Modèle évolutif d'un profil utilisateur
Application à la Recherche d'Information dans une bibliothèque numérique de thèses

Suela BERISHA-BOHE, Béatrice RUMPLER

INSA - Lyon, LIRIS
7, Avenue Jean Capelle,
Bâtiment 502 – Blaise Pascal
F69621 Villeurbanne Cedex
{suela.bohe; beatrice.rumpler}@insa-lyon.fr

RÉSUMÉ. *La prise en compte des besoins, des intentions et des spécificités cognitives, culturelles ou autres, qui caractérisent le profil d'un utilisateur constitue un élément déterminant pour améliorer la pertinence des réponses lors d'une session de Recherche d'Information dans de grandes bases de documents. La modélisation des profils et la manière de les adapter à différents utilisateurs qui n'ont pas une idée précise sur l'information qu'ils recherchent, nous permet d'offrir un accès personnalisé au contenu de documents scientifiques fondé sur l'exploitation du profil d'utilisateur. Nous proposons dans cet article, un modèle de l'utilisateur fondé sur les connaissances et un système implémentant le raisonnement à partir des cas, pour acquérir ces connaissances, les structurer et les faire évoluer.*

ABSTRACT. *Taking into account the needs, the intention and cognitive, cultural or various specificities to characterize the user profile, is a major challenge to improve the relevance of information retrieval systems. The users' models and the way to adapt them to different users (who need help to build their request during query processing) allow using a personalised access to scientific documents based on user profile. So, we propose a user model based on users' knowledge and users' preferences. We have defined a system based on CBR to capture users' preferences and knowledge, to structure them and to manage the user profile evolution. We validate some results by means of a prototype.*

MOTS-CLÉS : *Profil utilisateur, Modèle utilisateur, Recherche d'information, RaPC, Personnalisation de l'information.*

KEYWORDS: User profile, User model, Information Retrieval, CBR, Personalization of information

CORIA 2007, pp. 197–209
Saint-Étienne, 28–30 mars 2007

1. Introduction

Nous avons exploré diverses pistes pour permettre un accès pertinent au contenu des thèses scientifiques de la bibliothèque numérique de DOC'INSA de L'INSA de Lyon, mises en ligne par le système CITHER (http://csidoc.insa-lyon.fr/these/).

Premièrement, nous avons défini un modèle de document fondé sur l'intégration de balises sémantiques dans le corpus des thèses [Abascal, 2005] [Berisha-Bohé, 2005]. Ainsi, il est devenu possible d'extraire, grâce à ce balisage sémantique, les parties correspondant au mieux à un concept ou à une thématique recherchée.

En étudiant la structure d'un certain nombre de thèses, nous avons dégagé et hiérarchisé une base de concepts du domaine de l'informatique [Abascal, 2005]. Elle permet de structurer notre domaine d'étude. Les auteurs insèrent les concepts dans le contenu des thèses en tant que « tags sémantiques ». Les utilisateurs en recherche d'information les utilisent pour faciliter la construction de leurs requêtes.

Afin de mieux prendre en compte les besoins des utilisateurs, nous avons défini un modèle utilisateur permettant d'exploiter son profil et proposons un système de recherche d'information personnalisé, qui est l'objet principal de ce papier.

Donc, l'objectif est double. Il s'agit d'une part d'améliorer le système actuel, en offrant la possibilité d'accéder à plusieurs thèses à la fois, en récupérant des extraits pertinents et correspondant à une unité de corpus plus fine que le chapitre. D'autre part, nous souhaitons proposer un accès personnalisé et pertinent au contenu des thèses scientifiques diffusées et ainsi pouvoir donner des réponses pertinentes et adaptées à l'utilisateur en l'assistant dans la construction de ses requêtes.

Dans la section 2, nous commençons par un état de l'art sur la notion de profil utilisateur et continuons par une analyse des connaissances utilisateur et par la construction d'une arborescence de typologie de profil, dans la section 3. Ensuite, un modèle de l'utilisateur générique fondé sur l'arbre de typologie et sur le Raisonnement à Partir des Cas est décrit dans la section 4. L'accumulation des expériences de Recherche d'Information (RI), enrichit le modèle générique qui se spécialise pour permettre l'évolution du profil tel qu'indiqué dans la section 5. La section 6, présente un prototype qui intègre nos propositions. Nous terminons l'article avec un point sur les limites et les avantages de notre système.

2. Notion de « profil de l'utilisateur »

Selon [Gaussier 2003], « t*outes les variations qui caractérisent un utilisateur ou un groupe d'utilisateurs, peuvent se regrouper sous le terme de profil de l'utilisateur* ». Cette proposition, bien que générale, correspond à nos orientations.

Dans un premier temps, nous nous conformons à un des axes du projet de l'ACI « APMD » (http://apmd.prism.uvsq.fr/), qui porte sur la modélisation et l'évolution

des profils. Ce projet préconise une typologie de profil générique fondée sur six dimensions : d*onnées personnelles, domaines d'intérêt, qualité, préférences de livraisons*, s*écurité, historique d'exécution*. Une partie de ces dimensions se trouve au niveau 1 de notre modèle (Figure1), qui de plus, est fondé sur des approches, retenues lors de notre état de l'art (niveaux 2 à 5, Figure1). Ainsi, à partir de la littérature [Jeribi, 2001], 3 types d'approches ont attiré notre attention.

L'approche sociologique propose un modèle d'actions [Carberry 1994] ou de tâches, en construisant une hiérarchie de stéréotypes de l'utilisateur. « Un stéréotype est un trait (ou une caractéristique commune) partagé par plusieurs utilisateurs » [Rich 1979]. Parmi les utilisateurs de DOC'INSA nous pouvons identifier des stéréotypes comme : "doctorants", "étudiants", "enseignants" etc…

Un des objectifs de *l'approche cognitive* est d'apprendre le processus informationnel individuel et ensuite l'illustrer par un modèle [Bellkin 1987]. Ceci permet d'envisager un modèle de connaissances de l'utilisateur [Allen 1991], grâce à des corrélations entre les différents types de connaissances et de comportements de l'utilisateur face à une situation de recherche documentaire.

Enfin *l'approche de l'Intelligence Artificielle* (IA) [Pitrat 1990] tente de formaliser les modèles complémentaires issus des approches sociologiques et cognitives, pour exploiter et concevoir un système de RI répondant aux attentes de l'utilisateur et intégrant ses différentes caractéristiques. Les connaissances évoluent grâce à l'acquisition de l'expérience. Les tâches accomplies lors d'une RI documentaire suivent aussi cette évolution. C'est un autre aspect que l'IA peut gérer.

Notre approche est fondée sur la représentation et la formalisation des connaissances de l'utilisateur. Nous avons étudié les modèles formels de représentation d'un profil et les méthodes permettant son évolution.

3. Analyse des connaissances de l'utilisateur dans notre SRI documentaire

L'expérience montre qu'il est difficile d'anticiper toutes les caractéristiques d'un utilisateur en session de recherche pour l'aider dans tous les contextes possibles [Berisha-Bohe, 2005]. Nous avons retenu 5 types de connaissances adaptées à notre contexte et respectant la typologie définie par « APMD » (niveau1, Figure 1).

3.1. Connaissances générales (CG)

Elles renseignent l'identification civile, l'appartenance géographique, certaines particularités, et l'appartenance socioculturelle qui ici sera traduite par le statut de l'utilisateur. Ce dernier peut prendre des valeurs comme "auteur" d'une thèse, "bibliothécaire", "administrateur informatique" de la base des thèses, ou "invité".

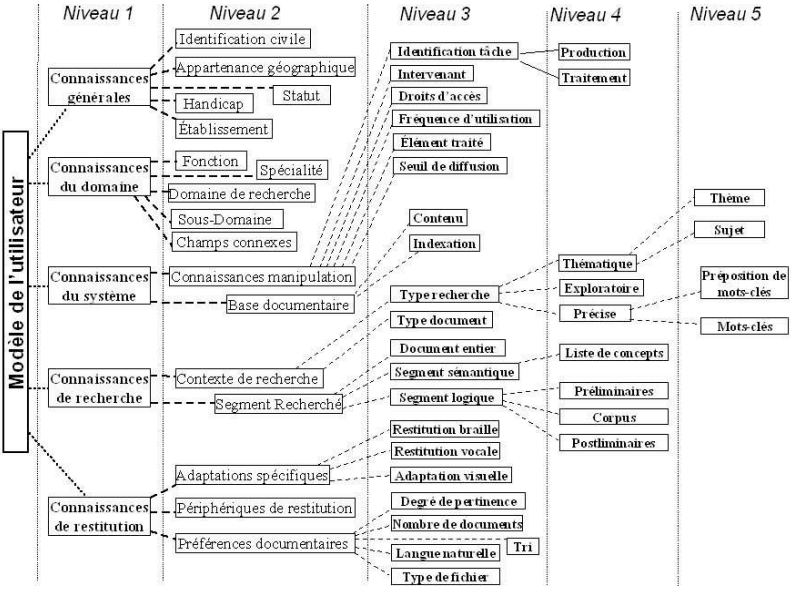

Figure 1. *Hiérarchie des connaissances de l'utilisateur du SI de DOC'INSA*

3.2. Connaissances du domaine (CD)

Ce type concerne les connaissances liées à la fonction de l'utilisateur (enseignant, étudiant, utilisateur RI), à ses domaines et sous-domaines de recherches scientifiques et à sa spécialité d'étude. Nous intégrons également des domaines connexes d'applications, comme par exemple les mathématiques peuvent l'être pour l'informatique. Les CD sont utilisées pour caractériser les stéréotypes de notre système. Par exemple, pour la fonction étudiant, nous avons dégagé les stéréotypes de : "étudiant_Master", "doctorant" et "ingénieur". En combinant la fonction et le domaine de recherche, le regroupement par catégorie d'utilisateurs est affiné.

3.3. Connaissances du système (CS)

En connaissant le système et en le manipulant aisément, l'utilisateur aura moins de difficultés à établir sa requête. Par exemple, si l'utilisateur connaît ses droits d'accès aux documents, il peut définir la requête en fonction des documents qu'il peut atteindre.

3.4. Connaissances de recherche (CRE)

Elles concernent les éléments nécessaires pour établir la requête de l'utilisateur (Figure 5), qui est définie par le segment recherché et le contexte de recherche. Un segment est une ou plusieurs entités lexicales comprises entre la marque du début et la marque de fin d'un tag sémantique inséré dans le document [Abascal 2005].

Le contexte comprend les types de documents traités comme les thèses, recueils bibliographiques, cours, et le type de recherche : thématique, exploratoire, précise.

Si l'utilisateur procède à une recherche thématique, le système va chercher les documents correspondant à ses domaines et sous – domaines, précisés dans l'écran de connexion (Figure 4). Si l'utilisateur procède à une recherche par sujet, le système va chercher les documents correspondant à sa spécialité et aux champs connexes. Pour la recherche exploratoire, le système s'appuie sur les résultats mémorisés par le stéréotype retenu. La recherche précise est effectuée à l'aide de mots-clés ou du texte intégral fourni par l'utilisateur durant une saisie libre. Elle agit sur un segment logique (chapitre, section, etc.), sémantique ou un document entier.

3.5. Connaissances de restitution (CRS)

Il s'agit ici des préférences documentaires, des adaptations nécessaires aux besoins spécifiques de l'utilisateur, et des types de périphériques de restitution. Ces renseignements concernent l'affichage des résultats, et le système ne les prend pas en compte durant la construction de la requête, mais juste avant de délivrer ses résultats.

Cette analyse de la typologie des connaissances mobilisées par un utilisateur lors de ses recherches, nous a aidés à construire un modèle générique d'utilisateur.

4. Modèle générique formel de l'utilisateur

L'IA fournit des techniques permettant de construire et de faire évoluer le profil par le système lui-même, en limitant l'intervention humaine.

4.1. Raisonnements fournis par l'IA

Nous avons rencontré différents types de raisonnements utilisés par les systèmes experts pour la résolution des problèmes [Benhamou et al 1993] [Giunchiglia et al., 1992]. Il s'agit du raisonnement fondé sur les règles logiques « si…alors », ou celui fondé sur les règles et les modèles ou encore le raisonnement fondé sur les ontologies. Le premier est bâti sur une forte composante procédurale ; la séparation entre les connaissances déclaratives et procédurales est absente. Le deuxième fournit une formulation sous-jacente des connaissances sous forme de contraintes et une solution identifiée par des moteurs de satisfaction de contraintes. Il ne peut s'appliquer que lorsque les modèles sont connus et facilement identifiables. Le

troisième est plus adapté à notre contexte. Il met en évidence l'intérêt de l'utilisation des ontologies pour la modélisation de l'utilisateur [Kay, 1999]. Dans ce cas, le domaine est modélisé comme une structure taxonomique de concepts en relations entre eux et chacun ayant ses attributs. En hiérarchisant les connaissances de l'utilisateur selon les typologies, nous nous orientons vers cette approche.

Nous procédons à la personnalisation de ce modèle global générique grâce aux valeurs affectées aux attributs durant chaque expérience de chaque utilisateur. Nous utilisons la technique du RàPC pour suivre l'évolution de ces valeurs, et par conséquent, l'évolution du ou des profils, ainsi que des groupes d'utilisateurs.

4.2. Raisonnement à Partir des Cas (RàPC ou CBR)

Définition

Cette méthode est fondée sur le principe des analogies entre les expériences précédentes et le problème courant. La généralisation se fait lorsque le problème concret doit être résolu, et donc quand les cas tiennent compte d'un contexte plus large que les règles heuristiques. « Le RaPC est une approche de résolution de problèmes qui utilise des expériences passées pour résoudre de nouveaux problèmes. L'ensemble des expériences forme une base de cas » [Lamontagne et al., 2002].

Principe

La base des cas est exploitée durant les phases de recherche, d'adaptation et d'apprentissage [Kolodner, 1993]. La ***recherche*** détermine les cas de la base les plus similaires au cas à résoudre. L'***adaptation*** modifie les réponses des cas retrouvés pour construire une nouvelle solution. **L'*apprentissage***, assure l'intégration des nouvelles solutions dans la base des cas et la modification des structures du système pour en optimiser les performances. Notre modèle de l'utilisateur est représenté par un cas qui affiche les connaissances mobilisées durant une session de RI. Nous estimons que l'évolution d'un profil est fonction de l'évolution de la base de cas correspondant à un utilisateur bien identifié.

Le cas : Structure de base

Typiquement, un cas décrit une situation diagnostique et contient en général la description des symptômes, du défaut, de sa cause et de la solution.. Pour notre étude, nous avons choisi une représentation des cas par une structure simple : une liste d'attributs-valeurs, les attributs étant les connaissances.

5. Modèle évolutif de l'utilisateur

Le modèle de l'utilisateur est donc formalisé par un cas composé de 3 parties (colonne 1, Figure 2) : la situation réelle de recherche, le diagnostic fait par le système, et la solution qu'il apporte.

La situation réelle de RI concerne les connaissances de l'utilisateur (CU), et la requête qu'il émet (RU). A partir de ces éléments, le système va chercher à améliorer la requête de l'utilisateur.

	Groupe	Sous - Groupe	Attribut	Poids	Type de valeur	Exemple de valeurs
Situation réelle	CU					
		Identification (CG)	Identifiant	2	Valeur précise	00001
			Handicap	1	Valeur précise	Néant
		Domaine (CD)	Fonction	3	Valeur précise	Étudiant
			Domaine principal	4	Valeur précise	Informatique
			Sous - domaine	4	Valeur précise	Connaissances et raisonnement
			Spécialité	3	Valeur précise	Donnée, Document, Connaissances
			Domaines Connexes	3	Valeur précise	Bibliothèques, Archivage
		Manipulation (CS)	Fréquence utilisation	2	Valeur précise	Fréquemment
	RU	(CRE)				
		Contexte de recherche	Type de document	2	Règles de similarité	Thèse
			Niveau de spécialisation	2	Règles de similarité	Débutant
			Exploration	3	Similarité d'arbre	NON
		Recherche thématique	Thème	3	Similarité d'arbre	Néant
			Sujet	3	Similarité d'arbre	Néant
		Recherche précise	Segment sémantique	3	Valeur précise	État de l'art
			Segment logique	1	Valeur précise	Néant
			Préposition mots-clés	1	Règles de similarité	Expression exacte
			Mots-clefs	3	Similarité d'arbre	Raisonnement à partir des cas
Diagnose	RA	Requête améliorée	Type de document			Thèse, *Master Recherche*
			Niveau de spécialisation			*Intermédiaire*
			Thème			*Informatique : Intelligence artificielle*
			Sujet			*Raisonnement à partir des cas*
			Segment sémantique			État de l'art, *méthodologie, raisonnement*
			Mots-clefs			*Raisonnement à partir de l'expérience*
			Note utilisateur			non renseigné
Solution	D		Titres de document			« Gestion des connaissances dans une base de documents multi - média », «Approche fonctionnelle générique des méthodes de segmentation d'images », « Modélisation de l'utilisateur pour la recherche d'information dans des bibliothèques numériques » …
			Auteur du document			« Egyed - Zsigmond Elod », «Zouaoui Tarik », « Berisha - Bohé Suela » …
			Contenu du document			«Fichier1.xml», «Fichier2.xml», «Fichier3.xml» …

Figure 2. *Structure générique du cas*

La requête améliorée (RA) constitue le diagnostic. Le choix des attributs retenus pour améliorer les requêtes obéit aux observations de nos expériences d'utilisateurs de CITHER. Nous pensons que l'amélioration porte sur le *contexte*, la *thématique* et la *précision de la requête*. Donc, notre système doit être capable d'agir sur ces éléments en comparant la « Situation réelle » entre nouveaux et anciens cas. Un exemple d'amélioration de requête est indiqué dans la partie « Diagnose » (Figure 2). Le système procède ensuite à la RI en fonction de la requête améliorée et propose une réponse. L'utilisateur a alors la possibilité d'en évaluer la pertinence par une note de 1 à 10, stockée parmi les descripteurs du diagnostic. La réponse, qui sert de solution du cas sera constituée des documents (D) retournés par le système (Figure 2) Dans l'exemple, le système trouve un certain nombre de documents stockables en format XML pour les mots-clés « raisonnement à partir des cas ».

La base est constituée des cas représentés sous la forme :

$$\left.\begin{array}{l} \text{Cas}_{\text{base}} = (CU, RU, RA, D) \\ CU = (CG, CD, CS) \\ RU = CRE \end{array}\right\} \Longrightarrow \text{Cas}_{\text{base}} = (CG, CD, CS, CRE, RA, D) \quad [1]$$

Un cas nouveau représentant la situation réelle de notre problème, est constitué des connaissances de l'utilisateur : $\text{Cas}_{\text{nouveau}} = (CU, RU) = (CG, CD, CS, CRE)$ [2]

Ensuite, le système procède à une amélioration de la requête en stockant un cas intermédiaire : $\text{Cas}_{\text{intermédiaire}} = (CU, RU, RA)$ [3]

Les attributs sont classés et pondérés au sein d'un même cas (colonne 5, figure 2) afin de faciliter le travail algorithmique du système durant le calcul des similarités. Nous avons utilisé une échelle de 1 à 4 pour l'évaluation et avons jugé que les descripteurs caractérisant le profil de l'utilisateur sont les plus importants, que ce soit au niveau des connaissances du domaine ou des connaissances de la recherche.

Dans l'avant-dernière colonne, trois types de similarités sont décrits :

- « valeur précise » où la comparaison se fait sur des chaînes de caractères identiques et donc le seuil de similarité est de 100% ;

- « similarité d'arbre » où il s'agit de calculer une distance entre concepts qui sont déjà catégorisés dans une base de concepts [Abascal, 2005]; et enfin,

- « règles de similarités », que nous avons définies manuellement car les valeurs des concepts ne sont pas nombreuses. Voici par exemple, des valeurs de similarités des types de documents : sim ("Thèse", "Mémoire CNAM")=7, sim("Master Recherche", "Mémoire CNAM")=8, sim("Thèse", "Master recherche")=9, sim("Thèse", "Cours")=5

En commençant par les documents les plus similaires, et en procédant par élimination, nous avons attribué des valeurs comprises entre 1 et 10. Ainsi, une "Thèse" et un "Cours", même s'ils traitent du même sujet, ne vont pas prendre en compte les mêmes informations et seront organisés différemment. A l'inverse, un document de Master Recherche ressemble plus à une thèse qu'à un cours. Une étude plus précise doit être effectuée pour définir ces règles de similarités. Au final, la similarité globale entre deux cas, est calculée par la formule

$$\text{Sim}(c_b, c_n) = \left(\sum p_i * \text{sim}(d_{ib}, d_{in}) / \sum p_i\right)$$

Où, $\text{Sim}(c_b, c_n)$ indique la similarité entre le nouveau cas (c_n) et un cas de la base (c_b), p_i indique le poids de l'attribut (ou descripteur) du rang i dans le cas, d_{ib} le descripteur du rang i du cas de la base et enfin d_{in} est le descripteur du rang i du nouveau cas. L'algorithme utilisé s'appuie sur le calcul du plus proche voisin [Berisha-Bohé, 2005].

Les données sur l'utilisateur (le stéréotype ou le groupe) sont stockées dans une base de données. Un de ses champs renvoie à un sous-ensemble de cas, constituant le profil de l'utilisateur. Il rassemble les cas correspondant à l'identifiant ou à la fonction (Figure 2) de l'utilisateur, ayant obtenu une note de 9 et/ou 10 par ce même utilisateur, et ayant un degré de pertinence supérieur à 90%. Ainsi, nous pouvons observer l'évolution d'un profil, d'un stéréotype ou d'un groupe d'utilisateurs.

6. Le prototype

Pour ce prototype nous avons développé deux parties. Une première qui fournit un espace de travail pour l'insertion des tags sémantiques dans le corpus d'une thèse, et qui ne va pas faire l'objet de ce papier.

La deuxième partie du prototype, qui fait l'objet de cet article, fournit une interface proposant de se connecter en choisissant une fonction, en saisissant un code d'identification, ou en renseignant les domaines sur lesquels l'utilisateur souhaite travailler (Figure 3). Une combinaison de ces choix est possible.

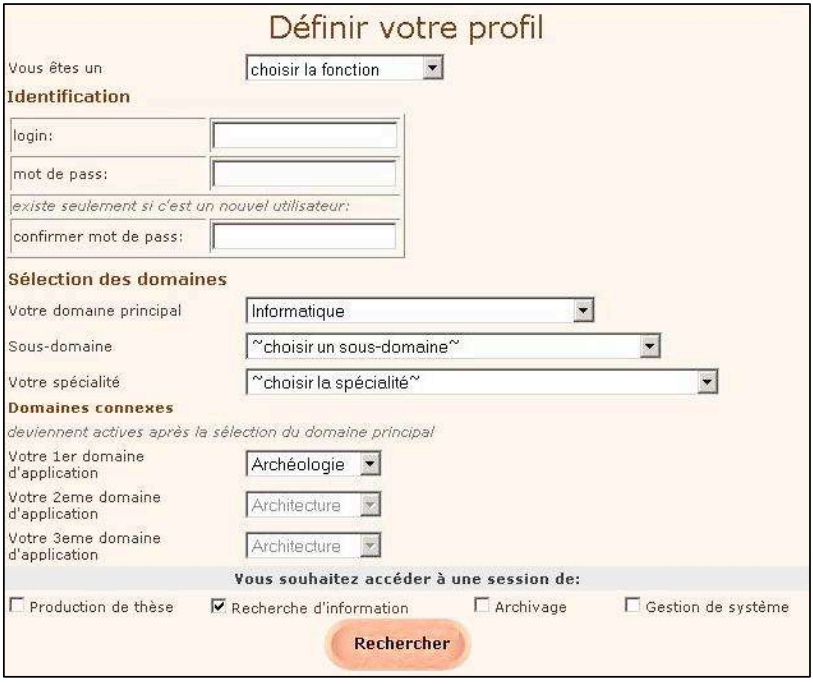

Figure 3. *Interface de connexion au système*

Dans cet écran, nous avons reporté un certain nombre de connaissances générales, de domaine et de système, identifiées dans la partie typologie des connaissances. Dans ce cas, un utilisateur RI est en cours de connexion pour une session de recherche.

La case « Recherche d'Information » (figure 3) est cochée par défaut et l'utilisateur accède à un écran de recherche approfondie (Figure 4). Il a la possibilité d'effectuer une recherche précise par mots-clés ou de définir des filtres de recherche comme la date, le sujet, le type de document, ou le segment sémantique ou encore de les combiner avec les renseignements de la page précédente. Il peut également, ne renseigner aucune préférence et lancer la recherche à partir des renseignements issus de l'interface de connexion. Il peut notifier des options de restitution. Après avoir lancé la recherche, le système stocke dans un nouveau cas les connaissances renseignées à partir de ces saisies et procède à la recherche selon le mécanisme du RàPC. L'outil utilise les préférences de restitution exprimées dans l'interface de la figure 4, pour afficher ses réponses.

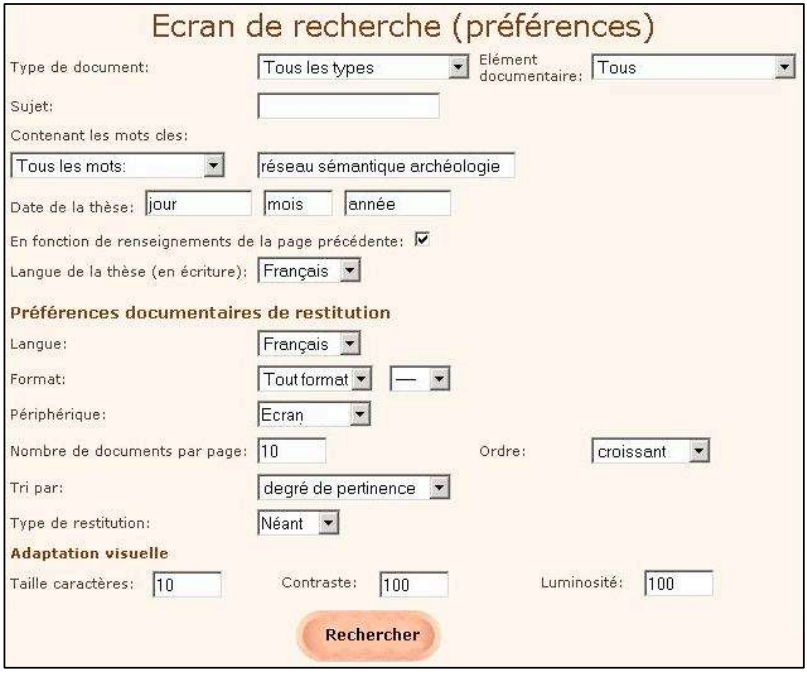

Figure 4. *Écran de recherche*

Nous travaillons actuellement sur un affichage des résultats mieux adapté à la recherche d'information par le contenu. Un exemple est celui de la figure 5, pour la recherche par le mot-clé « bibliothèque numérique », où l'utilisateur peut accéder à différents fragments de plusieurs thèses à la fois. Ceci est possible grâce à l'annotation du contenu par des tags sémantiques.

Figure 5. *Résultats de recherche par le mot-clé "Bibliothèque numérique"*

7. Discussion

Ce travail est une première piste d'étude pour l'accès pertinent au contenu de notre bibliothèque numérique. Nous avons défini les fondements de notre système encore expérimental et limité.

Une des limites provient du processus du RàPC. Il nécessite l'expérience des experts du domaine pour construire une base de cas initiale permettant le démarrage du système. Ainsi nous avons profité de notre expérience d'utilisateurs avertis, de l'expertise des responsables de DOC'INSA, et des administrateurs du SI de CITHER pour identifier des stéréotypes et leurs caractéristiques. Notre outil commence par un rapprochement de chaque nouveau cas à un sous-ensemble de cas, correspondant à un stéréotype et le système procède à une recherche traditionnelle de documents, après l'obtention d'une requête améliorée.

Seulement après avoir atteint un seuil consistant de volumétrie, les algorithmes traditionnels du RàPC prendront la main et piocheront les résultats de recherche dans les cas proches ou identiques, selon l'algorithme du K plus proche voisin. Actuellement, le seuil est défini à 100 cas. Mais une étude d'évaluation doit être effectuée dans le futur pour définir une valeur empirique de ce seuil. Ce procédé servira alors au système pour identifier et définir de nouveaux stéréotypes et groupes d'utilisateurs. Un autre seuil à évaluer est celui qui permet de choisir "LE"

cas le plus proche, fournissant les solutions. Le seuil actuel de 9 sur 10 obéit à notre bon sens, une évaluation doit être effectuée avec une base de cas plus consistante.

La deuxième limite concerne l'absence d'ontologies comme par exemple pour les "règles de similarités" et les "similarités d'arbres" (colonne 6, Figure 2). L'ontologie des règles de similarité entre différents types de documents, présentée à la fin de la section 5, "Modèle évolutif de l'utilisateur", et l'ontologie du domaine (comme informatique) qui remplacerait notre base de concepts sont prévues pour une exploitation opérationnelle de notre outil.

Le volume des documents traités est notre dernière limite. Nous préparons des tests dans un cadre plus large au sein de DOC'INSA.

Cependant, notre système offre déjà deux grands avantages. Tout d'abord, nous venons de lancer les bases d'une recherche d'information par le contenu, et personnalisée. Deuxièmement, nous avons déjà pu observer les avantages du système. Par exemple, lorsqu'une requête précise est lancée, dans CITHER, avec les mots clés « réseau sémantique », le système renvoie 9 documents, dont un qui constitue la seule réponse pertinente. Pour les mots « réseaux sémantiques archéologie », le système ne donne aucune réponse. Alors, qu'avec notre outil, le système nous a renvoyé une réponse pertinente.

A terme, le troisième avantage sera issu du mécanisme du RàPC. L'acquisition des connaissances sera rapide car, le système ne cherchera que dans les nouveaux documents ; le reste des réponses sera fourni par l'adaptation des résultats des cas proches ou similaires.

8. Conclusion

Ayant tiré profit des résultats de recherche effectuée précédemment au sein de notre équipe, nous avons exploré diverses pistes pour l'accès pertinent au contenu des documents de CITHER, comme l'amélioration de la description du contenu ainsi que l'adaptation des requêtes au profil de l'utilisateur.

À partir de la typologie des connaissances, nous avons créé un modèle générique formel d'utilisateur, stockable dans une structure de cas pour implémenter le mécanisme du Raisonnement à Partir de Cas. Les variations des valeurs de ce modèle générique, constituent l'évolution du profil d'un utilisateur ou d'un groupe d'utilisateurs. Cette démarche permet également d'adapter des résultats d'une recherche même lorsque l'utilisateur ne sait pas bien définir sa requête. En étudiant les groupes d'utilisateurs de DOC'INSA, nous avons relevé un certain nombre de stéréotypes et de caractéristiques qui permettent le démarrage du RàPC et interviennent dans l'amélioration de la requête.

Cette recherche a été partiellement soutenue par le Ministère délégué à la Recherche et aux Nouvelles technologies dans le programme ACI Masses de Données, projet MD-33

Bibliographie

Abascal R., Nouveau modèle de documents pour une bibliothèque virtuelle de thèses accessibles par leur contenu sémantique, Thèse de doctorat, INSA Lyon, 2005.

Abascal R., Rumpler B., Berisha-Bohé S. « Proposition d'une nouvelle structure de document pour améliorer la recherche d'information », *Proceedings of the CORIA'05*, ISBN: 2-9523810-0-3, IMAG, pp. 389-404, 2005.

Allen N., « Cognitive Research in information science : implication for design », *Annual review of Information science and technology*, 1991, Vol 26, p. 3-37.

Bellkin N.J., « Discourse analysis of human information interaction », *Canadian Journal of Information Science*, 1987, Vol 12, N°3/4, p. 31-42.

Benhamou F., Colmerauer A., Constraint Logic Programming: Selected Reearch, MIT Press, 1993, ISBN 0-262-02353-9

Berisha-Bohé S., Modélisation de l'utilisateur pour la recherche d'information dans des bibliothèques numériques, Master Recherche, INSA Lyon, 2005.

Berisha-Bohé S, Rumpler B., Abascal R. « A semantic structure to improve information retrieval using XML », *9th ICCC Elpub2005*, poster, Belgium, June, 2005.

Carberry S., Chu-Carroll J., « A plan-based model for response generation in collaborative task-oriented dialogues », *In AAAI-94: Proceedings of the Twelfth National Conference on Artificial Intelligence*, 1994, Seattle, volume 1, p. 799-805.

Gaussier E., Stefanini MH., *Assistance intelligente à la recherche d'informations*, Hermes Science, ISBN 2-7462-0726-5, 2003

Giunchiglia F., Toby Walsh: A Theory of Abstraction, Artificial Intelligence 57, 1992

Kolodner J., *Case based reasoning*. San Mateo, CA: Morgan Kaufman, 1993.

Lamontagne L.., Lapalme G., « Raisonnement à base de cas textuels-état de l'art et perspectives », *RSTI série RIA*, Volume 16-n°3, 2002, pages 339 à 366

Jeribi L., Aide à la recherche documentaire adaptée à l'utilisateur : approche par réutilisation d'expériences, Thèse de doctorat, INSA Lyon, Décembre 2001

Pitrat J., *Méta-connaissances, futur de l'intelligence artificielle*, Paris : Hermès, 1990, 401 p.

Rich E., « User Modeling via Stereotypes », *Cognitive Science*, 3, pp. 329-354, 1979

Sollenborn M., Funk « Category-Based Filtering and User Stereotype Cases to Reduce the Latency Problem in Recommender Systems », *In 6th European Conference on Case-Based Reasoning, ECCBR 2002*, pages 395-405, Aberdeen, Scotland, Springer, September 2002