Hybrid Profiling in Information Retrieval

Mandeep Pannu, Rachid Anane,Anne James
Faculty of Engineering and Computing, Coventry University,
Coventry,UK
{m.pannu, r.anane, a.james}@coventry.ac.uk

***Abstract* — One of the main challenges in search engine quality of service is how to satisfy the needs and the interests of individualusers. This raises the fundamental issue of how to identify and select the information that is relevant to a specific user. This concern over generic provision and the lack of search precision have provided the impetus for the research into Web Search personalisation. In this paper a hybrid user profiling system is proposed – a combination of explicit and implicit user profiles for improving the web search effectiveness in terms of precision and recall. The proposed system is content-based and implements the Vector Space Model. Experimental results, supported by significance tests, indicate that the system offers better precision and recall in comparison to traditional search engines.**

Keywords-Hybrid user profile, explicit, implicit, web search personalisation, Vector Space Model

I. INTRODUCTION

As the sources of information on the web and the number of web users are increasing, improvement to the quality of search results has become a crucial issue. The searching techniques used by search engines tend to retrieve both relevant and irrelevant information. As a result, there is a demand for advanced solutions for acquiring the information that meets users' needs [1]. In order to be able to provide documents with the information a user is searching for, there is a critical need to understand how people use the web, how they search for the information, and what techniques they are using to find documents that are relevant to them. The relation between a user query and web pages is problematical and it is driving the research in the field of information retrieval. Users have a variety of needs and the retrieval systems are often unable to offer the solution to fulfil the requirements of an individual user [2]. The potential mismatch between the user interests and the query interpretation by a search engine may have an adverse effect on the user experience – the reliance on keywords only can result in low quality of matches [3]. The linguistic implications ofkeywords are a major reason for the low retrieval accuracy [4]. One word may refer to multiple concepts, e.g. the word 'mission' may refer to an assignment, a group of people, or an organisation. Search engine results are based on average trends rather than needs of a single user as there are often not able to track the behaviour of individual users. Researchers have introduced and classified various schemes for web personalisation [5]. The personalised filtering process starts with individual users, their preferences and the generation of their profiles.

Two approaches are considered particularly useful for generating user profiles – explicit and implicit user profiling. In the explicit approach users create their profiles manually or provide some kind of feedback to a search system, while in implicit approach the system creates profiles based on observed search history and browsing behaviour.

The two approaches have formed the basis of many systems with mixed success. Used in isolation the explicit method may be accurate but intrusive; the implicit method on the other hand may be transparent to the user but less focused. This work is motivated by the need to overcome the inherent limitations of each method and to take advantage of their positive features in order to improve the search process. A hybrid profiling approach is proposed in this paper.

In this approach, both explicit and implicit profiles are generated independently and then combined into a single profile. Implicit profiling occurs in the background while the user is carrying out the task. After an initial learning period the profile is set and will henceforth help the user in the specific task in which he or she is engaged. The profiling should be associated with the task so that when the user next carries out the same task the relevant profile can be re-activated. The task-linked profile can update itself as often as required.

A system to support this approach has been implemented and evaluated. In the implemented prototype the determination of the similarity between user profiles and documents and the filtering process are realised with the Vector Space Model (VSM), in which a document and a profile are represented by term vectors. Each dimension of a term vector represents a single term (keyword) and the vector's value in that dimension determines the importance (weight) of that word.

The remainder of the paper is organised as follows. Section IIpresents related work on personalisation. Section III describes the proposed approach and the design and implementation of the system. Section IV gives an account of the experiments for evaluating the proposed approach in relation to traditional search engines. Section V offers a brief discussion, and Section VI concludes the paper.

II. WEB PERSONALISATION

The increasing amount of information and services available on the Web has a significant impact on users. The lack of user understanding of keyword search may have an adverse effect on the process of finding relevant results. An average user, with poorly chosen keywords, has to go through many returned documents to find the relevant ones. Although some customisation may help filter out irrelevant documents according to individual user preferences [6, 7], in search enginepersonalisation the focus of the results is on the users rather than on the submitted queries [8]. Research into profiling has been marked by the introduction of a variety of systems.

Syskill & Webert [9] is a system that makes recommendations of web pages based on theexplicit feedback of theuser. If a user has rated a hyperlink in a webpage, then the system recommends related pages that might be of interest to the user. Once a page is ranked as high, the system analyses the page content to learn about the information the user is interested in [10]. The system does not expose the user to new topics because it can only make recommendations based on the similarity to previously visited pages. If the user wants to change the area of interest then a new profile has to be created [5].

A number of methods have been used for implicit profile generation to improve the search results. Implicit generation requires observing user behaviour and capturing their search history [5, 11]. User actions that need to be observed include time spend on reading a web page, saving, printing, clicking, selecting text and bookmarking [12]. Aoidh et al. proposed a method of implicit profiling that involves capturing user mouse movements as well [13].

Lieberman [14] developed the Letizia system which creates implicit users profiles based on the analysis of the individual browsing behaviour. The system assumes that the user is interested in a document if it was saved or bookmarked, and that the user's interest is weak if the document was left without following the links inside the document. The system works by giving weight to the documents that were linked to the currently viewed document, and suggests linked documents that are similar and match the implicit profile. The system does not make use of any explicit data for the recommendations.

Gasparetti et al. [15] introduced a technique for building implicit user profiles with the help of the browsing history. Their algorithm relies heavily on the textual context of the links followed by users during browsing. One advantage of this technique is that it does not require any explicit user involvement.

Personalisation can be implicit or explicit. The creation of an explicit profile involves asking users for specific information in order to create an individual user profile. To learn about specific users needs, a large amount of information is required from users. The information regarding the interests of the user is usually gathered by collecting keywords or getting feedback on visited documents [16]. In general users are very reluctant to provide feedback [17]. Although this process is time consuming and increases the cognitive load on the users it can improve the search results and enhance the user experience [18].

In the explicit profile generation a user needs to directly provide the information in order to create an individual user profile. This approach may require pre-defined categorisation of user interests. Users may not be fully aware of their current and future needs. Furthermore, it is intrusive and can be time consuming and awkward for the user. It does offer however the user some direct control over the profiling process. The other approach – the implicit profile generation is transparent from the user point of view, but it is not trivial for an automated system to determine the relevance of a page that the user is viewing. The underlying assumption is that a user is expected to spend more time on relevant pages, and may wish to print or save them instead of merely reading them on-line. This assumption entails that sole reliance on the gathering of behavioural data during a browsing session may be open to different interpretations. The implicit method may not reflect accurately the current interests of the user or changes of their interests. Its main advantage however is that it is not intrusive. Changes in the interests or search area may not be reflected immediately in the results returned by the search engines with both explicit and implicit profiles; in general changes in explicit profile can be reflected in the search process within a short period.

III. PROPOSED APPROACH

This paper is concerned with the presentation of the architecture of a system based on hybrid user profiling, which combines explicit and implicit profiles. A hybrid system can enhance the flexibility of the profiling. The proposed system creates a context where user and system can collaborate in retrieving relevant documents. It has also the benefit of a clear identification of the factors that affect the search process.

A. Profile representation and filtering

The system makes use of the VSM for storing and combining the explicit and implicit user profiles, as well as for filtering the results. The VSM model is useful for effective information retrieval because weight values can be applied to each term in documents representations, the user query and the profile. With the normalisation of the vectors lengths, longer documents are not favoured over short ones, and because of the use of inverse document frequency vector, popular terms are not considered important while rare terms are promoted. In every case a user profile is represented in the VSM by a list of keywords with weights and stored as a term vector:

$$P = (<p_i, w_i>,\ldots,<p_i, w_i>\ldots, <p_n, w_n>)$$

A keyword is represented by p_i and its weight by w_i. The vector representation of a profile has the same representation for every kind of profile; however the way in which weights for each term are determined are different. A document in VSM is also represented as a term vector. Each word in a document is represented as a separate dimension of the vector.

$$D = (<d_i, w_i>, <d_i, w_i>\ldots, <d_n, w_n>)$$

For the purpose of the prototype, the vector representing a document is constructed from keywords that are extracted from the title and metadata of each document. VSM model can be applied to filter the documents by determining the degree of similarity between individual user profiles and each document representation. Each dimension of a vector represents a single word (keyword) and a weight value in that dimension determines the importance of that word. The similarity between a document and a query can be measured based on the weights of the corresponding terms. The cosine measure is used for this purpose. The cosine similarity function is given by the following formula:

$$Sim(D,P) = \frac{D \cdot P}{\|D\|\|P\|} = \frac{\sum_{i=1}^{m} d_i p_i}{\sqrt{\sum_{i=1}^{m} d_i^2}\sqrt{\sum_{i=1}^{m} p_i^2}}$$

$D = (d_i \ldots d_m)$ is the document vector and $P = (p_i \ldots p_m)$ is a profile vector. If vectors D and P are normalised then $||D|| = ||P|| = 1$ and the formula can be simplified to:

$$Sim\ (D,P)\ = \sum_{i=1}^{m} d_i p_i$$

The keywords that appear only in one of the two vectors are ignored (as the weight value for a keyword not present in a vector is equal to zero). For example, if the user profile P = (<science, 0.74>, <museum, 0.55>) – term “science” has a weight 0.74 and term “museum” has a weight 0.55, and all others terms weight will be consider as 0. For the document frequency vector D = (<museum, 0.82>, <history, 0.51>, <nature, 0.31>,) the similarity is equal to 0.55 • 0.82 (word 'museum') + 0 (other words from vector P not existing in vector D) + 0 (other words from vector D, not existing in vector P) which gives a similarity value of 0.451.

B. *Profile generation*

For the proposed prototype the creation of the explicit user profile is limited to asking the users to specify their interests in terms of keywords. All keywords are assumed

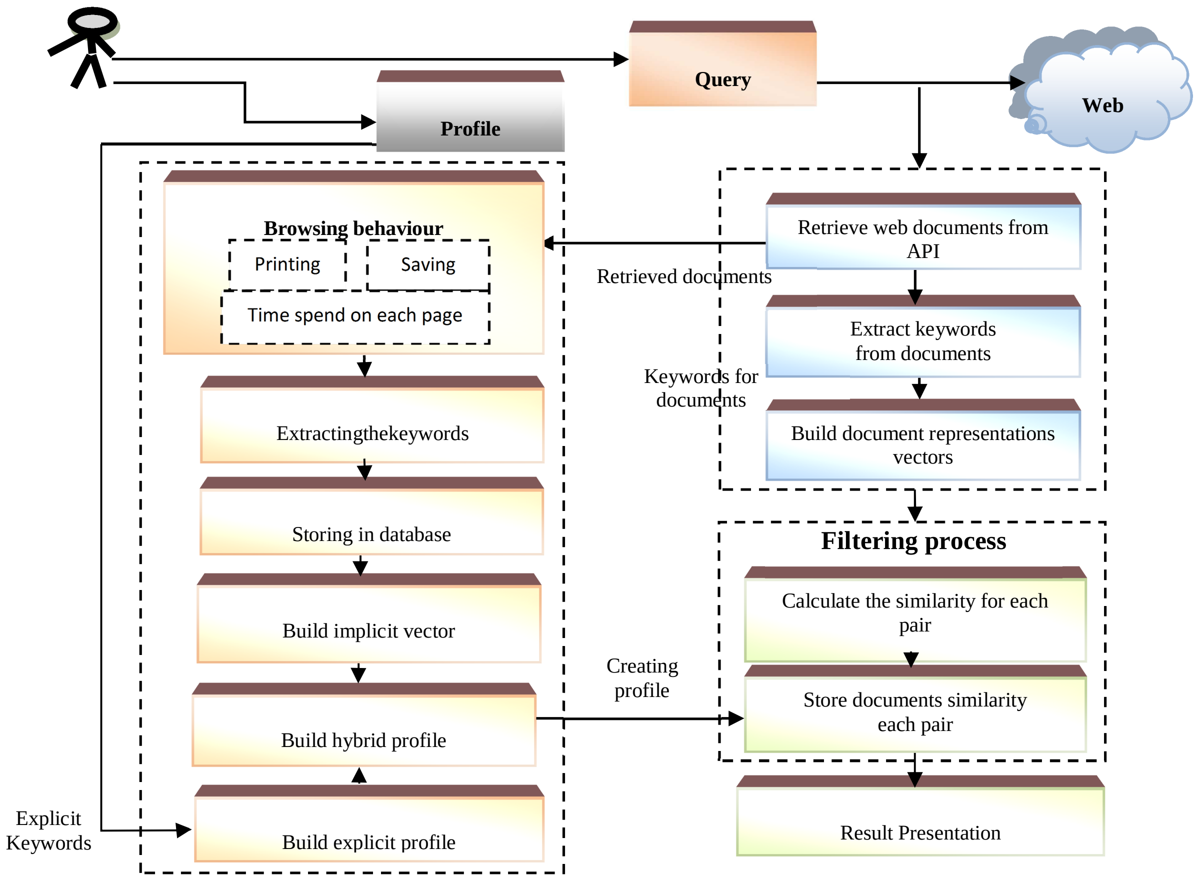

Figure 1: Hybrid system architecture

to be equally important and are given the same weight in the vector. The user profile can be modified later, at any time by adding, deleting or modifying keywords and therefore existing vectors. The explicit user profile is stored in a term vector for future use.

In order to create the implicit user profile, the system is constantly monitoring user's activities by storing the browsing history, the time spend on the each page, and additional actions like printing or saving.

The system extracts keywords from every visited document. In the prototype the keywords are extracted from the documents title and metadata (keywords and description). The keywords are given different weights depending on their position within the document – for instance keywords extracted from a document title are considered more important that those extracted from the description. The weight of a keyword is 0.5 if it is in the title, 0.3 if it is in the description and 0.2 if it is in the metadata.

In addition to the vector containing the keywords that describe the document content, the system also stores the activity type (whether the described document was viewed, printed or saved). Information about the time of the event is also stored – for activities such as printing or saving only the start time is provided, while for viewing both the start and end time are saved to allow for the calculation of the time for which a document was displayed. These values together with activity type are then used to calculate how important a document is, and therefore what weight should be applied to the representation of that document when the implicit vector is generated. After the collection of the information regarding the user browsing behaviour, the system is able to generate the implicit user profile in the form of keywords and weights.

Representations of documents that were opened for only a short time are ignored, while representations of documents that were saved or printed are considered especially important. The implicit profile vector is created by adding keywords from every included document representation after scaling them by the importance calculated for that document. After the summarised vector is created a number of keywords with the highest weights are used and returned as the implicit profile vector.

In the hybrid system the explicit profile and implicit profile are generated separately and combined into a single term vector. In the combination process vectors representing both explicit and implicit profiles are scaled, so that the weight of every explicitly entered keyword is equal to the highest weight of any keyword from the implicit profile, and then both vectors are added. If a keyword appears in both vectors, then its new weight is the sum of weights from both vectors. The combined vector is then normalised by the system and used for searching.

Figure 1 presents the essential components of the information filtering system for hybrid user profile. It incorporates the three main functions: explicit profiling and browsing behaviour, document retrieval and document filtering.

C. Implementation

The system utilises several components to perform a web search based on explicit profiling, implicit profiling and on hybrid user profile. In the explicit profile the users are required to explicitly specify their interests. In the implicit profile, the system generates it implicitly through the monitoring and the recording of the interaction of the user with documents. In the hybrid system, the explicit and implicit profiles are combined to improve the results of VSM filtering. Following is the pseudo code used for the implicit mode of operation.

```
1. When user is opening a webpage
a. calculate the keywords freq. vector
      i.    Read the title, keywords and
            description from the document
            metadata
     ii.    Scale each of the vectors by its
            importance
    iii.    Add terms from all these vectors to
            create one vector
     iv.    Remove keywords with lowest ranking
      v.    Normalise the vector
     vi.    Store the keywords in the database
2. When user is leaving a page
a. Store time of the visit in the database
3. When user is printing or saving
a. Store that event the database
```

The pseudo code responsible for creating the implicit user profile vector from stored information about the user behaviour is shown below:

```
1. For each document stored in the database
  a. Get the average time the user spend on each
             page from the database
b. For every document visited for longer
           than average
         i.      Get keywords (with weights)
c. For every other action (e.g. printing
                or saving)
         i.    Retrieve keywords associated
               with this action
        ii.    Add all retrieved keyword into
               one vector
                  a. Keyword weight is a sum of
                      weight from both  vectors
3. Normalise the vector
4.Return the vector so that it can now be used
for searching
```

To create the hybrid both explicit and implicit profiles are combined.

```
1.Get explicit keywords from the user
2.Create implicit user profile vector from the
browsing history
3.Create the combined vector
a. Get the highest keyword weight from the
     implicit vector
b. Scale the explicit vector by the highest
implicit weight (calculated in point a.)
c. Add explicit and implicit vectors
       i.      If  a  keyword  exist  in  both
               vectors, thenits new rating is a
               sum of rating from both vectors
4.Limit the number of keywords to ones with
highest weight
5.Normalise the vector
6.Return the normalised vector
```

At this stage a search with hybrid profile can be performed. Using the generated keywords, a number of documents are retrieved from a search engine API, followed by the application of the VSM to filter these documents with the hybrid profile.

D. User interface

The mediation system was developed as a stand-alone application, composed of a web browser, and user interface for searching. The user interface (Figure 2) displays the main functional components of the web browser, with additional facilities for searching. Users are identified inthe system by session names. In addition, there is an option for choosing a classical search engine, which is mediated by the application.

IV. EVALUATION

The effectiveness of the different systems was measured in terms of precision and recall. The experiment was conducted with respect to Yahoo! and Google web search APIs. In the experiment, the proposed hybrid system queries were submitted through both base web search APIs. The system then filtered the received results with the use of the hybrid profile.

The precision of a retrieval system for a given query is calculated as the number of relevant documents retrieved over the total number of retrieved documents. As a document can be classified as relevant, partially relevant or irrelevant, instead of using number of documents, a score (value) is assigned to each document as a reflection of the degree of its relevance. The precision (P)is then calculated as the total score assigned for all retrieved documents divided by the maximum score that would be given if all documents were fully relevant [19].

$$P = \frac{total\ score\ for\ relevant\ retrieved\ docs}{maximum\ score\ for\ all\ retrieved\ docs}$$

If most of the documents are assessed as irrelevant, then the precision is low, whereas if more documents match the expectations of the users then the precision is higher (for that particular query).

Recall (R) is the total score of all relevant document retrieved by a search engine over the total score for all relevant documents held in the database.

$$R = \frac{total\ score\ for\ relevant\ retrieved\ docs}{total\ score\ for\ all\ relevant\ docs}$$

Users should be able to view all relevant documents that may meet their information requirements. If the

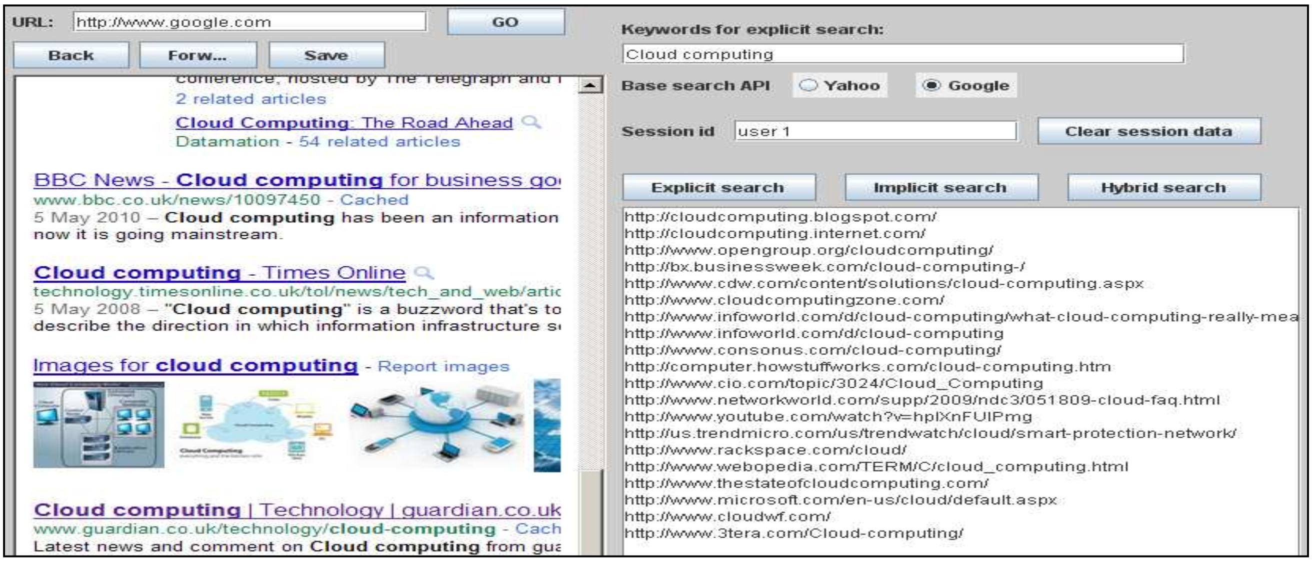

Figure 2. User Interface

relevance score from retrieved documents is close to the total score of all documents in the database, then the recall is high, otherwise it is low.

Recall is often nontrivial to measure because it is usually difficult to determine the number of relevant documents in the whole database. The issue is how to identify an acceptable pool of relevant documents. One approach is to combine all the relevant documents returned by more than one search engine and calculate the relative recall [19, 20]. Given that both systems are API-based, the score for all relevant documents for one system is:

$$R = \frac{total\ score\ for\ docs\ retrieved\ by\ evaluated\ system}{max.score\ for\ docs\ retrievd\ by\ both\ systems}$$

For example if two systems have to be compared – base API and hybrid system – then the hybrid system recall can be measured by dividing the total score for documents retrieved by the hybrid system by the total score for documents retrieved by both base APIs.

A. Experimenal methodology

The experiment was conducted in order to determine the effectiveness of the hybrid system in terms of precision and recall. For the mediated search the base API had to return 100 results, which were filtered by the implicit and the explicit systems to provide the highest rated 20 results for each system.

Users were instructed to select a set of keywords that represent their explicit profile and then to conduct the search process with different keywords throughout the experiment.

The experiment was performed with 30 users with their own choice of keywords as queries. Each user provided one set of keywords which gives a total number of 30 queries. To measure the system effectiveness the same evaluation was conducted first with the base APIs, Yahoo! and Google web search APIs and the results rated. For the implicit approach, the users were instructed to use a provided web browser for 15 minutes so that the browsing behaviour could be recorded in the database. The system recorded the time spent on each page and activities such as printing and saving of documents.

After the browsing session, users proceeded to enter the keywords for their explicit profile. Users were also instructed to use search keywords which were different from the keywords in the explicit profile. The same queries used in the implicit interaction were used to search again in the Yahoo! And Google web search APIs. During the evaluation the first 20 results from every search were taken into consideration.

In the hybrid mode of operation both implicit and explicit filtering were combined and the results rated. To ensure that all users are using the same scale of scores, they were presented with an indication on how to assess a page depending on whether it was relevant or not. Five categories were created to assess search results, these are “relevant”, “less relevant”, “irrelevant”, “links” and “no access” [19, 20]. Therating method in Figure 3 was provided to the user:

Category	Description	Score
Relevant	Related Conference paper, journal paper or web document fully related to the query	2
Less relevant	Document not fully concerned on to the query topic, but having the required information as part of its contents	1
URLs/Links	Page that provides a list of URLs where at least two URLs are redirecting to a page with the relevant information	0.5
Irrelevant	Documents totally irrelevant to the user intentions	0
No access	Web pages that for any reason cannot be accessed (e.g. ‘page not found error’).	Error (0)

Figure 3. Rating instructions

The results retrieved from the hybrid system and from the web search APIs were mixed together and presented to the user in a random order to ensure that the test was not affected by user’s opinion about any of the retrieval methods.

B. Experimental results

The experiment was conducted with 30 users. During every search each of the search systems returned 20 documents. The maximum allowed relevance score for a document is 2, which gives maximum total score for a search system of 1200. Figure 4 shows the score for the documents returned by base web search APIs and the hybrid system. The precision results for each API were calculated separately and then combined.

	APIs		**Hybrid system**	
Description	**Number of documents**	**Total score**	**Number of documents**	**Total score**
Relevant	453	906	528	1056
Less Relevant	212	212	218	218
URLs	131	65.5	137	68.5
Irrelevant	384	0	320	0
No access	20	0	5	0
Total	1200	1183.5	1200	1342
Precision	**0.49**		**0.56**	

Figure 4. Precision with base APIs and hybrid system

Similarly the hybrid system was tested once with each API and the results were combined. The precision of the base APIs is 0.49 and the precision of the hybrid system is 0.56.Figure 5presents the relative recall of the base search APIs and of the hybrid system. The 20 first results from each search API were combined as API results, and the results obtained with the hybrid system with both APIs were combined as hybrid search results.The relative recall was calculated for the two combinedsets.

Measurement	Base APIs	Hybrid System	Duplicated Docs
Document Score	989	1162.5	552.5
Recall	**0.62**	**0.73**	

Figure 5. Recall with base APIs and hybrid system

Figure 6 shows that the hybrid system has improved both the precision and the recall in relation to base search APIs. The hybrid system precision has been improved by 14% and the recall has been improved by over 17%.

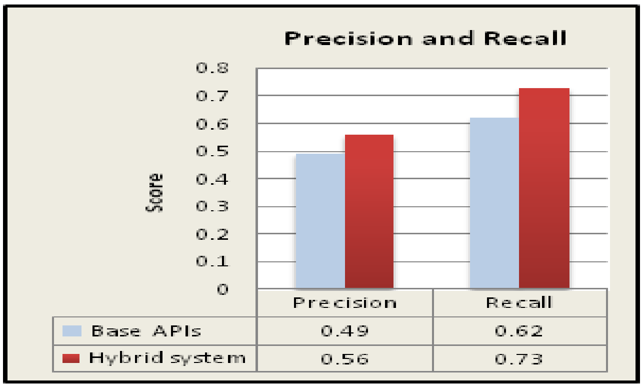

Figure 6. Precision and recall

The results were subjected to a t-test to determine their significance. Both precision and recall were significant with 99% confidence. It can be concluded therefore that more relevant documents are selected by the proposed system than the raw APIs, whereas the percentage of relevant documents from the pool of documents is more or less the same as base APIs. The results show that after a learning period there is a clear benefit in using the hybrid system.

V. DISCUSSION

Some studies have highlighted the fact that users prefer transparency and control in the systems they use. These studies also indicate that too much flexibility in the customisation process, such as editing profiles, can have an adverse effect on personalisation [21].

One of the key issues in the personalisation process is how to address 'the cold start problem'. The assumption that a significant amount of explicit feedback is required in order to build a profile has led to more emphasis on implicit feedback and on the synergy of user communities, rather than rely on explicitly formulated profiles [22]. Besides the dismissal of what is considered the 'brittle models' of the explicit profiles and their lack of relevance, many of the systems on user personalisation are increasingly relying on social networks to provide additional implicit information on user behaviour, and by implication pave the way for recommendation procedures [23]. Although this approach has the advantage of creating a richer context of interaction, it has the drawback of postulating the existence of a social network, an assumption that may affect its operation. Another disadvantage of this approach is the undue weight it gives to the implicitly generated user information.

One aspect that many controlled studies have reported is the correlation between the usefulness of documents to users and many of their interactive activities such as time spent viewing a document and other operations such as saving and printing them [24]. It was however pointed out that the information that a user is searching for has a significant impact on the usefulness of the implicit feedback [25]. Although explicit and implicit profiles have identified two extremes of profile generation in some studies many researchers have pointed out that they are complementary [26].

The proposed approach seeks to overcome the limitations of the two modes of operation and to capitalise on the complementary features. It also marks a departure from the 'feedback' related to explicit profiles, in order to minimise user intrusion and inconvenience. In contrast the focus is on the profile formulation by the user. This shift of emphasis means that the user has some control over the personalisation, while the concurrent implicit profile generation maintains the currency of the user interests. In the proposed approach, prominence is given to the user, the document and their interaction. This perspective is well served by a content-based approach rather than a collaborative approach. It provides focus, control and wider application. The content based approach allows the system to harvest relevant user information without the need of a community of users.

The novelty of the work lies in the seamless and balanced combination of discrete intervention and transparent implicit profile generation. No explicit feedback is required during the interaction with the documents such as, for example, rating the relevance of each document. Instead the user is allowed to state at the outset relevant interests in terms keywords. This is an on-going research programme. Work is currently being carried out on widening the semantic context of implicit and explicit profiling by incorporating ontologies. This will overcome the restrictions imposed by the exact matching of keywords. This work complements other work in information retrieval carried out by the authors in the areas of information retrieval [27], image processing [28] and recommendation systems [29].

VI. CONCLUSION

A hybrid profiling system which combinesexplicitly stated interests with observation of user behaviourwas presented in this paper. The experimental results indicate that the system with hybrid profiling has better and more accurate results than the APIs without profiling. These results indicate clearly that hybrid profiling can enhance the quality of Web search. The combination of explicit and implicit profiling in a content-based approach can offer an effective way of dealing with information overload. Overall hybrid profiling can be an important tool for enhancing search system performance in terms of precision and recall. The system has some limitations; if the interests of the user change, the hybrid system performance may be affected until a new profile is created.

REFERENCES

[1] Klusch, M.: Information Agent Technology for the Internet: A Survey. *Data & Knowledge Engineering, vol. 36*, 2001, pp337-372.

[2] Zigoris, P.& Zhang, Y. Bayesian Adaptive User Profiling with Explicit & Implicit Feedback.*Proceedings of the 15th ACM international conference on Information and knowledge management CIKM-2006, pp397--404.*

[3] Brin,S. & Page, L. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks, vol. 30*, 1998, pp.107-117.

[4] Brusilovsky,P. & Tasso, C. Preface to Special Issue on User Modeling for Web Information Retrieval. *User Modeling and User-Adapted Interaction, vol. 14,* 2004, pp147-157.

[5] Pazzani, M. J., & Billsus, D. Content-Based Recommendation Systems. *LNCS Vol 4321*,2007, pp325-341.

[6] Gauch, S., Chaffee, J., & Pretschner, A. Ontology-Based Personalized Search and Browsing. *Web Intelligence and Agent Systems, vol. 1*, 2003, pp219-234.

[7] Sieg, A., Mobasher, B. & Burke, R. Inferring User's Information Context: Integrating User Profiles and Concept Hierarchies. *Proceedings of the 2004 Meeting of the International Federation of Classification Societies, IFCS 2004*, 2004, pp563-574.

[8] Ferragina, P.& Gulli,A. A Personalized Search Engine Based on Web-Snippet Hierarchical Clustering.*Journal Software—Practice &Experience archive, vol. 38 Issue 2*, February 2008, pp189-225.

[9] Pazzani, M. J., Muramatsu, J.& Billsus, D. Syskill & Webert. Identifying Interesting Web Sites. *Proceedings of the Thirteenth National Conference on Artificial Intelligence, vol. 1*, 1996, Portland, Oregon, US, pp54-61.

[10] Garden, M., & Dudek, G. : Mixed collaborative and content-based filtering with user-contributed semantic features. *Proceedings of the 21st national conference on Artificial intelligence, vol. 2*, 2006, pp1307-1312.

[11]. Shen, X., Tan, B. & Zhai, C. Exploiting Personal Search History to Improve Search Accuracy. *Personal Information Management - A SIGIR 2006 Workshop*, pp94-97.

[12] Claypool, M., Gokhale, A. & Miranda, T. Combining Content-Based and Collaborative Filters in an Online Newspaper. *In Proceedings of the SIGIR-99 Workshop on Recommender Systems: Algorithms and Evaluation.* 1999.

[13] Aoidh, E. M., Bertolotto, M., Wilson, D. C. Implicit Profiling for Contextual Reasoning about Users Spatial Preferences. 271 *CaCoA, volume 271 of CEUR Workshop Proceedings, CEUR-WS.org*, 2007.

[14] Lieberman, H.: Letizia: An Agent That Assists Web Browsing. In: Mellish, C. (ed.), pp. 924-929. Morgan Kaufmann publishers Inc., San Mateo, CA ,1995.

[15] Gasparetti, F., & Micarelli, A. Exploiting Web Browsing Histories to Identify User Needs. *Proceedings of the 12th international conference on Intelligent user interfaces (IUI'07*), pp325-328. New York, USA, ACM Press.

[16] Salton, G., Singhal, A., Mitra, M. & Buckley, C. Automatic Text Structuring and Summarization. Information *Processing and Management, vol. 33,* 1997, pp193-207.

[17] White, R. W., Jose, J. M.& Ruthven, I. An Approach for Implicitly Detecting Information Needs. *Proceedings of the twelfth international conference on Information and knowledge management (CIKM'03*), 2003, pp504-507.

[18] O'Sullivan, D., Smyth,S.& Wilson, D. Explicit Vs Implicit Profiling - A Case-Study in Electronic Programme Guides. *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-03)*, Acapulco, Mexico, August 2003.

[19] Kumar, S., B.T.& Prakash, J. N. Precision and Relative Recall of Search Engines: A Comparative Study of Google and Yahoo. *Singapore Journal of Library & Information Management*, *vol. 38*, 2009, pp124-137.

[20] Shafi, S. M.& Rather, R. A. Precision and Recall of Five Search Engines for Retrieval of Scholarly Information in the Field of Biotechnology.*Webology, vol. 2, Number 2*, , 2005.

[21] Ahn, J. W., Brusilovsky, P., Grady, j., He, D.& Syn, S. Y. Open User Profiles for Adaptive News Systems: Help or Harm?. in *Proceedings of the Sixteenth International World Wide Web Conference*, 2007, Alberta, Canada.

[22] Zigoris, P.& Zhang, Y.Bayesian Adaptive User Profiling with Explicit & Implicit Feedback. *Proceedings of the 15th ACM International Conference on Information and Knowledge Management.* Arlington, Virginia, USA: ACM Press (2006)

[23] Cayzer, S.& Michlmayr, E. Adaptive User Profiles. HP Laboratories. 2008 [online] available from *http://www.hpl.hp.com/techreports/2008/HPL-2008-201.pdf*

[24] Fox, S., Karnawat, K., Mydland, M., Dumais, S.& White, T. Evaluating implicit measures to improve web search.. *Journal of ACM Transactions on Information Systems 23(2)*, 2005, pp147–168.

[25] Kelly, D., & Belkin N. J. Display time as implicit feedback: understanding task effects. *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, USA, 2004, pp377–384.

[26] Jawaheer, G., Szomszor, M. & Kostkova, P. Comparison of Implicit and Explicit Feedback from an Online Music Recommendation Service. *Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems*. New York, USA 2010.

[27] Pannu M., Anane R. and James A. The Impact of Modes of Mediation on the Web Retrieval Process. *The 23rd International Conference on Database and Expert Systems Applications (DEXA 2012), LNCS, Volume 7447,* Vienna, Austria, September 2012, pp297-304.

[28] Iqbal K, Odetayo M. & James A. Content-based image retrieval approach for biometric security using colour, texture and shape features controlled by fuzzy heuristics, *Computer and Systems Sciences, vol. 78 (4),* 2012, pp1258-1277.

[29] Buncle J., Anane R. & Nakayama M. A Recommendation Cascade for e-learning.*The 27th IEEE International Conference on Advanced Information Networking and Applications* (*AINA-2013),* Barcelona, Spain, March 25-28, 2013