency" being the exception; all categories mentioned by Fidel and Crandall are present in our study. Six of the 10 Barry and Shamber common categories are present in Fidel and Crandall study, but if their individual studies are considered, a complete overlap is found.

Non-overlapping categories: there are many possible reasons for dissimilarities among these studies. They range from differences in the formation of categories and interpretation of user statements by researchers; to different settings and methodologies of the empirical studies; to different domain, information systems, and data; to users' motivations for their participation. Below we list a few specific possible reasons:

Table 4
Overlap on relevance criterion categories

	Quiroga and Mostafa	Fidel and Crandall	Barry and Schamber
Characteristics of users	Demographic (age, gender, marital, family status, population)		Barry: Effectiveness
	Domain expertise (education, occupation, projects)	R: It was about a product or service that related directly to a project the participant was working on	(Barry: Background/ Experience/Ability to understand; Subjective accuracy/validity)
	Lifestyle (habits, hobbies; work, home environment and conditions)	R, NR: It was (not) relevant to the Boeing company NR: The participant had no influence over the issues the report raised	**Affectiveness**: (Barry: entertainment value) (Shamber: Geographic proximity)
	Own health status (status, illness, intolerance, (diagnosis)), being prone to (prevention, alert, detect, symptoms, immediateness, urgency, priority)	R: It had information that helped the participant keep up to date about a product with which he or she were familiar NR: The participant's group had already made a decision about the product or service that was addressed in the report	(Barry: time constraints)
	Relatives' and friends' concerns (demographic, health status, lifestyle)		
Characteristics of the documents. Topical factors	Orientation facets (e.g. side effects, procedures, aspects)		**Depth/Scope/Specificity**
	Specificity level (e.g. subtopic, kind of allergy, type of anxiety, kind of exercise)	NR: It was about a specific vendor NR: It was too basic or too general NR: It was too detailed or too technical	**Depth/Scope/Specificity**
	Combining topics (intersecting topics, facets, population; approach)		**Depth/Scope/Specificity**
Characteristics of the documents. Non-topical factors	Credibility	R: It confirmed or validated what the participant already knew	**Accuracy/Validity Verification**

Table 4 (*continued*)

	Quiroga and Mostafa	Fidel and Crandall	Barry and Schamber
Characteristics of the documents. Non-topical factors (*continued*)	Novelty (not know; known)	R: It was about new concepts, products or services NR: It was about something Boeing was already doing NR: The participant was not familiar with the product or service NR: It did not tell the participant anything he or she did not already know	(Barry: currency; Content, source, document novelty)
	Format (e.g. descriptive, reference book, statistics, facts, fact-sheets, equipment, press release)	R: It had hard data NR: It raised questions but gave no answers NR: It expressed opinions rather than presenting facts	**Tangibility**
	Availability	NR: It was about a technology that was not here yet	**Availability of information/Sources of information**; **Accessibility**
	Approach, comprehensibility (e.g. basic or applied research, technical paper)	R: It was a case study	(Barry: Background/ Experience/Ability to understand)
	Target audience (e.g. managers, specialist, researchers, technicians, doctors, patients, layman)	R: It was written on a non-technical level NR: It was completely non-technical NR: It took too long to understand what the report was about NR: It was too detailed or too technical	**Clarity** (Barry: Effectiveness; Background/Experience/ Ability to understand)
	Title appeal		**Affectiveness**
	Current affairs		
	Of general interest	R: It included background information or general information	
	Interesting but not relevant/ relevant but not interesting		**Affectiveness**

1. Different category labels for what looks like similar user criteria (e.g., clarity versus target audience; credibility versus accuracy; lifestyle versus affectiveness).
2. Same label for what looks like different categories (e.g., ''geographic proximity'' that could be considered as a mention of availability or topical preference).

3. Differences in the type of system being used might explain why "currency" did not appear in our study. Our subjects were using a filtering system simulating the arrival of new information by e-mail. Users might have assumed that all the information they received was current. Novelty might be a way for them to express the need for current information.
4. Again, differences in the type of system might explain categories not present in some studies: "Relatives' and friends' concerns" is more likely to appear in a filtering system because of the long term information need being modeled versus the short term need that might drive an information retrieval system.
5. Differences in points of view in the analysis: user characteristics versus document characteristics. This could be the reason for similar statements from different studies coded using the category "clarity" of the document or "domain expertise" of the user.
6. Differences in the emphasis given to positive or negative relevance. Fidel and Crandall would code statements referring to the technical level of the document in different categories, depending on the "relevant" or "non-relevant" judgment. In our study both statements go to same category.

A non-overlapping dimension in our research is the role of the information needs of family and relatives in the information preferences of individuals – personal profiles are influenced by profiles of family and friends. This notion of social or collaborative profiles is a lively topic in IF and IR research. Ardissono and Goy (1999) described an electronic shopping recommender system, SETA, that includes a set of "beneficiaries models", a model for each person for whom the shopper is selecting goods. Petrelli et al. (1999), in their research on personalized guides to museums, found that a visit to a museum could be a social or family event. They suggested a feature to support "family profiles" as part of individualized museum guides. Another example is an ongoing project by Raya Fidel and colleagues on Collaborative Information Retrieval (http://www.ischool.washington.edu/cir/).

Despite the differences, it is worth noting commonalties in comparing studies. Although we cannot claim generalizability of our results, this analysis seems to support Schamber's belief in the existence of a short set of categories of relevance applied by end-users regardless of type of situation. Findings are consistent with previous research on the role of context in the process of organization, retrieval, and filtering of information.

5. Conclusions

The first phase of this study investigated the influence of different profile acquisition modes on filtering performance. In some of these modes the profile is acquired using relevance feedback. The second phase attempted to understand the rationale of users providing feedback on a document. Through content analysis of thinking aloud protocols, several cognitive and highly situational dimensions emerged, establishing the importance context plays in relevance feedback assessments. Characteristics of the documents, but more important, characteristics of the user that influenced feedback were identified. Some of the user attributes related to preferences are more permanent and therefore easier to represent, such as the demographic data, while others are highly

situational, such as novelty and interestingness. In the middle range are semi permanent attributes such as work conditions, projects, and illness.

Our findings cannot be generalized because of the small sample and the specific domain. However, our findings are in accord with previous research on the situational nature of criteria that make a document relevant to an individual (Park, 1992; Barry, 1994; Schamber, 1994; Barreau, 1995; Fidel & Crandall, 1997; Wang, 1997; Bateman, 1998; Barry & Schamber, 1998). Those studies found a large set of situational criteria used by individuals when deciding to accept or reject information retrieved from an information system.

Our findings show how context influenced user-relevance judgments and help to explain the low performance of filtering systems when profiles are based solely on topics and fail to consider context and particular characteristics of users. This suggests that to improve filtering performance there is a need for better representation of user profiles, documents, and feedback mechanisms that incorporate dimensions that emerged from this research.

This study responds to calls for the advancement of IF and personalization research through empirical studies to test the effectiveness of filtering systems because in many cases only anecdotal evidence of their performance was provided (Oard, 1997). It also provides insights for developing a general typology of attributes that could be included in filtering profiles, as suggested by Fidel and Crandall (1997). In addition, Shapira, Shoval, and Hanani (1997), referring crucial knowledge in the design of filtering, cite the importance of finding out what the content of the profile in an IF system should be.

5.1. Future research

Our research on profile acquisition continues. To date, our qualitative analysis includes only the first step of analysis – eliciting categories that influence feedback. Next steps are coding and content analysis to learn whether relationships between these categories and certain types of documents, users, or information needs exist. We want to determine, for example, what kinds of documents receive more consistent feedback, and whether these documents and/or users have attributes in common. This more detailed analysis might give insights regarding potentially useful metadata for medical informatics. Several other aspects need more research, some examples of which we discuss below.

This study focused on consumer health information. *Other domains and environments* could be explored to refine the typology of categories that influence relevance feedback. For example, similar research could be conducted in settings that involve real information-seeking users and situations (e.g., users of Internet resources by regular citizens looking for everyday information on topics such as health, employments, entertainment, and education). Also, it would be interesting to analyze the literature on marketing and consumer studies that identify characteristics of consumers that may have value as components of profiles.

More research is needed on the value and feasibility of *representations of profiles*, *documents*, *and feedback mechanisms*, that incorporate topical and non-topical attributes. Dealing with topical attributes could be facilitated with a thesaurus-like structure of classes where different levels of specificity can reside, in which the facets and aspects of information can be distinguished, and where the relationships among topics can be maintained. Some non-topical attributes could

be included in document and profile representations that could be used to match the targeted audience of a document (e.g., gender, expertise, lifestyle and community) without ignoring privacy concerns. As for feedback mechanisms, a better understanding of interaction is needed. Subjects in our study wanted ways to express negative or partial relevance. Most studies focus on positive relevance feedback; they have been concerned with criteria applied to the best sources for a particular task. Negative feedback is not well supported in many IR and recommenders; for example, search engines implement mainly positive feedback by means of options such as ''more like this''. However, other studies suggest that negative feedback could be as important as positive (Fidel & Crandall, 1997; Spink, 1997; Bateman, 1998). Feedback mechanisms should also allow the user to explain not only why a document was relevant or not, but to express levels of relevance. Another issue that needs investigating is how to design a mechanism to collect feedback without overwhelming the user. Studies from HCI and interface design might provide guidelines.

Also needed are analysis of *models on interactive IR* such as the stratified interaction model proposed by Saracevic (1996), the document selection model of Wang (1997) and the interactive feedback by Spink (1997), to examine how these models, proposed in the context of IR, support the design of IF systems, user modeling and the representation of long term information needs. Finally, in our study users were concerned about information needs of relatives and friends. It might be of interest to examine the usefulness of providing sets of profiles. A ''family'' profile and a set of ''beneficiaries'' profiles would allow users to prioritize and distinguish their individual information preferences, but is not known how to implement this without risking information overload, an effect that IF seeks to combat.

References

Ardissono, L., & Goy, A. (1999). Tailoring the interaction with users in electronic shops. In J. Kay (Ed.), *Proceedings of the 7th international conference on user modeling* (pp. 35–44). Wien, NY: Springer. Available: http://www.cs.usask.ca/UM99/papers.shtml.

Balabanovic, M., & Shoham, Y. (1997). Content-based, collaborative recommendation. *Communications of the ACM, 40*(3), 67–72.

Barreau, D. K. (1995). Context as a factor in personal information management systems. *Journal of the American Society for Information Science, 46*(5), 327–339.

Barry, L. (1994). User-defined relevance criteria: an exploratory study. *Journal of the American Society of Information Science, 45*(3), 149–159.

Barry, C. L., & Schamber, L. (1998). Users' criteria for relevance evaluation: a cross-situational comparison. *Information Processing and Management, 34*(2/3), 219–236.

Bateman, J. (1998). Changes in relevance criteria: a longitudinal study. In C. M. Preston (Ed.), *Proceedings of the 60th annual meeting of the American Society for Information Science, Vol. 35* (pp. 23–32). Medford, NJ: Information Today.

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

Bowman, C. M., Danzig, P. B., Manber, U., & Schwartz, F. (1994). Scalable Internet resources discovery: research problems and approaches. *Communications of the ACM, 37*(8), 98–107.

Case, D. O. (1991). The collection and use of information by some American historians: a study of motives and methods. *Library Quarterly, 61*(1), 61–82.

Cool, C., Belkin, N. J., & Kantor, P. B. (1993). Characteristics of texts affecting relevance judgments. In M. E. Williams (Ed.), *Proceedings of the 14th national online meeting* (pp. 77–84). Medford, NJ: Learned Information.

Donovan, J. M. (1991). Patrons expectations about collocation: measuring the difference between the psychologically real and the really real. *Cataloging and Classification Quarterly, 13*(2), 23–43.

Dunlop, M. (1999). One-day workshop in relevance feedback – Glasgow. *IR-L Digest* [on-line], vol. xvi, n. 22, issue 458, June 7, 1999. Department of Computing Science, University of Glasgow, IR Group. Available: http://www.dcs.gla.ac.uk/rf.

Fidel, R., & Crandall, M. (1997). Users' perception of the performance of a filtering system. In N. J. Belkin (Ed.), *Proceedings of the 20th annual international ACM/SIGIR conference on research and development in information retrieval* (pp. 198–205). New York: ACM Press.

Foltz, P. W., & Dumais, S. T. (1992). Personalized information delivery: an analysis of information filtering methods. *Communications of the ACM, 35*(2), 51–60.

Kay, J., & Kummerfeld, B. (1994). *User models for customized hypertext* (pp. 47–69). Department of Computer Science, University of Sydney.

Korfhage, R. R. (1997). *Information storage and retrieval*. New York: Wiley.

Kwasnik, B. H. (1989). *The influence of context on classificatory behavior*. Unpublished Doctoral Dissertation, Rutgers, The State University of New Jersey, New Brunswick.

Lakoff, G. (1987). *Women, fire and dangerous things: what categories reveal about the mind*. Chicago: University of Chicago Press.

Malone, T. W. (1983). How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Office Information Systems, 1*(1), 99–112.

Morita, M., & Shinoda, Y. (1994). Information filtering based on user behavior analysis and best match text retrieval. In *Proceedings of the seventh annual ACM-SIGIR conference on research and development in information retrieval* (pp. 272–281). New York: ACM Press.

Mostafa, J., & Lam, W. (2000). Automatic classification using supervised learning in a medical document filtering application. *Information Processing and Management, 36*(3), 415–444.

Mostafa, J., Quiroga, L. M., & Palakal, M. (1998). Filtering medical documents using automated and human classification methods. *Journal of the American Society of Information Science, 49*(14), 1304–1318.

Oard, D. W. (1997). The state of the art in text filtering. *User Modeling and User-Adapted Interaction, 7*, 141–178.

Park, T. K. (1992). *The nature of relevance: An empirical study*. Unpublished Doctoral Dissertation, Indiana University, Bloomington, IN.

Petrelli, D., De Angeli, A., & Convertino, G. (1999). A user-centered approach to user modeling. In J. Kay (Ed.), *Proceedings of the 7th international conference on user modeling*. Wien, NY: Springer. Available: http://www.cs.usask.ca/UM99/papers.shtml.

Quiroga, L. (1999). *Personalized information organization: Acquisition and modeling of users' interest profiles in information filtering systems*. Unpublished Doctoral Dissertation, Indiana University, Bloomington, IN.

Quiroga, L., & Mostafa, J. (2000). Empirical evaluation of explicit versus implicit acquisition of user profiles in information filtering systems. In D. H. Kraft (Ed.), *Proceedings of the 63rd annual meeting of the American Society for Information Science and Technology, Vol. 37* (pp. 4–13). Medford, NJ: Information Today.

Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval*. New York, NY: McGraw-Hill.

Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. *Journal of the American Society of Information Science, 41*(4), 288–297.

Saracevic, T. (1996). Modeling interaction in information retrieval (IR): a review and proposal. In S. Harding (Ed.), *Proceedings of the 59th annual meeting of the American Society for Information Science, Vol. 33* (pp. 3–9). Medford, NJ: Information Today.

Schamber, L. (1991). *Users' criteria for evaluation in multimedia information seeking and use situations*. Unpublished Doctoral Dissertation, Syracuse University, Syracuse, NY.

Schamber, L. (1994). Relevance and information behavior. *Annual Review of Information Science and Technology, 29*, 3–48.

Self, J. A. (1990). Theoretical foundations of intelligent tutoring systems. *Journal of Artificial Intelligence in Education, 1*(4), 3–14.

Shapira, B., Shoval, P., & Hanani, U. (1997). Stereotypes in information filtering systems. *Information Processing and Management, 33*(3), 273–287.

Spink, A., & Losse, R. M. (1996). Feedback in information retrieval. In M. E. Williams (Ed.), *Annual review of information science and technology* (vol. 31) (pp. 33–78). Medford, NJ: Information Today.

Spink, A. (1997). Information Science: a third feedback framework. *Journal of the American Society for Information Science, 48*(8), 728–740.

Wang, P. (1997). The design of document retrieval systems for academic users: implications of studies on users' relevance criteria. In C. Schwartz, & M. Rorvig (Eds.), *Proceedings of the 60th annual meeting of the American Society for Information Science, Vol. 34* (pp. 162–173). Medford, NJ: Information Today.