Harnessing user contributions and dynamic profiling to better satisfy individual information search needs

Roman Y. Shtykh* and Qun Jin*

Networked Information Systems Laboratory
Faculty of Human Sciences
Waseda University, Japan
E-mail: roman@akane.waseda.jp
E-mail: jin@waseda.jp
*Corresponding authors

Abstract: In the situation of information overload we are experiencing today, conventional web search systems taking a one-size-fits-all approach are often not capable of effectively satisfy individual information needs. To improve the quality of web information retrieval, we propose a collaborative personalised search approach that makes an attempt to ‘understand’ and better satisfy the information needs for each and every searching user. We present a web information retrieval framework called Better Search and Sharing (BESS) that captures user-system interactions, profiles them and induces personal interests that changes over time with an interest-change-driven profiling mechanism that is also extensively used for the co-evaluation of documents found valuable inside a specific search context by users with similar interests.

Keywords: collaborative personalised search; relevance feedback; dynamic user profiling; subjective index; subjective concept-directed vertical search.

Reference to this paper should be made as follows: Shtykh, R.Y. and Jin, Q. (2008) ‘Harnessing user contributions and dynamic profiling to better satisfy individual information search needs’, Int. J. Web and Grid Services, Vol. 4, No. 1, pp.63–79.

Biographical notes: Roman Y. Shtykh is a Research Associate at Media Network Center, Waseda University, Japan, and a PhD candidate at the Graduate School of Human Sciences, Waseda University. He received his MSc in Computer Science and Engineering from the University of Aizu, Japan. His research interests include data mining, web search personalisation and mobile human-computer interaction.

Qun Jin is a Professor of Networked Information Systems at the Department of Human Informatics and Cognitive Sciences, the Faculty of Human Sciences, Waseda University, Japan. He has been engaged extensively in research works on computer science, information systems and internet computing. His recent research interests include cognitive information retrieval, human-computer interaction, human-centric ubiquitous services computing and computing for well-being. He seeks to exploit the rich interdependence between theory and practice in his works with interdisciplinary and integrated approaches. He is a member of Association for Computing Machinery (ACM) and Institute of Electrical and Electronics Engineers, Inc. (IEEE). See www.f.waseda.jp/jin/ for additional information.

Copyright © 2008 Inderscience Enterprises Ltd.
1 Introduction

The highly dynamic nature of the World Wide Web (the web) in terms of its constantly increasing size, the great diversity of web contents, the hyperlink structure variability and the changeability of users’ search needs (Ke et al., 2006) poses a tremendous challenge for Information Retrieval (IR) on the web. Nowadays, in the era of massive information overload, no longer is web search considered solely as a match of terms contained in a query and document collections – it is a blend of techniques used to enhance the existing IR approaches. Moreover, more and more attention has recently been focused upon an individual person as a creator and active evaluator of the shared web information. In other words, IR on the web is becoming more and more user-centric. This is greatly influenced by the growth of the popularity of web-based social networks that have showed again that there are no algorithms or a whole system capable to make decisions better than humans (Surowiecki, 2005).

Until recently, a number of techniques to improve the search experience and provide a better match of available information assets with users’ information needs were introduced. Among them, the web document categorisation (Open Directory Project,1 Yahoo! Directory2), information filtering (Belkin and Croft, 1992) and the clustering of search results (Vivisimo,3 Clusty4) may be named as those that have enjoyed considerable attention from the early days of web IR. Niche (or vertical, focused, specialised) search has gained increasing popularity over recent years. By gathering information into specialised document collections, it helps users to focus more easily on the information from a specific domain.

The described approaches increase the quality level of web information access and retrieval compared to conventional search methods, but they are not capable of coping efficiently enough with the tasks predetermined by rapidly changing individual information needs within constantly growing web information assets on their own. To better answer such information needs, IR systems must individualise the interaction with every user, i.e., abandon the approach to treat all alike.

To be user-centric, a web search system ought to understand a user’s search intents. Capturing search and post-search behaviour can possibly help to further understand search intents; the various forms of relevance feedback integrated into the web search system are being widely used recently. Relevance feedback, both explicit and implicit, is generally used in web IR research for the efficient collection of users’ behavioural data for further user behaviour analysis and modelling. Implicit relevance feedback in IR systems consists of a number of elements, such as a query history, a clickthrough history, the time spent on a certain page or a domain and others, that can be considered in general as a collection of the implicit behaviours of users interacting with the IR system. It is conducted without interruption of user activities, unlike an explicit one that requires direct user interferences; that is why many are showing keen interest in it. As an example, the Russian Internet advertisement company Begun5 states that behavioural advertisements based on watching user actions on certain websites and search portals bring about the same amount of profit as semantic advertisements. Begun creates user profiles that are later updated according to the user actions on specified websites. The main sources of information for profile formation are the user search queries leading to the advertised sites, the user routes from one site or page to another and the history of interactions with certain advertisements. Moreover, users are organised into groups and a
Harnessing user contributions and dynamic profiling

part of the prediction algorithm uses profile similarities to fill the gaps in other users’ profiles and do correct predictions. A similar research is being conducted by Microsoft AdCenter Labs.\textsuperscript{5} Taking the probabilistic approach, they are building user profiles based on page views, searches and other online behaviours for targeted advertisements. Furthermore, such profiles are clustered and segments of customers with similar interests are created.

In this paper, we shall briefly discuss the latest approaches to the personalisation of the web search and introduce an experimental approach to satisfy the individual information needs with the relevance judgements done by users interacting with a document collection and personalisation by multilayered profiles that change dynamically, together with the changes in users’ search intents inside a collaborative search and sharing system called BEtter Search and Sharing (BESS).

Here, we shall also discuss the concept of the subjective index (we refer to the index of conventional IR systems as the objective index) – a personal collection of textual indexed documents found valuable by users during their search for a particular piece of information and collected through the web search framework of BESS. Such a collection is especially valuable in the context of a particular search the user conducted, hence, the user’s analysed search and post-search behaviours or other information about the user’s search intentions must be kept together for further IR. For this, we propose the use of user profiles that capture and reflect the user’s information needs as closely as possible and can be used to influence upon document ranking, do query augmentation and discuss the concept-based interest-change-driven profile generation mechanism used for profile generation and subjective index data evaluation. Furthermore, we introduce the subjective concept-directed vertical search that reveals the collaborative nature of the proposed approach.

The rest of the paper is organised as follows. In Section 2, we review the related works. In Section 3, we define the ‘subjective index’ notion in the context of the collaborative personalised search. Section 4 describes the proposed system and discusses the profile modelling approach and profile-based contribution and retrieval. In Section 5, we touch upon system implementation issues and, finally, the conclusion and future work are given in Section 6.

2 Adapting the IR computational environment to every user

2.1 Recent web search personalisation approaches and the position of BESS among them

Reconsidering IR 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 personalisation have arisen recently. Nowadays, we are experiencing the much-anticipated breakthrough in personalised search efficiency by “actively adapting the computational environment – for each and every user – at each point of computation” (Pitkow \textit{et al.}, 2002).

To show the peculiarities of the existing web search personalisation systems and the position of BESS inside the web search personalisation approaches, we classify them as vertical and horizontal, individual-oriented and community-oriented based on the breadth of the search focus and the degree of collaborativeness they possess (see Figure 1; arrows denote the current trends in search personalisation).
Outride (Pitkow et al., 2002) and similar systems take a contextual computing approach trying to understand the information consumption patterns of each user and then providing better search results through query augmentation. Sugiyama et al. (2004) experiments with a collaborative approach, constructing user profiles based on collaborative filtering to adapt the search results according to each user’s information need. Almeida and Almeida (2004) harness the power of community to devise a novel ranking technique by combining the content-based and community-based evidences using the Bayesian Belief Networks. The approach shows good results, outperforming conventional content-based ranking techniques. Systems like Swicki, Rollyo and the Google Custom Search Engine correspond to the vertical and mostly community-oriented approach of search personalisation. They provide a community-oriented personalised web search by allowing communities to create personalised search engines around specific community interests. Unlike the horizontal (or broad-based) search systems mentioned above, such systems are considered personalised 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 the search. Although not web-based, we take tools like the Google Desktop Search as an example of individual-oriented vertical search systems. They search the contents of files, such as e-mails, text documents, audio and video files, etc., inside a personal computer. The absence (to our knowledge) of salient web-based systems of this kind can be explained by the increasing popularity of services on the web benefiting from community collaboration and favouring the fast transition of each person’s activities, from passive browsing to active participation.

As one can see from Figure 1, BESS is a community-oriented system having the features of both a horizontal and vertical search system. It performs searches on the information assets of both horizontal (objective index) and vertical (subjective index)
Harnessing user contributions and dynamic profiling

nature. The notion of the subjective index in our research is similar to the ‘social search’ of the vertical community-oriented systems presented above, but differ in the higher degree of personalisation for every user, the high granularity of the vertical search model (see subjective concept-directed vertical search in Subsection 4.2.2) and, finally, the way of collecting and (re)-evaluating the information pieces. Groups of users are created dynamically without a user’s interference based on the match of interests/expertise and the role of the community is indispensable for search quality improvement (see Section 4) and the system’s evolution in general.

2.2 User behaviour analysis as an important factor of web search personalisation

Relevance feedback is an essential element of any information filtering system and a significant part of the proposed system. It is exhaustively 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 and Teevan (2003) give a good classification and overview of the works on implicit feedback. Generally, user behaviour is considered to be an implicit feedback and its analysis is done for improving IR by predicting user preferences, re-ranking the web search results and disambiguating the queries (Agichtein et al., 2006a–b; Shen et al., 2005a).

Often, relevance feedback is used in an attempt to find out user behavioural patterns and generate individual user profiles reflecting current user interests. There exist many approaches for profile modelling. Nanas et al. (2003) discusses profiles made from concept hierarchies that are generated from user-specified documents and applied for information filtering. Profiles are divided into three layers by heuristic threshold, each of which determine the topic, subtopics and subvocabulary for the specified topic. The term weighting approach (Relative Document Frequency) is extensively used for hierarchy construction. Semeraro et al. (2005) uses a different approach for profile construction. As in Nanas et al. (2003), 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 (Fellbaum, 1998). Such an approach is prove to help infer more accurate user profiles. The user profile, defined in Koutrika and Ioannidis (2005), “treats query disambiguation and personalisation as a uniform term rewriting process”. Profile, a directed graph representing term relations, dictates the modifications of search queries.

BESS makes extensive use of both explicit and implicit relevance feedback for the construction of personal information assets and user profiles. Unlike the profiles in the above-mentioned approaches, the profiles in BESS are constructed with the main focus on the users’ interest change when searching and the concepts are loosely coupled and mobile (see Section 4).

3 Subjective index and collaborative personalised search

As we mentioned in the introduction, the role of an individual user in information creation and evaluation for further reuse is rapidly increasing. Such user activities are particularly productive and valuable inside communities with similar information
needs. In other words, collaboration is an important factor to increase the value of the information and the efficiency of its creation and reuse. The search experience we are trying to provide can be characterised as collaborative and personalised. Users’ searches and contributions have a personalised nature and the information pieces found valuable by every user in the context of his/her current information needs are shared among all users.

For a collaborative search, we distinguish two types of indexes: objective and subjective. The objective index is the indexed data of conventional web search engines. The subjective index (or social index) is created based on the information found valuable in the context of a specific information need and submitted by users (we specify the process of such a submission as a contribution) within the BESS web framework and implicitly evaluated by dynamically constructed user profiles. Collecting such personal information pieces gives us access only to highly selective information tied to a specific information need – without such a relation preserved, this information is not much different from that stored in conventional search systems. To understand the user’s search intents, we are profiling his/her contributions (explicit feedback) and interactions with the system (implicit feedback). Having the profiles, we can figure out the user’s expertise on a specific topic and evaluate the contributions accordingly. Furthermore, we find users with similar expertise or interests to collaboratively evaluate the contributions and improve the personalised search.

Literally, contributions are explicit feedback actions. Although explicit feedback can disrupt user search activities, it is important for subjective index creation and explicit measures in IR tasks are found to be more accurate than the implicit ones (Nichols, 1997).

Since we are interested in the results of the subjective search engine, BESS does not handle objective index data directly. Instead, it queries a conventional web search engine selected by a user to display its results in a separate frame and make it possible to browse and contribute valuable pages to the proposed collaborative search system.

4 User behaviour profiling for web search personalisation

To realise the collaborative personalised search presented in the previous section, we have built a BESS prototype system (Figure 2) that is capable of collecting and analysing user feedback, building profiles with interest-change-driven profiling in the centre, evaluating contributions and finally, creating subjective index data. The system captures authenticated users’ behaviours (see Subsection 4.1.1) in the ‘access control and data collection’ module. The captured data are analysed by the ‘data analyser’ and used as material for profile generation (see Subsection 4.1.3). Ontologies are optional here and can be used to achieve a more accurate profile analysis. The user contributions are analysed and evaluated on arrival based on the individual and community contribution statistics (see Subsection 4.2) and the information from user profiles, indexed and saved into the ‘subjective data’ repository to be searched on later. Besides searching on the subjective index, the system conducts searches on the objective index through major conventional search engines (‘conventional search engine’ in Figure 2).
4.1 Modelling user profiles

4.1.1 Data collection

As we have already mentioned in Section 3, profiles play one of the central roles in this collaborative search system. The building blocks of the user profiles are minimal user search and post-search behaviours, *i.e.*, user-system interactions while searching, browsing and contributing webpages captured by the system (Figure 2). Such behaviours give us ‘raw data’ for further basic analysis and profile generation. We are trying to use the minimal number of such elements and apply them to our profiling model to find out the methods for achieving the best match of user search (or possibly contribution) intents with documents that are subjectively indexed.

We use the following basic data for profile generation:

- User ID used for authentication
- Search query terms
- The Uniform Resource Locator (URL) of the page the user is interacting with
- Type: query, click or feedback
- Timestamp
- Session ID.
4.1.2 Concept as a principal profile component

The user profile can be defined as a structured representation of user preferences, interests and information needs. In the proposed system, it consists of concepts (semantic clusters) and each concept is the system’s piece of ‘knowledge’ about what the user is interested in, modelled as a cluster $c_i$ of $n$ document vectors $X = (\bar{x}_1, ..., \bar{x}_n)$ from the individual document set grouped by a specific ‘knowledge’ criteria.

Prior to concept extraction, the documents from the individual document collections are linearised by removing the HyperText Markup Language (HTML) and script tag data, non-content-bearing ‘stopwords’ are deleted and the document vectors are normalised to have a unit $L^2$ norm. Since the document term vectors are highly dimensional, we reduce their dimensionality by feature selection (Tang et al., 2005) to increase the clustering speed without losing the important content-bearing words. Then, concepts are computed using the Intelligent K-Means algorithm or iK-Means (Mirkin, 2005). iK-Means uses the anomalous pattern method as a procedure to meaningfully determine the number of initial seeds for K-Means that performs a disjoint partitioning of the document vectors and computes a centroid for each partition. Normalised centroids contain valuable semantic information about the partitions (concept), therefore, they are treated as representative components (vectors) of the concepts.

We applied the approach to the data produced by users interacting with a web social bookmarking service located in Japan. Social bookmarks were considered as explicit relevance feedbacks and used as a corpus for concept formation in the experiment. The following are examples of the sample concepts obtained in the experiment and their representative terms.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Concepts produced by simulation on web social bookmarking service data (terms are translated to English, since the original corpus is in Japanese)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept A</td>
<td>Concept B</td>
</tr>
<tr>
<td>Diary</td>
<td>Hyogo (region)</td>
</tr>
<tr>
<td>Book</td>
<td>Brand</td>
</tr>
<tr>
<td>Reader</td>
<td>Shop</td>
</tr>
<tr>
<td>Category</td>
<td>Confectionery</td>
</tr>
<tr>
<td></td>
<td>Author</td>
</tr>
</tbody>
</table>

4.1.3 Dynamic interest-change-driven profile generation

To distinguish the user interests by topic, we introduced concepts in the previous subsection. But user interests can differ not only topically. As it can be easily seen, they can differ temporally – some interests spark and decay and some stay for a lifetime. To introduce the temporal dimension into the user profiles proposed here, we divided them into four layers – static $p_{st}^a$, session $p_{ss}^a$, short-term $p_{sh}^a$ and long-term $p_{ln}^a$ (see Figure 3). Therefore, the profile of user $a$ can be defined as:

$$P_a = (p_{st}^a, p_{ss}^a, p_{sh}^a, p_{ln}^a).$$

(1)

A layer consists of concepts:

$$p_{st}^a = (X_1, ..., X_r).$$

(2)
where:

\[ l = \text{layer} \]
\[ k = \text{concept number}. \]

**Figure 3** User profile layers

The static layer is needed only at the beginning of the user-system interaction to know the basic preferences of a user when no other information is available. The latter three layers are also called the *dynamic layer group* that is generated after enough feedback data are accumulated and analysed. They are updated when a user is interacting with the proposed system, reflecting always-changing user information needs. The outer levels of the dynamic layer group change frequently by update, whereas the inner levels are rather inert and the analysed information must reach a specified threshold to modify them. All layers taken together influence upon the contribution evaluation, search and ranking of the resulting document set each by its fraction in common weight vector (boost value):

\[
W = (w_1, \ldots, w_n)
\]

(3)
given \(n\) profile layers.

After the users specify their interests at the beginning of participation in the proposed system, the static layer is formed and used until there are enough data for dynamic profile generation. The long-term dynamic one is generated in the process of user-system interactions, as will be explained later in this section. Both layers are interchangeable and play the same role in the profile model, reflecting the general long-time background and interests of the users. They both have a minimum of influence on every query in the search process or feedback in the contribution evaluation process. When used together, the two layers both have the proper fractions of the whole weight value common to both layers. As the data needed for dynamic profile generation grows, the static layer fraction in the influence on the query or feedback decreases.

The short-term layer is central to profile generation and the whole system. It consists of concepts (semantic clusters) generated within a specified period of time and its generation itself serves as an important factor for the feedback value definition mechanism described in Section 4.2. As shown in Figure 4, at the moment of user
participation initiation, it is substituted by a static profile layer, since there are no data for short-term layer generation. When data are accumulated and the predetermined threshold level for the dynamic layer is surpassed, this profile layer is generated and used henceforth. As users interact with the system, behaviour and feedback data grows and new concepts of the user interests/needs are detected. Concepts are grouped \( \{A, B, \ldots, F\} \in G_i \) and ranked. Those with a large number of quality terms and the highest document gain for a certain period of time (defined by the formula 4 in Section 4.2) form the short-term profile. When changes in the concept of group members or their rankings occur, a new short-term profile is defined.

**Figure 4** Short-term profile generation

The long-term profile layer is derived from the most frequent concepts in the short-term profile layer, starting from those concepts available at the moment of participation initiation to the concepts of the short-term layer included into the current concept group \( G \).

Finally, the session profile layer reflects the most recent information needs and have the highest degree of influence upon the current contribution evaluation, search and ranking. The session can be defined in several ways. The simplest way is to define it by network session abstraction or a single interaction with the service. For the current implementation, we employ the following method to specify its boundaries – a new session is detected on a change of interests defined as a change of search context. The concepts for this layer are derived from all terms in the current session. It is a lower concept level compared to the short-term profile level.
4.2 Profile-based contribution and retrieval

Contribution evaluation

As it was mentioned in Section 4.1, the short-time profile layer generation process serves as an important factor for the feedback value definition mechanism. In other words, its change-driven nature helps to determine the period of time not only for profile generation, but also for the amount of data used to evaluate the user contributions together with the user profiles. The period can be defined as:

\[
P(T_i, \Delta T) = \begin{cases} 
P(T_{i+1}, T - T_{i+1}), & G = G_i \\ P(T_i, T - T), & G = G_{i+1} \neq G_i. \end{cases}
\]  

(4)

where \(i\) is the number of interest changes, \(T_s\) – start time of evaluation period, \(\Delta T\) – the interval to the current position on the profile generation timeline, 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:

\[
P(T_{i+1}, \Delta T) = \begin{cases} 
P(T_{i+1}, T - T_{i+1}), & G = G_i \\ P(T, T - T), & G = G_{i+1} \neq G_i. \end{cases}
\]  

(5)

After the period is defined, we use profile concepts and the following data derived from the profiles and the collected basic behaviour data (see Subsection 4.1.1) within the period to determine the values of user feedback:

- overall number of contributions
- number of contributions by a particular concept
- overall value of the contributions
- value of the contributions by a particular concept
- identity of the contributions made by other users and their value
- evaluation of the importance of the contribution in a particular search context by other users.

These data show how much the user is involved in the topic the contributed document belongs to within a determined period of time by comparing the concepts possessed by the user and other users with similar interests and determining his/her role within the evaluation of documents of a particular topic. The values given to the document become its weight vector \(W\) (boost value) used together with the original feature vector of the document:

\[
F = (f_1, \ldots, f_m)
\]  

(6)

to rank retrieved documents. That is, the feature vector is modified to:

\[
F' = (f_1, \ldots, f_m, w_{m+1}, \ldots, w_{m+n})
\]  

(7)

to perform community-based personalised ranking.
4.2.2 Retrieval by concepts

When retrieving through the subjective search engine, the matching of concepts of the individual user profile with the concepts of the other users is done at first. 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, the relevant documents are retrieved by comparing query and document vector similarity (in cases when the vector space model of IR is employed).

By forming the search document space, we change the focus of the search, similar to what occurs in vertical search engines, but automatically detecting users with similar information needs. We ensure that 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 in the topic. That is, by drawing upon the community expertise and the similarity of information needs, we perform the subjective concept-directed vertical search.

The similarity between two concept-based profiles is computed as:

\[ \text{sim}(\mathbf{P}_u, \mathbf{P}_v) = \mathbf{P}_u \cdot \mathbf{P}_v, \]

with concept vectors normalised in each.

Figure 5 can give a better picture of how the search is done in the system, presenting it on the level of the concepts the user profile consists of, and described in Subsection 4.1.2.

Figure 5 Conceptual search
5 System implementation and use scenario

On a large scale, the proposed system consists of a contribution software component on the client side and the BESS collaborative IR framework, the architecture of which was introduced in Section 4.

The user contribution component is implemented as a Firefox browser extension (Figure 6). Other methods, such as contributing with the use of HTML elements placed into search result pages next to every result, can be considered.

Figure 6 Contribution widget

BESS, implemented as a server-side solution and composed of numerous components such as the web proxy, the user behaviour analyser, the profile construction engine and others, performs the rest of the job – it captures user search and post-search behaviours, analyses them together with the contributions to gather the statistics used for contribution evaluation, generates and updates profiles and does searches on the subjective index data.

In addition, the searches on objective index data from conventional web IR systems is done to show the generic search results along with the ‘recommended’ results, as opposed to other search personalisation systems like the one introduced by Google, the Google Personalized Search. Generic search results are also important for the proposed system, since they are the source for enriching subjective index collections through contributions. Figure 7 shows the generic results arranged in the left pane of the browser windows and the personalised search results in the right pane.

Figure 7 Personalised web search results
In constructing the system, we made use of the following technologies. For the contribution submission Firefox component development, we used Asynchronous JavaScript and XML (AJAX) and XML User Interface Language (XUL). To develop the server-side part, we employed a set of Java technologies like JavaServer Pages (JSP), the Java Servlet, the Spring Framework and others. For web document preprocessing, we used the CyberNeko HTML Parser and the basic functionalities of the Apache Lucene for document indexing and search.

### 5.1 Use scenario

To give a better idea how the 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 behaviours.

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

Let us assume that user A searches for pages explaining airconditioning in Mitsubishi cars and submits ‘Mitsubishi air conditioner’. The user does not know that Mitsubishi Electric produces airconditioning systems and has Mitsubishi cars on his/her mind. As a result of the query, the conventional search engine ranks as high the documents covering room airconditioning systems, whereas the proposed system will restrict results to the document collections of users with similar interests and rank them according to user evaluations, if available. In the case we are considering, user A’s interests matches those of user C and the query retrieves the documents explaining airconditioning in Mitsubishi cars (Figure 7).

### 6 Conclusion and future work

In the situation of information overload on the web and with the tremendously increasing popularity of web community services that lay a particular emphasis on an individual as a content producer and evaluator, considerable attention is given to web search personalisation.

In this paper, we propose a collaborative information search and sharing web system such as BESS that tries to better satisfy the search needs of every user within a particular search context. In the system’s design, we tackle the important personalisation problem posed in ‘Future Directions’ of Pitkow et al. (2002) – modelling a user’s changing interests over time – by providing interest-change-driven profiling for the construction of multilayered user profiles capturing and reflecting all the dynamics of change in the information needs of a specific user and serving as an important factor for the evaluation of the users’ contributions that form their subjective index. User profiles are constructed from the relevance feedback data collected by the proposed collaborative search and sharing framework. Such feedback can be purely implicit so that a user is not distracted from browsing and search activities. User profiles are used for the creation of subjective
index data – a personal collection of textual indexed documents found valuable by users during their search for a particular piece of information and evaluated based on the degree of each user’s involvement in a particular topic. Such an evaluation is done with community assessments in mind that emphasises the collaborative nature of subjective index creation in particular and information sharing in general. The search in the system is defined by introducing the notion of the subjective concept-directed vertical search that performs a profile concept match of users in the system and automatically focuses the search on the personal collections of users from the same interest domain.

We have implemented the major modules needed for full system functioning and integration work must be done in future to present the whole fully functional system. Future works also include performance and usability evaluation and the improvements of the algorithms. Additionally, we are considering investigating and implementing closer collaboration to have a more strongly expressed character of its ‘social’ aspect by the creation of groups with similar interests/needs. This can be done by the concept matching of the profile layers of every user.

The proposed system has the following ‘drawback’. Although it refocuses the search to provide results matching the individual information needs better than conventional search engines, it can suffer from the absence or a small number of contribution data if there are few users participating in the system. That is, even if the system directs a search to the subjective data matching the current information need of a particular user, the data can be not enough or outdated. However, this situation will not happen if the service is popular.

Any web personalisation system requires storing and analysing private information. This is seen to be a problem by privacy advocates (Privacy International), since there is never enough security to protect personal information. Client-side personalisation solutions (Shen et al., 2005b) can alleviate privacy concerns to some extent, but they are not feasible for community-oriented systems like ours. We have to consider how to ensure users’ privacy, probably by combining the client-side and server-side approaches for storing and processing user-sensitive information.

Acknowledgements

The work has been partly supported by the 2007 Waseda University Grants for Special Research Project No. 2007B-223 and No. 2007B-224.

References


Notes
Harnessing user contributions and dynamic profiling

19 Privacy International, ‘A race to the bottom – privacy ranking of internet service companies’,
   http://www.privacyinternational.org/.

Websites