A human-centric integrated approach to web information search and sharing

Roman Y Shtykh[*] and Qun Jin

* Correspondence: roman@akane.
waseda.jp; jin@waseda.jp
Networked Information Systems
Laboratory, Faculty of Human
Sciences, Waseda University, Japan

Abstract

In this paper we argue a user has to be in the center of information seeking task, as in any other task where the user is involved. In addition, an essential part of user-centrism is considering a user not only in his/her individual scope, but expanding it to the user's community participation quintessence. Through our research we make an endeavor to develop a holistic approach from how to harnesses relevance feedback from users in order to estimate their interests, construct user profiles reflecting those interests to applying them for information acquisition in online collaborative information seeking context. Here we discuss a human-centric integrated approach for Web information search and sharing incorporating the important user-centric elements, namely a user's individual context and 'social' factor realized with collaborative contributions and co-evaluations, into Web information search.

Keywords: human-centricity, user profile, search and sharing, personalization

1. User in the Center of Information Handling

1.1. Information Overload Problem

With the rapid advances of information technologies, information overload has become a phenomenon many of us have to face, and often suffer, in our daily activities, whether it be work or leisure. We all experience the problem whenever we are in need of some information, though "people who use the Internet often are likely to perceive fewer problems and confront fewer obstacles in terms of information overload" [1]. Any of us has experienced a situation when deciding to buy a certain product, say, a washing machine, and trying to figure out its characteristics, such as availability of delayed execution, steam and aquastop functions, we browsed the Web and encountered an excessive amount of information on the product. Then we had to filter out irrelevant information, categorize and analyze the remaining part to do the best choice. Many of those who work at office acquire, filter, analyze, conflate and use the collected information - the process which requires, today more than ever, special skills and software to cope with highly excessive and not always relevant information for proper decision making.

Despite of the public recognition of the problem and the great number of publications discussing and analyzing it, information overload is often a notion slightly differing in the contexts it is applied to and findings of researchers. The word itself has many synonyms, such as information explosion or information burden, and some

derivatives, such as salesperson's information overload [2], to name a few. So what is 'information overload'?

As in the example with the washing machine purchase, information overload is generally understood as the situation when there is much more information than a person is able to process. This definition is identical to that given by Miller [3] who considered human cognitive capacity to be limited to five to nine "chunks" of information. First of all, it is often mentioned when the growing number of Web pages and difficulties related to this are discussed. Considering the growing popularity of social network systems (SNS) and user-generated content, the Web is likely to remain the primary area of concern about information overload in future. Indeed, the amount of such content grows very fast (for instance, Twitter had about 50 million tweets per day in February 2010 [4]) and becomes even threatening for men - people are at the risk of being buried with tons of information irrelevant to a particular current information need. And since information technologies in general and the Web in particular are highly employed for most human activities today, the problems raises concerns in many other technology-intensive areas of human activities. However, the problem of information overload should not be considered with regard to growing information resources on the Web only - it is much wider and multidisciplinary problem encountered in sales and marketing, healthcare, software development and other areas.

Information overload is a complex problem. It is not just about effective management of excessive information but also, as Levy [5] argues, requiring "the creation of time and place for thinking and reflection". Himma [6] conducted a conceptual analysis of the notion in order to clarify it from a philosophical perspective and showed that although excess is a necessary condition for being overloaded, it is not a sufficient condition. The researcher writes: "To be overloaded is to be in a state that is undesirable from the vantage point of some set of norms; as a conceptual matter, being overloaded is bad. In contrast, to have an excessive amount of [entity] × is merely to have more than needed, desired, or optimal."

Thus, being overloaded implies some result on a person, and this result is of undesirable or negative nature. Generally, conception of information overload today implies such negative effects. For instance, conducting social-scientific analysis (in contrast to Himma [6]'s philosophical approach) Mulder et al. [7] define information overload as "the feeling of stress when the information load goes beyond the processing capacity."

The state of information overload is *individual*, in the sense it depends on personal abilities and experiences. As Chen et al. [8] point in their research on decision-making in Internet shopping, the relationship between information load and subjective state toward decision are moderated by personal proclivities, abilities and past relevant experiences. Also though information load itself does not directly influence an individual's decisions, its excess may negatively influence the decision quality. By conducting a series of non-parametric tests and logistic regression analysis, Kim et al. [9] determined factors which predict an individual's perception of overload among cancer information seekers. The strongest factors appeared to be education level and cognitive aspects of information seeking that proves again the individual nature of the information overload and emphasizes the importance of information literacy.

Information overload is a *multi-faceted* concept and have various implications to human activities, and society in general, many of them becoming known as new

researches are conducted. For instance, Klausegger et al. [10] found that information overload is experienced regardless of the nation, with its degree somewhat differing from nation to nation, - there is a significant negative relationship between the overload and work performance for all five nations the authors investigated. It was also found that the phenomenon negatively influence the degree of interpersonal trust, which is a critical component of social capital [1]. One of its plausible and severely harmful outcomes is information fatigue syndrome which includes "paralysis of analytical capacity," "a hyper-aroused psychological condition," "anxiety and self-doubt," and leads to "foolish decisions and flawed conclusions" [11]. Since the problem has a subjective nature, the first countermeasure is information literacy, efficient work organization and work habits, sufficient time and concentration [7] - again, one's strategy will depend on one's work tasks and subjective factors. Another, and not less important, countermeasure we put the focus in our research is technological. Till now a number of solutions as to how to reduce the negative effects caused by the phenomenon have been proposed. To name a few, in order to assure the quality of information and in this way reduce the problem in folksonomy-based systems, Pereira and da Silva [12] propose cognitive authority to estimate the information quality by qualifying its sources (content authors). To reduce excess of information in wiki-based e-learning, Stickel et al. [13] assume every link in the proposed hypertext system having a predefined life-time and use "consolidation mechanisms as found in the human memory - by letting unused things fade away" in order to remove unused links.

For more substantial information on the overload problem, interested readers are recommended to refer to [6,14]. But to summarize, though simplistically, we reflected the principal and essential components of the phenomenon in Figure 1:

- excessive amount of information;
- subjective and objective information processing capabilities conditioned by experience, proclivities, etc. and environment, situation, etc. respectively;
- individual's psychological and cognitive state.

Clearly, to alleviate the information overload for an individual, we can reduce the amount of information and/or increase our processing capabilities. Considering the fact that people with high organization skills and information literacy have less perceived information overload and usually require better tools to process information,
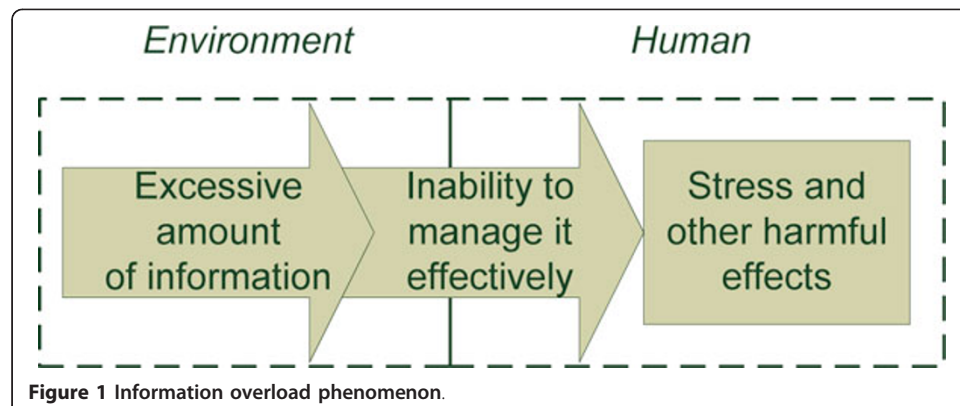

Figure 1 Information overload phenomenon.

and people with constantly perceived information overload requires better training as to how to manage it [15], probably the first step to alleviate the problem is providing information literacy and organization instructions prior to providing the tools. After such measures become ineffective due to the overwhelming amount of information, filtering, summarizing, organizing and other tools have to be applied. Certainly, there is no need for a separation of the approaches and normally they should be used together.

In this study we focus on the technological approach considering each and every individual's interests, preferences and expertise in order to provide selective information retrieval and access, thus expediting the acquisition of desired and relevant information. Section 1.3 will clarify the research questions and objectives, and give a further outline of the approach.

1.2. Growing Role of Human in Information Creation, Assessment and Sharing

In addition to the fact that information overload is a subjective phenomenon and it is a human who is affected by it and has to cope with it, it is easy to see that the phenomenon itself is largely caused by a human and his activities. It started to be particularly tangible with popularization of user-generated content (user-generated media, or user-created content) which, in turn, was enabled by new technologies, such as weblogging (or blogging), wikis, podcasting, photo and video sharing on the Web [16]. User-generated content is publicly available and produced by end-users, such as regular visitors of Web sites.

The motivations for people to share their time and knowledge are, as discussed by Nov [17] for the case of Wikipedia, 1) altruistic contribution for others' good, 2) increasing or sustaining one's social relationships with people considered important for oneself, 3) exercising one's skills, knowledge and abilities, 4) expected benefits in terms of one's career, 5) addressing one's own personal problems, 6) contributing to one's own enhancement (these six categories are closely related to the concept of *self-extension* we have outlined within social networking services [18]), 7) fun and 8) ideological concerns, such as freedom of information.

According to Nielsen//NetRatings [19], in July 2006 "user-generated content sites, platforms for photo sharing, video sharing and blogging, comprised five out of the top 10 fastest growing Web brands." Among them were ImageShack, Flickr, MySpace and Wikipedia - the brands that are also well-known nowadays to any more or less literate Web user. User-generated content sites continue growing by attracting new users of various ages and social groups. Particularly, such growth is strong in online social networks today. For instance, Twitter is reported to have about 270,000 new users per day [20]. Also, eMarketer reports that in 2011 half of Western Europe's online population will use social networks at least once a month, and 64.4% of Internet users in the region will be regular social network users [21].

With the emergence of user-generated content (UGC) concept, an individual's role as a creator and active evaluator of the shared Web information has become central, and perhaps will become critical in future. With increase of human activities on the Web, the percentage of information related to such activities grows; hence, it is becoming more and more *user-centric*. Such centricity becomes a cause of creation of excessive amounts of information, but, on the other hand, also can help people to overcome information overload problem with the wisdom of crowds [22]. People use the power

of user-generated content to make decisions on their daily activities, whether it be work or leisure, and researches are investigation on how to leverage it in order to benefit from it in a great number of work tasks. JupiterResearch [23] has found that 42 percent of online travelers using user-generated content trust the choices of other travelers and such UGC is very influential on their accommodation decisions. Exchange of user-generated content facilitates an enrichment of our life by creating new social ties and promoting interaction within communities, as, for instance, discussed in the study of enhancing a local community with IPTV platform to exchange user-generated audio-visual content conducted by Obrist et al. [24]. However, along with the virtues, such user-centricity of UGC brings new problems of trust, and quality and credibility of volunteered content that are transformed to adjust the UCG context. As an example, trust becomes a metric for identifying useful content and can be defined as "belief that an information producer will create useful information, plus a willingness to commit some time to reading and processing it" [25].

It should be noted that in our research we do not focus particularly on user-generated content, but, as everyone's Web experiences can show, the number of such content is great and its significance cannot be neglected. Although UGC has its specific problems, such as above-mentioned credibility and trust, to be solved, it shows the growing importance of every individual and proves the power of experience of online users taken altogether, which is an important pillar of our research. Generated by human, user-generated content is rapidly growing and influencing many aspects of human life. In other words, it can be named as a mechanism of indirect societal regulation by human, and this regulation is done by not a group of limited number of specialists, but by all interested people willing to participate. So the role of each and every individual in the modern society is growing and becomes more important than ever. Moreover, in the situation of information overload such an engagement is even essential to overcome the problems of excessive information that are, strictly speaking, created by the participants themselves. To reformulate this, nowadays we have to benefit from each other's expertise and this has to be enabled by appropriate technological solutions, which in turn ought to become as human-centric as possible to understand requirements to them in particular work task settings and employ all power of human expertise.

1.3. Research Objectives

The brief discussion of the problem of information overload and the importance of human to alleviate it take us to the research objectives of this research we will consider on two levels - macro and micro. Macro level will give us explanation of the objectives from the perspective of the presented concepts of information overload and user-centeredness of information creation, assessment and sharing on the Web. Micro level will help to outline the research questions and objectives we are working on in a closer perspective and domain of information retrieval (IR).

- *Alleviating Information Overload (macro level)*

In this work we tackle the problem of information overload primarily from technical perspective within which a consideration of situational and subjective nature of the problem is done. In other words, although we propose a technological solution

for the problem, we attempt to consider it as a problem lying also in a subjective dimension. We believe that no solution can be effective enough without considering a person's processing capabilities and information needs which are very individual, as we discussed above, and situational respectively.

- *Better Understanding and Satisfying Human Information Needs (micro level)*

IR is an important research and application area in the era of digital technology. Today information retrieval tools are essential for information acquisition. However, with information overload becoming more tangible every day, such tools reach their limits of providing information pertinent to users' information needs. This is a reason for revival of interest of scientists and enterprises to information filtering and personalization today. In order to perform effectively, an IR system has to understand a user's information needs in a particular situation, context, work task and settings, and only after such knowledge about the user is available (through inference or other methods) the search has to be done. The understanding of situational and contextual nature of seeking and endeavors to harness it for more effective seeking process stimulated the research of the cognitive aspects of IR, known today as cognitive information retrieval (CIR) [26,27]. Inferring the user's interests and determining his/her preferences is one of the useful techniques not only for CIR, but also for personalized IR (PIR). Since the difference between the two may be not clear-cut, we consider PIR as, though often considering the user's search context and situation, not making special focus on cognitive aspects of information seeking.

In our research we propose a collaborative information search and sharing framework called BESS (BEtter Search and Sharing) in attempt to incorporate the discussed user-centeredness into information seeking tasks. We present a holistic approach as to how to harnesses relevance feedback from users in order to estimate their interests, construct user profiles reflecting those interests and apply them for information acquisition in online collaborative information seeking context. The paper explains the notions of *subjective* and *objective* index in IR system, and demonstrates the methods for dynamic multi-layered profile construction changing with change of interests, evaluation of shared information with regard to each user's expertise, and *subjective concept-directed vertical search*.

1.4. Organization of the Paper

First of all, in Section 2 we discuss human-centric solutions for information seeking and exploration with main focus on personalization, its advances in academy and business, and speculate on user profiles as the core component of personalization. Further, we discuss BESS collaborative information search and sharing framework. Section 3 presents its conceptual basis, its model and architecture. Section 4 narrates about our original interest-change-driven modelling of user interests, discusses its role and position within the framework and compare with other profile construction approaches. Section 5 discusses shared information assessment and search in the framework. A demonstration of a search scenario is given to better reveal the concepts and

information seeking strengths of BESS. Finally, Section 6 concludes the paper with the summary of the presented research and outlines future research issues.

2. Enhancing Information Seeking and Exploration. Emphasis on User

Information overload problems have made a human to reconsider information retrieval process and IR tools that seemed to be effective to a certain point. It has become clear that the success of retrieval does not only consist in improving search algorithms, IR models and computational power of IR frameworks - new approaches to make information seeking closer to the end-user are needed. Such approaches include research in user interfaces better adapted to the user's operational environments, systems understanding the user's needs and whose intelligence spreads beyond an algorithmic query-document match seen in conventional "Laboratory Model" of IR discussed in [26]. This resulted, for instance, in the emergence of interactive TREC track and raise of great interest in user-centered and cognitive IR research. IR systems are seeking to incorporate the human factor in order to improve the quality of their results. Information seeking today is getting considered in dynamic context and situation rather than static settings, and a human is its essential and central part actively processing (receiving and interpreting) and even contributing information. Contextual information of the user is obtained from his/her behaviors collected by the system the user interacts with, organized and stored in user profiles or other user modeling structures, and applied to provide personalized information seeking experience.

In this section we introduce endeavors to improving Web IR by means of user interface improvements and support of exploration activities, and focus on personalization as the most wide-spread approach to user-centric IR. We discuss user profile (UP) as the core element of most personalization techniques, show its structural variety and construction methods.

2.1. Improving Web Information Retrieval

It is well known that alongside with search engine performance improvements and functionality enhancements one of the determinant factors of user acceptance of any search service is the interface. To build a true user-centric information seeking system, this factor must not be underestimated. Here we will show its importance considering mobile Web search, as the need for improvements are particularly tangible due to small screen limitations of handheld devices most of us possess today.

Landay and Kaufmann [28] in 1993 noted that "researchers continue to focus on transferring their workstation environments to these machines (portable computers) rather than studying what tasks more typical users wish to perform." In spite of all the advances of mobile devices, probably the same can be said about mobile Web search judging from its state today. Search today is poorly adapted to mobile context - often, it is a simplistic modification of search results from PC-oriented search services. For instance, many commercial mobile Web services, like those of Yahoo!, provide search results that consist of titles, summaries and URLs only. However, although all redundant information like advertisements is removed to facilitate search on handheld devices, users may still experience enormous scrolling due to long summaries. To improve the experience some services, like Google, reduce the size of summary snippets. However, this can hardly lead to the improvements and, quite the contrary, can

thwart the search. As shown in Figure 2, a mobile user searching for "fireplace" cannot know that the result page is about plasma and does not match his/her needs, and has to load the page to find it out. According to Sweeney and Crestani [29]'s investigation on the effects of screen size upon presentation of retrieval results, it is best to show the summary of the same length, regardless whether it is displayed on laptops, PDAs or smartphones.

Improvements to mobile Web search done in academia go further. For example, De Luca and Nürnberger [31] implement search result categorization to improve the retrieval performance and present the information in three separate screens: screen for search and presentation of the results in a tree, screen to show search results and bookmarks' screen. Church et al. [32] substitute summary snippets, which are coming with each result item, with the related queries of like-minded individuals - queries leading to the selection of a particular Web page in the search result list. The researchers argue that such queries can be as informative as summary snippets and using this approach they provide more search results per one screen.

In contrast to the existing approaches, Shtykh et al. [33] (see also [30]) do not make any modifications to the search results, but propose an interface to handle the results provided by any conventional search service. The approach abolishes fatigue-inducing scrolling while preserving "quality" summaries of PC-oriented Web search. The proposed interface, called slide-film interface (SFI), is a kindred of "paging" technique. Unlike most mobile Web search services that truncate summary snippets of the search result items to reduce the amount of scroll and in this way facilitate easier navigation through search results that often can lead to difficulties in understanding of the content of a particular result, (owing to the availability of one slide of a screen size for one search result) our approach has an advantage to provide the greater part of one slide screen to place the full summary without any fear to make the search tiresome. SFI was compared with the conventional method of mobile Web search and the experimental results showed that, though there was no statistically significant difference in search speed when the two interfaces are used, SFI was highly evaluated for its viewability of search results and ease to remember the interface from the first interaction.

Although such approaches to improve the search with focus on the user, his/her usability are very important and user-oriented, they treat the user regardless of his/her contextual and situational information. As we already mentioned and will discuss more

Figure 2 The same search result item for PC-oriented Web search (left) and mobile Web search (right) [30].
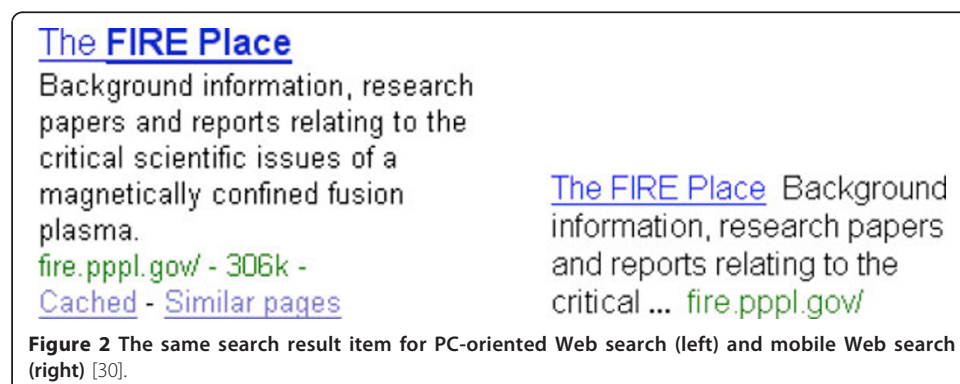
in Section 3, information need and human behavior are very contextual. Therefore peculiarities of information behavior, proclivities, preferences and everything that can give a better conception of the user, his/her behavioral patterns and needs must be considered in order to be able to provide a truly personalized information seeking experience. Although in the paper we focus on information seeking specifically, the application area of personalization spreads far beyond it. It is applied to Web recommendations and information filtering, user adaptation of Smart Home and wireless devices, etc.

Through our research we were particularly interested in personalizing and facilitating a human's interactions with various Web services. And search is not the only activity in Web information space users are engaged in. As empirical studies show [34], most of time users rediscover things they used to find in the past, and often they browse without any specific purpose discovering information space around them or with a particular purpose, such as learning miscellaneous information. To support such a discovery, we designed an exploratory information space [35] that makes use of human-centered power of bookmarking for information selection. The information space is built as a result of a search for something a user intends to discover, and serves as a place for rediscoveries of personal findings, socialization and exploration inside discovery chains of other participants of the system.

2.2. Personalization

Today personalization is the term we often relate to Web search personalization, such as in Google's iGoogle, recommendation system of Amazon.com, or contextual advertisements on Web sites. It is also about Decentralised-Me [36] of emerging Web 3.0 or is an essential part of Mitra [37]'s formula of Web 3.0 - *Web 3.0 = (4C + P + VS)*, where *4C* is Content, Commerce, Community, and Context, *P* is personalization, and *VS* is vertical search. However, the notion of personalization is much more diverse than that. It differs with regard to its application area and is being transformed over time and advances in its research. It is sometimes synonymous to *customization* and often to *adaptation*. It concurs with *information filtering* and *recommendation*.

In 1999 Hansen et al. [38] outlined two knowledge management strategies for business - *codification*, i.e., impersonalized storing knowledge in databases and its reuse, and *personalization*, which focuses on dialogue helping people to communicate knowledge. The authors claim that emphasizing the wrong strategy or pursing the both at the same time can undermine a business. However, today, in the situation of information overload, the both strategies often complement each other. Greer and Murtaza [39] define personalization as "a technique used to generate individualized content for each customer" and investigate the factors that influence the acceptance of personalization on an organization's Web sites. The research finds that ease of use, compatibility with an individual's value and his/her intents and expectations, and trialability ("the degree to which personalization can be used on a trial basis") are the key factors for personalization adoption. Monk and Blom [40] in their earlier works define personalization as "a process that changes the functionality, interface, information content, or distinctiveness of a system to increase its personal relevance to an individual," and Fan and Poole [41] extends this definition to "a process that changes the functionality, interface, information access and content, or distinctiveness of a system to increase its

personal relevance to an individual or a category of individuals" which serves as the working definition for the paper.

Such a great diversity in understanding of what personalization is results in difficulties to produce a holistic view on personalization, hurdles for sharing findings for researches of different fields and difficulties to compare approaches. And this is one of the conceivable reasons why the current approaches focus on "how to do personalization" rather than "how personalization can be done well," as Fan and Poole [41] has noted. Most personalization approaches on the Web are *system-initiated*, i.e., considering *adaptivity* which is the ability to adapt to a user automatically based on some knowledge or assumptions about the user. But another concept - of *adaptability*, which is a *user-initiated* (or *explicit* by Fan and Pool [41]) approach to modify the system's parameters in order to adapt its functionalities to his/her particular contexts, - is also important when considering personalization. Monk and Blom [40] emphasized that people always personalize their surroundings, and their Web environment is not an exception, and presented their theory of user-initiated personalization of appearance.

Personalization has a lot of advantages over impersonalized approaches, some of which are obvious and some of which are hidden and have to be empirically proven. For instance, Guida and Tardieu [42] prove that personalization, similarly to long-term working memory, helps to overcome working memory limitations, expanding storage and processing capabilities of human-beings. Although the discussed personalization is considered as a creation of the situation of individual expertise that is generally not exactly what modern personalization systems can provide, such approach indicates the need in better considering context and situation in order to fully employ its merits.

2.3. Modeling User Interests

In order to be user-centric, a service has to know each user it interacts with. This is the task personalization attempts to fulfill with a variety of methods in various work task and environmental settings. Personalization systems extract the user's interests, infer his/her preferences, update and rely on knowledge about the user accumulated and structured in user profiles that differ by the data used for their definition, their structure and complexity, and construction approaches.

At this point we have to note that in modeling user interests we do not make a distinction between Web search personalization, recommendation or information filtering because the differences in their methods and goals are very subtle. All such approaches utilize a certain scheme to know the user's preferences to adapt to his/her future interactions with the system and information it provides, and constructing user profiles (or user modeling) is the most popular method. It has been extensively used from days of first information filtering systems, for instance as a user-specified profile or a bag-of-words extracted from the documents accessed by the user, and today it takes many richer and diverse forms to meet the requirements of the variety of information systems.

2.3.1. Relevance Feedback as a Modeling Material

As the reader can see from the above discussions, use of relevance feedback for personalization is very important and widely utilized. Let us see what types of feedback exists and what kinds of data are used for feedback.

Feedback Types Relevance feedback is extensively used in Web IR for efficient collection of user behavioral data for further user behavior analysis and modeling. Relevance feedback can be *explicit* (provided explicitly by the user) or *implicit* (observed during user-system interaction). The first form of relevance feedback is high-cost in terms of user efforts and the latter one is low-cost but requires a thorough analysis to reduce the noise it normally contains. Implicit relevance feedback in IR systems consists of a number of elements, such as a query history, a clickthrough history, time spent on a certain page or a domain, and others, that can be considered in general as a collection of implicit behaviors of users interacting with the information retrieval system. It is conducted without interruption of user activities, unlike explicit one that requires direct user interferences, that is why many are showing keen interest in it. Interested readers are referred to [43] for survey on the use of classic relevance feedback methods and [44] for extensive bibliography of papers on implicit feedback, or any modern information retrieval (IR) textbook for the detailed introduction of relevance feedback.

With emergence of social network, new types of feedback become available. Thus, social bookmarking and tagging, as described in [45], are sui generis mixture of both implicit and explicit relevance feedback. On one hand, bookmarking is an explicit action done by a user and not monitored for by the system, on the other hand, in contrast to explicit feedbacks, it is normally not a burden for the user. We would classify such a feedback as *motivated explicit feedback*, since it is motivation that removes burdens from the explicit nature of the feedback.

Another emerging type of relevance feedback that is worth mentioning is *contextual relevance feedback* which shows again an increasing attention to context for personalization. As a matter of fact, it is often of no difference from many other approaches based on user profiles. Thus, in [46]'s approach contextual relevance feedback is a feedback to a search result list to filter it based on user-collected document piles. Another example is contextual relevance feedback architecture by Limbu et al. [47] which, in addition to profiles, utilizes ontologies and lexical databases.

Types of Data for Relevance Feedback As to the types of data used for profile construction, their choice depends on the application domain of the system to be personalized. For IR systems, relevance feedback is normally documents, queries, network session duration and everything related to information search process on the Web and beyond. For instance, Teevan et al. [48] extend the conventional relevance feedback model to include the information "outside of the Web corpus" - implicit feedback data is derived from not only search histories but also from documents, emails and other information resources found in the user's PC. With the change of the application domain the type of data differs. For instance, mobile device features and location can be considered for profile construction in nomadic systems [49], and user interests can be learnt from TV watching habits, as in [50]. Naturally, any user behavior can be considered as a source for inference of his/her interests and further user profiling, and there are as many selection decisions in regard to use of a particular feedback type as there are systems that utilize them. Fu [51] proposes to examine a variety of behavioral evidences in Web searches to find those that can be captured in a natural search settings and reliably indicate users' interests.

2.3.2. Modeling Methods

With the afore-mentioned data, user interests can be inferred and user profiles (models) can be created in a number of ways and various methods. Most of them use vector-space and probabilistic modeling approaches, some of them are based on neural networks or graphs. It is hard to clearly classify all of them, since many of them are very domain-data-dependent and thus their methods are very specific. Often user interest modeling is done specifically for the system it is applied to with regard to its application domain and based on the specific data that can be obtained from user-system interactions of this particular system. Consequently, modeling methods for user interests will be constrained to that type of systems, in contrast to other generic modeling approaches.

For instance, the personalized peer-to-peer television system by Wang et al. [49] is interested in user interests inferred from TV watching habits. For user u_k the interest in program i_m is calculated as

$$x_k^m = \frac{WatchedLength(m, k)}{OnAirLength(m) \cdot freq(m)} \tag{2.1}$$

where *WatchedLength(m, k)* is the duration of program i_m in seconds watched by user u_k, *OnAirLength(m)* is the full duration of program i_m, and *freq(m)* denotes the number of times its has been broadcast. Models in e-learning, in addition to interests, often consider learning styles and performance, cognitive aspects of a learner, etc. They are complex and require explicit directives and assessments of an instructor. For instance, student profile in [52] consists of four components: 1) cognitive style, 2) cognitive controls, 3) learning style and 4) performance. It is created by a student registering to the course and complemented by the instructor's and psychological experts' surveys on the user's cognitive and learning styles. It is updated with the student's feedback, monitored performance and the instructor's decisions based on the user's learning history.

2.3.3. Structural Components

There is a great variety of profile structure types. The simplest and most widespread one is to represent user interests learnt from relevance feedback with document term vectors for each interest's category. Shapira et al. [53] enhance such vectors with sociological data (profession, position, status). Profiles in Sobecki [54] are attribute-value tuples, where the attributes characterize usage such as visited pages or past purchases, or demographic data such as name, sex, occupation, etc. In Ligon et al. [55]'s agent-based approach user profiles are a combination of information categories and a preference database containing search histories related to the categories.

User profiles become more elaborate and complex trying to reflect the dynamics of constantly changing user context and interests. For instance, Bahrami et al. [56] distinguish static and dynamic user interests for profile construction in their information retrieval framework. Barbu and Simina [57] distinguish Recent and Long-Term continuously learnt user profiles and apply them to information filtering tasks. Further, information systems utilized by mobile devices often extend the notion of user profile in conventional IR systems bringing specific contextual information into it. For instance, Carrillo-Ramos et al. [48], in attempt to adapt information to a nomadic user by taking context of use into consideration, introduce Contextual User Profile which

consists of user preferences and current context (location, mobile device features, access rights, user activities) of use. Ferscha et al. [58] propose context-aware profile description language (PPDL) expressing mobile peers' preferences with respect to a particular situation. Finally, some attempts to provide more holistic approaches to profile structuring, such as Gargi [59]'s Information Navigation Profile (INP) defining attributes for characterizing IR interfaces, interaction and presentation modes, are made resulting in complex profiles that consist of multiple search criteria.

2.3.4. On User Contexts

As we already noted, personalization with better focus on user contexts and situations is the topic to be better investigated in the near future. As personalization depends much of the intents of and results expected by a user, it is essential to accurately assess his/her contextual characteristics.

In spite the fact that a number of personalization approaches today use the notion of context, such 'context' is usually derived from queries and retrieved documents and/or inferred from user actions. They are not likely to accurately capture the situation and the context which includes far more factors than taken in such approaches. Furthermore, the definition differs from one solution to another. And, naturally, the diversity grows in mobile and ubiquitous personalization approaches because of context peculiarities. For instance, while context of a user is being learnt, for instance, from documents and ontologies [60], multiple context attributes like environmental and other properties (time, location, temperature, space, speed, etc.) are considered in [61] to define context-aware profiles. And probably because of such differences related to application domains, there is very little exchange of verified practices among researchers working on personalization in different areas and, despite available similarities in various domains, the one-sided views on context are not rare. There are endeavors to utilize context and situation in a holistic fashion (e.g., [26]), however they are mostly on the level of theory. We believe that accurately and timely estimated contextual information will greatly contribute the field of personalization, therefore further endeavors to characterize, methods to capture and systematize knowledge about it should be continued, deepened and corroborated with empirical studies.

3. User-Centric Information Search and Sharing with BESS

3.1. Being User-Centric by Knowing User's Preferences through Contexts

One of the main driving forces of human information behavior is information need that is recognition of one's knowledge inadequacy to satisfy a particular goal [62], or "consciously identified gap" in one's knowledge [26]. Therefore its understanding is crucial for systems that are supposed to facilitate information acquisition. However, in many cases capturing and correctly applying individual information needs is extremely difficult, even impossible. For instance, in IR systems a user's input cannot usually be considered as a correct expression of his/her information needs - that results in invalidity of many traditional relevance measures [63]. And this happens not only in IR, but in any system when context, in which an information need was developed, is lost.

Then, the following question arises. From the discussion to this point in the paper, we can define user-centric system as a system that "understands" (is able to capture) the user's information need in order to satisfy it effectively. But *how can the system be*
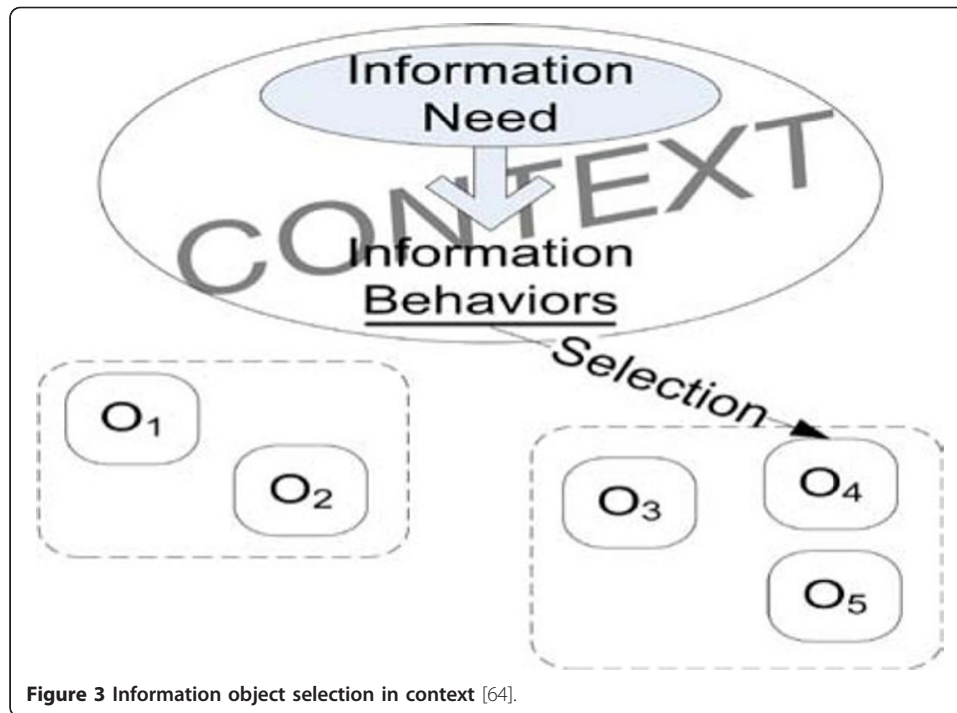
Figure 3 Information object selection in context [64].

user-centric and satisfy sufficiently the user's information need without being able to capture it?

Information need emerges in one's individual context, and both context and information need are evolving over time. Information behaviors happening to satisfy the information need and leading to an information object selection also take place in the same particular context (Figure 3). Therefore, although knowing particular contexts does not give us the full understanding of a particular user's information needs, such knowledge can give us some conception (or a hint) of conceivable information a user tries to obtain in a particular context, i.e., lead us to the potentially correct object selection. As shown in Figure 3, particular information need in a particular context leads to information behaviors which, in their turn, result in object selections from, for instance, two groups of similar objects. Knowing information behavior patterns (and their contexts) resulting in particular object selections, in our research we try to induce a user's current preferences for a particular object without clear knowledge of current information need. Such knowledge gives a chance for a service to identify user contexts during user-service interaction and help with correct information object selection. Further, by matching context information of one particular user with contexts of other users that utilize the same service, we can try to foresee a situation new to the user (an unknown context) and facilitate his/her information behavior.

Essentially, context can be considered as a formation of many constituents - an individual's geographical location, educational background, emotions, work tasks and situations, etc. With the advances of spatial data technologies, ubiquitous technologies and kansei engineering we are likely be able to collect a large part of them in the near future, but this task is still very challenging. Even more challenging is the task to effectively utilize all these constituents in various user-centric services. Moreover, the need

in some particular constituent of the whole context depends on the task one particular system is trying to facilitate.

In information seeking tasks we are studying, as in most tasks that support information activities today, it is impossible to collect all contextual information, so the contexts considered here have a *fragmentary nature* - basically consisting of information behaviors obtained from users' explicit and implicit relevance feedback [65]. Generally, it is a feedback of textual, temporal or behavioral information with regard to the resources a user interacts with.

3.2. User-Centrism in BESS: Main Concepts of the Proposed Approach

In the proposed approach we attempt to utilize acquired user contexts as much as possible to make the services of BESS user-centric and consequently help users with effective acquisition of information pertinent to their particular contextual and situational information needs. The main concepts for achieving such user-centeredness after having appropriate contextual information are

  1) concept;
  2) multi-layered user profile;
  3) interest-change-driven profile construction mechanism;
  4) subjective index creation and its collaborative assessment;
  5) subjective concept-directed vertical search.

3.2.1. Determining and Organizing Personal Interests

Information seeking, as any information behavior, is done in a context determined by situation, interest, a person's task, its phase and other factors. In the process, some user interests tend to change often influenced with temporal work tasks and personal interests, and some tend to persist. Capturing them gives us a fragmentary understanding about current user contexts and can be used to induce a general understanding about the user. In our research such interests are inferred from relevance feedback information provided by the user and are a set of conceivably semantically-adjacent terms. Therefore they are called *concepts*.

However, such concepts are not much of interest when they are not organized by some criterion that helps an IR system to understand their tendency to emerge and change. In order to organize user interests and have the whole contextual picture, we chose user profile construction based on the temporal criterion. As a result, user profiles in BESS are *multi-layered* - each of layers reflecting user interests temporally, corresponding to long-lasting, short-term and volatile interests. Furthermore, they are generated with *interest-change-driven profile construction mechanism* which relies entirely on dynamics of interest change in the process of profile construction and determination of current user interests (see Section 4).

Obviously, for inference of interests we have to handle a user's relevance feedback separately from all information resources available at the system. Therefore, each user has its own *subjective index* data which is generated from his/her relevance feedback. It distinguishes from index data of conventional search engines, which we call *objective index*, by its social nature - it is created based on the information found valuable in the context of a specific information need and submitted by users, in contrast to objective index which is collected by crawlers or specialists without any particular

consideration of context, situation or information need. Collecting such personal information pieces gives us access only to highly selective information tied to a specific context - without such a relation preserved, this information is not much different from that stored in conventional search systems.

3.2.2. From I-Centric to We-Centric Information Search and Sharing

Determining and organizing a user's personal interests is very helpful to further facilitate user-system interactions in general, and information seeking tasks in particular. However, would such facilitation be fully user-centric without collaboration of all members of the system? Probably, it would be. But, as we discussed in Section 1, such an approach would not benefit from "wisdom of crowds" [22] of other users and loose much predictive power it could draw upon other users' experiences. In addition, personalization that is oriented on one individual will lead to different experiences among community of users and can increase problems of transparency and interpretation [66], but sharing information with others creates new possibilities for discovery and reinterpretations. Recognizing this, BESS is designed as a highly collaborative information search and sharing system. It harnesses collective knowledge of its users who share their personal experiences and benefit from experiences of others. In other words, this is *We-Centric* part of the system, in contrast to *I-Centric* one harnessing solely personal experiences.

To emphasize the collaborative nature of relevance feedback submitted by users explicitly, it is called a *contribution* in our research. Although explicit feedback can disrupt search user activities, it is important for subjective index creation, and explicit measures in information retrieval tasks are found to be more accurate than implicit ones [67]. Together with implicit feedback it forms subjective index of each user which in turn is used for concept creation. As we already mentioned, concepts correspond to user interests, and, placed into user profiles, they are used to assess each user's expertise with regard to a concept of the relevance feedback the user contributes. These assessments are an important mechanism to estimate the value of a particular piece of information based on the contributor's expertise, which is induced from dynamically changing user profiles, and help to find relevant information to people with similar interests and work tasks through *subjective concept-directed vertical search*, which is discussed in detail in Section 5.

To summarize, the search experience we are trying to provide can be characterized as collaborative and personalized. Users' searches and contributions have a personalized (I-Centric) nature, and information pieces found valuable by every user in context of his/her current information needs are shared among all users (We-Centricity).

3.3. Position of BESS among Modern Web Personalization Systems

Reconsidering information retrieval in the context of each person is essential to continue searching effectively and efficiently. That is why so much attention is paid to this problem and consequently a number of approaches to Web search personalization have emerged recently. Nowadays we are experiencing the much anticipated breakthrough in personalized search efficiency by "actively adapting the computational environment - for each and every user - at each point of computation" [68].

To show the peculiarities of existing Web search personalization systems and the position of BESS inside Web search personalization approaches we classify them as

vertical and *horizontal*, *individual-oriented* and *community-oriented* based on breadth of search focus and degree of collaborativeness they possess (see Figure 4; arrows denote current trends in search personalization).

Outride [68] and similar systems take a contextual computing approach trying to understand the information consumption patterns of each user and then provide better search results through query augmentation. Matthijs and Radlinski [69] construct an individual user's profile from his/her browsing behaviour and use it to rerank Web search results. On the other hand, Sugiyama et al. [70] experiments with a collaborative approach constructing user profiles based on collaborative filtering to adapt search results according to each user's information need. Almeida et al. [71] harnesses the power of community to devise a novel ranking technique by combining content-based and community-based evidences using Bayesian Belief Networks. The approach shows good results outperforming conventional content-based ranking techniques. Systems like Swicki, Rollyo, and Google Custom Search Engine correspond to vertical and mostly community-oriented approach of search personalization. They provide community-oriented personalized Web search by allowing communities to create personalized search engines around specific community interests. Unlike horizontal (or broad-based) search systems mentioned above, such systems are considered personalized in the sense that available document collections are selected by a group of people with similar interests and the systems can be collaboratively modified to change the focus of search. Although not Web-based, we take tools like Google Desktop Search as an example of individual-oriented vertical search systems. They search contents of files, such as e-mails, text documents, audio and video files, etc., inside a personal computer. The absence (to the best of our knowledge) of salient Web-based systems of this kind can be explained by the increasing popularity of services on the Web benefiting from

Figure 4 Search personalization services and BESS.
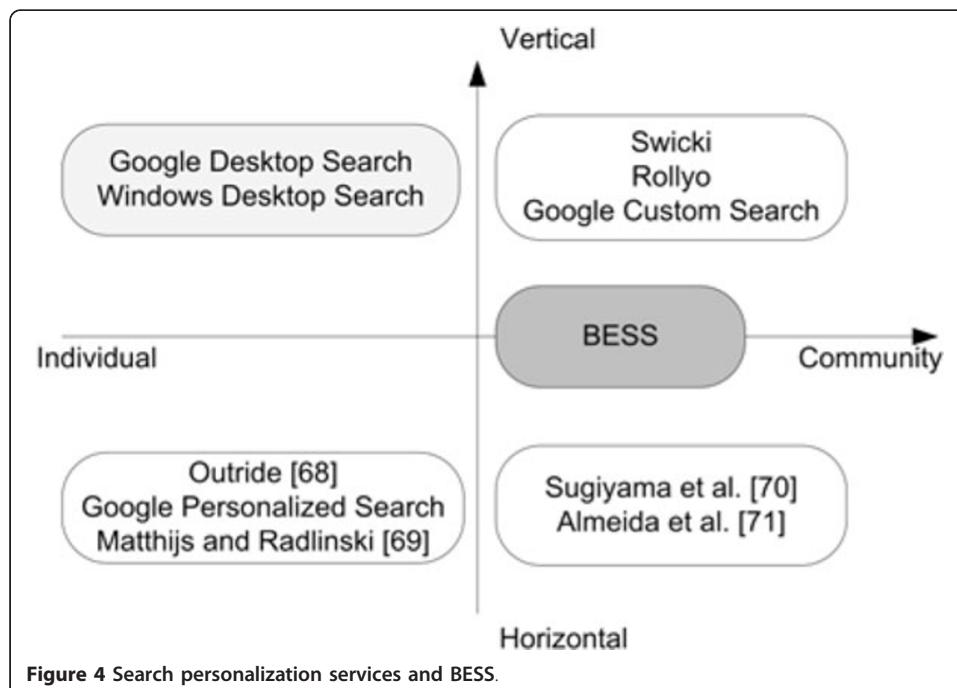
community collaboration and favoring fast transition of each person's activities from passive browsing to active participation.

As it is shown in Figure 4, BESS is a community-oriented system having the features of both horizontal and vertical search system. It performs search on information assets of both horizontal (objective index) and vertical (subjective index) nature. The notion of *subjective index* in our research is similar to 'social search' of vertical community-oriented systems presented above, but differ in higher degree of personalization for every user, high granularity of vertical search model (see *subjective concept-directed vertical search* in Section 5) and, finally, the way of collecting and (re-)evaluating information pieces. Groups of users are created dynamically without a user's interference based on match of interests/expertise, and the role of community is indispensable for search quality improvement and the system's evolution in general.

3.4. Architecture and System Overview

BESS is a complex system that consists of several components for relevance feedback collection, analysis and evaluation, online incremental clustering, user profile generation, indexing and a few elements realizing several search functionalities.

As we have already discussed, the main purpose of BESS is to realize collaborative personalized search. And to achieve the assigned tasks, first of all, our collaborative search and sharing system has to be capable of distinguishing users, and collecting and analyzing their personal feedback. "Access control and data collection" module of BESS is responsible for this. A user is authenticated when accessing the system, so we know whom it is used by. After that, his/her interactions with the system are logged. To have an understanding of the user's interests we are primarily interested with contributions (explicit feedback), done through the contribution widget of a Web browser, and implicit feedback, collected by monitoring the clickthrough. All the interaction data is stored in "Activity data" database, as shown in Figure 5. Then, this 'raw' data is processed and clusters (concepts) reflecting the user's interests are created by "Data analyzer." Existing concepts are incrementally updated. At this moment the interests are inferred and known, but are of little interest because they say nothing about their temporal characteristics. As a result, some concepts can be outdated, others can be recent and topical.

In order to organize the concepts, "Profile generator/analyzer" generates a user profile using interest-change-driven profile construction mechanism, as described in Section 4, and it is stored. We have to note that, as it is also discussed in the next section, user profile is very central for the system functioning in general. As it is shown in Figure 5, user expertise, together with expertise of other users, with regard to a particular topic (concept) is used for assessing his/her feedback, which is then indexed and stored in the "Subjective data" repository for further retrieval. This personal and 'collectively evaluated' feedback becomes a piece of the user's subjective index data.

Now, when we have data to be searched on, let us consider search.

On logging in, the user has an opportunity to search both with conventional search engines and the search engine provided by BESS. Essentially, both are used when a search request is issued. The results of the conventional one are shown in "Objective search results area" and the results of the one provided by BESS are shown in "Hidable subjective search results area" (see Figure 6). The user can select his/her favorite Web

Figure 5 General system architecture.
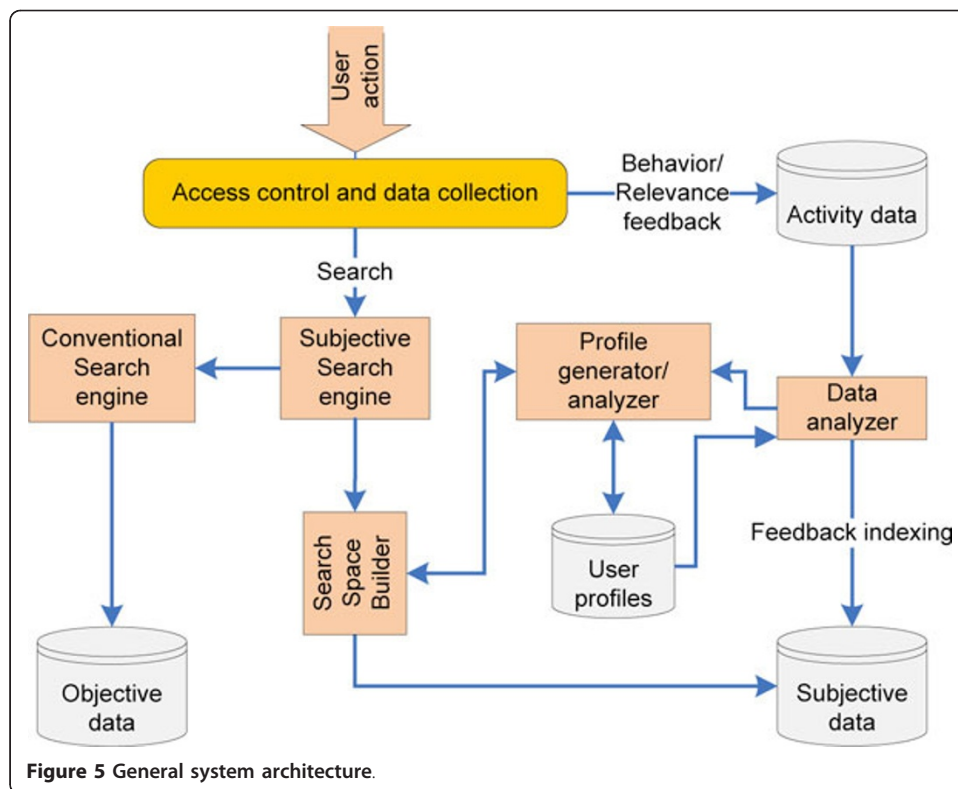
search service from "SE Switch" and hide "Hidable subjective search results area" if there is not enough subjective contributions for the topic in concern, or he/she is simply not interested in collaboration temporarily and wants to concentrate on objective search only. In any case, the user is enriching his/her personal subjective index, and consequently all shared subjective index.

Search on the subjective index data is normally done in the all-shared mode, when the subjective index of all users is searched on. In this case, query-document matching is performed, and all matched documents are retrieved and listed according to the ranking algorithm. However, the user has another option - to search on the subjective index data of the users whose user profiles are conceptually close to his/her current user profile by switching with "Search mode switch." This is what we mentioned as

Figure 6 User interface schematically.
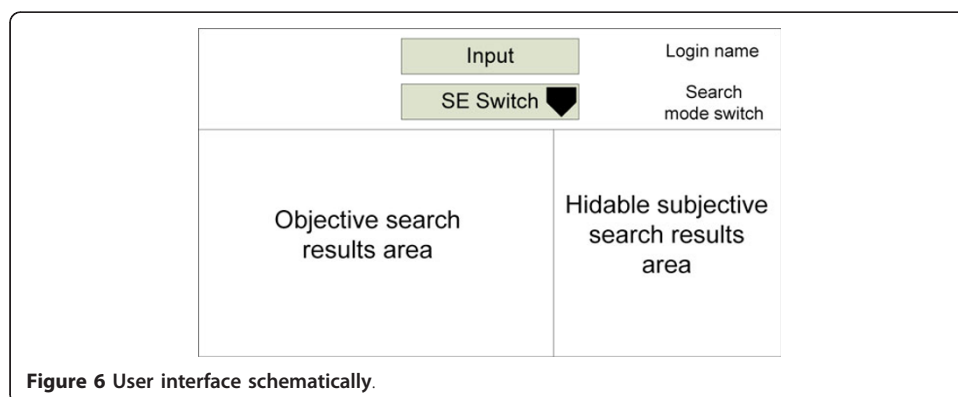
subjective concept-directed vertical search already (Detailed discussion of the ranking algorithm and *subjective concept-directed vertical search* is given in Section 5).

3.5 Notes on Implementation Technologies

In order to realize all the described functionalities, BESS employs a number of technologies, such as online incremental clustering, indexing and search. Indexing and search is done with help of customized Apache Lucene. User profile construction is a module set implemented according to the methods described in the following section of the paper, and online incremental clustering is described in [72] in every detail. All implementation is done with Java, using JSP (Java Server Pages), Java Servlet, Spring Framework and other Java technologies. For the development of contribution submission Firefox component, we used AJAX (Asynchronous JavaScript and XML) and XUL (XML User Interface Language).

4. Constructing Interest-Change-Driven User Profile

As we have discussed in Section 2, there are many different ways to construct and organize a user's interests using user profiles. The organization structure usually depends on what characteristics of the user a user profile is designed to capture. User profiles in BESS are designed to timely and effectively capture the user's interests, to update his/her profile in regard with its temporal, and transitively interest-involvement-degree, characteristics, and to be used for collaborative contribution evaluation and information retrieval. User profiles are composed from concepts which serve as representatives of the user's interests. They are *multi-layered* with layers reflecting temporal characteristics of user contexts. Furthermore, they are dynamically updated to precisely reflect changes in interests using *interest-change-driven profile construction mechanism* presented further in this section.

4.1. The Role and Position of User Profile

User profiles play a key role in our BESS information retrieval framework. The framework is developed in attempt to capture information needs and information seeking contexts of every individual, and better facilitate information seeking activities by identifying and providing information resources pertinent to every individual's needs. This is achieved by modeling a user's changing interests from relevance feedback (explicit feedback, called contributions, and observed user behavior, such as clickthrough information) over time and using the models

- to evaluate the feedback by considering the contributor's expertise and his/her past experiences with the concept the user feedback belongs to, and
- to change the focus of search, similarly to what occurs in vertical search engines, but automatically, detecting users with similar contexts and using their concepts.

These steps ensure the search is done on highly selective documents evaluated by the users with similar interests taking into account their expertise, or the degree of their involvement into a particular topic.

Figure 7 is a schematic fragment of the system architecture describing the position and the role that user profiles have inside the system. First, the analyzed relevance feedback is used to update a user's profile with a newly created or updated concept.
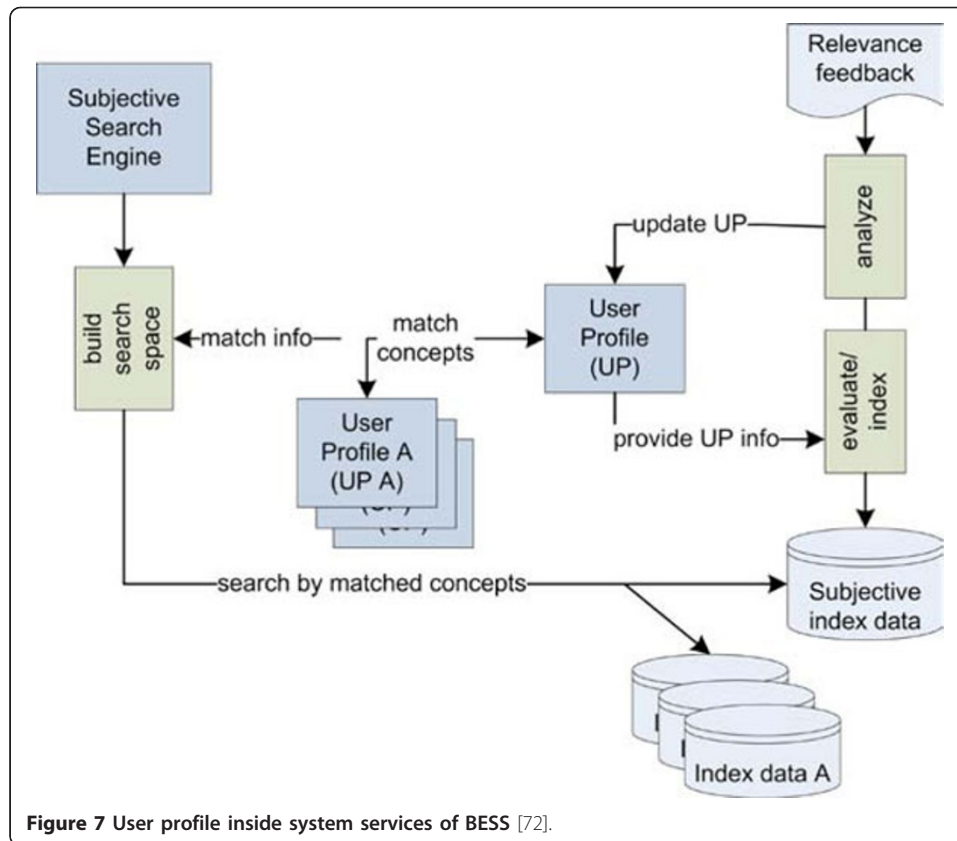
Figure 7 User profile inside system services of BESS [72].

Then, the updated profile and its concepts' peculiarities are used for the evaluation of the same relevance feedback item (details of the evaluation mechanism are given in Section 5). And finally, the feedback is indexed in every individual's subjective index repository which is shared among all users of the system. When a user searches, he/she can search on the multiple sets of information assets evaluated according to each user's expertise or narrow his/her search to the resources of those users whose interests (concepts in user profiles) are similar to his/her own.

As it can be seen from this short description, the position of user profile in the system operations is central and the quality of the profile is of vital importance not only to information seeking experiences of one user but to the experiences of all users of the system. Therefore, in this paper we pay the particular attention to the profile construction and to the quality of the concepts, which are the constituents of user profiles and indicators of user interests, in particular.

4.2. Concept as a Principal Profile Component

Relevance feedback is an essential element of any information filtering system and a significant part of the proposed system. It is extensively researched in its various forms. Explicit feedback often disrupts normal user activities; therefore another form of feedback that can be collected with no extra cost to the user - implicit - is used widely. Sometimes these two forms are combined to get better insight about a user's peculiarities. Kelly et al. [44] gives a good classification and overview of works on implicit feedback. In many cases, user behavior is considered to be an implicit feedback, and its analysis is done for improving

information retrieval by predicting user preferences, re-ranking Web search results and disambiguating queries.

Often relevance feedback is used in attempt to find out user behavioral patterns and generate individual user profiles reflecting current user interests. There exist many approaches for profile modeling. Nanas et al. [73] has profiles made from concept hierarchies that are generated from user specified documents and applied for information filtering. Profiles are divided into three layers by a heuristic threshold each of which determines the topic, subtopics and subvocabulary for the specified topic. Term weighting approach (Relative Document Frequency) is extensively used for hierarchy construction. Matthijs and Radlinski [69]'s profiles are built with the emphasis on users' browsing behavior, therefore, in addition to terms, a list of visited URLs, the number of visits to each, a list of past search queries and pages clicked for these search queries are used for profile construction. Semeraro et al. [74] uses a different approach for profile construction. As in [73], profiles consist of concepts, but the approach employs ontologies where semantic user profiles are built with the use of content-based algorithms extended using WordNet [75]. Such an approach is proved to help infer more accurate user profiles.

BESS makes extensive use of both explicit and implicit relevance feedback for the construction of personal information assets and user profiles. Unlike profiles in the above-mentioned approaches, profiles in BESS are constructed with the main focus on users' interest change when searching, and concepts in them are loosely coupled and dynamic.

User profile in BESS is a structured representation of user contexts which are in turn consist of preferences and interests of a user. It consists of concepts (semantic clusters), and each concept is the system's piece of 'knowledge' about what the user is interested in. Each concept is modeled as a cluster c_i of n document vectors $X = (x_1, ..., x_n)$ from the individual document set grouped by a specific 'knowledge' criteria. Concepts are extracted from minimal user search and post-search behaviors (user-system interactions while searching, browsing and contributing Web pages). The system is configured to capture the following data:

- user ID used for authentication;
- search query terms;
- URL of the page the user is interacting with;
- type: query, click or feedback;
- timestamp;
- session ID.

Prior to concept extraction, documents from individual document collections are linearized by removing HTML and script tag data, non-content-bearing 'stopwords' are deleted and document vectors are normalized. Then, a classification method is used to extract concepts from the document vectors. Virtually, any method can be applied for this.

4.3. User Profile Structure

Information seeking, as any information behavior, is done in the context determined by situation, interest, person's task, its phase and other factors. In the process of seeking information, needs and their contexts are changing even within the same seeking task.

Recognizing this fact, we introduce a temporal dimension to user profiles by splitting and combining (generalizing) all concepts on a time line. For this, we make user profiles in BESS multi-layered - each layer reflects user interests within a certain period. It consists of four layers - static $pr^{(st)}$, session $pr^{(ss)}$, short-term $pr^{(sh)}$ and long-term $pr^{(ln)}$ (Figure 8). Thus, profile of user a can be defined as

$$\text{Pr}_a = (pr_a^{(st)}, pr_a^{(ss)}, pr_a^{(sh)}, pr_a^{(ln)}) \tag{4.1}$$

Each layer consists of concepts which are the components of profiles representing user contextual information by topics:

$$pr_a^{(l)} = (C_{a1}, ..., C_{ak}) \tag{4.2}$$

where l is a layer and k is a concept number.

Each layer has a pool of concepts that characterize best a user's seeking context for the layer's time span. The static layer is defined at the start of user-system interaction to solve so-called "cold start" problem when the system has no information about the user and cannot facilitate his/her activities or can even damage the whole interaction. Other three layers can be classified as dynamic layers, since they are dynamically constructed and changed along with changing user information needs and their contexts.

The session layer contains the fragmentary context of the current information behavior of a particular user. It is a highly changeable layer and defined by a concept that best matches one of the concepts available in the short-term layer or a newly created concept. In other words, the session layer is the indicator of context switch at the lowest level. The short-term layer is a central layer of the whole system - it consists of concepts formed in all user-system interaction sessions within a specified period of time, and its generation itself serves as an important factor for collaborative feedback evaluation mechanism. And finally, the long-term layer is derived from the most frequent concepts of the short-term layer, as discussed in the profile construction section,
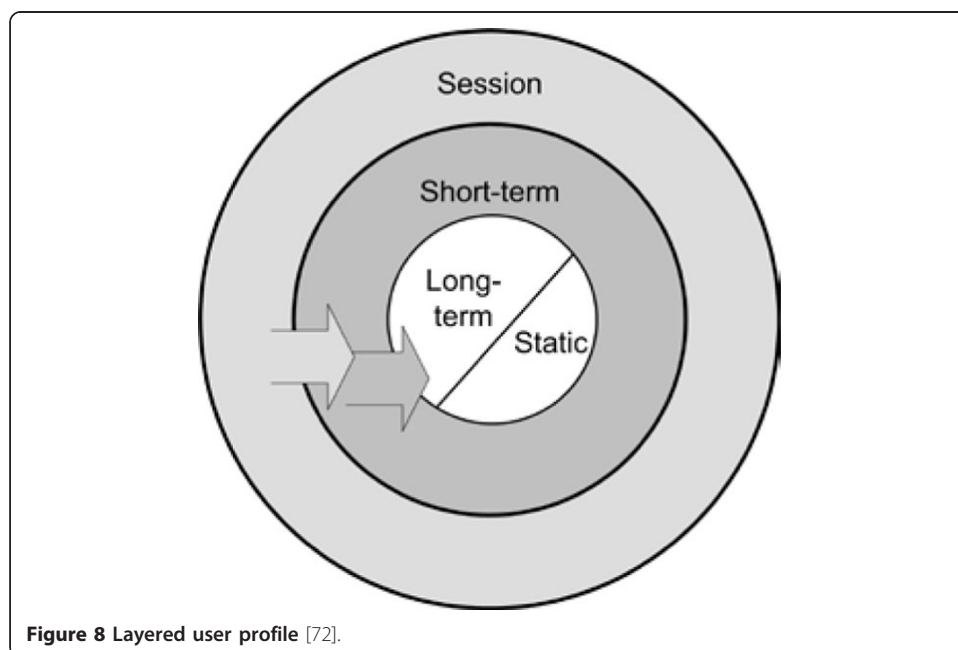

Figure 8 Layered user profile [72].

and reflects general user context of interaction with the system. When there is enough information for its formation, it is created and gradually supersedes the static layer. The profile layer construction mechanism is further described in the next subsection.

4.4. Dynamic Interest-change-driven Profile Construction

As we have already described, user profile plays an important and central role in BESS for collaboratively evaluating documents contributed to the community and for adjusting the focus of search. Therefore user profiles have to be precise and accurate, and this is achieved by correctly specifying and evolving their concepts in the online and incremental fashion. Moreover, profiles ought to timely reflect the changeability of user interests while maintaining the steadiness of persistent preferences. In our *interest-change-driven* model for dynamic user profile generation we proposed in [65] we adopt *recency*, *frequency* and *persistency* as the three important criteria for profile construction and update.

Once we have concepts extracted from a user's feedback, we can detect the change of a user's context and set the latest one as the current context (*recency criterion*), which is the session layer in multi-layered user profiles. By observing concept creation dynamics we can set some to be the short-term layer according to the following (*frequency* and *recency*) rule:

For n concepts in the latest clustering output, choose newly-created and already existing concepts whose input item growth is high in a reverse order (newness) of the output sequence.

And finally, the long-term layer is formed from *n* most frequent concepts which have also been observed in the short-term layer.

Thus, concept extraction method produces $C_a = \{C_{a1}, ..., C_{an}\}$ set of *n* concepts which are ordered by recency criterion, i.e., a concept that is newly created or most recently updated appears at the top of the recency list. C_{a1} is the most recent concept and considered to be the current context and the session layer of the profile of user *a*, i.e., $pr_a^{(ss)} = C_{a1}$.

The short-term layer consists of *m* most frequently updated and used concepts, which are, in their turn, chosen from *r* most recent (top) concepts in the concept recency list. In other words, these are the concepts that are frequently used and still of some interest for the user. Figure 9 explains how the short-term layer is created.

The goal of the long-term profile layer is to find persistent user interests. Therefore its construction is based on *persistency* criterion and, indirectly, on frequency and recency considered for the short-term layer creation - the layer is derived from the concepts of the short-term layer which were most frequently observed as the layer's components. To determine the concepts matching the afore-mentioned criteria, in addition to concept update frequency $freq_c$, we introduce frequency measure $freq_s$ for the number of times the concept was a component of the short-term layer and find *m*

---

1   Fetch *r* most recent concepts from the recency list.
2   Sort them by update frequency in the descending order.

3   Take top *m* concepts to create the short-term layer $pr_a^{(sh)} = \{C_a^1, ..., C_a^m\}$, $m <= r$. The concepts are ranked by frequency.

Figure 9 Short-term layer creation procedure.

concepts whose *persistency factor PF* is high. Persistency factor is a measure to infer the user's continuous interests by combining a concept's frequency count with its evidence of being a user's short-term layer's constituent.

$$PF_{C_{ai}} = \alpha \frac{freq_c(C_{ai})}{\max freq_c(C_a)} + (1 - \alpha) \frac{freq_s(C_{ai})}{\max freq_s(C_a)} \qquad (4.3)$$

where α is set experimentally. C_{ai} is a concept of the set of concepts C_a produced from relevance feedback of user a.

The concepts for the long-term layer are found by the procedure shown in Figure 10.

All the layers dynamically created at time t form concept-based *interest-change-driven model* of user a, and are the representation of the user's interests at t. A change of the concepts in terms of their ranking in the short-term profile layer signifies a change of user interests and emergence of a new model of user a. The model update is not constrained with the predefined parameters, such as fixed time period after which the update occurs, and driven by natural dynamics of changing user interests. This mechanism is used to find a user's n past profiles and their concepts to determine the areas of expertise of the user to be used in his/her feedback evaluation mechanism, as described in Section 5.

4.5. User Profile Construction: An Example

To demonstrate profile construction using the proposed profile construction scheme and show the rationality of the chosen approach, we give an example of profile construction and discuss its peculiarities.

First, we implemented the profile construction system where every user relevance feedback was processed one by one and the extracted concepts were used to create user profiles according to the scheme described in Section 4.4. Then, we prepared relevance feedback obtained from 12 users with ages from the mid-20's to the mid-40's during one of our experiments for observing users' Web search behavior, which lasted two weeks and resulted in average 320 records collected per participant. The data was processed sequentially using H2S2D (High-Similarity Sequence Data-Driven) clustering method we proposed in [72] with 0.1 threshold, which was proven to produce concepts of reasonably good quality fast and in online and incremental fashion. As a result, in overall 20 concepts were created.

Here we show typical user profile construction results for one user. Since the session profile layer is simple - consisting of one currently used concept - and very frequently changed with the change of the user's current interests and needs, we skip it to illustrate the dynamics of short-term and long-term layers. Figure 11 shows how the user's short-term profile layer is being generated during concept extraction process. "Processed items" axis refers to the number of relevance feedback items processed by H2S2D method. So, for instance, label "288" indicates 288 items processed one by one and it is a point of change of user interests - literally, change of rank of concepts C1,

1 Calculate $PF_{C_{ai}}$ for each concept.

2 Take m concepts with the highest values to create the long-term layer $pr_a{}^{(\ln)} = \{C_a^1, ..., C_a^m\}$.
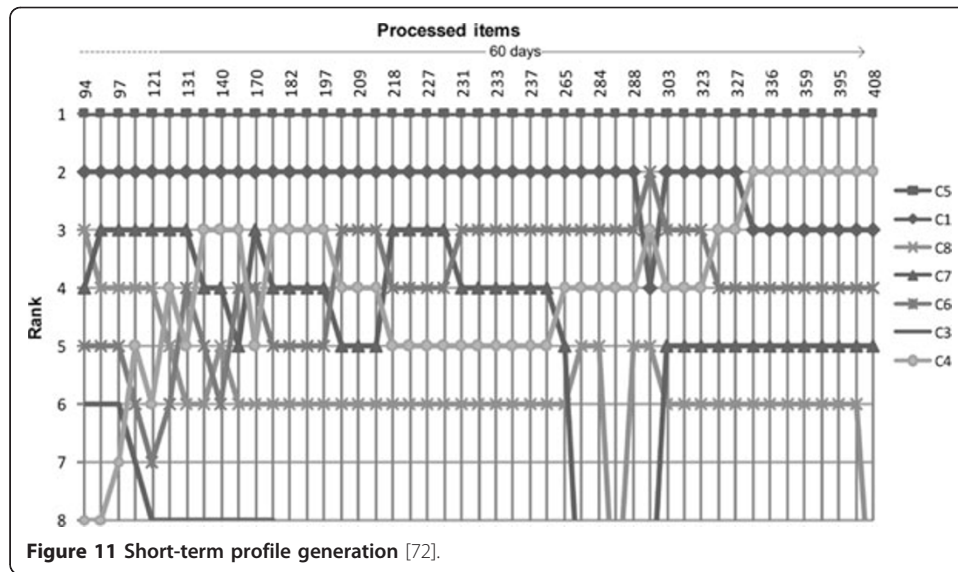
Figure 10 Long-term layer creation procedure.

Figure 11 Short-term profile generation [72].

C6, C4 and C8 in the short-term profile layer. "Rank" axis refers to the rank of concepts in the layer (explained in subsection 4.4), where only top m items are considered being the layer's concepts and the others are given to show the change dynamics of concepts during the period the user supplies relevance feedback data when interacting with the system. As shown in the figure, the rank of the concepts, as it can be expected, tend to change often initially and become more stable when more feedback, and accordingly better-quality concepts, is available. In fact, we do not expect the scheme to produce an unchangeable layer, since this layer has to reflect short-term user interests and the change in concept rank indicates emergence of a new user model. This model is called *interest-change-driven model*, since a new profile generation/update is caused not by some predefined settings, such as days, hours, etc., but by the dynamics of model generation itself (concept rank change). For instance, if we choose *three* most frequent of r most recent concepts ($m < = r$) in Figure 11 to be in the short-term layer, we can see that concept rank change, and accordingly a new model definition, occurs after item 94, 131 and so on are processed. The most highly ranked concept C5 keeps its top position since the user keeps working on plugin implementation for Firefox browser. The third highly ranked concept changes often from interests in news to travel and conference-related topics. See Table 1 for simple explanation of concepts presented as a number of terms representative of the concept topics.

Figure 12 shows the generation process of the long-term profile layer which was being constructed alongside with the short-term layer. Again m concepts with the

Table 1 Five discriminative keywords for concepts in Figure 11 and 12 [72]

C5	C1	C8	C7	C6	C3	C4
Firefox	real estate	software	Japanese news	Conference	Massage	travel
customize	square meters	hardware	company	Workshop	Tokyo	tour
plugin	Area	computer	business	Mobile	Stress	Malta
colored scrollbar	Rent	forum	politics	multi-media	traditional	ticket
browser	Floor	talks	finance	IR	Nerves	hotel

Figure 12 Long-term profile layer generation [72].
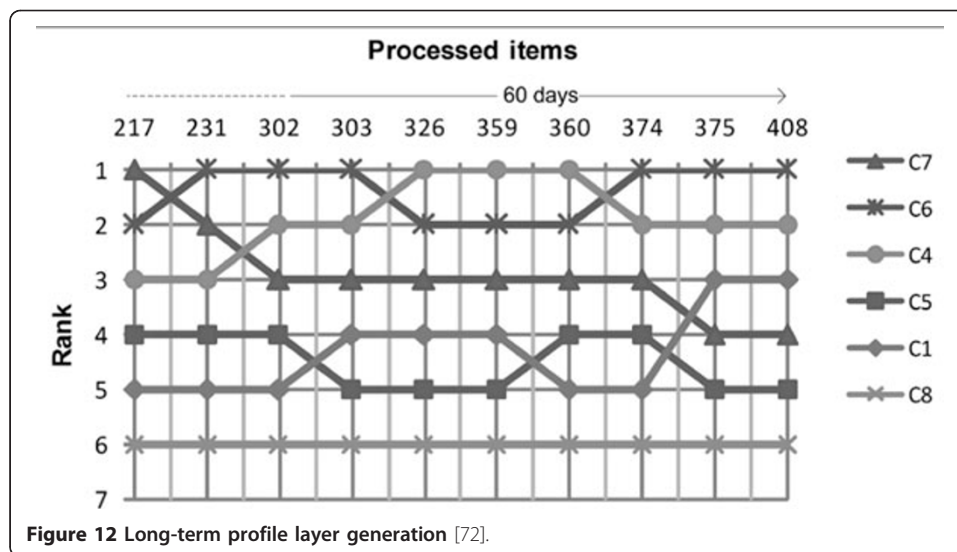
highest *persistency factor PF* (with α = 0.1) are chosen to be the long-term layer. But, in contrast to the short-layer concepts, they are not ranked and only the fact of their belonging or not belonging to the top *m* concepts is important for the definition of the layer. The concept ranks in Figure 12 are shown to indicate the change dynamics of the concepts' *PF* value and their less frequent changeability in comparison to the short-term layer's concepts. In other words, if we choose *three* concepts with the highest *PF* to be the long-term layer's concepts, the layer remains the same from the formation time (after item 217 is processed) - $Pr_a^{(ln)}$ = {C6, C7, C4}. It changes only after concept C1 gains *PF* value that is higher than the value of C7 - $Pr_a^{(ln)}$ = {C6, C4, C1}. *Persistency factor* ensures that only those concepts that have a tendency for being long-term interests gain higher value. For instance, concept C1 is ranked highest in the short-term layer and there is no evidence that this is not a long-lasting temporary interest and the user will return to it in future, therefore its *PF* value is not one of the highest. However, if such evidence will be available in future when the user returns to interests reflected by C1, its *PF* value will increase and the chances to be a concept of the long-term profile layer will grow.

As it can be seen from the figures and their explanations, the constructed layers meet our expectations and requirements to reflect a user's current interests for the session profile layer, to be a representation of both recent and frequent interests (i.e., the recent and vivid interests lasting for some time) for the short-term layer, and to collect persistent interests for the long-term layer.

5. On Sharing and Search

The collaborative spirit and deeper user-centeredness of the proposed approach reveals through information sharing and search we present in this section. Here we show how the dynamics of interest-change-driven user profile, its contextual information and its construction process are employed for contribution co-evaluation and search. A reader will be able to see the central role of user profiles we described in Section 4 and understand how information search and sharing are done in BESS.

5.1. Contribution Co-evaluation

One of the main characteristics of the proposed framework is community-empowered user-centeredness realized through users' contributions. Together with implicit relevance feedback they are used for the users' interest inference and formation of dynamically changing user profiles. In contrast to implicit feedback, which does not give a clear understanding of a user's interests before it is analyzed, contributions are clear indicators of the user's interests and can be used for promotion of contributed documents with regard to his/her expertise to all community members of the information sharing and search system. In BESS, contributions and the user's expertise drawn from his/her user profile are used for co-evaluation of information resources the search is done on.

In order to 'promote' documents contributed by a particular user a, first of all, his/her user profile, and specifically concepts in it, has to be known. Clearly, not all concepts can be used for various reasons - an involvement of too many concepts into the computation process is costly, or many concepts can be considered unsuitable in terms of the current user context and situation. Therefore, we consider concepts of the short-term layer of UP as a highly dynamic and, at the same time, rather consistent layer reflecting most of the recent and significant concepts of user a. As we discussed in Section 4, a change of the concepts in terms of their ranking in the short-term profile layer signifies a change of user interests and emergence of a new model of user a. The short-time profile layer generation process serves as an important factor for feedback value definition mechanism. In other words, its change-driven nature helps to determine the period of time for amount of data used to evaluate user contributions - to determine the concepts that will be used for contribution assessments.

The layer generation process is illustrated in Figure 13. Note that concept re-ranking is done according to the criteria described in Section 4.

i - number of interest changes. G_i - group of dynamic profile concepts. A, B, C, D, E, F - interest concepts. $\{A, B,..., F\} \in G_i$

T_i - moment of concept change in profiles. G and T - current concept group and moment respectively.

Then, the period to be used for contribution assessments can be defined as

$$P(T_s, \Delta T) = \begin{cases} P(T_{i-1}, T - T_{i-1}), & G = G_i \\ P(T_i, T - T_i), & G = G_{i+1} \neq G_i \end{cases} \tag{5.1}$$

where i is the number of interest changes, T_s - start time of evaluation period, ΔT - the interval to the current position on profile generation time line, and G and T are the current concept group and moment respectively.

For instance, we define the evaluation period with the last interest changing point at T_2 as

$$P(T_s, \Delta T) = \begin{cases} P(T_1, T - T_1), & G = G_2 \\ P(T_2, T_3 - T_2), & G = G_3 \neq G_2 \end{cases}$$
$$= \begin{cases} P(T_1, \Delta T_2 + \Delta T'_2), & G = G_2 \\ P(T_2, \Delta T_3), & G = G_3 \neq G_2 \end{cases} \tag{5.2}$$
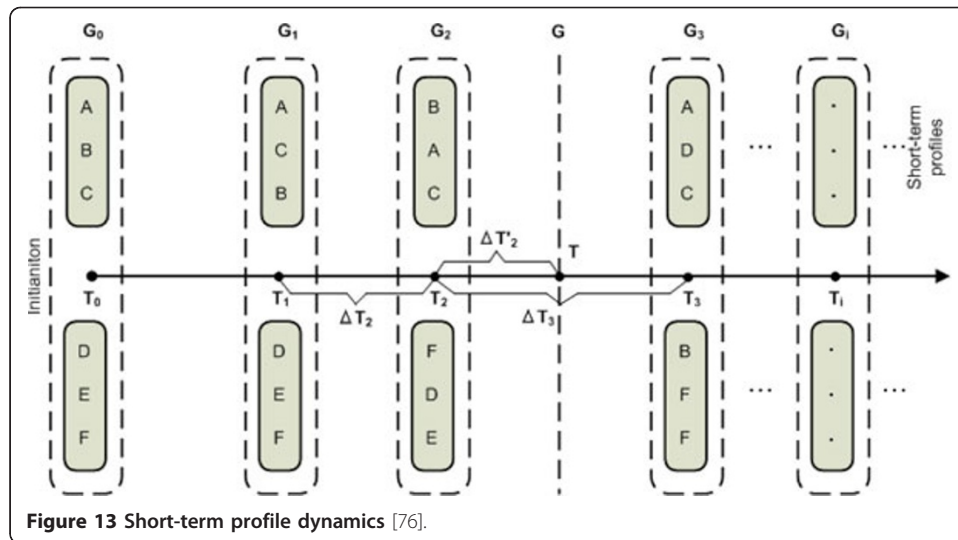
Figure 13 Short-term profile dynamics [76].

After the period is defined, a contribution can be evaluated according to the following criteria with regard to the available concepts of this particular user and all concept space of the system - concepts of all users interacting with the system:

- contribution activeness;
- contribution popularity.

Contribution activeness (*CA*) is defined as the ratio of the contribution number of user *a* to the number of the most active contributor with regard to concept *i*. To rephrase, it is an indicator of how active user *a* is in a particular area of expertise compared to the most active

$$CA_{aCi} = \frac{|C_{ai}|}{\max |C_{ui}|} \tag{5.3}$$

Contribution popularity (*CP*) is defined as the ratio of the sum of all contributions to concept *i* to all contributions done in the framework.

$$CP_{Ci} = \frac{\sum_{u=1}^{U} |C_{ui}|}{\sum_{i=1}^{I} |C_i|} \tag{5.4}$$

where $C_i \in C = \{C_1,..., C_I\}$ and $C_{ui} \in C_u = \{C_{u1},..., C_{uI}\}$.

Using the two criteria, the contribution of user *a* to document *d* which belongs to concept *i* is estimated as

$$cntr_{aCi} = \alpha \cdot CA_{aCi} \cdot CP_{Ci} \tag{5.5}$$

where α is a constant regulating the power of different contribution types - for instance, if implicit feedback is considered as a contribution, its estimated value has to be weakened because it is less indicative of the user's interests compared to explicit feedback.

Since a document often includes several concepts, we have to consider them in the estimation as

$$cntr_{aCi} = \frac{\sum\limits_{Cai} \alpha \cdot CA_{Cai} \cdot CP_{Ci}}{|C_d|} \qquad (5.6)$$

where $|C_d|$ is the number of (top n) concepts d belongs to.

Evaluating contributions in this way we give an assessment of a user's expertise by his/her involvement degree (CA) and potential value the user brings to the whole community by his/her contribution (CP). More specifically, these criteria take into account:

- contributions by user a to a particular concept;
- contributions by all users to a particular concept;
- potential value of contributions with regard to a particular concept;
- activeness of the whole community with regard to a particular concept.

As the same contribution can be done by multiple users, its score is re-assessed according to the following formula

$$cntr_d = \frac{\sum\limits_{u=1} cntr_d}{|U_d|} \qquad (5.7)$$

and stored for document ranking in retrieval process described next.

5.2. Collaborative Search

In Section 3 we mentioned that besides subjective search the framework provides objective search functionality - searching through conventional Web search services. Objective search is the functionality each of us uses daily and, therefore, not of much interest in our discussion. On the other hand, subjective search of BESS is collaborative and includes two modes using a custom rank function as shown in Table 2.

Both modes are labeled as collaborative. Search in normal mode is collaborative in the sense it is performed on the collection of contributed and shared documents and co-evaluations of the documents are considered for search result re-ranking. Focused search is even more collaborative since it detects search space from subjective document collections of users with similar contexts.

5.2.1. Normal Subjective Search

Normal subjective search mode is the basic mode to search on subjective index data of BESS. When using the mode, all index data becomes the search target. Furthermore, in contrast to objective search where query-document match is done regardless of a user's search context, BESS provides research results ranked with regard to the user's degree of involvement in the query context using conceptual information in user profiles.

When retrieving documents from subjective index repositories, first a simple impersonal query-document match, as in objective index search, is done to find documents objectively relevant to the user's query. Then the system attempts to personalize the retrieved results by considering the user's search contexts found in UP using the following formula and rank them with maximum sim_L values at the top.

Table 2 Subjective search modes

Mode	Normal	Focused
Ranking	Custom Rank	Custom Rank

$$sim_L = \sum_{k=1}^{K} \alpha_k \cdot sim(d, l_k) \tag{5.8}$$

where L - user profile, l_k - layer k of UP, d - document, and α_k - ratio in the mixture which is set experimentally and $\sum_{k=1}^{K} \alpha_k = 1$. Normally, a_{ss} has to be considered as the minimum and a_{ln} the maximum values, since they represent the most volatile and persistent concepts respectively.

5.2.2. Subjective Concept-directed Vertical Search

Another method to search in BESS is concept-based *subjective concept-directed vertical search*, or *focused search*. It is realized by matching of concepts of the individual user profile with concepts of other users when retrieving through the subjective search engine. This operation forms the target search document space for the query by finding users with similar interests and reaching their subjective index repositories. After the search document space is determined relevant documents are retrieved by comparing query and document vector similarity.

By forming the search document space, we change the focus of search, similarly to what occurs in vertical search engines, but automatically detecting users with similar information seeking contexts. In this way we ensure the search will be done on highly selective documents evaluated by the users with similar interests taking into account their expertise, or the degree of their involvement into the topic. That is, by drawing upon community expertise and similarity of information needs, we perform *subjective concept-directed vertical search*.

Figure 14 shows how the search is done in the system, presenting it on the level of concepts user profile consists of.

Users with similar profiles, and hence their subjective index collections, are found by comparing reference points of concepts in each layer of UP one by one. The following formula is used for the estimation.

$$\sum_{k=1}^{K} sim_L(l_{ak}, l_{bk}) = \sum_{k=1}^{K} \frac{\sum_{j=1}^{J}\sum_{i=1}^{I} sim(Rp_j, Rp_i)}{JI} \alpha_k \tag{6.9}$$

When n top users with similar contexts are found, their collections form the search document space. Then, the search is done as in normal subjective search mode.

5.2.3. Re-ranking with Contribution Assessments

Regardless a search mode, re-ranking of retrieved document is done by listing them with maximum sim_L values at the top. But this is not the only value considered for re-ranking. Another value is a contribution assessment value $cntr_d$ described in Section 5.1. Re-ranking is done by sorting according to the following rules:

1) documents with $cntr_d$ values exceeding a certain contribution threshold and similarity threshold (0.5 and 0.6 respectively in Table 3) are considered as high priority in ranking and sorted by $cntr_d$ field;

2) all other documents are sorted by sim_L field.

Figure 14 Search by concepts [65].
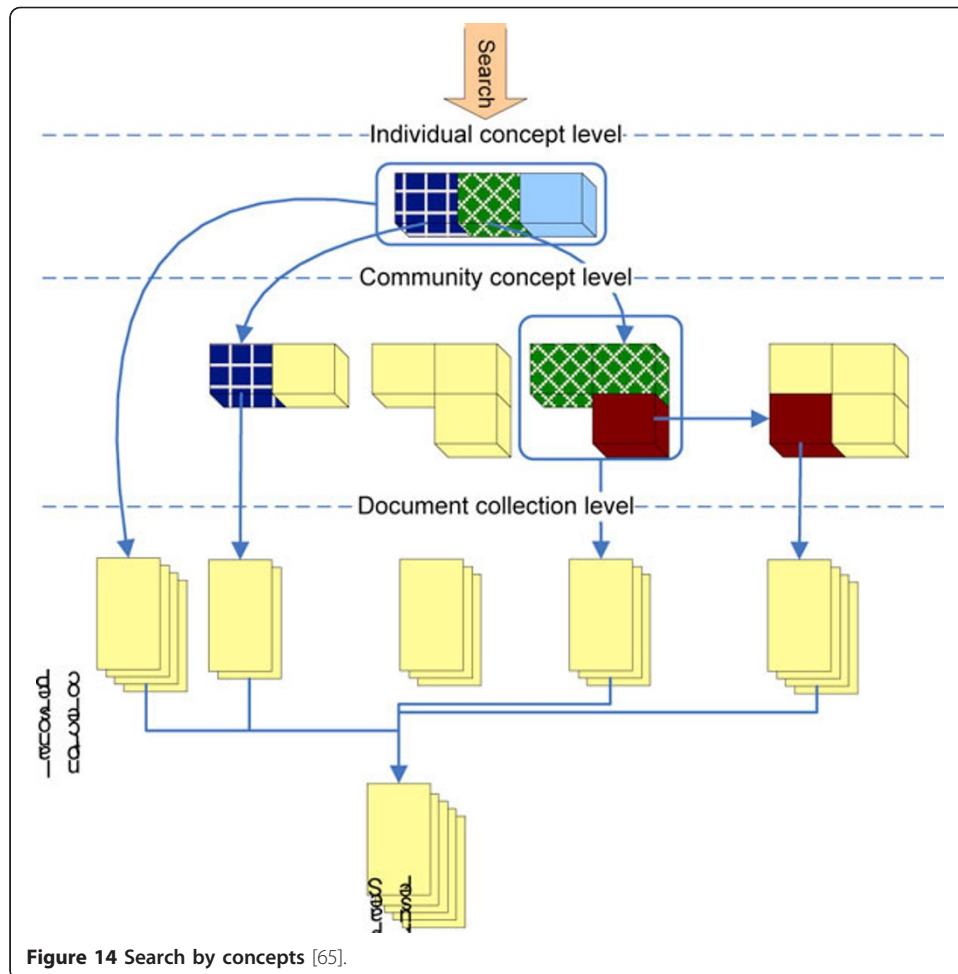
5.3. Search Scenario

To give a better idea how, for instance, subjective concept-directed vertical search works, let us consider several users of the system having the following rather abstract concepts generated from their contributions, and search and post-search behaviors.

- User A: cars, politics, golf;
- User B: action movies, Disney cartoons, blogging;
- User C: online shopping, cars, travel;
- User D: computer games, messenger.

Let us assume that *user A* searches for the pages explaining air-conditioning in Mitsubishi cars and submits "Mitsubishi air conditioner" query. The user does not know

Table 3 Re-ranking example [75]

Document	sim_L	$cntr_d$
d_1	0.69	0.65
d_2	0.81	0.59
d_3	0.56	0.35
d_4	0.42	0.21
d_5	0.3	0.52
d_6	0.3	0.51

that Mitsubishi Electric produces room air-conditioning appliances, but knows about Mitsubishi cars and is interesting particularly in this topic. As a result of the query, conventional search engine ranks high the documents covering room air-conditioning systems (which are unrelated to the user's search intents), whereas the proposed system will restrict results to document collections of users with similar interests and rank them according to user evaluations if available. In this particular case, *user A*'s interests match those of *user C* and the query retrieves documents explaining air-conditioning in Mitsubishi cars (Figure 15).

6. Conclusions

6.1. Summary

With the exponential growth of information on the Web and, as a result, failures to manage and process it effectively and efficiently, solutions beyond conventional information organization and filtering are being sought. Realizing the subjective nature of information overload and witnessing the fast proliferation of user-generated media, academia and business turn to the human more and more often today - human problems are proposed to be solved directly by humans, rather than being mediated by information systems. Algorithms and systems to facilitate information access and acquisition are still very important and their role cannot be diminished, however they have to get better understanding whom they are used by in order to do their job with the expected efficacy and efficiency - that is to say, they have to be *user-centric*. Having the complete knowledge of each particular individual might be considered sufficient to call a system user-centric, but, in our understanding, such a definition of user-centrism is one-sided and lacks a concept of community as an important part of human context. An approach considering an individual user only will not benefit from "wisdom of crowds" [22] of other users, loose much predictive power it can draw upon other users' experiences, and will not be able to collect all contextual information necessary for complete understanding of the user. Thus, an essential part of user-centrism is
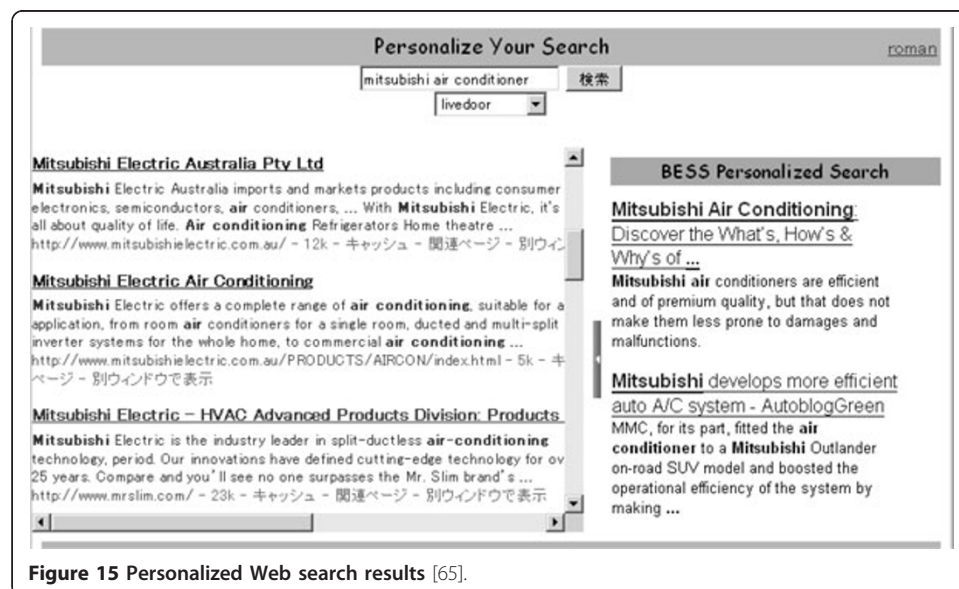

Figure 15 Personalized Web search results [65].

considering a user not only in his/her individual scope, but expanding it to the user's community participation quintessence.

Recognizing this, we designed and implemented BESS (BEtter Search and Sharing) as a highly collaborative information search and sharing system. In addition to learning individual interests of each particular user, BESS harnesses collective knowledge of its users who share their personal experiences and benefit from experiences of others. We made an endeavor to develop a holistic approach of how to harnesses relevance feedback (both explicit and implicit) from users in order to estimate their interests, construct user profiles reflecting those interests to applying them for information acquisition in online collaborative information seeking context. To emphasize the collaborative nature of relevance feedback submitted by users explicitly, it is called a *contribution* in our research. In comparison to implicit feedback, contributions are better indicators of a user's interests and therefore have bigger values for shared information assessment.

The central part in the system is allotted to user profiles claimed to contain fragmentary user contexts and used for contribution evaluation and search. Each user profile consists of a set of concepts representing user interests and inferred with *High Similarity Sequence Data-Driven clustering* method thoroughly discussed in [72], which is proven to be easy to implement and produces concepts of reasonably good quality fast and in online and incremental fashion. To ensure that a user profile is always updated and reflects current, recent and long-term interests, it is designed multi-layered and changing with change of the user's interests as a result of estimations done with *interest-change-driven profile construction mechanism*. The mechanism adopts *recency*, *frequency* and *persistency* as the three important criteria for profile construction and update.

As mentioned, user profiles are important for both relevance feedback evaluation and search. And they are especially important to realize a unique mechanism of personalized search - *subjective concept-directed vertical search*, or *focused search*. It is realized by matching of concepts of the individual user profile with concepts of other users when retrieving from the subjective search engine.

Furthermore, in order to distinguish between impersonal data of conventional search systems and individualized data of the members searching and sharing in the proposed system, we introduced the notions of *subjective* and *objective* index in IR system. Such distinction aims to enrich users' experiences with both conventional and personalized search results.

All these enables tight coupling of information sharing and search in the personalization cycle of *contribution submission through search* → *user modeling from contributions* → *contribution assessment based on the user's expertise inferred from user models* → *search result ranking based on the user's individual and community expertise (inferred from user models)* → *(new) contribution submission through search and so on* (Figure 16), which is, together with separation of subjective value-added resources from impersonal objective data of conventional search services (creation of 'social' subspace for subjective information accumulation and sharing through search) is the major difference of our approach from many modern conventional personalization solutions on the conceptual level.

Figure 16 Search and Sharing Integration in BESS [76].
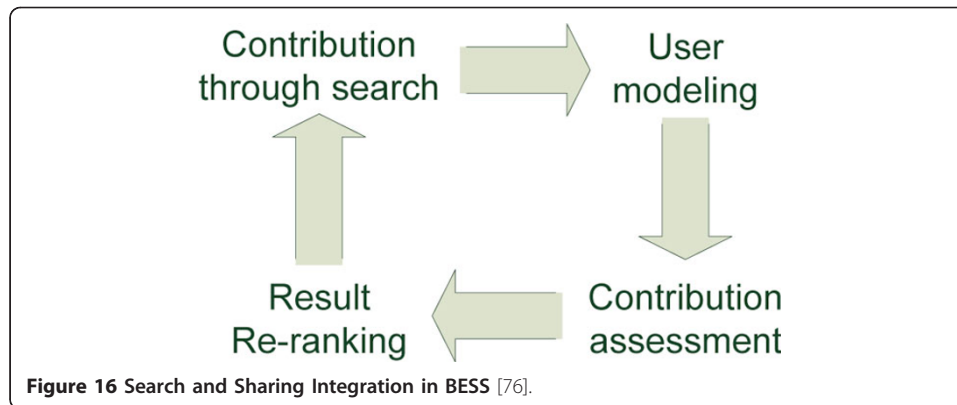
6.2. Future Research Directions

In order to achieve highly user-centric experience, the discussed research approach puts user profiles in the center of search and sharing. Profiles contain concept-based information about users inferred from relevance feedback and designed to dynamically reflect user contexts. However, contexts handled in this research (as in all today's research) are of fragmentary nature, reflecting only the part of user interests. Recognizing the importance of contextual knowledge about a user for personalization, additional methods to define, capture and organize contextual information effectively, extension of the architecture and algorithms to include more information from a user and assessments of effects of such inclusion on personalized experience are one of the future directions of our research.

Another direction is a closer examination of motivational factor for relevance feedback in social bookmarking and tagging, and search as a means to collect more precise information about a user.

References
1. Beaudoin CE (2008) Explaining the Relationship between Internet Use and Interpersonal Trust: Taking into Account Motivation and Information Overload. Journal of Computer-Mediated Communication 13(3):550–568
2. Hunter GL, Goebel DJ (2008) Salespersons' Information Overload: Scale Development and Validation and its Relationship to Salesperson Job Satisfaction and Performance. Journal of Personal Selling and Sales Management 28(1):21–35
3. Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review 63:81–97
4. Twitter blog, Measuring Tweets.http://blog.twitter.com/2010/02/measuring-tweets.html. Accessed 21 August 2011
5. Levy DM (2005) To Grow in Wisdom: Vannevar Bush, Information Overload, and the Life of Leisure. Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital libraries 281–286
6. Himma KE (2007) The concept of information overload: A preliminary step in understanding the nature of a harmful information-related condition. Ethics and Information Technology 9(4):259–272
7. Mulder I, de Poot H, Verwij C, Janssen R, Bijlsma M (2006) An information overload study: using design methods for understanding. Proceedings of the 2006 Australasian Computer-Human Interaction Conference (OZCHI 2006) 245–252
8. Chen YC, Shang RA, Kao CY (2007) The Effects of Information Overload on the Outcomes of On-Line Consumption Behavior. Proceedings of International Conference on Wireless Communications, Networking and Mobile Computing, 2007 (WiCom 2007) 3791–3794
9. Kim K, Lustria MLA, Burke D, Kwon N (2007) Predictors of cancer information overload: Findings from a national survey. Information Research 12(4)

10. Klausegger C, Sinkovics RR, Zou HJ (2007) Information overload: a cross-national investigation of influence factors and effects. Marketing Intelligence & Planning 25(7):691–718

11. Reuters Ltd (1996) Dying for information: an investigation into the effects of information overload in the USA and worldwide. London: Reuters Limited

12. Pereira R, da Silva SRP (2008) The Use of Cognitive Authority for Information Retrieval in Folksonomy Based Systems. Proceedings of the Eighth International Conference on Web Engineering (ICWE '08) 325–331

13. Stickel C, Holzinger A, Ebner M (2008) Useful Oblivion Versus Information Overload in e-Learning: Examples in the context of Wiki Systems. Proceedings of the ITI 2008 30th International Conference on Information Technology Interfaces 171–176

14. Griseé ML, Gallupe RB (1999) Information Overload: Addressing the Productivity Paradox in Face-to-Face Electronic Meetings. Journal of Management Information Systems 16(3):157–185

15. Janssen R, de Poot H (2006) Information overload: Why some people seem to suffer more than others. Proceedings of the 4th Nordic Conference on Human-Computer Interaction 397–400

16. Wikipedia, User-generated content.http://en.wikipedia.org/wiki/User-generated_content. Accessed 21 August 2011

17. Nov O (2007) What Motivates Wikipedians? Communications of the ACM 50(11):60–64

18. Shtykh RY, Jin Q, Nakadate S, Kandou N, Hayata T, Ma J (2008) Mobile SNS from the Perspective of Human Self-Extension. In Handbook of Research on Mobile Multimedia, vol XLVIII., 0th edn.IGI Global

19. Nielsen//NetRatings (2006) User-Generated Content Drives Half of U.S. Top 10 Fastest Growing Web Brands, According to Nielsen//NetRatings. http://www.nielsen-netratings.com/pr/PR_060810.PDF. Accessed 21 August 2011

20. Blodget H (2010) Twitter Is Adding A Spectacular 370,000 New Users A Day – But Mostly Outside The US. http://www.businessinsider.com/twitter-users. Accessed 21 August 2011

21. eMarketer, The State of Social Networking in Western Europe.http://www.emarketer.com/mobile/article_m.aspx?R=1008257. Accessed 21 August 2011

22. Surowiecki J (2005) The Wisdom of Crowds, Anchor; Reprint edition.

23. JupiterResearch (2008) US Online Travel Consumer Survey, 2008". http://www.jupiterresearch.com. Accessed 21 August 2011

24. Obrist M, Beck E, Kepplinger S, Bernhaupt R, Tscheligi M (2008) Local Communities: Back to Life (Live) through IPTV. Changing Television Environments, Lecture Notes in Computer Science 5066:148–157

25. Golbeck J (2008) Weaving a Web of Trust. Science 321(5896):1640–1641

26. Ingwersen P, Jarvelin K (2005) The Turn: Integration of Information Seeking and Retrieval in Context (The Information Retrieval Series). Springer-Verlag New York, Inc., Secaucus, NJ

27. Spink A, Ed, Cole C, Ed (2005) New Directions in Cognitive Information Retrieval, Springer.

28. Landay JA, Kaufmann TR (1993) User Interface Issues in Mobile Computing. Proceedings of the Fourth Workshop on Workstation Operating Systems 40–47

29. Sweeney S, Crestani F (2006) Effective search results summary size and device screen size: is there a relationship? Information Processing and Management 42, 4:1056–1074

30. Shtykh RY, Jin Q (2008) Improving Mobile Web Search Experience with Slide-Film Interface. Proceedings of SITIS2008/KARE2008 (First International Workshop on Knowledge Acquisition, Reuse and Evaluation, in conjunction with the Fourth IEEE International Conference on Signal-Image Technology & Internet Based Systems) 659–654

31. De Luca EW, Nürnberger A (2005) Supporting information retrieval on mobile devices. Proceedings of the 7th international conference on Human computer interaction with mobile devices & services 347–348

32. Church K, Keane MT, Smyth B (2005) An Evaluation of Gisting in Mobile Search. In: Losada DE, Fernández-Luna JM (eds) ECIR 2005, LNCS 3408, 546–548

33. Shtykh RY, Chen J, Jin Q (2008) Slide-Film Interface: Overcoming Small Screen Limitations in Mobile Web Search. In: MacDonald C et al (ed) ECIR 2008, LNCS 4956, 622–626

34. McKenzie B, Cockburn A (2001) An empirical analysis of web page revisitation. Proceedings of the 34th Annual Hawaii International Conference on System Sciences 128–137

35. Shtykh RY, Jin Q (2006) Design of Bookmark-Based Information Space to Support Exploration and Rediscovery. Proceedings of CIT2006 30

36. O'Brien R (2007) The next thing after 2.0! http://www.outofrhythm.com/2007/04/14/the-next-thing-after-20/. Accessed 21 August 2011

37. Mitra S (2007) Web 3.0 = (4C + P + VS). http://www.sramanamitra.com/2007/02/14/web-30-4c-p-vs/. Accessed 21 August 2011

38. Hansen MT, Nohria N, Tierney T (1999) What's Your Strategy for Managing Knowledge? Harvard Business Review 106–116

39. Greer TH, Murtaza MB (2003) Web Personalization: The Impact of Perceived Innovation Characteristics on the Intention to Use Personalization. Journal of Computer Information Systems 43(3):50–55

40. Monk AF, Blom JO (2007) A theory of personalization of appearance: quantitative evaluation of qualitatively derived data. Behaviour and Information Technology 26(3):237–246

41. Fan H, Poole MS (2006) What is personalization? Perspectives on the design and implementation of personalization in information systems. Journal of Organizing Computing and Electronic Commerce 16(3):179–202

42. Guida A, Tardieu H (2005) Is personalization a way to operationalise long-term working memory? Current Psychology Letters 1:1–17

43. Ruthven I, Lalmas M (2003) A survey on the use of relevance feedback for information access systems. The Knowledge Engineering Review 18(2):95–145

44. Kelly D, Teevan J (2003) Implicit Feedback for Inferring User Preference: a Bibliography ACM SIGIR Forum. 37(2):18–28

45. Noll MG, Meinel C (2008) Web Search Personalization Via Social Bookmarking and Tagging. The Semantic Web, Lecture Notes in Computer Science 4825:367–380

46. Harper DJ, Kelly D (2006) Contextual Relevance Feedback. Proceedings of the 1st International Conference on Information Interaction in Context 129–137

47. Limbu DK, Connor A, Pears R, MacDonell S (2006) Contextual Relevance Feedback in Web Information Retrieval. Proceedings of the 1st International Conference on Information Interaction in Context 138–143

48. Teevan J, Dumais ST, Horvitz E (2005) Personalizing Search via Automated Analysis of Interests and Activities. Proceedings of the 28th Annual ACM Conference on Research and Development in Information Retrieval (SIGIR '05) 449–456

49. Carrillo-Ramos A, Villanova-Oliver M, Gensel J, Martin H (2007) Profiling Nomadic Users Considering Preferences and Context of Use. On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops, Lecture Notes in Computer Science 4805:457–466

50. Wang J, Pouwelse J, Fokker J, de Vries AP, Reinders MJT (2008) Personalization on a peer-to-peer television system. Multimedia Tools and Applications 36:89–113

51. Fu X (2007) Evaluating Sources of Implicit Feedback in Web Searches. Proceedings of ACM Recommender Systems 2007 (RecSys '07) 191–194

52. Santally MI, Alain S (2006) Personalisation in Web-Based Learning Environments. International Journal of Distance Education Technologies 4(4):15–35

53. Shapira B, Hanani U, Raveh A, Shoval P (1997) Information filtering: a new two-phase model using stereotypic user profiling. Journal of Intelligent Information Systems 8(2):155–165

54. Sobecki J (2004) Hybrid adaptation of web-based systems user interfaces. Computational Science - ICCS 2004, Lecture Notes in Computer Science 3038:505–512

55. Ligon GL, Balachandran MB, Sharma D (2006) Personalization of Web Search: An Agent Based Approach. Knowledge-Based Intelligent Information and Engineering Systems, Lecture Notes in Computer Science 4253:1192–1200

56. Bahrami A, Yuan J, Smart PR, Shadbolt NR (2007) Context-Aware Information Retrieval for Enhanced Situation Awareness. Military Communications Conference (MILCOM 2007) 1–6

57. Barbu C, Simina M (2003) A probabilistic information filtering using the profile dynamics. IEEE International Conference on Systems, Man and Cybernetics 5:4595–4600

58. Ferscha A, Hechinger M, Riener A, Schmitzberger H, Franz M, dos Santos Rocha M, Zeidler A (2006) Context-aware profiles. International Conference on Autonomic and Autonomous Systems 48

59. Gargi U (2005) Information Navigation Profiles for Mediation and Adaptation. Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05) 515–520

60. Sieg A, Mobasher B, Burke R (2004) Inferring User's Information Context: Integrating User Profiles and Concept Hierarchies. Presented at the 2004 Meeting of the International Federation of Classification Societies

61. Ferscha A, Hechinger M, Riener A, Schmitzberger H, Franz M, dos Santos Rocha M, Zeidler A (2006) Context-aware profiles. International Conference on Autonomic and Autonomous Systems 48

62. Case DO (2002) Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior. Amsterdam: Academic Press

63. Kagolovsky Y, Moehr JR (2003) Current Status of the Evaluation of Information Retrieval. Journal of Medical Systems 27(5):409–424

64. Shtykh RY, Jin Q (2008) Capturing User Contexts: Dynamic Profiling for Information Seeking Tasks. Proceedings of I-CENTRIC 2008 (International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services) 365–370

65. Shtykh RY, Jin Q (2008) Harnessing user contributions and dynamic profiling to better satisfy individual information search needs. Int J Web and Grid Services 4(1):63–79

66. Smeaton AF, Callan J (2005) Personalization and recommender systems in digital libraries. International Journal on Digital Libraries 5(4):299–308

67. Nichols DM (1997) Implicit ratings and filtering. Proceedings of the 5th DELOS Workshop on Filtering and Collaborative Filtering 31–36

68. Pitkow J, Schütze H, Cass T, Cooley R, Turnbull D, Edmonds A, Adar E, Breuel T (2002) Personalized search. Communications of the ACM 45(9):50–55

69. Matthijs N, Radlinski F (2011) Personalizing web search using long term browsing history. WSDM '11: Proceedings of the fourth ACM international conference on Web search and data mining

70. Sugiyama K, Hatano K, Yoshikawa M (2004) Adaptive web search based on user profile constructed without any effort from users. Proceedings of 13th international conference on World Wide Web, ACM Press 675–684

71. Almeida RB, Almeida VAF (2004) A Community-Aware Search Engine. Proceedings of the 13th International Conference on World Wide Web, ACM Press 413–421

72. Shtykh RY, Jin Q (2009) Dynamically constructing user profiles with similarity-based online incremental clustering. International Journal of Advanced Intelligence Paradigms 1(4):377–397

73. Nanas N, Uren V, De Roeck A (2003) Building and applying a concept hierarchy representation of a user profile. Proceedings of the ACM Conference on Research and Development in Information Retrieval (SIGIR) 198–204

74. Semeraro G, Lops P, Degemmis M (2005) Personalization for the Web: Learning User Preferences from Text. In: Hemmje M, Niederée C, Risse T (eds) Lecture Notes in Computer Science, vol 3379. From Integrated Publication and Information Systems to Virtual Information and Knowledge Environments, pp 162–172

75. Fellbaum C (1998) WordNet: An Electronic Lexical Database, MIT Press.

76. Shtykh RY, Jin Q (2009) Integrating Search and Sharing: User-Centric Collaborative Information Seeking. Proceedings of Eighth IEEE/ACIS International Conference on Computer and Information Science 388–393