PROPOSITION A L'INTEGRATION DES PROFILS DANS LE PROCESSUS DE RECHERCHE D'INFORMATION

Anis Benammar
Institut de Recherche en Informatique de Toulouse
118 Route de Narbonne, 31062 Toulouse Cedex, France
benammar@irit.fr

Gilles Hubert
Institut de Recherche en Informatique de Toulouse
118 Route de Narbonne, 31062 Toulouse Cedex, France
hubert@irit.fr

Josiane Mothe
Institut de Recherche en Informatique de Toulouse
118 Route de Narbonne, 31062 Toulouse Cedex, France
mothe@irit.fr
Institut Universitaire de Formation des Maîtres Midi-Pyrénées
http://www.irit.fr/~Josiane.Mothe

Résumé : La formulation du besoin d'information est un des éléments clés pour obtenir des résultats pertinents dans un processus de recherche. Pour aider à cette formulation, des travaux proposent d'introduire la notion de profil d'un utilisateur. Les profils regroupent des informations qui permettent d'enrichir ou d'affiner l'expression des besoins d'un utilisateur mais également de filtrer des documents parmi les résultats de recherche. Les travaux présentés dans cet article s'inscrivent dans ce cadre. Nous proposons un système qui intègre des profils exploités au cours des différentes étapes d'un processus de recherche. Un profil correspond à l'expression d'un besoin en information. Il évolue en fonction des résultats de recherche obtenus en appliquant des mécanismes issus des méthodes de reformulation de requêtes. Les profils sont mémorisés et peuvent être partagés au sein d'un groupe d'utilisateurs. Ils peuvent ainsi être réutilisés et enrichis en tirant profit des recherches d'utilisateurs concernés par le même besoin d'information. Notre système de profils illustre une variante de la recherche collaborative.

Abstract : The formulation of the user's information need is a key step to get relevant results. In order to assist users to better express their information need, many systems introduce the user profile concept. Profiles include several information in order to improve the expression of the user's need and also to filter the retrieved results. The system presented in this paper joins this framework. Our system integrates the use of the profiles in all the steps of the information retrieval process. A profile corresponds to the user's information need. It evolves

along the search sessions based on query reformulation mechanisms. Profiles are memorized and shared by many users. Doing so, profile can be reused and improved by benefiting from researches of users concerned by the same information need. Our profile system illustrates a variant of the collaborative search.

Mots-clés : recherche d'information, profil d’interrogation, profile d’identification, reformulation automatique, recherche collaborative permettre

Keywords : information retrieval, querying profile, identification profile, automatic reformulation, collaborative search.

Proposition à l'intégration des profils dans le processus de recherche d'information

INTRODUCTION

Les systèmes de recherche d'information (SRI) ont pour objectif de restituer les documents correspondant au mieux aux besoins des utilisateurs. Différents éléments influent sur l'efficacité des mécanismes mis en œuvre. Parmi ces éléments, nous pouvons citer la représentation des documents (ou indexation), la représentation du besoin d'information de l'utilisateur et la mise en correspondance de ces deux éléments. Différents travaux s'intéressent à personnaliser la représentation du besoin d'information afin d'améliorer la satisfaction de l'utilisateur [Korfage97]. L'utilisation des profils utilisateurs entre dans ce cadre [Lieberman95], [Jeribi01]. Deux types d'utilisation de profil se distinguent. D'une part, un profil peut être utilisé dans une étape de pré-recherche pour aider l'utilisateur à formuler ou reformuler son besoin [Jeribi01]. Il peut s'agir par exemple d'affiner l'expression d'une requête proposée par l'utilisateur en fonction de son (ses) profil(s). D'autre part, un profil peut également être utilisé dans une étape post-recherche pour filtrer les résultats d'une recherche [Lainé99].
Différents formats de profils ont été étudiés [Korfhage97]. Un profil peut être simple [Baudisch97] ou étendu [Lieberman95]. Un profil simple se présente sous la forme d'un ensemble de mots-clés et éventuellement des poids associés. Un poids traduit l'importance de chaque terme dans le profil. Un profil étendu inclut, en plus des mots-clés et de leur poids, une série d'informations qui décrivent le contexte de la recherche (caractéristique de l'utilisateur, but de la recherche, thème de la recherche, etc.).
Nos travaux s'intéressent à la définition d'un système qui intègre les profils dans le but d'assister l'utilisateur tout au long d'un processus de recherche d'information. Selon notre approche, un profil représente l'expression du besoin en information d'un utilisateur. Ce profil évolue au fur et à mesure des interrogations du système en exploitant les résultats successifs au travers de mécanismes issus des techniques de reformulation de requêtes [Croft00]. Par ailleurs, les profils estimés efficaces peuvent être sauvegardés. Ces profils peuvent alors être partagés et réutilisés au sein d'un groupe permettant ainsi aux utilisateurs de tirer mutuellement profit de leurs recherches. Nous favorisons ainsi une forme de recherche d'information collaborative [Kurzke98]. La recherche d'information dans ce cas, n'est plus seulement orientée vers un unique utilisateur mais tient compte des besoins d'un groupe d'utilisateurs. De plus, notre système cherche à alléger la tâche de l'utilisateur au niveau de la gestion des profils. Son intervention est dans la plupart des cas optionnelle afin de minimiser son interaction avec le système. Elle reste cependant obligatoire dans certains cas afin que les profils puissent être utilisés de façon optimale.
Cet article détaille le système de profils que nous proposons. Dans la deuxième section nous examinons la problématique à laquelle nous nous intéressons à travers l'utilisation des profils. Dans la troisième section, nous spécifions les différentes composantes d'un profil dans notre système. Dans la dernière section, nous décrivons les fonctionnalités de notre système de profils.

1 – PROBLEMATIQUE ET OBJECTIFS

Un objectif des systèmes de recherche d'information et de permettre à l'utilisateur de retrouver l'information la plus pertinente pour lui. Parmi les d'études menées dans le cadre de la recherche d'information, différents travaux proposent d'introduire la gestion des profils des utilisateurs pour améliorer les recherches. Les profils sont utilisés pour filtrer un ensemble de documents afin de ne restituer à l'utilisateur que ceux qui correspondent le mieux à son (ses) profil(s). Alternativement, ils sont utilisés pour affiner l'expression des besoins en information de l'utilisateur. Deux types de gestion de profils ont été proposés à travers les différents systèmes existants : directe ou indirecte.
Dans la plupart des systèmes, la gestion des profils est indirecte [Lieberman95], [Jeribi01], c'est à dire qu'elle est transparente à l'utilisateur. L'avantage de ce type de gestion est qu'elle permet de minimiser l'interaction de l'utilisateur avec le système et ainsi de ne pas alourdir sa tâche de recherche. Cependant elle implique que le système soit capable de détecter les changements des intérêts de l'utilisateur au cours d'une même connexion. En effet, dans la gestion indirecte des profils, l'utilisateur n'ayant pas connaissance de l'existence des profils, il n'indiquera pas au système qu'il change de contexte de recherche. La difficulté est dans ce cas de déceler les changements d'intérêt de l'utilisateur pour modifier le profil de recherche.
A l'opposé, dans la gestion directe des profils, comme par exemple dans l'éditeur de profil de Baudish [Baudish97], l'utilisateur doit intervenir dans toutes les étapes du processus de recherche pour gérer ses profils. Cette stratégie permet à l'utilisateur d'avoir plus de contrôle sur ses profils puisqu'il les gère directement depuis

leur création. La gestion directe implique cependant de lourdes charges pour l'utilisateur qui peuvent constituer un frein à l'utilisation des profils.
Pour limiter les inconvénients introduits par les deux types de gestion précédemment décrits, nous proposons une gestion de profils intermédiaire qui exploite les avantages de chaque solution. Dans certaines étapes du processus de recherche des mécanismes automatiques sont mis en œuvre par défaut, tout en laissant à l'utilisateur la possibilité d'intervenir directement. Dans d'autres étapes, l'intervention de l'utilisateur est obligatoire. Ces éléments sont détaillés dans la section 4.

Le but du système que nous proposons est d'aider l'utilisateur dans son processus de recherche d'information. L'aide est mise en œuvre via des fonctionnalités permettant à l'utilisateur :

- d'enrichir et d'affiner l'expression de son besoin en exploitant des résultats de recherche successifs et ainsi concentrer progressivement les résultats autour du contexte de la recherche,
- de réutiliser les recherches passées propres à un utilisateur ou partagées par un groupe d'utilisateurs et validées comme fournissant un résultat satisfaisant. Les utilisateurs peuvent d'une part garder trace de leurs recherches et d'autre part échanger leurs expériences de recherche avec les autres.

La réutilisation des recherches passées permet à l'utilisateur de développer un même besoin en information d'une manière incrémentale sur plusieurs sessions de recherche. Plusieurs utilisateurs peuvent aussi collaborer pour rechercher une information puisque les profils peuvent être utilisés successivement par différents utilisateurs. Les profils regroupent ainsi les expériences de recherche d'un ou de plusieurs utilisateurs pour un même besoin d'information. Dans ce cas, l'utilisation des profils est une forme de la recherche collaborative [Kurzke98].
Dans les sections suivantes, nous détaillons les différentes composantes d'un profil ainsi que l'utilisation des profils dans un processus de recherche d'information.

2 - MODELISATION DES PROFILS

Nous avons défini des formats de profils spécifiques permettant de mémoriser les recherches effectuées par différents utilisateurs. Un profil contient les informations suivantes :

- un sous-profil d'identification de l'utilisateur.
- un sous-profil d'interrogation : il correspond à la description du contexte de recherche d'un utilisateur pendant une seule session de recherche (court terme) ou plusieurs sessions de recherche (long terme).

Chaque utilisateur maintient un seul sous-profil d'identification et un ensemble de sous-profils d'interrogation qui traduisent son contexte de recherche par thème d'étude.

2.1 Le sous-profil d'identification

Cette première composante du profil sert à identifier un utilisateur à travers une série d'informations. Ce sous-profil spécifie des informations sur l'utilisateur ainsi que sur ses droits d'accès aux profils d'interrogation des autres utilisateurs dans le même groupe de travail. Ces informations, gérées par chaque utilisateur, permettent de réglementer l'échange des profils d'interrogation dans un groupe d'utilisateurs. Chaque utilisateur spécifie les droits d'accès qu'il attribue à ses propres profils pour les autres membres du groupe. Ainsi, le profil d'identification est défini à la première connexion au système de profils et est mis à jour par incrémentation à chaque création d'un profil d'interrogation.

2.2 Le sous-profil d'interrogation

Un sous-profil d'interrogation peut être assimilé à une requête. Il traduit le besoin en information de l'utilisateur. Dans les systèmes existants, une recherche est généralement caractérisée uniquement par un ensemble de mots-clés éventuellement pondérés. Elle est ainsi séparée de son contexte (le but de la recherche, son thème, etc.). Dans le but d'optimiser la réutilisation des profils et faciliter la compréhension des profils, nous optons pour l'association de la recherche à son contexte. Nous incluons ainsi dans la structure du sous-profil d'interrogation, en plus des mots-clés, des informations telles que la description courte et détaillée de la recherche, le thème de la recherche, le lien avec les autres profils et la notion de durée de vie d'un profil. Ainsi, nous distinguons les profils d'interrogation à court terme qui correspondent à une seule session de recherche (même utilisateur, même besoin) et qui ne sont pas réutilisables et les profils d'interrogation à long terme qui peuvent être issus de plusieurs sessions de recherche et qui peuvent être réutilisés. Ces deux types de profils sont liés puisqu'un profil à court terme peut être sauvegardé sous la forme d'un profil à long terme s'il correspond à un besoin d'information récurrent ou utile pour d'autres utilisateurs.
Le schéma suivant, illustre les liaisons entre les deux types du profil d'interrogation.

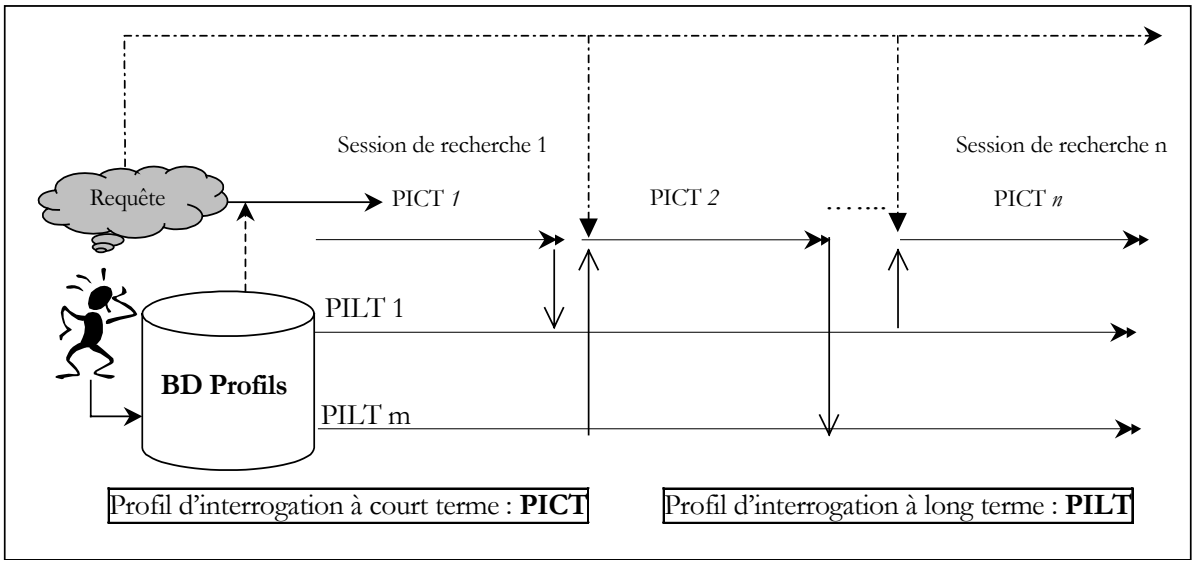

Figure 1 : Interaction entre les composantes du profil d'interrogation

2.2.a Le sous-profil d'interrogation à court terme

Initialement, un sous-profil d'interrogation à court terme correspond soit à une requête soumise au système de recherche par l'utilisateur soit à un profil d'interrogation à long terme choisi par l'utilisateur parmi ceux existant dans le système de profil.
Un profil à court terme est employé tout au long d'une session de recherche dans le processus de recherche. Son contenu évolue durant la session pour aboutir à une expression du besoin de l'utilisateur plus adaptée au contenu de la collection interrogée. Cette évolution est basée sur des mécanismes dérivés des processus de reformulation automatiques de requêtes par réinjection de pertinence [Rocchio71], [Croft00].

2.2.b Le sous-profil d'interrogation à long terme

Le sous profil d'interrogation à long terme correspond à la mémorisation d'un profil à court terme ou à une nuance ou évolution d'un profil à long terme existant. Dans ce cas, il peut s'agir de développer et de faire évoluer un même besoin en information dans le temps. Le lien d'antécédence ou de dérivation entre profils à long terme est mémorisé dans le système de profil.
La création ou mise à jour d'un profil d'interrogation à long terme intervient à la fin de chaque session de recherche. Ainsi, les profils d'interrogation à long terme peuvent être utilisés dans les sessions futures pour relancer de nouveaux processus de recherche.

3. - EXPLOITATION DES PROFILS AU COURS D'UN PROCESSUS DE RECHERCHE

La gestion des profils est assurée tout au long du processus de recherche :

- au début du processus de recherche : l'utilisateur peut se baser sur un profil d'interrogation (à long terme) existant. Cette aide à l'expression du besoin correspond à une réutilisation directe des profils qui permet à l'utilisateur de débuter une recherche à partir d'un profil qui correspond ou qui est proche de son besoin de recherche.
- au cours du processus de recherche : des procédures de reformulation automatiques sont proposées à l'utilisateur pour mettre à jour le profil d'interrogation qu'il utilise [Benammar02]. Elles sont basées sur les principes de réinjection de pertinence aveugle (totalement automatique) ou de réinjection des jugements de pertinence des documents retrouvés spécifiés par l'utilisateur.
- à la fin du processus de recherche : le profil utilisé lors de la recherche peut être sauvegardé pour d'éventuelles utilisations futures.

3.1 Au début du processus de recherche

Durant cette étape, notre but est d'assister l'utilisateur dans la formulation de son besoin en information en lui proposant de réutiliser des profils de recherche existants. Cette aide à l'utilisateur est fondamentale dans un système de recherche d'information compte tenu des difficultés que l'utilisateur peut rencontrer pour formuler d'une manière efficace son besoin en information.

La réutilisation des profils présente un grand intérêt pour l'utilisateur au moment où il formule son besoin en information. En effet, dans certains cas l'utilisateur cherche une information dans un domaine sans avoir d'idées précises de ce qu'il cherche et sans savoir comment l'information pertinente est représentée dans la collection interrogée.
Pour un utilisateur, les formulations passées mémorisées dans ses propres profils ou dans ceux des autres utilisateurs peuvent faciliter l'expression de son besoin en information. L'utilisateur peut partir d'un profil existant pour le développer ou bien pour en créer un nouveau en s'inspirant des profils existants.

Figure 2 : Spécification du besoin en information de l'utilisateur

En revanche, il peut s'avérer fastidieux pour un utilisateur de consulter tous les profils existants. Des critères de sélection définis par l'utilisateur (figure 3) lui permettent de limiter le nombre de profils à consulter. Les critères de sélection que l'utilisateur peut appliquer sont les suivants :

- utilisateur : limitation aux profils d'un utilisateur donné,
- thème de recherche : limitation aux profils qui correspondent au thème de recherche de l'utilisateur,
- date de dernière modification : limitation aux profils modifiés (utilisés) depuis une certaine date.

Figure 3 : Recherche des profils existants

Le résultat de sélection des profils (figure 4) détaille à l'utilisateur les profils (description, thème associé, liste des mots-clés, etc.).
Le système offre également à l'utilisateur la possibilité de naviguer entre les profils. En effet, pour garder un historique de l'évolution des profils, ces derniers sont liés entre eux par une relation de type prédécesseur/successeur. Un profil peut posséder un prédécesseur et un ou plusieurs successeurs. L'utilisateur peut exploiter cette relation d'enchaînement entre les profils pour bien comprendre l'évolution du besoin en information et effectuer le bon choix du profil d'interrogation à réutiliser. La relation de prédécesseur/successeur peut dans certains cas être assimilée à une relation de généralisation/spécialisation. En effet, un profil d'interrogation est réutilisé pour affiner ou inversement pour développer un besoin en information. Dans certains cas, l'utilisateur peut trouver que le contexte de recherche rattaché à un profil est assez générique. En partant sur les successeurs de ce même profil, l'utilisateur pourra retrouver des spécialisations qui correspondent à son contexte de recherche précis. Inversement, les prédécesseurs d'un profil peuvent intéresser l'utilisateur parce qu'ils présentent le cadre général de son besoin.

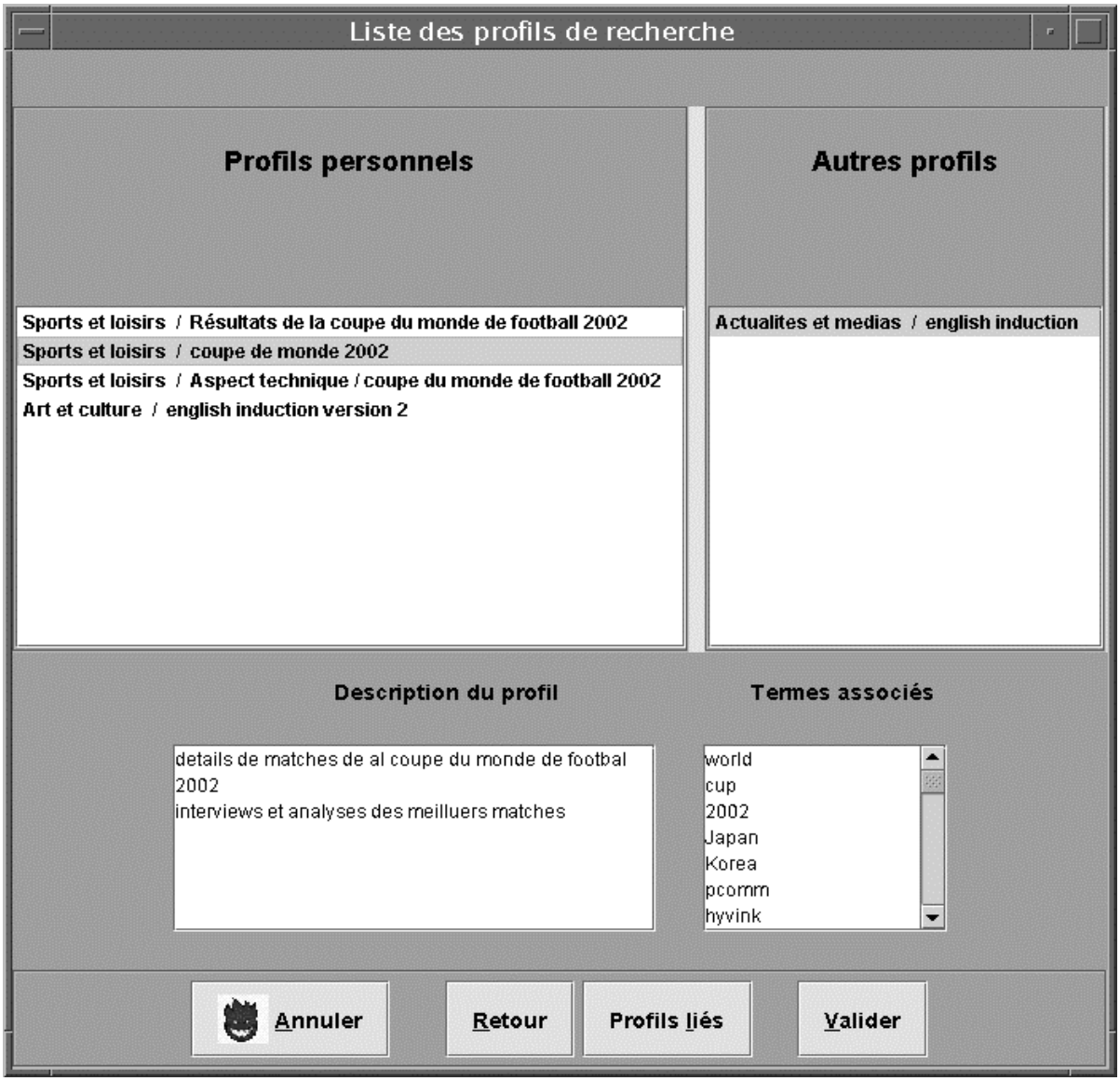

Figure 4 : Résultat de recherche des profils

3.2 Au cours du processus de recherche

Durant cette étape, l'expression du profil d'interrogation est affinée afin d'aboutir à une meilleure expression du besoin de l'utilisateur. Nous avons adapté les mécanismes de réinjection de pertinence habituellement utilisés pour la reformulation automatique de requête à la modification automatique des profils d'interrogation [Benammar02].

Dans notre cadre d'étude, l'intervention de l'utilisateur au cours du processus de reformulation est optionnelle. Dans le cas où l'utilisateur précise des jugements par rapport au résultat de la recherche (documents restitués), la procédure de reformulation est exécutée sur la base des documents jugés pertinents par l'utilisateur. Dans le cas contraire, une procédure permet d'estimer les préférences de l'utilisateur en terme de documents pertinents. Cette procédure analyse les documents restitués et applique un filtrage pour ne garder que ceux qui sont susceptibles d'être les plus pertinents. La procédure de reformulation automatique du profil est alors exécutée sur la base de ces documents.

Dans les deux cas, la procédure de reformulation se base sur une analyse de cooccurrence des termes dans les documents pertinents et dans l'expression du sous-profil d'interrogation.

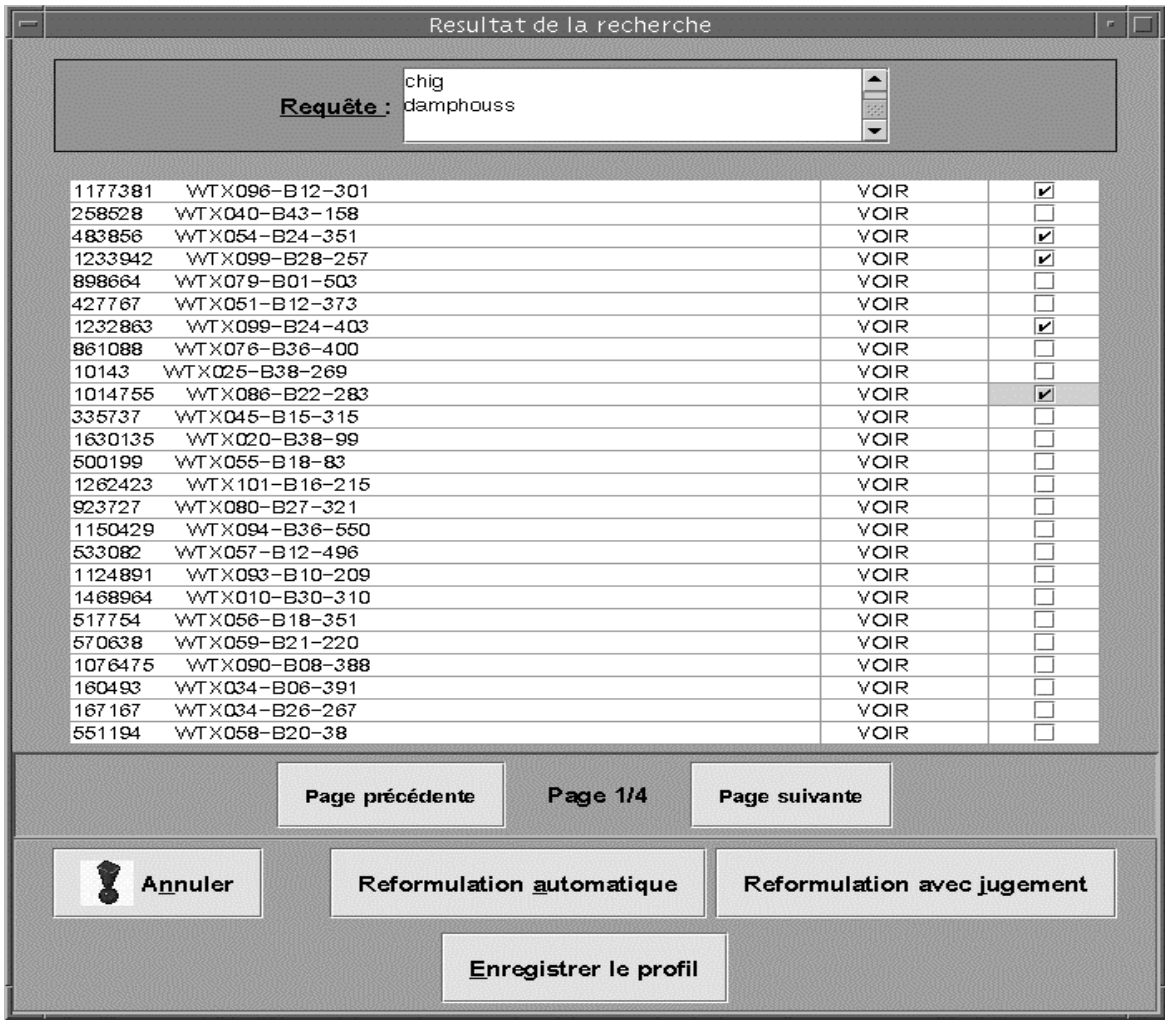

Figure 5 : affichage des résultats de recherche

3.3 A la fin du processus de recherche

Durant la phase finale d'une session de recherche, l'intervention de l'utilisateur est obligatoire afin de faire face au problème de changement des intérêts. En effet, durant une même connexion, l'utilisateur peut changer à plusieurs reprises son contexte de recherche. A chaque changement de contexte de recherche, l'utilisateur doit l'indiquer au système en annulant la recherche en cours, ou en sauvegardant le profil associé. Ainsi, le profil d'interrogation employé est modifié dès que l'utilisateur change de direction de recherche. L'intervention de l'utilisateur durant cette étape a pour but de valider le résultat de recherche qu'il vient d'obtenir. L'utilisateur doit préciser si le profil obtenu à la fin d'une recherche permet de retrouver l'information qui correspond à son besoin. Si ce n'est pas le cas, profil ne pourra pas être sauvegardé. En revanche si l'expression du besoin courante s'est avérée efficace, le profil d'interrogation correspondant pourra être sauvegardé afin d'être utilisé lors des prochaines interrogations. Dans le cas où il est issu d'un autre profil à long terme (choisi au début de la session), cette sauvegarde pourra soit correspondre à une mise à jour du profil initial, soit aboutir à la création d'un nouveau profil, lié à son "père". Nous traduisons ainsi, l'intérêt de l'utilisation des profils comme étant des expériences de recherche dont l'utilisateur pourrait tirer profit dans ses recherches futures.

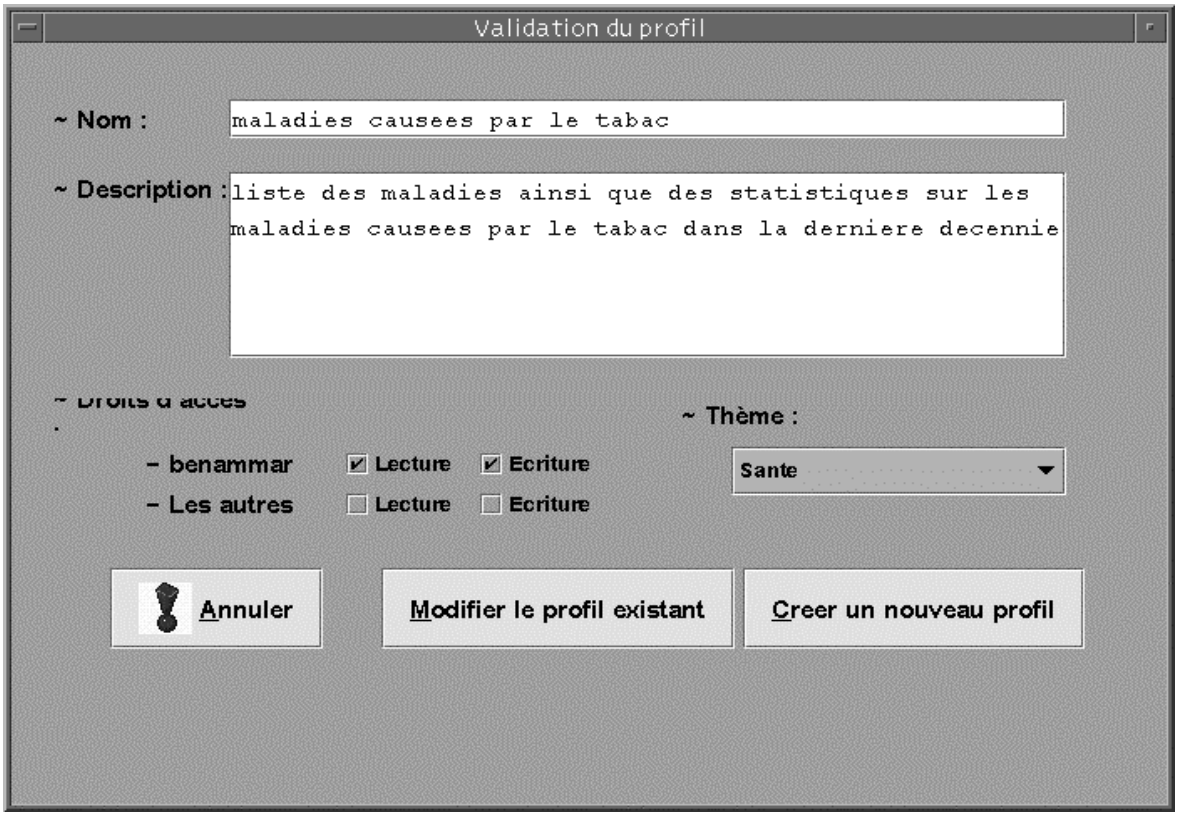

Figure 6 : Enregistrement des propriétés d'un profil

CONCLUSION

Dans cet article, nous avons proposé un système de gestion de profils pour la recherche d'information. L'objectif de ce système est d'aider l'utilisateur à retrouver les documents les plus pertinents tout en mutualisant les expériences de recherche.

Un profil représente un besoin en information. Ce profil peut évoluer au fur et à mesure des interrogations effectuées au cours d'une session (même contexte de recherche, même utilisateur) ou au cours de différentes sessions. Cette évolution est basée sur des techniques de reformulation de requêtes automatiques. Ainsi, le besoin en information exprimé initialement est progressivement enrichi ou affiné. D'autre part, le système permet à un utilisateur de sauvegarder les profils considérés comme fournissant des résultats pertinents. Ces profils peuvent être partagés entre différents utilisateurs ayant les mêmes besoins d'information. Chaque utilisateur peut donc profiter des expériences validées par d'autres utilisateurs afin d'améliorer les résultats de ses recherches.

Dans notre système, un profil est composé de deux éléments : le sous-profil d'identification et le sous-profil d'interrogation. Le sous-profil d'identification permet d'identifier un utilisateur à travers un ensemble d'information comme son nom et ses droits d'accès par rapport aux profils des autres utilisateurs. Le sous-profil d'interrogation décrit le besoin en information de l'utilisateur. Nous distinguons deux niveaux de profil d'interrogation. Le sous-profil d'interrogation à court terme correspond au besoin courant de l'utilisateur et sa durée de vie est celle de la session de recherche. Un sous-profil d'interrogation à long terme décrit un besoin d'information issu de plusieurs sessions de recherche.

Le système employé permet d'assister l'utilisateur durant toutes les étapes de son processus de recherche. L'intervention de l'utilisateur selon les étapes est soit facultative, soit obligatoire. Elle est facultative par exemple lors des reformulations de profil où le système utilise les jugements de l'utilisateur lorsqu'il les spécifie. Elle est obligatoire lorsque l'utilisateur juge l'efficacité du sous-profil d'interrogation à court terme par rapport au résultat de recherche retrouvé, afin de sauvegarder son évolution en tant que sous-profil d'interrogation à long terme.

Les travaux actuels concernent notamment les aspects de reformulation automatiques des profils d'interrogation en exploitant les liens hypertextes qui peuvent exister entre les documents. Un autre aspect concerne la stabilité des profils à long terme et la gestion des avis divergents des utilisateurs concernant la pertinence des documents par rapport à un profil.

REFERENCES BIBLIOGRAPHIQUES

{Croft00} Croft W.B, Xu J., *Improving Effectiveness of information retrieval with local context analysis.* ACM Transaction on Information systems Volume 18, Number 1, pages 79–112, 2000.

{Benammar02} Benammar A., Mothe J., Hubert G., *Automatic profile reformulation using a local document analysis*, European colloquium on IR research , Glasgow. Springer-Verlag , pages 124-134, 2002.

{Baudisch97} Baudisch P., *The Profile Editor: designing a direct manipulative tool for assembling profiles*, In Proceedings of Fifth DELOS Workshop on Filtering and Collaborative Filtering, pages 11-17, Budapest, November 1997.

{Jeribi01} Jeribi L., Rumpler B., Pinon J. M., *Système d'aide à la recherche et à l'interrogation de bases documentaires, fondé sur le réutilisation d'expériences*. XIXème Congrès INFORSID, pages 443-463, Genève, 2001.

{Korfhage97} Korfhage R., *Information storage and retrieval.* Wiley Computer Publishing 0-471-14-338 3, 1997.

{Kurzke98} Kurzke C., Galle M., Bathelt M., *WebAssist : a user profils specific information retrieval assistant*, 1998.

http://decweb.ethz.ch/WWW7/1903/com1903.html

{Lainé99} Lainé-Cruzel S., *ProfilDoc : filtrer une information exploitable,* Bulletin des Bibliothèques Françaises (BBF), T. 44, Numéro 55, pages 60-64, 1999.

{Lieberman95} Lieberman H., *Letizia: An agent that assists web browsing*, Proceedings of the fourteenth International Joint Conference on Artificial Intelligence, pages 924-929, 1995.

{Rocchio71} Rochio J. J., *Relevance Feedback in Information Retrieval*, in «The SMART retrieval system- Experiments in Automatic Document Processing » Prentice Hall Inc., pages 313-323, 1971.