



HAL
open science

Architecture de Recherche et de Recommandation d'Information à base de Profils Définitions, acquisitions, usages de profils

Pascaline Tchienehom

► **To cite this version:**

Pascaline Tchienehom. Architecture de Recherche et de Recommandation d'Information à base de Profils Définitions, acquisitions, usages de profils. 22ième Congrès National Inforsid'04, May 2004, Biarritz,, France. p. 143-159. hal-01580564

HAL Id: hal-01580564

<https://hal.science/hal-01580564>

Submitted on 1 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Architecture de Recherche et de Recommandation d'Information à base de Profils

Définitions, acquisitions, usages de profils

Pascaline Tchienehom

IRIT – SIG

118 route de Narbonne

31062 Toulouse Cedex 4

Pascaline.Tchienehom@irit.fr

RÉSUMÉ. La pertinence de l'information est au centre des problématiques des techniques d'accès à l'information. Une solution à l'amélioration de cette pertinence est la personnalisation ou l'adaptation des réponses fournies aux utilisateurs. C'est à cette fin que nous proposons, dans cet article, une architecture de recherche et de recommandation dans laquelle tout élément (ou composant) de l'architecture est décrit de façon détaillée par un profil. L'originalité de cette architecture se situe au niveau de son aspect générique et des nombreuses possibilités d'interactions et de complémentarités entre profils instanciés à partir d'un modèle générique de profil. Par la suite, nous décrivons des méthodes générales d'acquisition et des règles générales d'usage de profils. Nos différentes propositions sont illustrées par des exemples.

ABSTRACT. The relevance of information is in the center of problems of access to information. 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 (or component) of the architecture 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 and complementarities between profiles derived from a generic model of profile. Thereafter, we describe general methods of acquisition and general rules of use of profiles. Our various proposals are illustrated by examples.

MOTS-CLÉS : Recherche et Filtrage d'Information, Profil, Personnalisation.

KEYWORDS: Information Retrieval and Filtering, Profile, Personalization.

1. Introduction

Avec l'expansion d'Internet et du Web, on assiste à une multiplication des documents mis à disposition et à celle des utilisateurs de ces documents. Les documents et les utilisateurs sont caractérisés par une très grande hétérogénéité. Rechercher, retrouver et exploiter une information dans ce contexte s'avère être une tâche ardue. Plusieurs outils d'accès à l'information (moteurs de recherche, systèmes de recommandation) ont été développés pour aider l'utilisateur dans cette tâche.

A travers ces outils, la question de la *pertinence système*¹ des résultats restitués aux utilisateurs a fait l'objet d'une réflexion très approfondie. Cependant, une autre question qui a été beaucoup moins approfondie est celle qui consiste à évaluer si ces résultats sont réellement adaptés à l'utilisateur, relativement à un certain nombre de critères. Il s'agit de s'assurer entre autre que les résultats obtenus sont compréhensibles par l'utilisateur, qu'ils correspondent aux buts et aux préférences de ce dernier, qu'ils sont compatibles avec son environnement logiciel et matériel, qu'ils sont effectivement utilisables par l'usager, etc. L'objectif est d'améliorer la pertinence en essayant de rapprocher le plus possible la *pertinence système* de la *pertinence utilisateur* (ou attentes de l'utilisateur). Pour cela, les résultats restitués doivent être personnalisés c'est-à-dire adaptés à chaque usager. Cette personnalisation des résultats restitués dépend de plusieurs facteurs liés à différentes caractéristiques relatives aux utilisateurs mais également aux informations recherchées. Il faut donc décrire avec le plus de détails possibles les éléments ou composants amenés à interagir au travers de processus d'accès à l'information. Ces descriptions sont appelées profils. L'objet des travaux entrepris est de proposer une architecture de recherche et de recommandation d'informations pertinentes et personnalisées. Cette architecture se base sur des profils adaptables et évolutifs et exploite leur complémentarité pour la restitution d'informations adaptées aux usagers.

Le présent article s'organise de la façon suivante : la section 2 présente un état de l'art sur la recherche et le filtrage (ou recommandation) d'information d'une part et sur la notion de profil d'autre part, en mettant en avant la nécessité de fournir des informations pertinentes et adaptées aux usagers. La section 3 présente l'architecture de recherche et de recommandation à base de profils proposée ainsi que le modèle générique de profil associé. Ensuite, nous présentons des méthodes générales qui permettent d'obtenir des valeurs pour les différents critères descriptifs de profils et notamment l'apprentissage à partir de profils existants. Nous y expliquons également les règles générales d'usage des critères de profils pour la restitution d'informations pertinentes et adaptées à l'utilisateur, notamment au travers d'une combinaison d'appariements entre critères comparables de profils. Nos différentes propositions sont illustrées par des exemples et nous concluons dans la section 4, sur les perspectives de notre travail.

¹ La pertinence système est le degré de similitude entre le contenu effectif d'un document et les besoins (centres d'intérêt ou requête) en information de l'utilisateur.

2. Etat de l'art

Le développement du Web a créé un besoin de techniques nouvelles pour aider les utilisateurs à trouver ce qu'ils recherchent mais aussi pour faire savoir qu'une information existe. Dans cette section, nous présentons dans un premier temps, une synthèse des techniques d'accès à l'information qui représentent notre contexte d'étude et dans un second temps, la notion de profil telle qu'elle est utilisée dans ces différentes techniques.

2.1. Techniques d'accès à l'information

Les techniques d'accès à l'information permettent à un individu d'obtenir des informations répondant à ses besoins. Nous pouvons les regrouper en deux grands groupes :

- celles qui reposent sur une approche « *service au comptoir* » ou « *pull* » : qui consistent à renvoyer des informations répondant à une demande explicite d'un individu. C'est le cas de la *Recherche d'Information* ;

- celles qui reposent sur une approche « *service à domicile* » ou « *push* » : qui consistent à renvoyer automatiquement à un individu des informations qui pourraient l'intéresser, sans qu'il n'en ait fait explicitement la demande. C'est le cas du *Filtrage (ou Recommandation) d'Information*.

Les sections suivantes présentent ces deux types d'approches (*push* et *pull*) au travers des processus de recherche et de filtrage d'information.

2.1.1. Recherche d'Information

Le processus de Recherche d'Information repose sur l'expression du besoin d'un individu au travers d'une requête formulée dans un langage libre plus ou moins structuré. En réponse à cette requête, un appariement est réalisé entre les termes (ou mots-clés) d'indexation de la requête et ceux des informations pré-indexées par le système. La recherche d'information est principalement basée sur le principe d'un appariement optimal, de type vectoriel ou probabiliste (Baeza-Yates *et al.*, 1999), (Rijsbergen, 1979). Enfin, le système propose traditionnellement à l'individu les informations pertinentes sous forme d'une liste ordonnée selon leur degré de pertinence décroissant.

Cependant en Recherche d'Information, l'intention réelle de l'utilisateur n'est pas toujours évidente dans sa manière de formuler sa requête et cela peut générer des ambiguïtés au niveau du sens des mots qu'elle contient. De nombreuses solutions existent pour préciser le sens d'une requête et on peut citer en particulier : l'utilisation des jugements de pertinence d'un utilisateur sur un ensemble d'informations pour reformuler sa requête et affiner ainsi la recherche. C'est la méthode de réinjection de pertinence ou *relevance feedback* (Boughanem *et al.*, 1999) ; l'utilisation du profil (des besoins) de l'utilisateur ou mots-clés représentant les centres d'intérêt de l'individu et de la notion de contexte de l'utilisateur, pour aider à

l'interprétation de ses requêtes afin de ré-évaluer et de ré-ordonnancer les résultats d'une recherche (Bottraud *et al.*, 2003) ; etc.

2.1.2. Filtrage d'Information

Alors que la Recherche d'Information (RI) est une tâche très interactive, celle du Filtrage d'Information (FI) est relativement passive (Belkin *et al.*, 1992) car l'utilisateur ne formule pas explicitement ses besoins au travers d'une requête (ou expression d'un besoin ponctuel) comme c'est le cas en RI. En Filtrage d'Information, on utilise plutôt une représentation de l'utilisateur appelé profil utilisateur pour lui envoyer des informations. Ces informations proviennent généralement d'un flux dynamique ou sont obtenues grâce à un agent. Elles sont ensuite comparées aux différents profils disponibles pour déterminer ceux auxquels elles correspondent.

Il existe plusieurs méthodes de filtrage (Montaner *et al.*, 2003) :

- le *filtrage cognitif ou basé sur le contenu* qui utilise la description du contenu des informations pour déterminer à quels profils utilisateurs elles correspondent (Pazzani *et al.*, 1996). Le profil utilisateur, en filtrage cognitif, décrit les centres d'intérêt durables ou récurrents de l'individu qui sont représentés communément par une liste de mots-clés pondérés (Korfhage, 1997). Dans cette méthode de filtrage, on peut aussi retrouver les approches réclusives (Yager, 2002) qui sont basées sur la recherche d'une similarité entre objets en comparant leur description respective ;

- le *filtrage social ou collaboratif* qui utilise les jugements (ou feedback) d'un ensemble d'utilisateurs concernant un ensemble d'informations pour effectuer des recommandations. On utilise une mesure de similarité entre individus et/ou jugements d'individus pour déterminer si une information correspond à un individu donné (Goldberg *et al.*, 1992). La description du contenu réel des informations est ignorée ;

- le *filtrage démographique* qui utilise les données démographiques de l'utilisateur (âge, profession, ville d'origine, etc.) pour lui faire des recommandations. Pour cela, on se base sur une catégorisation des informations en fonction des données démographiques des individus (Krulwich, 1997).

Ces approches ne sont pas exclusives et différentes méthodes hybrides, combinant ces différents types de filtrage, ont été développées (Good *et al.*, 1999), (Pazzani, 1999). L'utilisation des approches hybrides permet d'améliorer la pertinence des résultats des systèmes de filtrage en palliant certaines limites des types de filtrage présentés précédemment (Balabanovic *et al.*, 1997) comme : la sur-spécialisation en filtrage basé sur le contenu ; l'obtention des jugements qui est une tâche coûteuse pour les utilisateurs ; etc.

Pour résumer, les différentes techniques d'accès à l'information partagent le même objectif qui est d'aider l'utilisateur à obtenir des informations pertinentes. Cependant, nous pouvons souligner la nécessité grandissante de réunir le maximum de critères descriptifs sur les informations manipulées par les processus de recherche

et de recommandation d'information afin de garantir une meilleure adaptation des réponses renvoyées aux utilisateurs. Mieux les informations sont décrites, plus il est aisé de satisfaire les utilisateurs en s'adaptant aux caractéristiques particulières de chacun d'eux à travers une personnalisation de l'information restituée par ces processus. Cette description des informations est désignée sous le nom de profil.

2.2. Notion de profil

De façon générale, le profil d'un objet est un ensemble de caractéristiques permettant de l'identifier ou de le représenter. Plusieurs types de profils sont exploités dans les différentes techniques d'accès à l'information présentées dans la section précédente :

- *le profil utilisateur* : il s'agit de la description des caractéristiques d'un utilisateur (données démographiques, centres d'intérêt, préférences, etc.). Plusieurs approches d'acquisition des éléments d'un profil utilisateur existent et peuvent être regroupées en approches manuelles et en approches automatiques ou semi-automatiques. Parmi les approches automatiques ou semi-automatiques, nous pouvons distinguer le profiling (Cho *et al.*, 2002) et les approches par stéréotypes (Shapira *et al.*, 1997). Le profiling consiste à scruter, enregistrer et analyser les actions et successions d'actions d'un utilisateur lors de différentes sessions de recherche pour déterminer son profil. En revanche, l'approche par stéréotype est fondée sur l'identification de groupes d'utilisateurs et la détermination des caractéristiques clés de chaque groupe ;

- *le profil de document* : il correspond à la description d'un document qui est souvent réduite, en RI ou FI, à une liste de mots-clés pondérés décrivant le contenu sémantique du document. Les mots-clés et leurs poids sont obtenus en général par une opération d'indexation (Rijsbergen, 1979). Plusieurs travaux permettent actuellement de décrire les documents en utilisant également d'autres critères que ceux liés à leur contenu effectif. On peut citer par exemple les métadonnées du Dublin Core² pour la description de ressources : Titre, Auteur, Sujet, etc. Nous pouvons également citer les travaux de Lainé-Cruzel (Lainé-Cruzel, 1999) qui permettent de définir des propriétés liées à l'ensemble d'un document (profession de l'auteur, type de document, etc.) ainsi que celles relatives à des parties de documents (type d'unité documentaire, forme discursive, style, etc.) afin de restreindre les documents pertinents (du point de vue du sujet dont ils traitent) aux seuls documents exploitables et réellement utilisables. De même, une liste non exhaustive de métadonnées pour l'annotation qualitative de documents est donnée par Laure Berti-Equille (Berti-Equille, 2002) dans le contexte de la recommandation multi-critères.

A travers la notion de profil, les techniques d'accès à l'information tentent d'améliorer la pertinence des réponses renvoyées aux utilisateurs. A l'heure actuelle, la tendance est à la prise en compte de métadonnées (ou propriétés), obtenues par annotation, qui permettent d'améliorer la pertinence des résultats restitués (Berti-

² Le Dublin Core est constitué de quinze éléments de métadonnées décrits à l'adresse : <http://dublincore.org/documents/dces/>.

Equille, 2002), (Lainé-Cruzel, 1999).

En résumé à notre état de l'art, nous remarquons que dans les approches de recherche et de filtrage d'information classiques, les systèmes se focalisent principalement sur les besoins des utilisateurs (requête ou centres d'intérêt) et sur le contenu effectif des informations mises à disposition pour répondre aux attentes des usagers. Cependant, d'autres caractéristiques comme : le format des informations, l'environnement des usagers, les préférences ou goûts de ces derniers..., ne sont pas prises en compte. Or, elles peuvent jouer un rôle important dans la satisfaction des besoins des utilisateurs. Certains travaux, présentés dans la section précédente, essaient de résoudre ce problème notamment par la définition de métadonnées sur des documents.

Nos travaux se situent également dans ce cadre mais nous nous inscrivons plus dans une démarche générique que spécifique. Pour cela, nous proposons une architecture à base de profils adaptables et évolutifs dont l'originalité se situe au niveau de son aspect générique et des nombreuses possibilités d'interactions entre profils. Ces interactions se traduisent par : des méthodes d'acquisition de valeur de profils ; des appariements entre critères comparables de profils différents ; la combinaison de différents profils pour retrouver et renvoyer des réponses qui correspondent davantage aux utilisateurs. L'objectif de tout ceci est d'améliorer la qualité des résultats qui sont restitués par les processus de recherche et de recommandation.

Par ailleurs, malgré l'existence de standards comme le Dublin Core, chaque application choisit ses critères de description de profils en fonction de son contexte et des buts qu'elle veut atteindre. Dans cet article, nous proposons également un modèle générique de profil utilisable par différentes applications pour décrire des profils.

3. Architecture de recherche et de recommandation à base de profils

Dans cette section, nous présentons notre proposition d'architecture de recherche et de recommandation à base de profils. Nous décrivons également le modèle générique de profil à partir duquel nous pouvons dériver différents profils. Ensuite, nous expliquons les méthodes générales d'acquisition des valeurs des critères descriptifs de profils ainsi que les règles générales d'usage de ces derniers pour la restitution d'informations pertinentes et adaptées aux utilisateurs. Nous illustrons nos différentes propositions par des exemples.

3.1. Présentation de l'architecture

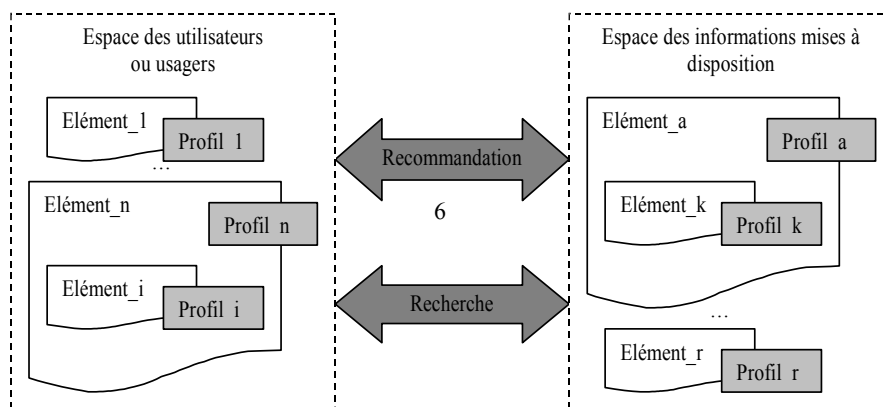


Figure 1. *Architecture de recherche et de recommandation à base de profils*

Le schéma de la figure 1 présente l'architecture de recherche et de recommandation à base de profils que nous proposons. Cette architecture est assez générale pour servir de modèle à différentes applications. Elle résulte de l'analyse de la mise en œuvre de différents systèmes de recherche et de recommandation. Chacun de ces systèmes a été conçu pour atteindre des objectifs particuliers en fonction des spécificités propres de son contexte : recommandation de pages web en fonction des signets (Rucker *et al.*, 1997), filtrage de mails (Goldberg *et al.*, 1992), commerce électronique (Cho *et al.*, 2002), etc.

Notre architecture ne s'applique pas à un cadre pré-défini. Elle est constituée de d'un ensemble d'éléments qui interviennent dans la mise en œuvre d'un système de recherche et de recommandation et c'est à chaque application de sélectionner dans cette architecture les éléments qui l'intéressent. Elle peut être utilisée comme point de départ pour toute construction de systèmes de recherche et de recommandation.

Sont mis en évidence, dans cette architecture, les processus de recherche et de recommandation ainsi que la structure générique des éléments manipulés par ces derniers. Ces éléments sont regroupés en deux grands groupes : ceux qui sont relatifs à l'espace des utilisateurs et ceux relatifs à l'espace des informations mises à disposition. Notons qu'à chaque type d'élément on associe un profil qui le décrit de façon détaillée et qui est exploitable par les processus d'accès à l'information. De plus, ces éléments peuvent aussi être composés d'un ou de plusieurs sous éléments eux-mêmes décrits par des profils.

Les éléments **liés à l'espace des utilisateurs** peuvent être, par exemple :

- *les informations d'usage* des différents utilisateurs en session de recherche : historique des requêtes, des sites visités, des informations (sites, documents, etc.) jugées, des informations sauvegardées, des informations transférées, etc. ;

- *les informations sur l'environnement de travail* des utilisateurs : structure et contenu de leur espace de travail (signets, courriers, documents divers, environnement logiciel et matériel, etc.) ;

- *les informations sur les utilisateurs* : données démographiques, centres d'intérêt, préférences, etc. Ces informations sont obtenues soit manuellement, soit par apprentissage (*cf.* section 2.2 et section 3.3) ; *etc.*

Notons que les éléments de l'espace des utilisateurs peuvent être définis par individu ou par groupe d'individus. Les profils de ces éléments peuvent donc être combinés pour décrire des individus ou groupe d'individus afin de constituer leur profil qui peut être : de court terme ou de long terme (Widyantoro *et al.*, 1999), positif ou négatif (Hoashi *et al.*, 2000), etc.

De même, les éléments **liés à l'espace des informations mises à disposition** peuvent être, par exemple :

- les *informations mises à disposition* qui peuvent être : des documents, des parties ou granules de documents (chapitres, paragraphes, sections, etc.), des collections de documents, des sites web, etc. Ces informations sont éventuellement *annotées* par des utilisateurs ou par des experts ou auteurs ;

- les *informations sur les ressources physiques* (supports d'information, serveurs, etc.) ; etc.

La figure 2 présente un exemple d'architecture dérivée de la figure 1.

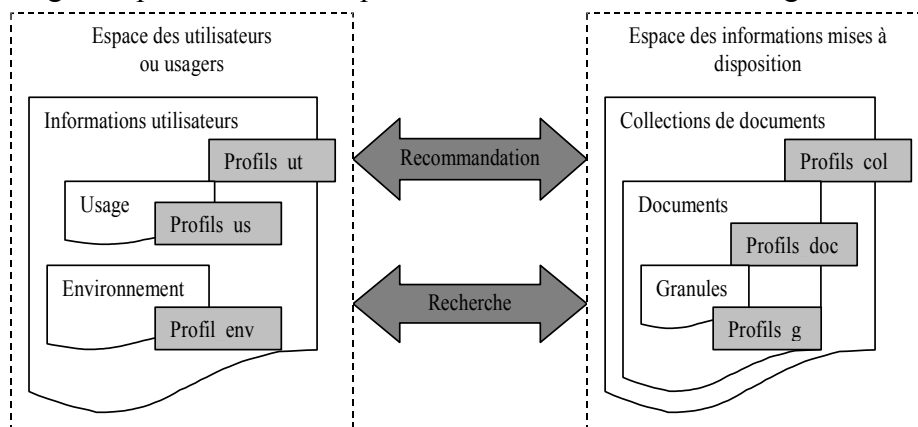


Figure 2. Exemple d'Architecture à base de profils

L'objectif de cette architecture est la restitution d'informations personnalisées à travers l'usage de profils. Ces profils sont issus d'un modèle générique que nous présentons dans la section suivante.

3.2. Définition de profils pour la recherche et la recommandation

Afin de pouvoir définir différents profils qui soient adaptables et évolutifs, nous avons défini un modèle générique de profil.

3.2.1. Modèle générique de profil

Le modèle générique de profil de la figure 2a présente la structure générale d'un profil. Cette structure est sous la forme d'une hiérarchie de catégories de critères permettant de caractériser un profil. Cette hiérarchie est une forêt ou un ensemble d'arbres dont les nœuds sont des catégories ou sous catégories de critères et les feuilles sont tout simplement des critères auxquels on peut affecter des valeurs. Un profil peut donc être soit une forêt, soit un arbre, soit une végétation (ou liste) de critères. Ainsi, si P est un type de profil : $Structure(P) \in \{forêt, arbre, végétation\}$.

L'organisation des différents critères par catégorie permet de regrouper les critères similaires dans une même classe et de définir ainsi une nomenclature (ou taxinomie) des critères. A partir du modèle générique de profil, nous pouvons

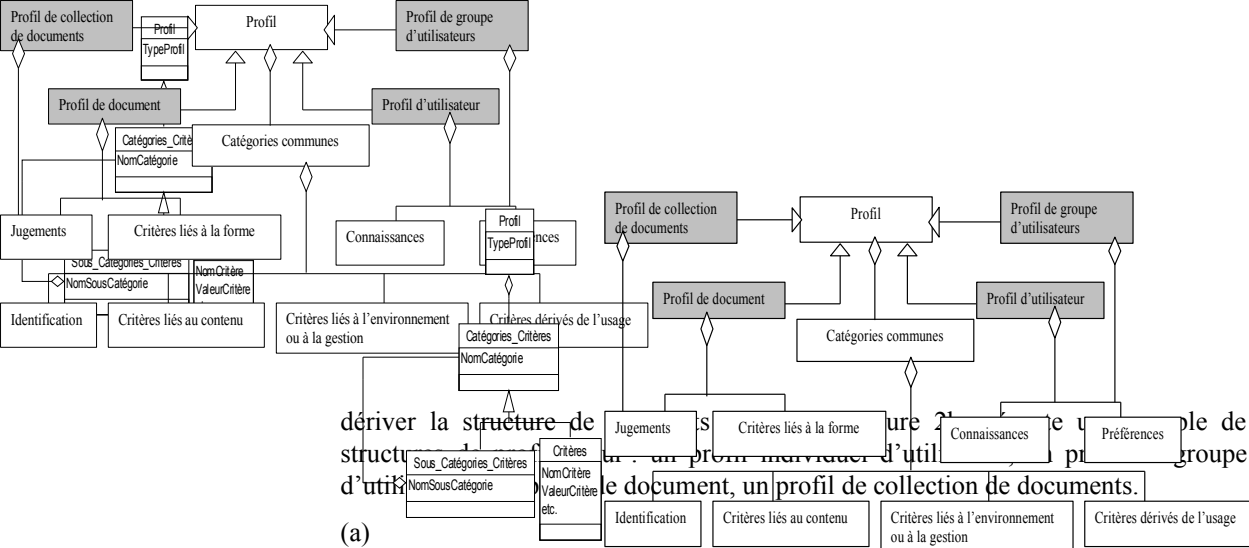


Figure 2. (a) *Modèle générique de profil* (b) *Exemple de structures de profils*

L'intérêt de l'utilisation d'un profil générique pour définir un type de profil donné est que la structure de base qu'il propose peut être utilisée par tout type d'application afin de définir tout type de profils. Ici, nous ne nous intéressons pas à des critères particuliers mais plutôt à la mise en place d'un cadre pour la définition de ces critères. Notons que pour chaque critère il faut définir un certain nombre de propriétés (du critère) pour faciliter son usage et son acquisition (*cf.* sections 3.3 et 3.4) : *nom, valeur, type, représentation, sémantique, sécurité, etc.*

Dans la section suivante, nous définissons des exemples d'instances de profils dérivés de notre modèle générique.

3.2.2. Exemples d'instances de profils dérivés du modèle générique

Le modèle générique de profil de la figure 2a va nous permettre de dériver la structure de différents profils sous la forme d'une hiérarchie de catégories ou classes de critères. Dans le contexte de la recherche et de la recommandation d'informations textuelles nous pouvons envisager, par exemple, les modèles de profils illustrés par le tableau 1 et décrivant une taxinomie pour les profils individuels d'utilisateurs, de groupes d'utilisateurs, de documents et de sources ou collections de documents.

Soulignons que toutes les catégories ne sont pas forcément renseignées pour un profil donné. De plus, en fonction de l'application qui l'utilise toutes les catégories d'un profil ne sont pas forcément prises en compte. Chaque application choisit les éléments à considérer dans un profil en fonction de l'objectif qu'elle veut atteindre.

Dans le tableau 1, chaque catégorie est composée d'une ou de plusieurs sous catégories. Ces listes ne sont pas exhaustives et on peut ajouter de nouvelles catégories ou sous catégories selon le besoin. On peut également effectuer d'autres décompositions sur les sous catégories. La structuration des profils en hiérarchie de catégories permet de modéliser avec le plus de détails possibles les informations ou éléments que l'on souhaite décrire. L'objectif est d'aider à trouver une information qui corresponde à l'utilisateur ou à faire savoir qu'elle existe.

Catégories de critères	Profil individuel d'utilisateur	Profil de groupe d'utilisateurs	Profil de document	Profil de source
Identification	Login	Nom du groupe	Adresse	Adresse
	Noms	Membres du groupe	Auteurs	Propriétaires
	Date de naissance		Titre	
	Profession			

Critères liés au contenu	Centres d'intérêt	Centres d'intérêt communs	Mots-clés	Thèmes
Critères liés à l'environnement ou à la gestion	Logiciels	Logiciels communs	Accessibilité	Accessibilité
	Matériels	Matériels communs	Disponibilité	Disponibilité
			Sécurité	Sécurité
Critères dérivés de l'usage			Popularité	Popularité
			Fraîcheur	Fraîcheur
Jugements	Jugements : pertinence, etc.	Jugements divers sur des informations	Crédibilité	Crédibilité
			Public cible	
Critères liés à la forme			Taille	
			Langue	
			Format	
Connaissances	Langues	Langues communes		
	Spécialités	Spécialités communes		
Préférences	Auteurs	Auteurs communs		
	Formats	Formats communs		

Tableau 1. Exemple de taxinomie des critères de description d'un profil individuel d'utilisateur, d'un profil de groupe d'utilisateurs, d'un profil de document et d'un profil de source ou collection de documents.

La généralisation de la structure d'un profil quelconque constitué de plusieurs critères (critères de type feuille) pas toujours tous renseignés, permet de garder le même profil pour un élément donné de l'architecture. Un profil peut donc être partagé et enrichi par différentes applications : ce sont des profils adaptables. Par exemple, un profil utilisateur peut être considéré comme une « carte d'identité » de la personne. Quelle que soit l'application, l'utilisateur peut être reconnu avec le même profil (critères de type feuille et contenu).

Dans la section suivante, nous présentons des méthodes générales pour acquérir les valeurs des critères de profils.

3.3. Méthodes d'acquisition des valeurs des critères de profils

Les profils de l'architecture proposée ne sont pas statiques et peuvent évoluer. Pour cela, leurs critères vont nécessiter une combinaison de différentes méthodes pour la mise à jour de leurs valeurs. Ces méthodes sont pour la plupart des extensions de méthodes existantes car elles vont se baser principalement sur l'usage des profils existants des éléments de l'architecture. Nous les avons classifiées comme suit :

- les *annotations* : elles vont permettre d'obtenir manuellement la valeur de certains critères. Notons qu'il existe des critères pour lesquels l'annotation est

indispensable : date de naissance d'un usager, auteur(s) d'un document, etc. ;

- l'*indexation* : elle permet de déterminer les mots-clés (avec leur poids ou degré de représentativité) représentant le contenu effectif des documents ou des besoins (requête ou centres d'intérêt) des utilisateurs. Notons que pour déterminer les centres d'intérêt d'un utilisateur on procède à l'analyse des profils, ou directement à celle du contenu brut, de certaines informations : requêtes fréquentes, sites fréquemment visités, signets, documents jugés ou sauvegardés ou transférés, etc. ;

- le *profiling* : il permet l'apprentissage de profils utilisateurs individuels en analysant les profils, ou en analysant directement le contenu brut, des informations d'usage ou des informations de l'environnement de travail (*cf.* section 3.1) de ces derniers. Le but est de pouvoir détecter des critères récurrents (requêtes fréquentes, etc.) ou relativement invariants (matériels, logiciels) ;

- les *approches par stéréotypes* : elles permettent l'apprentissage automatique de profils de groupe (utilisateurs, documents, collections de documents, etc.) à partir de profils individuels de même type via la notion de similarité entre critères de profils ;

- l'*apprentissage inter-profils* : il permet d'apprendre certains critères de profils à partir des critères d'autres types de profils. Par exemple : nous pouvons apprendre les centres d'intérêts d'un utilisateur à partir des mots-clés des documents qu'il a jugés pertinents ou des mots-clés de ses signets ; nous pouvons déterminer les thèmes abordés en majorité dans une collection de documents en analysant les mots-clés décrivant le contenu des documents de cette collection ; certains critères d'un profil de document comme l'accessibilité ou la sécurité vont dépendre de la valeur de ces mêmes critères dans le profil de source de ce document ; etc.

Nous décrivons dans la section suivante, les règles générales d'utilisation des critères des différents profils dans les processus de recherche et de recommandation d'information afin de garantir des résultats plus adaptés aux usagers.

3.4. Règles d'usage des profils

Les règles d'usage des profils sont définies par la complémentarité entre profils des éléments de l'architecture notamment à travers : des méthodes d'acquisition de valeurs faisant usage de profils (*cf.* section 3.3) ; des appariements de profils ; etc.

3.4.1 Appariements de profils

Nous avons identifié différents appariements de critères de profils. La combinaison de certains de ces appariements va permettre de sélectionner les résultats correspondant aux utilisateurs ou tout simplement de réordonner ces résultats. Pour appairer des critères de profils différents il faut que ces critères soient de même type, qu'ils soient représentés dans le même référentiel et aussi qu'ils aient la même sémantique. Nous avons classifié ces appariements comme suit :

1. *appariement multi-termes de critères de valeur textuelle* : ce type d'appariement est utilisé lorsque au moins un des critères à appairer est multivalué. Dans ce cas, nous représentons les différents critères à appairer dans un même espace

vectorel à la dimensionnalité donnée par la taille du vocabulaire. A chaque vecteur de termes $d=(t_1, t_2, \dots, t_n)$ est associé un vecteur de poids réel ou booléen $p=(w_{d,t_1}, w_{d,t_2}, \dots, w_{d,t_n})$ qui permettra de calculer une mesure de similarité entre critères à appairer.

Exemples

- *correspondance aux besoins de l'usager* : il s'agit d'une mesure de similarité entre le vecteur des poids des termes représentant les besoins (requête éventuellement reformulée ou centres d'intérêt) de l'utilisateur et celui des termes représentant le contenu d'une information (document, granule de document, collection de documents, etc.). La définition d'un seuil va permettre de décider de la sélection de l'information. Notons que l'utilisation de profil de collection de documents va permettre de repérer les collections de documents spécialisées dans des domaines spécifiques et aider ainsi à résoudre le problème de la localisation de sources pertinentes (Gravano, 1999). De même, la définition d'un profil de granule (ou partie) de document va permettre de renvoyer des granules pertinents ou de renvoyer aux usagers des documents jugés globalement non pertinents mais qui auraient des granules qui correspondent aux besoins de ces derniers ;

- *compatibilité avec la langue* : il s'agit de s'assurer que l'utilisateur possède les connaissances requises en langue pour comprendre le document. Ainsi, soit x la langue d'un document et $L=\{\text{ensemble des langues compréhensibles par l'utilisateur}\}$, il faut vérifier que $\exists y \in L : y=x$;

- *correspondance avec le public cible* : il s'agit de s'assurer que l'utilisateur (âge, profession, spécialité, connaissances, etc.) correspond au public cible (âge minimum, profession, spécialité, etc.) d'un document ;

- *compatibilité à l'environnement de l'utilisateur* : il s'agit de garantir que l'utilisateur peut effectivement exploiter le document qui lui est renvoyé. Pour cela, on procède à la vérification de la compatibilité entre les caractéristiques physiques de l'information à restituer (format du document, etc.) et l'environnement matériel et logiciel de l'utilisateur. Ainsi, soient L et M respectivement l'ensemble des logiciels et l'ensemble des matériels nécessaires à l'exploitation d'une information, il faut vérifier que : $\forall y \in L, \forall z \in M, y \in L1 = \{\text{Environnement logiciel de l'utilisateur}\}$ et $z \in M1 = \{\text{Environnement matériel de l'utilisateur}\}$; etc.

2. *appariement mono-terme de critères de valeur textuelle ou numérique* : ce type d'appariement est utilisé lorsque les critères à appairer ont une seule valeur ; On peut donc effectuer un appariement terme à terme ou terme à valeur.

Exemples

- *compatibilité aux préférences de l'utilisateur* : il s'agit de s'assurer qu'une information correspond bien aux critères de préférences d'un utilisateur. Pour cela, des comparaisons sont effectuées entre les critères des préférences de l'utilisateur et les critères correspondant de description de l'information comme par exemple : langue, auteur, format, popularité, taille, etc. ;

- *sélection multi-critères de type simple* : il s'agit d'un appariement des critères de type simple (texte ou numérique) de description d'une information avec des valeurs pré-définies de ces critères. Ainsi, $\forall x \in L = \{\text{listes de critères de type simple}\}$, des relations de la forme $x \theta \alpha$ peuvent être définies où $\theta \in \{\neq, \leq, \geq, =, <, >\}$ et α est un nombre ou une chaîne de caractères ; etc.

3. *combinaison d'appariements* : pour combiner différents types d'appariements, ces derniers doivent tout d'abord être effectués séparément. Ainsi, chaque résultat d'appariement (ou combinaison de résultats d'appariements) va représenter un critère potentiel de sélection ou d'ordonnement des informations. Un exemple de liste de critères de sélection ou d'ordonnement des informations peut être : *correspondance aux besoins, compatibilité avec la langue, compatibilité au critère de préférence en taille, etc.* On peut donc décrire les informations recherchées sous forme de listes de critères $d = (c_1, c_2, \dots, c_n)$. A chaque liste on va associer un vecteur de poids $p = (w_{d,c_1}, w_{d,c_2}, \dots, w_{d,c_n})$ relatif au pouvoir discriminant des critères les uns par rapport aux autres. Le vecteur de poids p peut être utilisé pour : ordonner les résultats obtenus selon une sous liste de d ; sélectionner des informations en calculant à partir de p un degré de correspondance de d à l'utilisateur et il va falloir, dans ce cas, définir un seuil pour la sélection de d .

L'algorithme ci-dessous résume les différentes étapes pour la restitution d'informations adaptées à un utilisateur.

Algorithme de restitution d'informations à un utilisateur

1. *Choix des profils à utiliser décrivant des éléments de l'architecture*
2. *Choix des critères utilisables pour les appariements en fonction de la liste de critères de description des profils sélectionnés*
3. *Effectuer les différents appariements*
4. *Combiner les résultats des différents appariements*
5. *Restituer les informations sélectionnées*

Dans la section suivante, nous illustrons une application de notre algorithme à travers un exemple.

3.4.2 *Illustration de la personnalisation sur un exemple de Recherche d'Information*

Dans cette illustration, nous utilisons deux types de profils : profil individuel d'utilisateur et profil de document. Soit la requête « bases de données », formulée par les utilisateurs U1, U2 et U3 de profils respectifs :

<i>Profils utilisateurs</i>	Login	Profession	Spécialité	Langue	Environnement
U1	emanuel	étudiant	informatique	anglais, français	word, excel, gsvie, etc.
U2	carmand	étudiant	biologie	anglais, français	word, excel, access, etc.
U3	jmarcel		Informatique	français	word, excel, gsvie,

					etc.
--	--	--	--	--	------

Soient les profils de documents D11, D12, D21 et D22 suivants :

<i>Profils de documents</i>	Adresse	Mots clés	Langue	Format	Public cible
D11	X1	Design, relational, database, Firebird	anglais	html	Profession=étudiant
D12	X2	Evolution, bases, données, méthodes, outil	français	html	Spécialité=informatique
D21	Y1	Tutoriel, bases, données, relationnelles	français	ps	
D22	Y2	Systèmes, gestion, bases, données, relationnelles	français	html	Profession=étudiant

Le choix des critères utilisables pour les appariements va permettre de définir la liste de critères $d=(c_1, c_2, c_3, c_4)$ pour représenter les documents. Les c_i représentent respectivement et par ordre de pouvoir discriminant : *la correspondance aux besoins, la compatibilité avec la langue, la correspondance avec le public cible, la compatibilité à l'environnement de l'utilisateur*. Dans cet exemple, le principe de recherche consiste à sélectionner les documents selon les résultats des appariements c_1 et c_2 et à ordonnancer les résultats sélectionnés selon les critères de d .

Les résultats, dans l'ordre, restitués aux différents utilisateurs suite à leur requête sont :

- l'utilisateur U1 recevra les documents D11, D12, D21 et D22 ;
- l'utilisateur U2 recevra les documents D11, D22, D12 et D21. Le document D12 est en troisième position parce l'utilisateur ne correspond pas au public cible de ce document. Le document D21 est en fin de liste car l'environnement de l'utilisateur doit être mis à jour, par une installation de gsvie par exemple, pour pouvoir exploiter ce document qui est au format *ps* ;
- l'utilisateur U3 ne recevra que les documents D12, D21 et D22 parce qu'il ne comprend pas l'anglais qui est la langue dans laquelle est écrit le document D11.

Par cet exemple, nous avons illustré le fait que la combinaison de différents appariements permet de restituer des réponses mieux adaptées aux usagers.

4. Conclusion

Dans cet article, nous présentons une architecture de recherche et de recommandation à base de profils adaptatifs et évolutifs qui sont dérivés d'un modèle générique. La généralité de l'approche proposée garantit une coopération et une complémentarité maximales entre tout élément interagissant dans le cadre d'un même processus. Nous montrons également que la combinaison de différents appariements permet d'améliorer théoriquement la qualité des résultats renvoyés à un individu. Cette architecture peut servir de base pour toute conception d'applications visant à faciliter l'accès à l'information : Recherche d'Information,

Filtrage d'Information, aide à la découverte d'informations, etc.

En terme de perspectives à notre travail nous comptons : valider l'architecture proposée par des expérimentations et des tests sur des applications de recherche et/ou de recommandation ; proposer un modèle de profil intégrant différents aspects comme : le respect des données privées, la sémantique des critères, etc.

L'utilisation de différents appariements permet de fournir des réponses mieux adaptées aux usagers c'est-à-dire des informations pertinentes et personnalisées. L'objectif en premier lieu a été de travailler sur la personnalisation dans le cadre de la recherche et de la recommandation d'information. Il reste néanmoins à vérifier, par expérimentations, l'impact réel de celle-ci sur les résultats restitués.

Bibliographie

- Baeza-Yates R., Ribeiro-Neto B., *Modern Information Retrieval*, First edition, Addison Wesley, ISBN 0-201-39829-X, 1999.
- Balabanovic M., Shoham Y., Fab : Content-Based, Collaborative Recommendations, *Communications of the ACM*, vol. 40, n° 3, pages 66-72, 1997.
- Belkin N. J., Croft W. B., Information Filtering and information Retrieval : Two Sides of the same Coin?, *Communications of the ACM, Information Filtering*, vol. 35, n° 12, pages 29-38, 1992.
- Berti-Equille L., Annotation et recommandation collaboratives de documents selon leur qualité, *RSTI série ISI-NIS Recherche et filtrage d'information*, vol. 7, n° 1-2, pages 125-155, 2002.
- Boughanem M., Chrisment C., Soulé-Dupuy C., Query modification based on relevance backpropagation in adhoc environment, *Information Processing & Management Journal*, Elsevier Science, vol. 35, pages 121-139, 1999.
- Bottraud J. C., Bisson G., Bruandet M. F., An Adaptative Information Research Personal Assistant, *In proceedings of Workshop AI2IA (Artificial Intelligence, Information Access and Mobile Computing) IJCAI 2003*, 2003.
- Cho Y. H., Kim J. K., Kim S. H., A personalized recommender system based on web usage mining and decision tree induction, *Expert System with Applications*, vol. 23, n° 3, pages 329-342, 2002.
- Goldberg D., Nichols D., Oki B. M., Terry D., Using Collaborative Filtering to weave an Information Tapestry, *Communications of the ACM, Information Filtering*, vol. 35, n° 12, pages 61-70, 1992.
- Good N., Schafer J., Konstan J., Borchers A., Sarwar B., Herlocker J., Riedl J., Combining Collaborative Filtering with Personal Agents for Better Recommendations, *In Proceedings of AAAI*, vol. 35, pages 439-446, AAAI Press, 1999.
- Gravano L., Garcia-Molina H., Tomasic A, GLOSS : Text-Source Discovery over the Internet, *ACM transactions on Database systems*, vol. 24, n° 2, pages 229-264, 1999.
- Hoashi K., Kazunori M., Naomi I., Hashimoto K., 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, 2000.
- Korfhage R.R., *Information storage et retrieval*, Wiley computer publishing, ISBN 0-471-14-338-3, 1997.
- Krulwich B., LifeStyle Finder : Intelligent User Profiling Using Large-Scale Demographic Data, *AI Magazine*, vol.18, n° 2, pages 37-45, 1997.
- Lainé-Cruzet S., ProfilDoc : Filtrer une information exploitable, *Bulletin des Bibliothèques de France*, n° 5, pages 60-65, http://www.enssib.fr/bbf/bbf-99-5/10_lainecruzet.pdf, 1999.
- Montaner M., Lopez B., Rosa J. L. D. L, A Taxonomy of Recommender Agents on the Internet, *Artificial Intelligence Review*, vol. 19, pages 285-330, Kluwer Academic Publishers, 2003.
- Pazzani M., Muramatsu J., Billsus D., Syskill & Webert : Identifying interesting web sites, *In Proceedings of the Thirteenth National Conference on Artificial Intelligence*, Pages 54-61, 1996.
- Pazzani M., A Framework for Collaborative, Content-Based and Demographic Filtering, *Artificial Intelligence Review*, 1999.
- Rijsbergen C. J. V, *Information Retrieval*. Second edition, Butterworths, 1979.
- Rucker J., Polanco M. J., Siteseer : Personalized Navigation for the Web, *Communications of the ACM*, vol. 40, n° 3, pages 73-75, 1997.
- Shapira B., Shoval P., Hanani U., Stereotypes In Information Filtering Systems, *Information Processing & Management*, vol. 33, n° 3, pages 273-287, 1997.
- Widyantoro D. H., Ioerger T. R., Yen J., 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, 1999.
- Yager R. R, Fuzzy logic methods in recommender systems, *Fuzzy sets and Systems*, vol. 136, pages 133-149, Elsevier, 2002.