A PROFILE-BASED ARCHITECTURE FOR A FLEXIBLE AND PERSONALIZED INFORMATION ACCESS

Max Chevalier, Chantal Soulé-Dupuy, Pascaline Tchienehom

A PROFILE-BASED ARCHITECTURE FOR A FLEXIBLE AND PERSONALIZED INFORMATION ACCESS

Max Chevalier[*,], Chantal Soulé-Dupuy[*], Pascaline Laure Tchienehom[*]**

[*]*IRIT – SIG, 118 route de Narbonne, 31062 Toulouse Cedex 4, France*
[**]*Laboratoire de Gestion et de Cognition (LGC), 115 route de Narbonne, 31062 Toulouse, France*

{Max.Chevalier, Chantal.Soule-Dupuy, Pascaline.Tchienehom}@irit.fr

ABSTRACT

Information relevance is an important issue in information access techniques. A solution for improving this relevance is the personalization or adaptation of the answers provided to users. For this purpose we propose, in this article, an information retrieval and recommendation architecture in which any element is described in detail by a profile. The originality of this architecture is at the level of its generic aspects and the numerous possibilities of interactions between complementary profiles derived from a profile generic model. Thereafter, we propose a flexible use of profiles to evaluate the personal relevance of a user or a users group by combining various matchings between profiles criteria.

KEYWORDS

Information Retrieval and Filtering, Profile, Personalization

1. INTRODUCTION

Internet and Web expansion have led to multiplication of both information placed at disposal and users of this information. Information and users are characterized by their heterogeneity. To seek, find and exploit information in this context prove to be a very difficult task. The aim of the work undertaken is to propose an architecture of retrieval and recommendation for a flexible access to information. This architecture is based on re-usable and adaptable profiles and exploits profiles complementarities for the restitution of personalized information. The objective is to improve the relevance by trying as much as possible to bring the *system relevance* closer to the *personal relevance* (or waitings) of user or users group: needs, preferences, goal, etc.

This article is organized as follows: section 2 presents a state of art about information retrieval and filtering (or recommendation) and also about the profile concept. Section 3 describes the profile-based retrieval and recommendation architecture proposed as well as the associated profile generic model. Then, we explain general rules of matching profiles criteria for the restitution of relevant and adapted information, in particular through combination of different comparisons between profiles criteria.

2. LITERATURE REVIEW

Access techniques to information allow an individual to obtain information that meets his needs. We can gather them in two main groups: the *pull* technique, which needs an explicit request of an individual and the *push* technique, which does not need an explicit demand to return information to users.

Information Retrieval (IR), which is a pull technique, rests on need expression of an individual through a query formulated in a more or less structured free language (Baeza-Yates *et al.*, 1999), (Rijsbergen, 1979). However, in Information Retrieval, the real intention of the user is not always obvious in his manner of formulating his query and that can generate ambiguities on the sense of words that it contains. Many solutions exist in order to precise the sense of a query: the method of *relevance feedback* (Boughanem *et al.*, 1999) that uses the user relevance judgements on information to reformulate his query and thus to refine his research; the use of *long term profile* concept and *short term profile* (or context) concept of a user, to interpret his queries in order to re-evaluate and re-order the search results (Bottraud *et al.*, 2003); the use of contextualisation and individualization concepts for a personalized search (Pitkow *et al.*, 2002); etc.

Information Filtering (IF), which is a push technique, is a relatively passive task (Belkin *et al.*, 1992) because the user does not explicitly formulate his needs through a query, as it is the case in IR. In Information Filtering, we rather use a representation of the user called *user profile* to send information to him. There are several methods of filtering (Montaner *et al.*, 2003): *cognitive or content-based filtering* which uses the description of information contents for determining to which users profiles it corresponds (Pazzani *et al.*, 1996), (Korfhage, 1997); *social or collaborative filtering* which uses the users judgements concerning a set of information to carry out recommendations (Goldberg *et al.*, 1992), (Konstan et al., 1997); *demographic filtering* which uses users demographic data (age, profession, etc.) to make recommendations (Krulwich, 1997).

These filtering approaches are not exclusive and various hybrid methods, were developed (Good *et al.*, 1999), (Pazzani, 1999). Filtering hybrid approaches improve the relevance of filtering results by mitigating some limit of filtering methods presented previously (Balabanovic *et al.*, 1997).

For the restitution of results to users, a description of information handled by processes of information retrieval and recommendation is made. This information description is called profile. An object profile is a whole of characteristics that identifies or represents it. Profiles used in information access techniques are of varied nature (user profile, document profile, etc.) and structure (mono or multi criteria). In information access, profiles can be related only to contents of information or extended with criteria like: demographic data or preferences for users and metadata for documents (Berti-Equille, 2003).

In this article, we use the concept of profile to define a flexible approach for evaluating personal relevance of users. For that, we propose an architecture based on profiles. Our architecture originality is at the level of its generic aspect and its numerous possibilities of interactions between profiles that it offers. These interactions appear during combinations of various complementary profiles in order to restitute appropriate answers to each user or users group through matching of profiles.

3. A FLEXIBLE AND PERSONALIZED INFORMATION ACCESS

In this section, we present our generic framework for information access which consists in: an architecture and a generic profile model. Then, we explain the general matching rules of profiles.

3.1 Generic framework

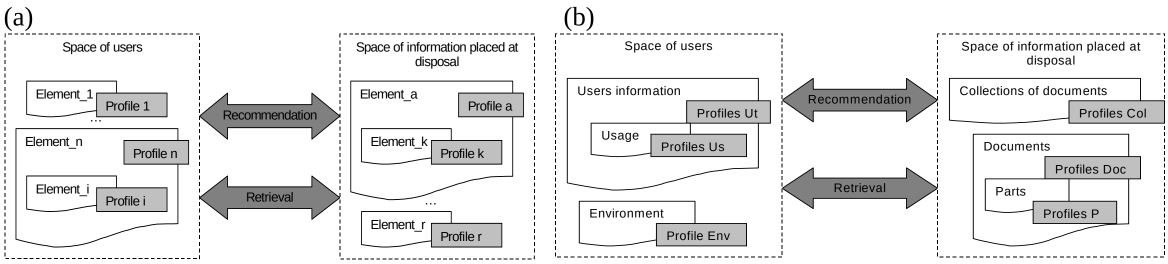

Figure 1. *(a) A general architecture based on profiles (b) Examples of architecture elements*

The diagram of the figure 1a presents the architecture of retrieval and recommendation based on profiles that we propose. This architecture is enough general to be used as a model for various applications. It results from the analysis of various retrieval and recommendation systems. Each one of these systems was conceived to achieve particular goals according to specificities of their context: recommendation of Web pages according to bookmarks (Rucker *et al.*, 1997), emails filtering (Goldberg *et al.*, 1992), electronic trade (Cho *et al.*, 2002), etc. The figure 1b presents examples of architecture elements described in figure 1a.

Our architecture is not applied to a predefined framework. It consists of a set of elements that play a role in processes of retrieval and recommendation. Each application is in charged of selecting in this architecture the elements that interest it. Our architecture can be used as a starting point for the construction of retrieval and recommendation systems.

Are highlighted, in this architecture, the processes of retrieval and recommendation as well as the generic structure of the elements handled by the processes. The elements are gathered into two groups: those that are related to the space of users and those that are related to the space of information placed at disposal. For each element type we associate a profile that describes it in detail and that is used by the processes of access to information. Moreover, these elements can also be composed by one or several elements that are themselves described by profiles. Note that usage information can be: queries, visited sites, judged information, safeguarded information, transferred information, etc. Moreover, elements of the space of users can be defined for individuals or groups of individuals. The profiles of this space can be of long term or short term (Widyantoro *et al.*, 1999), positive or negative (Hoashi *et al.*, 2000), etc.

The objective of the architecture suggested is the restitution of personalized information through the use of profiles. These profiles are derived from the generic model of figure 2a.

The profile generic model of the figure 2a presents the general structure of a profile. This structure is in the form criteria categories hierarchy that characterize a profile. This hierarchy is a forest or a set of trees in which nodes are categories of criteria and leaves are simply criteria to which we can affect values. A profile can thus be either a forest, or a tree, or a vegetation (or lists) of criteria. Thus, if *P* is a profile: *Structure (P)* $\in$ *{forest, tree, vegetation}*. The reflexive aggregation on the class *"profile"* denotes the fact that a sub-tree of a profile can have the structure of another existing profile. Thus, the structure of some profiles can be re-usable. For instance, a user profile can be composed of an existing environment profile.

The organization of the various criteria by category allows to gather the similar criteria in the same class and to define a criteria nomenclature (or taxonomy). From the generic model of profile, we can derive the structure of various profiles by applying decomposition rules on criteria categories. The figure 2b presents examples of profiles structures and taxonomies for: a user profile, a document part profile, a document profile, a collection of documents profile. Those profiles are also illustrated in figure 1b.

(a) (b)

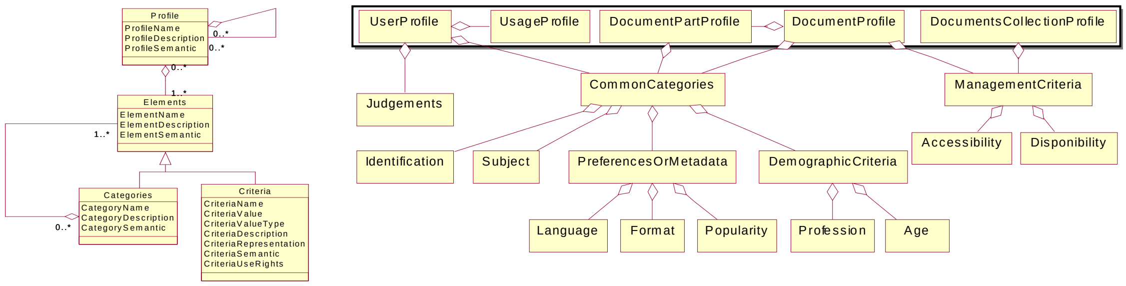

Figure 2. *(a) Generic model of profile (b) Examples of profiles structures and taxonomies*

The interest in using a generic profile to define a given profile is that any application, in order to define any type of profile, can use the basic structure it proposes. Here, we are not interested in particular criteria but rather in the modelling of a framework for the definition of these criteria.

All the categories are not always well informed for a given profile. Moreover, according to the application, all profile categories are not inevitably taken into account. Each application chooses the elements to be considered in a profile according to the objective that it wants to achieve. The generalization of the structure of a given profile made up of several criteria not always all well informed, allows to keep the same profile for a given element of the architecture. This gives an adaptable dimension to our profiles. Thus, a profile can be shared and enriched by various applications. For example, whatever the application, the user can be recognized with the same profile (criteria and content).

The architecture and the model of profile suggested will enable us to define profiles for a flexible access to information. The combination of various profiles will allow an optimal exploitation of the complementarities between profiles through matching rules of profiles criteria.

3.2 Matching rules of profiles criteria for a flexible access to information

We identified various matching types of profiles criteria. The combination of those similarity measures (or matching) will allow the selection of results that correspond to users or simply the reordering of these results. We classified similarity measures as follows:

1. *mono-term criteria matching*: this matching is used when the criteria to compare are mono-valued;

A matching example of this type can be the evaluation of *compatibility to mono-valued user criteria*. The evaluation of this matching consists in making comparisons between mono-valued user criteria and the corresponding descriptive criteria of information such as: popularity, size, target profession, target age, etc;

2. *multi-terms criteria matching:* this type of matching is used when at least one of the criteria is multi-valued. In this case, we represent the various criteria to be compared in the same vector space with the dimensionality given by the size of the vocabulary. To each vector of terms $d=(t_1,t_2,\ldots,t_n)$ is associated a real or boolean vector of weights $p_d=(w_{d,t1},w_{d,t2},\ldots,w_{d,tn})$ which will allow to calculate a measure of similarity between criteria: $sim(p_{di},p_{dj})$;

Examples

- *correspondence to user needs:* we evaluate a measure of similarity between the weighted vector of terms representing the needs (reformulated query or centers of interests) of a user and the weighted vector of terms representing the contents of information (document, parts of document, collection of documents, etc). The weights, in this case, are generally calculated with the formulas of *tf* or *tf.idf* and the similarity with the *cosine* formula;

- *compatibility to multi-valued user preferences*: we measure the similarity of an information for a given criterion (language, format…) to the preferences of the user for this criterion. Table 2 illustrates an example for the criterion *language*; etc.

Table 2. *Example of compatibility measure to languages preferences: f*

Criterion *language*	**Spanish (t_1)**	**English (t_2)**	**French (t_3)**	**Similarity**
Document weight p_d	$w_{d,t1}$=0	$w_{d,t2}$=1	$w_{d,t3}$=0	$sim(p_d,p_u)=V_{d,u,f}=\Sigma(w_{d,ti}.w_{u,ti})=0.5$
User weight p_u	$w_{u,t1}$=1	$w_{u,t2}$=0.5	$w_{u,t3}$=0.25	

3. *profiles matching (combination of matchings):* the matching of two profiles is a combination of the matching results between certain profiles criteria (as seen before). Thus, each matching result (or combination of matching results) may represent a selection or ranking factor. We can thus base this profiles matching on a factors list $m=(f_1,f_2,\ldots,f_n)$ to which correspond a matching results vector between two profiles *u* and *d*: $p_{d,u}=(v_{d,u,f1},v_{d,u,f2},\ldots v_{d,u,fn})$. We can note that a sublist of *m* could be used for the selection and another sublist (or the same one) for the ranking of information. An example list of factors can be: *correspondence to user needs (relevance of document, granules, collection), compatibility to user preferences in languages, compatibility to user environment, etc*. Moreover, we associate a vector of weights $p_{m,x}=(w_{f1},w_{f2},\ldots,w_{fn})$ to each user or to each users group or to the whole population of users. w_{fi} represents the discriminating power of the factor f_i. In order to determine the w_{fi} values, orders of preferences must be given for all elements of $p_{m,x}$. Let us consider, in table 3, an example of preferences orders for a factors list given by a user. The evaluation method for elements of vector $p_{m,x}$ is given by: $\alpha_i=\beta\sum_{j>i}\alpha_j$, where α_1 and β are pre-defined, α_i is the weight assigned to factors having the preference order *i*. Thus, if there are *k* preferences orders, there will be *(k-1)* equations with *(k-1)* unknown to solve by using the method of *Gauss pivot* for example. We can thus calculate *a selection weight* (in IR or IF) and/or *a ranking weight* (in IR) for each information which will be evaluated with $p_{d,u}$ and $p_{m,x}$ using the *weighted mean* formula *fw*: $fw(p_{d,u},p_{m,x})=\frac{\sum_i v_{d,u,fi}.w_{fi}}{\sum_i w_{fi}}$, for instance. For the selection of information, it will be necessary to define a threshold that will help to decide if the *correspondence* of information to user is significant enough.

Table 3. *Preferences orders and factors weights*

Factors vector *m*	f_1	f_2	f_3	f_4	f_5	f_6	f_7	…	f_n
Preferences orders *i*	1	1	2	3	3	4	5	…	k
$P_{m,x}$	$W_{f1}=\alpha_1$	$W_{f2}=\alpha_1$	$W_{f3}=\alpha_2$	$W_{f4}=\alpha_3$	$W_{f5}=\alpha_3$	$W_{f6}=\alpha_4$	$W_{f7}=\alpha_5$	…	$W_{fn}=\alpha_k$

The algorithm of table 4 summarizes the stages to be followed for the restitution of personalized information to each user or to each users group.

Table 4. *Algorithm for a flexible and personalized access to information*

1. Choice of profiles to be used that describe elements of the architecture;
2. Determination of various matchings according to the descriptive criteria lists of the selected profiles;
3. Evaluation of the various matchings; *4. Combination of the various matching results*; *5. Restitution of information*

4. CONCLUSION

In this article, we present an architecture of retrieval and recommendation based on re-usable and adaptable profiles which are derived from a generic model. The genericity and the flexibility of the suggested approach guarantee a maximum co-operation and complementarities between any element interacting within the framework of the same process. We have shown that the combination of various matchings allow theoretically to improve the quality of results restituted to an individual. This architecture can be used as a basis in designing applications for information access.

Our future work will consist in: validating our proposals by experiments and tests on an application of retrieval and/or recommendation; proposing a model of profile that integrates various aspects of the semantic web like: semantics of criteria, ontology, etc.

The use of various matchings provides adapted answers to users. The initial objective was to work on personalization within the framework of information retrieval and recommendation. It remains nevertheless to check, by experiments, the real impact of this personalization on the restituted results.

REFERENCES

Baeza-Yates R., Ribeiro-Neto B., 1999. *Modern Information Retrieval.* First edition, Addison Wesley, ISBN 0-201-39829-X.

Balabanovic M., Shoham Y., 1997. Fab : Content-Based, Collaborative Recommendations. *Communications of the ACM*, vol. 40, n°. 3, pages 66-72.

Belkin N. J., Croft W. B., 1992. Information Filtering and information Retrieval : Two Sides of the same Coin? *Communications of the ACM, Information Filtering*, vol. 35, n° 12, pages 29-38.

Berti-Équille L., 2003. Quality-based recommendation of XML documents. *Journal of Digital Information Management*, vol. 1, n° 3, pages 117-128.

Boughanem M., Chrisment C., Soulé-Dupuy C., 1999. Query modification based on relevance backpropagation in adhoc environment. *Information Processing & Management Journal*, Elsevier Science, vol. 35, pages 121-139.

Bottraud J. C., Bisson G., Bruandet M. F., 2003. An Adaptative Information Research Personnal Assistant. *In proceedings of Workshop AI2IA (Artificial Intelligence, Information Access and Mobile Computing) IJCAI 2003*.

Cho Y. H., Kim J. K., Kim S. H., 2002. A personalized recommender system based on web usage mining and decision tree induction. *Expert System with Applications*, vol. 23, n° 3, pages 329-342.

Goldberg D., Nichols D., Oki B. M., Terry D., 1992. Using Collaborative Filtering to weave an Information Tapestry. *Communications of the ACM, Information Filtering*, vol. 35, n° 12, pages 61-70.

Good N., Schafer J., Konstan J., Borchers A., Sarwar B., Herlocker J., Riedl J., 1999. Combining Collaborative Filtering with Personal Agents for Better Recommendations. *In Proceedings of AAAI*, vol. 35, pages 439-446, AAAI Press.

Hoashi K., Kazunori M., Naomi I., Hashimoto K., 2000. Document filtering Method using non-relevant information profile. *In proceedings of the 23rd Annual International ACM-SIGIR Conference on research and development in information Retrieval*, pages 176-183.

Konstan J. A., Miller B. N., Maltz D., Herlocker J. L., Gordon L. R. and Riedl J., 1997. Grouplens: Applying Collaborative Filtering to Usenet News. *Communication of the ACM*, vol. 40, n°3, pages 77-87.

Korfhage R.R., 1997. *Information storage and retrieval*. Wiley computer publishing, ISBN 0-471-14-338-3.

Krulwich B., 1997. LifeStyle Finder : Intelligent User Profiling Using Large-Scale Demographic Data. *AI Magazine*, vol.18, n° 2, pages 37-45.

Montaner M., Lopez B., Rosa J. L. D. L, 2003. A Taxonomy of Recommender Agents on the Internet. *Artificial Intelligence Review*, vol. 19, pages 285-330, Kluwer Academic Publishers.

Pazzani M., Muramatsu J., Billsus D., 1996. Syskill & Webert : Identifying interesting web sites, *In Proceedings of the Thirteenth National Conference on Artificial Intelligence*, Pages 54-61.

Pazzani M., 1999. A Framework for Collaborative, Content-Based and Demographic Filtering. *Artificial Intelligence Review*.

Pitkow J., Schütze N., Cass T., Cooley R., Turnbull D., Edmonds A., Adar E. and Breuel T., 2002. Personalized Search : A contextual computing approach may prove a breakthrough in personalized search efficiency. *Communications of the ACM*, vol. 45, No. 9, Pages 50-55.

Rijsbergen C. J. V, 1979. *Information Retrieval*. Second edition, Butterworths.

Rucker J., Polanco M. J., 1997. Siteseer : Personalized Navigation for the Web. *Communications of the ACM*, vol. 40, n° 3, pages 73-75.

Widyantoro D. H., Ioerger T. R., Yen J., 1999. An Adaptative Algorithm for Learning Changes in User Interests. *In Proceedings of the Eighth International Conference on Information and Knowledge Management (CIKM'99)*, Pages 405-412, New York, ACM Press.