

# Reformulation automatique des profils utilisant un ensemble local de documents

**Anis BENAMMAR (\*)**, **Gilles HUBERT (\*)**, **Josiane MOTHE (\*,\*\*)**  
benammar@irit.fr , [hubert@irit.fr](mailto:hubert@irit.fr) , mothe@irit.fr

(\*) IRIT/SIG, université PAUL SABATIER 118 Route de Narbonne, 31062 Toulouse cedex 04

(\*\*) Institut universitaire de formation des maîtres, Toulouse 1

## **Mots clefs :**

Recherche d'information, profil utilisateur, réinjection de pertinence, analyse locale, mesures de similarité.

## **Keywords:**

Information retrieval, user profile, relevance feedback, local analysis, similarity measure.

## **Résumé**

Pour assister l'utilisateur à récupérer l'information pertinente, de plus en plus d'agents intelligents emploient les profils utilisateur. Notre approche vise à mettre en place un système de profils personnalisés dans le cadre d'un processus de recherche d'information. Dans cet article, nous nous focalisons sur les mécanismes d'adaptation utilisés pour mettre à jour le contenu d'une composante principale des profils. Le processus d'adaptation est composé de deux étapes. Dans la première étape, les documents résultants d'une recherche initiale sont reclassés en utilisant des mesures de similarité pour construire un ensemble de documents. Dans la deuxième étape, nous analysons cet ensemble de documents pour déduire les termes à ajouter à l'expression du profil. Des expérimentations ont été réalisées sur un extrait de la base de données OHSUMED pour évaluer l'efficacité du processus d'adaptation.

# 1 Introduction

Dans le domaine de la recherche d'information, les profils représentent des besoins d'informations permanents de l'utilisateur [KORFHA97]. Ces profils sont généralement représentés par un ensemble de mots clés et des poids associés [BELKIN92]. Un poids représente l'importance d'un terme dans le profil. Les profils sont actuellement de plus en plus employés par les serveurs et les agents intelligents dans le but de remédier aux problèmes associés aux [GURTHER99] :

- ✓ Larges espaces de données. La masse d'information manipulée ne cesse de s'accroître, l'utilisateur se sent 'perdu' dans l'espace d'information. Le profil est utilisé pour éliminer toute information qui n'est pas en accord avec le contexte de recherche.
- ✓ Variations des langues dans lesquels sont écrits les documents. Le profil est utilisé pour faire face aux problèmes liés au croisement de langues.
- ✓ Difficultés de formuler d'une manière exacte un besoin en information. En effet, les utilisateurs éprouvent dans la plupart des cas des difficultés pour trouver les mots clés qui expriment d'une manière exacte leur besoin en information. Le profil est utilisé dans ce cas pour enrichir les expressions des requêtes émises par l'utilisateur.

Nos travaux visent à développer un système de profils qui s'intéresse plus particulièrement au troisième problème. Le système mis en place intègre les spécificités suivantes :

- Un profil se compose d'un sous-profil utilisateur et de différents sous-profils d'interrogation. Le sous-profil utilisateur contient des informations sur l'identification de l'utilisateur et sur ses droits d'accès par rapport aux sous-profils d'interrogation des autres utilisateurs. Le sous-profil d'interrogation correspond à une description d'un besoin en information de l'utilisateur (ensemble de mots clés et de poids associés). Il est utilisé pour stocker de façon permanente et évolutive un besoin de l'utilisateur. Un sous-profil d'interrogation est défini à court terme (une seule session de recherche) ou à long terme (plusieurs sessions de recherche),
- Les sous-profils d'interrogation sont périodiquement adaptés pour suivre l'évolution des besoins en information de l'utilisateur,
- Chaque utilisateur dispose d'un sous-profil utilisateur et d'un certain nombre de sous-profils d'interrogation traduisant son contexte de recherche par thème d'étude,
- Les utilisateurs au sein d'un même groupe de travail peuvent mettre en commun leurs sous-profils d'interrogation ce qui permet de tirer mutuellement profit des recherches réalisées par d'autres utilisateurs.

Dans ce document, nous présentons les stratégies d'adaptation des sous-profils d'interrogation à court terme. En effet, le sous-profil d'interrogation est régulièrement mis à jour pour suivre l'évolution des besoins de l'utilisateur. Notre processus est composé de deux étapes. En première étape, nous nous basons sur des mesures de similarité pour construire un *ensemble local de documents*. Cet ensemble local sert à regrouper les documents qui sont les plus en accord avec le contexte de recherche en cours. L'ensemble local est ensuite utilisé pour déduire les meilleurs termes à ajouter à l'expression du sous-profil d'interrogation.

Le papier est organisé comme suit : la section 2 présente les systèmes employant les profils et les techniques existantes dans le cadre de la reformulation automatique de requêtes. La section 3 présente en détail le processus d'adaptation des sous-profils d'interrogation que nous proposons. La section 4 décrit la méthodologie expérimentale, la collection de test utilisée et les résultats obtenus.

## 2 Travaux existants

Les profils sont de plus en plus employés par les systèmes de recherche d'information. Un profil contient des informations qui reflètent le besoin de l'utilisateur. Il s'agit dans la plupart des cas d'un ensemble de mots clés et de poids associés [KORFHA97]. Dans la partie qui suit, nous présentons certains travaux employant les profils dans un processus de recherche d'information.

*WebAssist* [KURZKE98] est un système qui utilise des bases de données de profils pour classer et filtrer l'information. Les bases de profils sont générées en supervisant les activités de l'utilisateur

naviguant sur le Web. Lorsqu'un utilisateur rejoint le système WebAssist, l'ensemble des mots clés ainsi que les URLs de toutes les pages Web qui l'intéressent sont collectées dans la base de données. D'autres informations comme les fréquences des mots clés ou la date de la dernière visite ou de modification sont aussi incluses dans la base.

*Syskill & Webert* [PAZZAN97] est un agent intelligent qui utilise le modèle booléen pour représenter les profils et un classifieur bayésien pour déterminer si une page est pertinente par rapport à un contexte de recherche. L'adaptation des profils est réalisée à travers des mécanismes de réinjection de pertinence en utilisant des informations explicites (jugements de pertinence spécifiés par l'utilisateur par rapport aux documents retrouvés). Une fois établis, les profils sont utilisés :

- Pour suggérer de nouveaux liens à l'utilisateur,
  - Pour construire une nouvelle requête qui sera adressée au système de recherche LYCOS.
- Néanmoins, *Syskill & Webert* se limite à l'utilisation des jugements de pertinence explicites pour adapter le contenu du profil ce qui représente une lourde charge à l'utilisateur. Pour chaque recherche, l'utilisateur doit spécifier son jugement par rapport aux documents retrouvés.

Le système *LETIZIA* [LIEBER95] est une interface intelligente qui assiste l'utilisateur dans ses différentes étapes de recherche sur le Web. *LETIZIA* ne prend jamais le contrôle de l'interface, il se limite à fournir des suggestions de navigation à l'utilisateur. Les suggestions de navigation se présentent sous la forme de liens à suivre. Lorsqu'un utilisateur examine une page, le système déduit ses préférences et cherche ensuite dans la structure des liens les pages qui pourraient correspondre à son contexte de recherche.

Au fur et à mesure des recherches réalisées par l'utilisateur, le contenu du profil doit évoluer pour aboutir à une meilleure expression du besoin en information de l'utilisateur. Les poids des termes du profil peuvent être modifiés et de nouveaux termes peuvent être ajoutés. Les mécanismes de réinjection de pertinence sont utilisés pour mettre à jour le contenu d'un profil. Ces mécanismes sont généralement utilisés pour la reformulation automatique des requêtes. Ils peuvent être classés en trois catégories [BAYEZA99] :

- Les techniques globales : elles sont basées sur des données issues de toute la collection de données. Les techniques globales se basent sur les hypothèses d'association [RISJBE79]. L'une des premières techniques d'analyse globale est le regroupement des termes [SPARCK71]. Il s'agit de regrouper les termes en fonction de leurs cooccurrences dans toute la collection des documents. L'idée de base de l'analyse globale, comme utilisée dans le système *PhraseFinder* [CROFT98], est l'utilisation des concepts (un simple mot clé ou bien deux ou trois mots adjacents dans le texte) à la place de simples mots clés et les paragraphes (suite de 300 termes par exemple) à la place des documents entiers. Les concepts sont utilisés pour éliminer le problème d'ambiguïté. L'utilisation des paragraphes améliore les temps d'exécution parce que l'analyse est limitée aux parties les plus intéressantes d'un document [ATTAR77], [BAYEZA99], [CROFT96].
- Les techniques de réinjection de pertinence : elles sont basées sur des informations fournies par l'utilisateur. Dans un cycle de réinjection de pertinence utilisateur, ce dernier examine le résultat de la recherche pour marquer les documents pertinents. L'idée générale est de retenir les termes et les expressions attachés aux documents qui sont jugés pertinents par l'utilisateur par rapport à la requête [ROCCHI71].
- Les techniques locales : elles sont basées sur un sous-ensemble de documents du résultat initial. Les techniques locales sont similaires aux mécanismes de réinjection de pertinence utilisateur sauf qu'elles se déroulent sans intervention de la part de l'utilisateur. Il s'agit d'un processus automatique. Nous distinguons deux techniques d'analyse locale : la réinjection de pertinence locale [ATTAR77] et l'analyse locale du contexte [CROFT96], [MOTHE01]. Pour la première technique, il s'agit de construire des structures globales comme les matrices d'association qui quantifient les corrélations entre termes. Les termes les plus corrélés avec l'expression de la requête sont ajoutés à cette dernière. Le problème avec cette technique se présente lorsque la recherche initiale ne ramène pas des documents pertinents, dans ce cas le processus de réinjection de pertinence locale n'arrive pas à améliorer les résultats puisque les termes qui sont ajoutés à l'expression de la requête parviennent de documents non pertinents. La deuxième technique d'analyse locale du contexte réutilise les concepts de base de l'analyse globale (concepts et paragraphes) et les applique sur des ensembles locaux de documents. Les concepts à ajouter à l'expression de la requête sont sélectionnés à partir des meilleurs documents en se basant sur leurs valeurs de cooccurrence avec la requête entière.

L'approche que nous développons pour adapter les sous-profil d'interrogation intègre les aspects suivants :

✓ Les documents qui sont utilisés pour réaliser l'analyse et déduire les modifications à apporter au sous-profil d'interrogation appartiennent au résultat initial de la recherche. Il s'agit d'une approche similaire à la technique de réinjection de pertinence locale où la modification est basée uniquement sur le résultat de la recherche. Dans nos traitements, nous ajoutons une étape de calcul supplémentaire qui consiste à reclasser les documents initialement retrouvés. Le but de cette étape est d'améliorer la qualité des documents à utiliser pour modifier l'expression du sous-profil d'interrogation. En effet, plus les documents à utiliser pour la modification sont pertinents (ou à la limite proches du contexte de la recherche), plus la modification est efficace. L'étape de reclassement est basée sur les mesures de similarité entre les premiers documents du résultat et le sous-profil en cours.

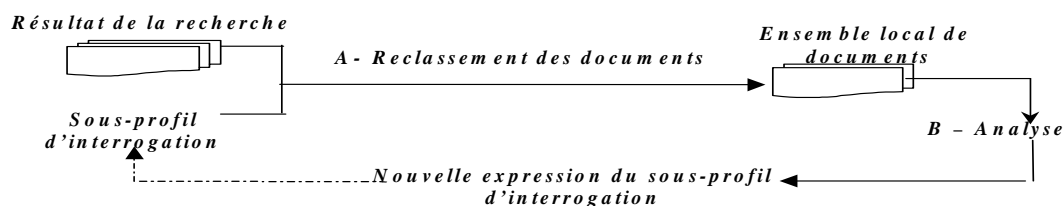
✓ L'analyse des documents de l'ensemble local est basée sur des mesures de cooccurrence entre les termes des documents et l'expression du sous-profil d'interrogation. Le but de cette analyse est de déterminer les termes à utiliser pour étendre l'expression du sous-profil d'interrogation.

Dans la section suivante, nous présentons avec plus de détails notre processus d'adaptation.

### 3 Adaptation des sous-profil d'interrogation à court terme

Initialement le sous-profil d'interrogation correspond soit à la requête proposée par l'utilisateur soit au sous-profil d'interrogation à long terme élaboré durant des sessions de recherche antérieures. Tout au long d'une session de recherche, il est utilisé en tant que requête. Il est d'autre part automatiquement adapté en se basant sur le résultat de recherche associé.

Le processus d'adaptation utilise une approche comparable à celle de la réinjection de pertinence. Le schéma suivant illustre les différentes étapes de ce processus.



**Figure 3-1** : Schéma général d'adaptation des sous-profil d'interrogation à court terme

La première étape du processus est la construction de l'ensemble local de documents. Dans les autres approches de réinjection de pertinence, les premiers documents de la recherche sont automatiquement utilisés pour mettre à jour le contenu de la requête. Dans notre processus, nous ajoutons une étape supplémentaire dans le but d'inclure plus de documents pertinents dans l'ensemble qui servira de base pour l'étape suivante d'analyse. La deuxième étape du processus est l'analyse des documents de l'ensemble local pour déterminer les termes à utiliser pour étendre l'expression du sous-profil d'interrogation.

#### 3.1 Construction de l'ensemble local

Les techniques habituelles de réinjection de pertinence utilisent les premiers documents du résultat de recherche pour créer la nouvelle expression de la requête. La qualité des modifications à apporter à la requête dépend de la qualité de ces documents. Or les premiers documents de la recherche ne sont pas toujours pertinents. Des résultats de TREC prouvent que les performances des techniques qui utilisent les premiers documents d'une recherche sont irrégulières [CROFT96]. L'efficacité de ces techniques peut être sérieusement dégradée si quelques documents rangés parmi les premiers résultats de la recherche initiale ne sont pas pertinents.

Pour améliorer la qualité de l'ensemble des documents qui seront utilisés pour modifier l'expression du sous-profil d'interrogation, nous employons une étape de calcul supplémentaire. Cette étape consiste à construire, à partir du résultat de la recherche initiale, un ensemble local de documents. Le but est de rechercher dans l'ensemble des documents initialement retrouvés ceux qui correspondent le

plus au contexte de la recherche. Il s'agit en fait, de reclasser les documents par rapport au sous-profil d'interrogation et de prendre les meilleurs au lieu de partir directement sur les premiers documents du résultat initial. Les documents sont reclassés en fonction de leur similarité avec le sous-profil d'interrogation à court terme. En effet, dans plusieurs cas, les documents pertinents se trouvent mal classés dans le résultat de la recherche initiale. Pour reclasser les documents nous employons des mesures de similarité (produit scalaire ou Cosinus ou mesures logiques) entre les documents et le sous-profil d'interrogation à court terme représentés tous par des vecteurs définis dans l'espace des termes. Les documents les plus similaires au sous-profil d'interrogation forment un ensemble local. Un document est représenté par un vecteur dans l'espace des termes. A chaque terme est associé un poids calculé comme suit :

$$wt(td_i) = \log_{10} \left( \frac{N}{N_{t_d}} \right)$$

Où  $td_i$  est un terme du document,  $N$  est le nombre de documents dans la collection et  $N_{t_d}$  est le nombre de documents dans la collection qui contiennent le terme  $t_d$ .

Pour le sous-profil d'interrogation, le poids d'un terme est calculé comme suit :

$$wt(tp_j) = \frac{wp(t_p)}{wp_{\max}}$$

Où  $tp_j$  est un terme du sous-profil,  $wp(tp_j)$  est la fréquence du terme dans le sous-profil d'interrogation et  $wp_{\max}$  est la fréquence maximale dans le sous-profil d'interrogation.

La valeur de similarité entre un vecteur document  $\vec{D}$  et un vecteur profil  $\vec{P}$ , en utilisant par exemple la mesure *Cosinus*, est calculée comme suit :

$$Sim(D, P) = Cos(\vec{D}, \vec{P}) = \frac{\sum_{i=1}^N wt(td_i) wt(tp_i)}{\sqrt{\left( \sum_{i=1}^n wt(td_i) \right)^2} \times \sqrt{\left( \sum_{i=1}^n wt(tp_i) \right)^2}}$$

### 3.2 Analyse de l'ensemble local de documents

L'analyse contextuelle locale est réalisée sur la base de l'ensemble local de documents ainsi construit. Le but de cette analyse est d'extraire automatiquement les corrélations qui existent entre les termes des documents et l'expression du sous-profil d'interrogation. Les termes les plus corrélés avec l'expression du sous-profil en cours sont utilisés pour l'étendre.

La corrélation entre un terme ( $t_i$ ) d'un document et l'expression d'un sous-profil d'interrogation  $P$  est exprimée comme suit :

$$Correlation(t_i, P) = \sum_{j \in P} cooccurrence(t_i, p_j)$$

Où  $cooccurrence(t_i, p_j)$  représente le degré de cooccurrence entre un terme  $t_i$  du document et un terme  $p_j$  du sous-profil d'interrogation. Elle est exprimée comme suit [CROFT00] :

$$cooccurrence(t_i, p_j) = \log_{10} (Coocc(t_i, p_j) + 1) idf(t_i) / \log_{10}(n)$$

$$Coocc(t_i, p_j) = \sum_{d \in S} tf(t_i, d) tf(p_j, d)$$

- $tf(t_i, d)$  et  $tf(p_j, d)$  représentent respectivement les fréquences des termes  $t_i$  et  $p_j$  dans le document  $d$ ,
- $idf(t_i) = \log_{10} (N / N_{t_i})$
- $S$  est l'ensemble local des documents,
- $n$  est le nombre de documents de l'ensemble local,
- $N$  est le nombre de documents de la collection,
- $N_{t_i}$  est le nombre de documents de la collection qui renferment le terme  $t_i$ .

## 4 Expérimentations

Le but de nos expérimentations est de tester l'efficacité du processus d'adaptation présenté dans la section 3. La procédure peut être résumée comme suit :

- 1- Réaliser une recherche initiale (Moteur de recherche MERCURE [BOUGHA98]). Le profil d'interrogation à court terme initial correspond à la requête de recherche utilisée,
- 2- Construire l'ensemble local de documents.
- 3- Procéder à une analyse locale pour déduire les nouveaux termes à ajouter à l'expression du profil. Le nombre de termes à ajouter est relatif au nombre de documents de l'ensemble local. Nous réalisons un classement global de tous les termes des documents en même temps et nous sélectionnons les  $N$  meilleurs termes.
- 4 - Ajouter ces nouveaux termes à l'expression du profil et l'utiliser pour réinterroger le système de recherche.

L'expérimentation a été menée sur un extrait de la base de documents médicaux (OHSUMED). Cette base a été utilisée dans le cadre du programme TREC. Le tableau suivant (Figure 4-1) montre les caractéristiques de la base OHSUMED.

Collection OHSUMED (1987)	
Taille de la collection	54710 documents
Nombre de termes	46095 termes
Taille Moyenne d'un document	130 termes environ

**Figure 4- 1** : Les caractéristiques de la collection OHSUMED

Nous utilisons un ensemble de 63 requêtes. Des jugements de pertinence accordés à chaque requête sont utilisés pour évaluer l'évolution des résultats de recherche retrouvés.

Pour évaluer les résultats de recherche, nous utilisons les programmes *Trec\_Eval*. La précision est mesurée à plusieurs niveaux. Soient les niveaux  $p_5$ ,  $p_{10}$ ,  $p_{15}$ ,  $p_{20}$ ,  $p_{30}$  et  $p_{100}$  qui représentent respectivement le nombre de documents pertinents parmi les 5, 10, 15, 20, 30 et 100 premiers documents du résultat de la recherche. Le système d'évaluation calcule aussi la précision moyenne (AvgPr) sur l'ensemble des documents restitués.

Différents paramètres ont été évalués :

- Nombre de documents utilisés initialement pour construire l'ensemble local,
- Nombre de termes ajoutés à l'expression du profil,
- Mesure utilisée pour définir les similarités entre les documents (Cosinus, Produit Scalaire ou mesures logiques).

### 4.1 Résultats des tests

Sur l'ensemble de 63 requêtes utilisées initialement, 18 requêtes (23%) présentent des résultats pour lesquels aucun document pertinent n'est retrouvé initialement. Dans ces cas particuliers, nous ne pouvons pas améliorer les résultats de la recherche en utilisant notre procédure d'adaptation puisque tous les documents initialement ramenés ne sont pas pertinents et par conséquent les termes de ces documents sont aussi considérés non pertinents. Nous nous intéressons donc uniquement aux 48 requêtes restantes. L'évaluation est réalisée selon deux axes :

- Comparaison du résultat de la recherche initiale au résultat obtenu après avoir appliqué le processus d'adaptation. Le but est de l'évaluer l'efficacité du processus d'adaptation employé. Nous utilisons un ensemble local composé de 5 documents. Nous utilisons aussi une valeur seuil lors du calcul des valeurs de similarité. En effet, seuls les documents dont la valeur de similarité dépasse le seuil sont inclus dans l'ensemble local. Le seuil est utilisé essentiellement pour éviter d'inclure dans l'ensemble local des documents qui ne correspondent pas au contexte de la recherche. Le nombre de termes à ajouter à l'expression du sous-profil d'interrogation est fonction du nombre des documents de l'ensemble local. Soient  $n$  le nombre de termes à ajouter à l'expression du sous-profil d'interrogation et  $i$  le nombre de documents de l'ensemble local. La valeur  $n$  est calculée comme suit [BELKIN99] :

$$n = 5 * i + 5$$

- Comparaison du résultat obtenu lorsque nous appliquons le processus d'adaptation entier et le résultat lorsque nous utilisons directement les meilleurs documents sans passer par l'étape de reclassement. Le but est d'évaluer l'importance de l'étape de reclassement induite dans le processus d'adaptation.

### 4.1.1 Analyse locale VS Recherche initiale

Les graphiques suivants montrent l'évolution des valeurs de précision moyenne (*AvgPr*) et exacte (*Exact*) entre le résultat d'une recherche initiale et le résultat après avoir appliqué le processus d'adaptation.:

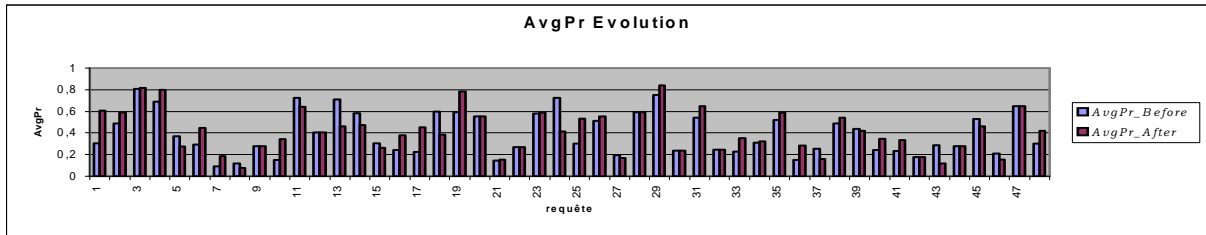


Figure 4-2: Evolution des valeurs *AvgPr*

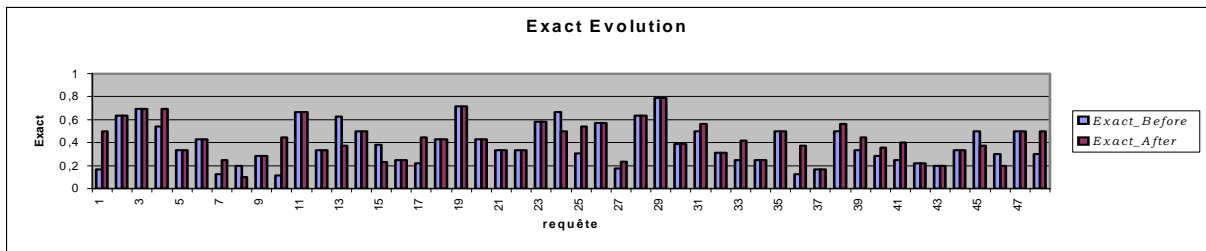


Figure 4-3: Evolution des valeurs *Exact*

Le processus d'adaptation permet d'améliorer les résultats dans une majorité des cas (26 requêtes parmi l'ensemble des 48 requêtes utilisées). Dans 10 autres cas, le résultat demeure invariant. Dans ces cas particuliers, l'ensemble local initialement construit reste vide parce que nous ne retrouvons pas des documents assez proches du sous-profil d'interrogation à court terme, quoique le résultat de la recherche initiale compte des documents pertinents. Les moyennes globales montrent une amélioration de 6,1% pour les valeurs *AvgPr* et de 8,75% pour les valeurs *Exact*.

### 4.1.2 Construction de l'ensemble local

Les graphiques suivants montrent l'évolution des valeurs de précision moyenne et exacte lorsque le processus d'adaptation est appliqué directement sur les premiers documents de la recherche et lorsque nous faisons intervenir l'étape de construction de l'ensemble local.

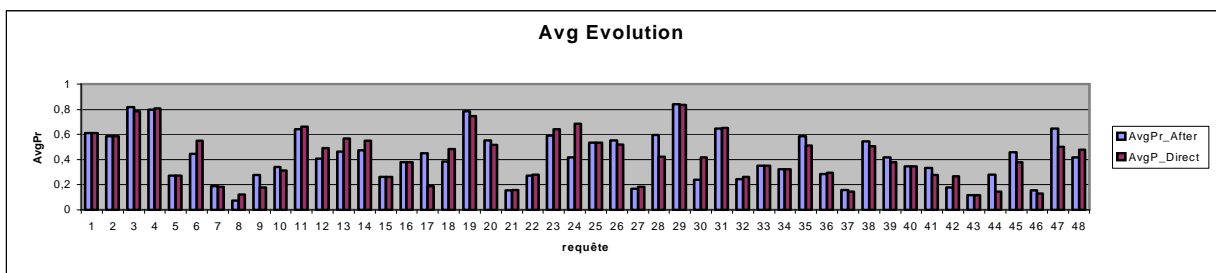
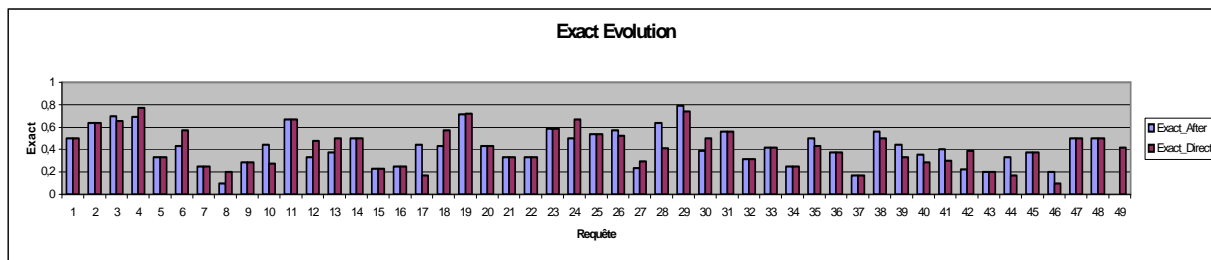


Figure 4-4: Evolution des valeurs *AvgPr*



**Figure 4-5:** Evolution des valeurs *Exact*

Les expérimentations montrent que le processus d'adaptation utilisant l'étape préliminaire de reclassement des documents améliore les résultats plus que le processus utilisant directement les premiers documents du résultat de la recherche. En effet, l'étape de reclassement permet d'améliorer la qualité des documents à utiliser lors de l'analyse. Les documents qui sont inclus dans l'ensemble local sont plus proches du contexte de la recherche. Dans ces cas, l'analyse fournira de meilleurs résultats.

## 4.2 Conclusions des expérimentations

Notre but à travers les expérimentations était d'évaluer l'efficacité du processus d'adaptation et l'importance de l'étape de construction de l'ensemble local. Les expériences effectuées sur la base de données d'OHSUMED 1987 mettent en évidence les conclusions suivantes :

- Le processus d'adaptation améliore les résultats de recherche dans la majorité des cas (le pourcentage d'améliorations est de 55%). Dans 10 autres cas, les résultats restent les mêmes. Dans ces cas, l'analyse n'est pas prise en compte parce que l'ensemble local est vide.
- La construction de l'ensemble local est une étape importante puisqu'elle permet d'améliorer dans plusieurs cas les résultats de recherche.

## 5 Conclusion

Dans cet article, nous présentons les mécanismes d'adaptation utilisés pour mettre à jour le contenu du sous-profil d'interrogation à court terme qui représente une composante de base du profil. Le processus d'adaptation mis en place est basé sur une analyse de cooccurrence réalisée sur la base des documents initialement retrouvés. Une étape préliminaire du processus d'adaptation consiste à construire un ensemble local de documents. L'objectif de cette étape est d'inclure plus de documents pertinents dans l'ensemble local qui sera utilisé lors de l'étape d'analyse. Plusieurs expérimentations ont été menées sur un extrait la base de données OHSUMED 1987 utilisée dans le cadre du programme TREC. Plusieurs critères ont été évalués (nombre de documents qui forment l'ensemble local, nombre de termes à ajouter à l'expression du sous-profil par modification, ...). Dans 55% des cas, le processus d'adaptation améliore le résultat de la recherche. Dans 20% des cas, la performance du résultat demeure la même. L'étape de construction de l'ensemble local s'avère aussi importante. En effet, elle permet d'améliorer la qualité des documents à utiliser lors de l'analyse.

D'autres pistes restent à explorer pour améliorer les 25% des cas restant où nous enregistrons une dégradation des résultats. L'axe principal de recherche sera l'étape de construction de l'ensemble local. En effet, de cette étape dépend l'amélioration du résultat.

## Bibliographie

- [ATTAR77] Attar, R., and Frankel, A. S, *Local feedback in Full-Text Retrieval Systems*. Journal of associations for Computing Machinery, 24 (3), 397-417, 1977.
- [BAYEZA99] Baeza-Yates, R., Ribeiro-Neto, B., *Modern Information Retrieval*, Addison-Wesley Ed., ISBN 0-201-39829-X, 1999.
- [BELKIN92] Belkin, N., J., Croft W.B., *Information retrieval and information Filtering: two sides of the same coin*, CACM, Pages 29-38, 1992.
- [BELKIN99] Belkin N. J., *Relevance Feedback versus Local Context Analysis as term suggestion devices*, Rutgers' TREC-8 Interactive Track Experience, Proceedings of Trec-8, Pages 565- 574, November 16-19, 1999



- [BOUGHHA98] Boughanem M., Dkaki, T., Mothe, J., Soulé-Dupuy, C. *Mercure at Trec-7*. 7th International Conference on Text REtrieval TREC7, Harman D.K. (Ed.) SP 500-236, November 11-17, NIST Gaithersburg, 1998.
- [GURTHE99] Gurthet A., [http://www.biermans.com/culminating/fall\\_1999.html](http://www.biermans.com/culminating/fall_1999.html), fall 1999.
- [CROFT98] Jing, Y., & Croft, W.B. *Corpus-Based Stemming Using Co-occurrence of Word Variants*. Transactions On Information Systems Volume 16, number 1 pp 61-81, 1998.
- [KORFHA97] Korfhage Robert R., *Information storage and retrieval*. Wiley Computer Publishing 0-471-14-338 3, 1997.
- [KURZKE98] Kurzke, C., Galle, M., Bathelt, M., *WebAssist : a user profiles specific information retrieval assistant* 1998, <http://decweb.ethz.ch/WWW7/1903/com1903.html>
- [LIEBER95] Lieberman H., *Letizia: An agent that assists web browsing*, Proceedings of the fourteenth International Joint Conference on Artificial Intelligence, Pages 924-929, 1995.
- [MOTHE01] Mothe J. and al, *Local document analysis in order to re-order a retrieved document list*, submitted to SIGIR 2001
- [PAZZAN97] Pazzani Mickael J. and Billsus Daniel, *Learning and Revising user profiles: the identification of interesting Web sites*, Machine Learning, Volume 27, Number 3, Pages 313-331, 1997.
- [ROBERS00] Robertson S. D. Hull, *The TREC-9 filtering track final report*, TREC-9, 2000
- [ROCCHI71] J. J. Rocchio, *Relevance feedback in information retrieval*, In G. Salton, editor, *The Smart retrieval System, Experiments in Automatic Document processing*, Prentice Hall Inc., Engelwoods Cliffs, NJ, 1971.
- [SPARCK71] Sparck Jones, K., *Automatic Keywords Classification for Information Retrieval*, Buterworths, London, 1971.
- [RISJBE79] Van Risjbergen, *Information Retrieval*, 2nd ed., Butterworths, 1979.
- [CROFT96] Xu, J., Croft, W.B., *Query Expansion using local and global document analysis*. Proceeding of the 19<sup>th</sup> Annual International ACM SIGIR Conference on research and development in Information retrieval (SIGIR 96', Zurich, Switzerland, August 18-22, )1996.
- [CROFT00] Xu, J., Croft, W.B, *Improving Effectiveness of information retrieval with local context analysis*. ACM Transaction on Information systems Volume 18, Number 1, January 2000, Pages 79 – 112.