User Profile Modeling and Applications to Digital Libraries

Giuseppe Amato and Umberto Straccia

Istituto di Elaborazione dell'Informazione - C.N.R.
Via S. Maria 46, I-56126 Pisa, Italy
{amato,straccia}@iei.pi.cnr.it
http://faure.iei.pi.cnr.it/~{amato,straccia}

Abstract. The ultimate goal of an information provider is to satisfy the user information needs. That is, to provide the user with the right information, at the right time, through the right means. A prerequisite for developing personalised services is to rely on *user profiles* representing users' information needs. In this paper we will first address the issue of presenting a general user profile model. Then, the general user profile model will be customised for digital libraries users.

1 Introduction

It is widely recognised that the internet is growing rapidly in terms of the number of users accessing it, the amount of information created and accessible through it and the number of times users use it in order to satisfy their information needs. This has made it increasingly difficult for individuals to control and effectively seek for information among the potentially infinite number of information sources available on the internet. Ironically, just as more and more users are getting online, it is getting increasingly difficult to find relevant information in a reasonable amount of time, unless one knows exactly *what* to get, *from where* to get it and *how* to get it. New emerging services are urgently needed on the internet to prevent computer users from being drowned by the flood of available information.

Typical information sources on the internet, like search engines, digital libraries and online database (*e.g.*, [1, 6, 12, 13], just to mention some), provide a search and retrieval service to the web community at large. A common characteristics of most of these retrieval services is that *they do not provide any personalised support to individual users*, or poorly support them. Indeed, they are oriented towards a generic user. In fact, they answer queries crudely rather than, for instance, learning the long-term requirements idiosyncratic to a specific user. Moreover, they seldom select and organise information for users accordingly, *e.g.*, assisting in the selection of books or other archived documents from libraries, news items from press agencies, television station and journals, or documents from administrative bodies. Providing personalized information search and delivering services, as additional services to the uniform and generic information search offered today, is likely to be the first step to make *relevant* information available to people in the appropriate *form*, *amount* and level of *detail*, at the *right time* through the *right means*, and with *minimal* user effort.

A prerequisite for developing systems providing personalised services is to rely on *user profiles*, *i.e.* a representation of the preferences of any individual user. Roughly, a user profile is a structured representation of the user's needs through which a retrieval system should, *e.g.*, act upon one or more goals based on that profile and autonomously, pursuing the goals posed by the user (irrespective of whether the user is connected to the system).

It is quite obvious that a user profile modeling process requires two steps (which constitutes the user profile modeling methodology). We have to describe

- *what* has to to represented, that is which information pertaining to the user has to represented, and
- *how* this is information is effectively represented.

The topic of this paper is to describe both steps. We will show that the first one can be described in a quite general and application independent way, while the second one depends on a particular application. In order to be concrete, we will propose a user profile model which can be used in the context of the NCSTRL digital library (Networked Computer Science Technical Reference Library) [13]. Essentially, using profiles, users will be able to create their own customised scientific interest representation. This allows the digital library to provide a "notification service", by *e.g.* e-mailing the users when documents (like technical reports and articles) matching their scientific interest become available in the digital library.[1] The interesting point is that simple modifications to the existing architectures are sufficient in order to provide this service.

We proceed as follows. In the next section we will introduce those concepts which have to be taken into account in a quite general user profile modeling process. In Section 3 we will apply these concepts to a special case: digital libraries (like NCSTRL). In Section 4 we will present two solutions for extending an existing search service in a retrieval system in order to take into account user profiles. Section 5 concludes and describes further work.

2 User Profile Modeling

The topic of this section is to describe some general concepts involved in the user modeling process. In particular, we will describe *what* has to represented in a user profile from the users point of view.

2.1 The General Information Retrieval Scenario

The general concern of a user is the retrieval of relevant information that pertains to its information needs. So, let us first introduce a global and general information seek scenario (see Fig. 1).

[1] Similar features are promised within the ACM Digital Library [6] as a forthcoming service.

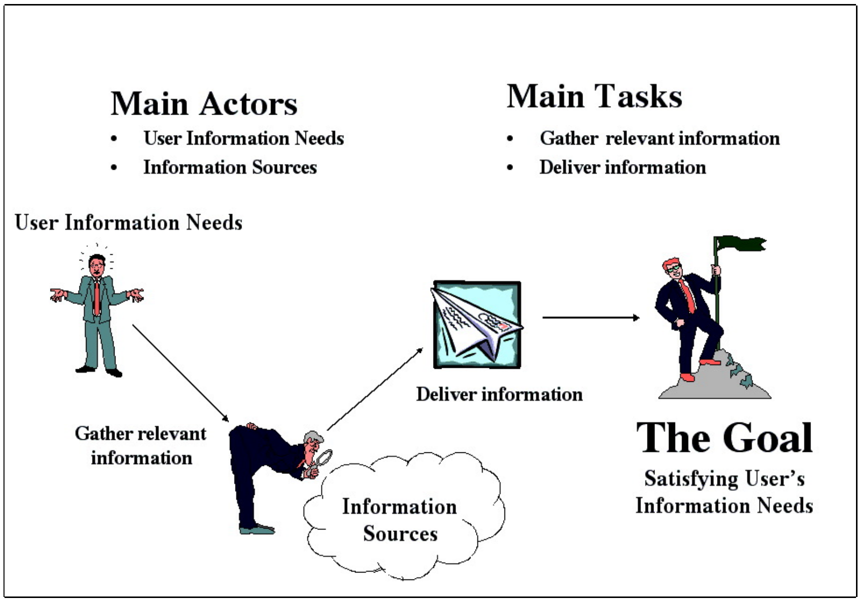

Fig. 1. The information seek scenario

We can distinguish two main actors in it: the *user information needs* and the *information sources*.

User Information Needs. With user information needs we mean "what" a user is really looking for. Examples of user information needs may be

1. "I'm looking for journal articles about computer networking, published not later than 1996. I want to pay less than 2$ for each."
2. "I'm looking for news concerning the latest trend about stock quotes of High-Tech companies."
3. "I'm looking for MPEG videos about Formula One races, downloadable from the web in less than 2 hours."
4. "I'm looking for hike tours in the Alps."

In the following, we will consider the information needs described by Point $1.-4.$

The first observation is that a user information need may be quite different w.r.t. information *type* and *content*. With respect to the type, in cases $1. - 4.$ we are looking for journal articles, news, MPEG videos and images, respectively. These describe the type of information we are looking for. On the other hand, w.r.t. the content, in cases $1. - 4.$ we are also looking for information which is about computer networking, about stock quotes of High-Tech companies, about Formula One races and geographical maps with hike trails. These describe what information we are looking for from an information content point of view.

The second observation is that not only different users have heterogeneous information needs, but there may be a heterogeneity in between the needs of a single user too. That is, the information needs $1.-4.$ may belong to four different users or may be four different needs (from a type and content point of view) of the same user.

A third and final observation is that a user information need may be a *short term* user information need or a *long term* user information need. In the former case we refer to an ad-hoc, occasional user information need, whereas in the latter case we refer to a user information need which is of interest during a relevant time period. It is easily verified that in fact our daily information seek process involves both temporary needs as well as long term interests. Of course, whether an information need is a short term or a long term interest depends on the user. For instance, if an economist (say, John) is planning to hike in the mountains next weekend (an event that seldom happens for John), then he is looking for some site map and tour and may express its need through Point 4. above. This is a short term information need of John. But, John, as a serious economist, is also interested in any kind of news related to stock exchange quotes. He may express his information need through Point 2. above. Of course, this is a long term information need. On the other hand, Point 2. may be a short term interest of a computer scientist (say, Tom), whereas Point 1. may be his long term interest.

In summary, information needs may differ w.r.t. their type, their content and their duration (short term and long term). Moreover, information needs are heterogeneous among users and in between users. All these aspects have to be taken into account during the user profile modeling process.

Information Sources. With information sources we mean all the heterogeneous digital information providers distributed over the Internet, which make available any kind of information which might be of interest to Internet users. Examples of information sources are web sites, online databases, news groups, news agencies, search engines, digital libraries, etc. Essentially, they differ in *what kind of information they provide, what services they provide and which users they address.*

The ultimate goal of an information provider is to satisfy user information needs, that is *to provide the user with the right information, at the right time, through the right means.* It is easily verified that this requires the execution of two separate tasks: *to gather relevant information* and *to deliver* them. The first tasks, typically the hardest one, is that of gathering the information which is thought to be of interest to the user. Once the information has been collected, it has to be delivered to the user, according to his preferences (second task). Examples of delivery modalities may be web pages (this is the usual case for which most of us are familiar with), e-mail, phone, fax (*e.g.*, a user wants to receive stock quotes by phone, e-mail or fax), or surface mail (*e.g.*, a user wants to receive the proceedings of a conference by surface mail).

As far as our work concerns, we will concentrate on user information needs. In particular, in the next section we will refine the concepts involved in the user information need modeling process. An equally important topic is the modeling of information sources, which we will not address in this paper.

2.2 The Data Categories of a User Profile

In this section we will a present a general *user profile model* through which we may represent user's preferences and needs.

By relying on the discussion of the previous section (see also Fig. 1) it is quite clear that in a user profile we have to represent at least

– *what* has to be *gathered*, and
– *how* the gathered information has to be *delivered* to the user.

We will show in the following that the information to be represented about the users is not only restricted to the two categories above, but may be classified in fact into (at least) five data categories. These categories are *the personal data category*, *the gathering data category*, *the delivering data category*, *the actions data category*, and *the security data category*. In the following we will describe these five categories in detail.

The Personal Data Category. The *personal data category* is a collection of user's personal identification data. Under this category we consider data like user's name, birth date, gender, identity certificate, employer, home contact information, business contact information, etc. (see *e.g.*, [14] as a concrete case).

The Gathering Data Category. The *gathering data category* collects preferences and restrictions about the documents a user is looking for. These preferences and restrictions may be classified into three distinct subcategories, each addressing orthogonal document dimensions. These subcategories are:

– *the document content category*: specification of *what* has to be gathered. Under this category we consider preferences on document's properties that relate to the *content* a user is looking for: the document language and its aboutness. For instance, "I'm looking for documents talking about computer networking, written in English.".
– *the document structure category*: user's specification of all those properties of a document he/her is looking for which relate to the *structure* of a document, like its format (text formats, image formats, audio formats, video formats), its type (article, technical report, proceedings, news, novel, poem, www home page), its creation date, its cost and dimension. For instance, "I'm looking for GIF images created today."
– *the document source category*: specification of where to gather *from*. In this category we collect all the user's restrictions on the *source* from which he/she would like to receive information, like a restriction on the URL (*e.g.*, "I want only documents from `http://www.w3.org/`"), the specification of publishers

(*e.g.*, "I want only news from Reuters"), series (*e.g.*, "I want only articles from the Lecture Notes in Computer Science"), author's (*e.g.*, "I'm looking for audio records of Giuseppe Verdi.").

In summary, in a user profile we should allow the representation of *what* to gather (in terms of the structure and the content of a document) and where to gather *from*.

The Delivering Data Category. Under the *delivering data category* the user specifies preferences on the delivery modes of the gathered information. These preferences may be classified into two distinct subcategories, each of them addressing orthogonal delivering dimensions. These subcategories are:

1. *the delivery means category*: specification of *how* to deliver. In this category we consider user preferences regarding the delivery means, like phone, fax, web and e-mail, that should be used in order to deliver the information the user requested for.
2. *the delivery time category*: specification of *when* to deliver. In this category we consider user preferences regarding the delivery time, like interval (*e.g.*, "deliver me the news I'm interested in each morning at 9 am, except during the weekend.") and as soon as possible (*e.g.*, "deliver me the news I'm interested in as soon as you gather it.", or "deliver me the stock quote exchange rate I'm interested in as soon as it looses more than 5%.").

In few words, in order to represent the user's delivery preferences we should represent *how* to deliver and *when* to deliver.

Actions Data Category. A personalised service should be highly responsive to the needs of the user. In particular, long term information needs involve repeated interactions with the user. Assuming that a lot of the user actions are consistent, a retrieval service should match increasingly better his/her needs over time. Furthermore, since the interaction could extend over a long period of time, it cannot be assumed that the users interests will remain constant. The change in interest could be anything from a slight shift in relative priorities to completely losing interest in some domain and gaining interest in another. In general, a system must be able to detect or must allow the user to indicate the change in interests and should respond by adapting to these changes. The system must be able to explore newer domains and prospect for interesting information. To summarise, personalised service should be capable not only of dealing with the currently known needs of the user, but also exploring different domains to find documents of potential interest to the user. Thus, it should be specialised, adaptive and exploratory.

In order to provide a service with the above capabilities, under the *actions data category*, we collect a set of actions, not necessarily taken only by the user him/her self. The actions data generally contains the recording of the user's interaction with retrieval systems and navigation data. Typical actions data may be URLs of visited web pages, read documents and user's relevance judgements. By relying on techniques based on user relevance feedback [2, 4, 8, 9, 11, 15–17], as well as on collaborative feedback (in this case the user is usually member of an interest group) [5, 10], the actions may be profitably be used for refining the user's gathering data specification.

Security Data Category. Users wants to express their privacy practices. The *security data category* is a collection of user preferences establishing the conditions under which the data represented in the user profile may be accessed. These preferences may regard all the previous categories (the personal data, the gathering data, the delivering data and the actions data categories). Typically, user's may establish different privacy practices for each of the services they access to. An extensive work about privacy preferences can be found in [14].

An Example. We end this part with an example, illustrating the concepts introduced in the sections above. Suppose a user's profile is as follows:

1. "I'm John Smith, 34 year old and I'm looking for
2. video sequences, dated after than April, 1st, for which I don't want to pay for,
3. which are about Michael Schumacher driving his Ferrari and
4. published by FIA (Federation Internationale de l'Automobile).
5. Deliver me as soon as possible
6. an audio summary message of the top ranked video I'm interested in and a SMS message containing the source URL, at my cellular phone, +39.0347.593404.
7. I have already seen `http://www.fia.com/news/news1.mov` and consider `http://www.ukmotorsport.com/news/ferrari.html` as relevant to what I'm looking for.
8. I do not allow to access to my personal data.".

According to the user profile schema resumed in Table 1, Point 1. pertains to the personal data category, Point 2. pertains to the document structure category, Point 3. pertains to the document content category, Point 4. pertains to the document source category, Point 5. pertains to the delivery time category, Point 6. pertains to the delivery means category, Point 7. pertains to the actions data category and Point 8. pertains to the security data category.

3 A Profile Schema for Digital Library Users

As for documents there exists several ways to represent them (like the vector space model, Dublin Core, MARC, etc.), similarly, there may be different, application dependent, user profile representations. In this section a profile schema tailored for digital library users is discussed. The proposed profile schema, while

Table 1. User profile schema summary

User Profile:
- **personal data category**
- **gathering data category:**
  - **document content category**
  - **document structure category**
  - **document source category**
- **delivering data category:**
  - **delivery means category**
  - **delivery time category**
- **actions data category**
- **security data category**

remaining within the general user profile model presented in the previous sections, tries to capture typical aspects that can be required by a digital library user. These features, if well exploited, can significantly help an advanced digital library to automatically search for documents relevant to the user. For instance, a particularly interesting case concerns digital libraries users having long term interest, as, *e.g.* scientist. In this case, a digital library (like NCSTRL and ACM [6, 13]) may notify the user as soon as a new article, technical report or the like has been made available and matches his/her research interests.

We will first describe the general structure of the user profile schema, then particular attention will be paid to the *gathering data* category.

3.1 The Profile Schema

Users that want to exploit the retrieval capabilities of a digital library are supposed to subscribe to the service. As consequence of the subscription, a personalized profile is created for the user. The profile is identified by a unique *profile identifier*. This can be formalized as follows:

$$Profiles = ProfID \rightarrow UserProfile\ . \tag{1}$$

As we have seen, user profile data may be classified into five categories. We formalise this with

$$UserProfile = (PersData \times GathData \times DeliData \times ActData \times SecData)\ . \tag{2}$$

In the following we will formally describe each of these categories.

Personal Data. The personal data category contains information about the user identity. For complying a standard, we propose to rely on the P3P "user" schema [14] for the $PersData$ specification.

Gathering Data. This category of the user profile specifies what documents a user is interested in. In Section 2.1 we have seen that a certain user may have at the same time several different interests. In the user's profile all the user's interests should be described separately so that different types of preferences can be specified for different interests. We call *topic* a single user information need. In order to capture the fact that the user may have several interests, the proposed user profile is associated with a set of topics. This is formalized as follows:

$$GathData = TopicID \rightarrow Topic \ . \tag{3}$$

Each topic is identified by a *topic identifier* that should be unique for a given user profile, *i.e.* the pair $(ProfID, TopicID)$ is unique. The complete definition of a topic will be given separately in Section 3.2.

Delivery Data. Different users may have different delivery modalities. In order to take into account the delivery means and the delivery time, we formalise $DeliData$ as

$$DeliData = (DelMode \times TimeMode) \ . \tag{4}$$

The delivery mode contains the specification of which means should be used to deliver information, how the delivered information should look like and the destination address. More formally:

$$DelMode = (DelMeans \times Layout \times Destination) \ . \tag{5}$$

The user can choose to be notified using one of the delivery means available, like e-mail, web page, phone, fax, etc. Since the potential users of a digital library like NCSTRL are scientist, e-mail or web page is adequate. In the former case an e-mail, formatted accordingly the layout preferences, is sent to the address specified in the destination field. In the latter case, retrieved documents are published in the web page identified by the destination field and formatted accordingly the layout preferences. The layout specification may contain preferences about *e.g.* the colors and fonts to be used, and preferences about the information to be included for each relevant document found (*e.g.* title, abstract, authors, keywords, etc.). The destination specification is an address identifier that depends on the delivery means.

The time mode specifies when to deliver. We will consider basically a condition like "new document found" and "updated document" associated with a delivery time. The deliver time can be a fixed time interval (e.g. every day at 9 am) or "as soon as possible". Formally we have:

$$\begin{aligned}
&TimeMode = (NewDoc \times UpdatedDoc \times Time) \\
&NewDoc = (\mathtt{yes} + \mathtt{no}) \\
&UpdatedDoc = (\mathtt{yes} + \mathtt{no}) \\
&Time = (TimeInterval + \mathtt{asap}) \ ,
\end{aligned} \tag{6}$$

where $TimeInterval$ is defined accordingly to the Unix OS *crontab* file.

It is worth noting that the formalisation of *DeliData* establishes that a user has an *unique* delivery modality. We may enhance the schema by allowing a delivery modality, for each user topic – *i.e.* user interest. This is, for instance, required to model cases like "*upload* the proceedings of *ECDL* to my ftp server, while send me the abstracts of papers about *filtering* systems by *e-mail*". In order to take this into account we may formalise *DeliData* as follows:

$$DeliData = TopicID \rightarrow (DelMode \times TimeMode) \ . \tag{7}$$

Actions Data. The actions data category is a sequence of pairs each representing an action performed on a certain document. As actions are typically used for coding user's relevance feedback within one of his topics of interest, we formalise *ActData* as follows:

$$\begin{aligned} &\mathit{ActData} = TopicID \rightarrow (\mathit{Action} \times \mathit{DocumentID})^* \\ &\mathit{Action} = (\texttt{read} + \texttt{relevant} + \texttt{notrelevant}) \end{aligned} \tag{8}$$

and $DocumentID$ is the identifier of a document notified to the user (a URI). The action identifiers will be used according to the following meanings:

read: the user looked at the full text of the document
relevant: the user judged the document as relevant
not relevant: the user judged the document as not relevant

Of course, other actions may be included as well. It's beyond the scope of this paper to further detail how the actions data may be used for relevance feedback analysis.

Security Data. As users subscribing to a service agree on its privacy practices (see, *e.g.* [14]), the simple privacy maintenance mechanism we adopt is to specify in the security data category which on-line services may access the user's information. It is basically a list of the hosts that are authorized to ask for information contained in the user's profile.

$$SecData = (HostName)^* \ . \tag{9}$$

3.2 Topics

As specified in Section 3.1, a user may have several topics of interest and a topic specifies *what* to gather. Accordingly to Section 2.2, we define $Topic$ as

$$Topic = (TopicName \times DocContent \times DocStruct \times DocSource) \ . \tag{10}$$

The document content category contains the information that allows the system to recognise documents relevant to a topic from a document content point of view. In digital libraries, the content of a document is described by means of its title, its textual description (the abstract or summary), a list of relevant keywords and a list of standard categories (*e.g.* the ACM categories) and its language. We formalise this as

$$DocContent = (Title \times TextualDescr \times Keywords \times Categories \times Languages)\ . \quad (11)$$

The document structure category contains information that allows the system to eliminate candidate documents according to their structural properties. Documents in digital libraries can be stored in different file formats and they can be different types of documents (*e.g.* book, technical report, scientific article). Moreover, documents have a publication date and, in case of a paying service, they can have a price. The *DocStruct* category is defined as follows:

$$\begin{aligned} &DocStruct = (FileFormat \times Type \times PublicationDate \times Price) \\ &FileFormat = (\texttt{all} + \texttt{postscript} + \texttt{pdf} + \texttt{html} + \ldots)^* \\ &Type = (\texttt{all} + \texttt{book} + \texttt{technicalreport} + \texttt{journalarticle} + \ldots)^*\ . \end{aligned} \quad (12)$$

The `all` is used in order to specify all file formats or all document types.

A system should eliminate unwanted documents by considering the information about the source of a document. The document source category is intended to provide a conceptual information that specifies from where to gather documents. We model *DocSource* as follows:

$$\begin{aligned} &DocSource = (AllowSources \times DenySources) \\ &AllowSources = (Sources) \\ &DenySources = (Sources) \\ &Sources = (Collection^* \times Publisher^* \times Series^* \times Author^*)\ . \end{aligned} \quad (13)$$

The `all` value is used to indicate all collections, all publishers, all series or all authors. The deny list contains sources that should not be considered while the allow list those that should be considered.

4 Architecture

In this section we will show how an existing digital library may simply be extended in order to provide a new service: to alert automatically a user when a new document, matching the user's profile, is available in the digital library.

There are basically two different possibilities to implement the above functionality which we call *pull modality* and *push modality*. In the former case the profile is used in order to generate a query based on it and submit the query to the native information retrieval engine of the digital library. In the latter case, any new incoming document is matched against all available profiles in order to select those which the document is relevant for. The two possible corresponding architecture are depicted in Fig. 2 and Fig. 3.

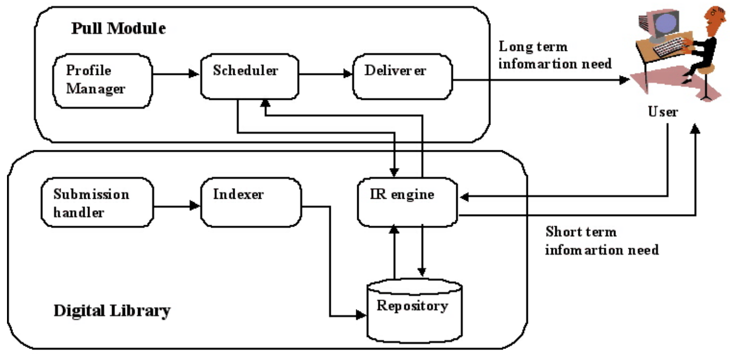

Fig. 2. Pull modality

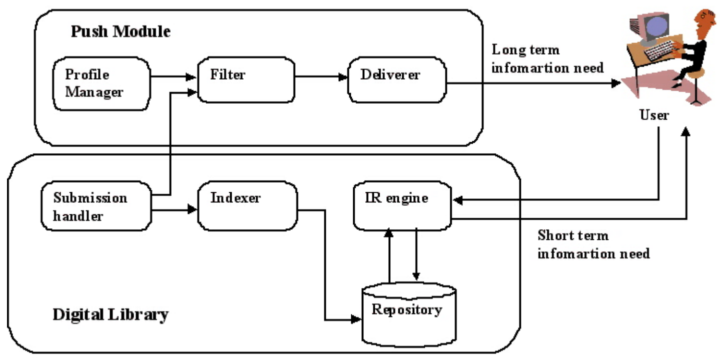

Fig. 3. Push modality

The *profile manager* and the *deliverer component* are common to both approaches. The profile manager mainly maintains the user profiles. It allows users and authorized components to modify a user profile and send profiles or portion of it to authorized components that request them. The deliverer component is responsible for delivering according to user's delivery preferences.

In the pull modality, the *scheduler component* at scheduled times, depending on the profile preferences, generates queries based on the profiles content and submit them to the built-in information retrieval engine of the digital library. From the result list we have to consider only those documents which have not yet been delivered to the user. The obtained list is returned to the delivery component. It is worth noting that this solution may be applied to any existing digital library as ideally no modifications are needed to existing systems. The pull module could be customized for different digital libraries just defining for the scheduler component ad-hoc wrappers that translate profiles into queries.

However, the pull modality allows to implement an approximated "as soon as possible" functionality only. In fact, this functionality is implemented by timely asking the digital library whether there are new documents available (scheduler's job). While this solution is unpracticable for rapidly changing information sources, like news feeds, it seems to be quite feasible in the context of digital libraries. For instance, in the case of NCSTRL, is it more than enough to query the library once per day and deliver a message for each satisfied profile.

In the push modality, as soon as a new document is available in the digital library, the document is sent to the filter component that matches it against all profiles. For all matching profiles, the deliverer component sends a notification. It is quite clear that the push schema allows "as soon as possible" notifications to be effectively handled. Another advantage is that, since only new documents are checked for profile matching, it is guaranteed, unlike the pull modality, that only new documents are delivered to the user. Unfortunately, in order to be implemented, some modifications to the existing digital library are needed.

It is worth noting that a system which, among others, uses both of the solutions above, together with a quite similar profile schema as proposed in this paper, has been implemented within the *Eurogatherer* Project [7]. In vew words, the Eurogatherer system is a personalised gathering and delivering system which offers user profiling, gathering from heterogenous information sources and all the delivery modalities described in this paper.

5 Conclusions

There are three contributions in this paper. First, we have presented a quite abstract user profile model in the context of an information seek scenario. We discussed what information has to represented into a user profile. Second, we have described a user profiles for digital library users. Third, we have provided two simple architectural solutions in order to extend digital libraries to cope with user profiles.

Actually, we are planning to provide a personalised search service, by relying on user profiles, within the ETRDL (ERCIM Technical Reference Digital Library) [3] in which our institute is involved.

Acknowledgements

This work is funded by the European Community in the context of the *Eurogatherer* Telematics Information Engineering Project Number 8011 [7].

References

1. Altavista. Altavista Home Page: `http://www.altavista.com`, WWW.
2. Gianni Amati, Fabio Crestani, and Flavio Ubaldini. A learning system for selective dissemination of information. In *Proc. of the 15th Int. Joint Conf. on Artificial Intelligence (IJCAI-97)*, pages 764–769, Nagoya, Japan, 1997.

3. Antonella Andreoni, Maria Bruna Baldacci, Stefania Biagioni, Carlo Carlesi, Donatella Castelli, Pasquale Pagano, and Carol Peters. Developing a European Thechnical Reference Digital Library. In *Third European Conference on research and advanced technology for Digital Libraries*, Paris, September 1999. Lecture Notes in Computer Science.
4. Krishna Bharat, Tomonari Kamba, and Michael Albers. Personalized, interactive news on the web. *Multimedia Systems*, (6):349–358, 1998.
5. *Berkeley Workshop on Collaborative Filtering*, 1996. `http://www.sims.berkeley.edu/resources/collab/`.
6. ACM Digital Library. ACM Digital Library Home Page: `http://www.acm.org/dl`, WWW.
7. *Eurogatherer* Telematics Information Engineering Project Number 8011. Eurogatherer Home Page: `http://pc-erato2.iei.pi.cnr.it/eurogatherer/`, WWW.
8. Larry Fitzpatrick and Mei Dent. Automatic feedback using past queries: Social searching? In *Proceedings of the 20th International Conference on Research and Development in Information Retrieval (SIGIR-97)*, pages 306–313, Philadelphia,PA, July 1997.
9. A. Goker and T.L. McCluskey. Towards an adaptive information retrieval system. In Zbigniew W. Ras and Maria Zemenkova, editors, *Proc. of the 6th Int. Sym. on Methodologies for Intelligent Systems (ISMIS-91)*, number 542 in Lecture Notes In Artificial Intelligence, pages 349–357. Springer-Verlag, 1991.
10. David J. Goldberg, David Nichols, Brian M. Oki, and Douglas Terry. Using collaborative filtering to weave information tapestry. *Communications of the ACM*, 35(12):61–70, 1992.
11. Toshiki Kindo, Hideyuki Yoshida, Tetsuro Morimoto, and Taisuke Watanabe. Adaptive personal information filtering system that organizes personal profiles automatically. In *Proc. of the 15th Int. Joint Conf. on Artificial Intelligence (IJCAI-97)*, pages 716–721, Nagoya, Japan, 1997.
12. Medline. Medline Home Page: `http://igm.nlm.nih.gov/`, WWW.
13. NCSTRL. Networked Computer Science Technical Reference Library Home Page: `http://www.ncstrl.org`, WWW.
14. P3P. Platform for Privacy Preferences P3P Project Home Page: `http://www.w3.org/P3P/`, WWW.
15. Gerard Salton and J. Michael McGill. *Introduction to Modern Information Retrieval*. Addison Wesley Publ. Co., Reading, Massachussetts, 1989.
16. Amit Singhal and Mandar Mitra Buckley. Learning routing queries in a query zone. In *Proceedings of the 20th International Conference on Research and Development in Information Retrieval (SIGIR-97)*, pages 25–32, Philadelphia,PA, July 1997.
17. Bienvenido Vélez, Ron Weiss, Mark Sheldon, and David K. Gifford. Fast and effective query refinement. In *Proceedings of the 20th International Conference on Research and Development in Information Retrieval (SIGIR-97)*, pages 6–15, Philadelphia,PA, July 1997.