

---

# Proposition des cadres d'évaluation adaptés à un système de RI personnalisé

**Mariam Daoud, Lynda Tamine-Lechani**

*Laboratoire IRIT, Université Paul Sabatier  
118 Route de Narbonne, F-31062 TOULOUSE CEDEX 9  
{daoud, lechani}@irit.fr*

---

*RÉSUMÉ. L'évaluation d'un système de recherche d'information (RI) personnalisé consiste principalement à mesurer ses performances. Les cadres d'évaluation classiques en RI basés sur les approches orientées laboratoire méritent d'être étendus et révisés vu que le contexte de recherche de l'utilisateur n'est pas considéré dans le protocole d'évaluation et les collections de test. Nous présentons dans ce papier des cadres d'évaluation adaptés à un système de RI personnalisé basés sur l'enrichissement des collections TREC par des contextes/profils utilisateur simulés. Plus précisément, un protocole issu de TREC adhoc consiste à construire des profils utilisateur à partir des domaines d'intérêts prédéfinis dans TREC adhoc. Le protocole issu de TREC HARD consiste à construire le profil à partir des sessions de recherche simulées par les sujets des requêtes de la collection. Les résultats obtenus confirment la stabilité de la performance de notre modèle de RI personnalisé selon les cadres proposés sur des collections de test différentes.*

*ABSTRACT. The evaluation of a personalized search systems aims at measuring its performance. Traditional laboratory-based evaluation of the information retrieval systems is challenged because of its lack of user evidence expressing his search context in the data test sets. In this paper, we present evaluation protocols of personalized search system based on enhancing TREC collections with simulated contexts/user profiles. An evaluation protocol based on TREC adhoc consists of building the user profile across simulated search sessions designed by the TREC domains. Another evaluation protocol based on TREC-HARD consists of building the user profile across simulated search defined by the topics of the collection. Experimental results show the stability performance of our personalized search system according to the proposed evaluation protocols using different test collections.*

*MOTS-CLÉS : Système RI personnalisé, cadre d'évaluation, profil utilisateur*

*KEYWORDS: Personalized IR system, evaluation protocol, user profile*

---

## 1. Introduction

L'objectif de l'évaluation d'un SRI est de mesurer ses performances vis à vis du besoin de l'utilisateur. Les protocoles d'évaluation largement adoptés en RI sont basés sur une approche de type laboratoire (*laboratory-based model*) initiée par *Cleverdon* (Cleverdon, 1967) dans le cadre du projet *Cranfield project II*. Ce modèle fournit une base d'évaluation comparative de l'efficacité de différents algorithmes, des techniques et/ou des systèmes moyennant des ressources communes : des collections de test contenant des documents, des requêtes préalablement construites et des jugements de pertinence associés construits selon la technique de *pooling*, des métriques d'évaluation essentiellement basées sur le rappel-précision. L'émergence de la RI orientée utilisateur a cependant remis en cause la viabilité de ce modèle pour l'évaluation de systèmes interactifs ou de manière générale, les systèmes d'accès contextuel à l'information (Ingwersen *et al.*, 2005). Ces systèmes ont pour objectif de délivrer de l'information pertinente et appropriée au contexte de l'utilisateur qui a émis la requête. Ceci a motivé les réflexions autour de méthodologies d'évaluation adaptées à une recherche d'information contextuelle. Nous proposons dans ce papier des méthodologies d'évaluation mettant en jeu des contextes utilisateurs simulés représentés par leur profil dans le but d'évaluer notre modèle d'accès personnalisé à l'information. Notre approche de base consiste à exploiter un profil utilisateur à court terme construit sur une session de recherche dans la chaîne d'accès à l'information. Les cadres d'évaluation proposés sont basés sur l'enrichissement des ressources TREC (Voorhees, 2001) par des contextes utilisateurs simulés. Le profil utilisateur est intégré comme étant une composante principale de la collection de test. La notion de session est intégrée dans la stratégie d'évaluation du modèle de RI et est soit simulée par le domaine d'annotation des requêtes dans le cas du cadre d'évaluation issu de TREC *ad hoc*, soit simulée par des sous-requêtes d'une même requête dans le cas du cadre d'évaluation issu de HARD TREC.

Ce papier est organisé comme suit. La section 2 présente les cadres d'évaluations proposés en RI personnalisée. La section 4 est dédiée à la description des cadres d'évaluation TREC *ad hoc* et TREC-HARD. La section 5 conclut et présente les perspectives de nos travaux dans le domaine.

## 2. Évaluation des modèles d'accès personnalisé à l'information

A ce jour, il n'existe pas un cadre standard pour l'évaluation de l'efficacité d'un modèle d'accès personnalisé à l'information. Les approches classiques d'évaluation des performances en RI sont des approches de type laboratoire. Ce type d'évaluation révèle des limitations d'évaluation des SRI avec l'émergence de la RI contextuelle.

### **2.1. Limites de l'évaluation orientée laboratoire**

L'évaluation orientée laboratoire d'un SRI est principalement fondée sur l'utilisation d'une collection de test où les requêtes sont les seules ressources clés qui traduisent le besoin en information de l'utilisateur. Les principales limitations de ce cadre d'évaluation sont liées à l'inadéquation des collections de tests pour l'évaluation de la recherche d'information en contexte (Tamine *et al.*, 2009, Ingwersen *et al.*, 2005). D'une part, les utilisateurs directs ayant émis ces requêtes, leurs centres d'intérêt et interactions avec le SRI ne font pas partie intégrante de la collection. D'autre part, les jugements de pertinence sont thématiques et indépendants des situations et du contexte de recherche. Or, il a été bien montré dans (Borlund, 2003, Mizzaro, 1998) que la notion de pertinence est plus complexe, couvrant des niveaux divers liés à la situation de recherche en cours : pertinence cognitive, pertinence affective, pertinence situationnelle, etc.

Les premières tentatives faites dans le cadre de l'évaluation des systèmes de RI en contexte ont été proposées dans TREC à travers les tâches Interactive et HARD. Ces tâches ont permis l'intégration des métadonnées concernant l'utilisateur dans le processus de RI afin d'augmenter la performance du système pour des requêtes difficiles. Les métadonnées utilisateurs concernent des critères tels que la familiarité, le genre du document, la langue, etc. Étant très spécifiques, ces tâches ne permettent pas d'évaluer un système de RI personnalisé intégrant des dimensions du contexte plus large tels qu'un profil utilisateur à centres d'intérêts multiples, un profil utilisateur mobile, etc. Ceci a conduit l'émergence des approches d'évaluations fondées sur l'utilisation des contextes de recherche simulés ou des contextes réels.

### **2.2. Emergence des approches d'évaluation orientée contexte**

#### **2.2.1. Évaluation par simulation de contextes**

L'évaluation d'un système de RI personnalisé par simulation de contextes consiste à définir des contextes de recherche qui simulent des utilisateurs et interactions hypothétiques. Le protocole d'évaluation présenté dans (Tamine *et al.*, 2007, Daoud *et al.*, 2008) repose sur l'extension des cadres d'évaluation TREC via l'enrichissement de la collection de test par des centres d'intérêts ou profils utilisateurs simulés. Ces contextes sont créés sur la base des interactions hypothétiques fournies par les jugements de pertinence de TREC. L'approche dans (Sieg *et al.*, 2007) utilise une collection de test issue d'une ontologie Web prédéfinie et l'enrichit par des requêtes utilisateurs et des contextes de recherche simulés. Le contexte de l'utilisateur est créé comme étant son profil traduisant un centre d'intérêt et est représenté par un concept de l'ontologie. La requête est générée automatiquement par les termes représentatifs du concept. Quant aux jugements de pertinence, un document retourné par le système est considéré comme pertinent s'il est classifié dans le concept/centre d'intérêt simulé. Les mesures d'évaluation de performance du système sont principalement basées sur les mesures classiques de rappel et précision (Tamine *et al.*, 2008, Daoud *et al.*, 2008)

dans le cas d'une collection de test prédéfinie (e.g. TREC). Le référentiel d'évaluation d'une requête issue d'un concept/centre d'intérêt de l'ontologie dans (Sieg *et al.*, 2007) est construit sur la base des documents Web classifiés sous ce concept. Généralement, l'évaluation par simulation de contextes présente deux avantages majeurs : (1) n'est pas coûteuse en temps puisqu'elle n'implique pas des utilisateurs réels (2) permet d'effectuer une évaluation comparative.

### 2.2.2. Évaluation par utilisation des contextes réels

Ce type d'évaluation est basé sur une étude de cas de vrais utilisateurs dans des contextes de recherche réels permettant d'intégrer les interactions utilisateur-SRI.

Deux types d'évaluation par utilisation des contextes réels peuvent être adoptés ; le premier type est basée sur une approche hybride (Shen *et al.*, 2005) qui intègre une collection de test prédéfinie (e.g. TREC) dans un protocole d'évaluation impliquant des vrais utilisateurs. Cette approche étend les collections TREC par des contextes représentés par l'historique de recherche de l'utilisateur englobant les requêtes et les documents *cliqués*. Chaque requête de la collection TREC est reformulée par les utilisateurs dans le but de définir des requêtes reliées à un même besoin en information et par la suite définir une session de recherche.

Le deuxième type d'évaluation utilise une collection Web via une interface (telle que Google API) et des vrais utilisateurs qui formulent leur requête selon un besoin spécifique. Les interactions des utilisateurs (tels que les *clics*, temps passé sur une page, etc.) seront enregistrés dans un fichier log exploité dans l'étape de l'évaluation de performance du système. L'approche dans (Challam *et al.*, 2007) adopte ce type d'évaluation où les données de *clics* sur un document sont un indice de pertinence du document. Ces documents *cliqués* sont utilisés afin de construire le profil utilisateur en tant que hiérarchie de concepts pondérés traduisant le contexte de recherche.

Les mesures d'évaluation de performance du système peuvent être des mesures classiques de rappel et précision et leur dérivée. Ceci est faisable dans l'évaluation par utilisation des contextes réels grâce à la construction d'un référentiel de pertinence sur la base de l'ensemble des documents jugés pertinents par les différents utilisateurs pour une même requête (Liu *et al.*, 2004). D'autres mesures orienté rang telles que le rang moyen (Challam *et al.*, 2007) et DCG (Discounted Cumulative Gain) (Jarvelin *et al.*, 2002) sont également utilisées et se basent sur la position des documents pertinents dans la liste de résultats. Dans ce cas, les utilisateurs jugent les Top  $N$  documents retournés par le système. Les principales limitations de ce type d'évaluation sont liées au coût en temps et à la reproductibilité des résultats et par conséquent à l'inconsistance de l'évaluation comparative.

## 3. Proposition des cadres d'évaluation d'un système de RI personnalisé à base de sessions

Vu qu'il n'existe pas un cadre standard d'évaluation d'un système de RI personnalisé à base de sessions, nous proposons des cadres d'évaluation issus de collections TREC adaptées à un accès personnalisé à l'information par simulation de contextes.

Dans ces cadres d'évaluation, le contexte simulé représente le profil utilisateur construit sur une session de recherche. Le cadre d'évaluation issu de TREC *ad-hoc* est proposé dans le but d'évaluer un modèle de RI personnalisé avec une délimitation prédéfinie des sessions de recherche. Le cadre d'évaluation issu de HARD TREC est proposé dans le but d'évaluer un modèle de RI personnalisé en l'absence des relations de corrélation entre les requêtes de la collection.

### 3.1. Le cadre d'évaluation issu de TREC *ad-hoc*

Ce cadre d'évaluation a été initialement défini pour l'évaluation de l'accès personnalisé guidé par le profil utilisateur, basé mots clés (Tamine *et al.*, 2008). Nous étendons ce même cadre pour supporter un profil utilisateur basé sur un graphe de concepts issu d'une ontologie Web prédéfinie.

#### 3.1.1. Collection de test

##### A. Requêtes

Le choix des requêtes la collection TREC 1 est guidé par le fait qu'elles sont annotées d'un champ particulier noté " **Domain**" qui décrit un domaine d'intérêt traité par la requête. C'est à juste titre, cette métadonnée qui sera exploitée pour simuler des utilisateurs hypothétiques avec des profils issus de ces domaines.

##### B. Collection de documents

La collection de test de la campagne d'évaluation TREC 1 *ad-hoc* utilisée dans ce cadre, est celle des disques 1, 2 et 3. Les documents de cette collection sont issus de différents articles de presse tels que *Associate Press (AP)*, *Wall street journal (WJS)*, *Financial times*.

##### C. Le profil utilisateur

Le profil utilisateur est un élément intégré dans la collection de test selon un algorithme de simulation qui le génère à partir des requêtes du même domaine décrit comme suit :

1) pour chaque domaine  $k$  de la collection (noté  $Dom^k$  avec  $k = (1..6)$ ), nous sélectionnons, parmi les  $n$  requêtes associées à ce domaine, un sous-ensemble de  $n - 1$  requêtes qui constitue l'ensemble des requêtes d'apprentissage,

2) à partir de cet ensemble d'apprentissage, un processus automatique se charge de construire le profil utilisateur selon la méthode propre du système de RI personnalisé.

#### 3.1.2. Stratégie d'évaluation

La stratégie de validation consiste en un scénario qui se base sur la méthode de la validation croisée et ce, pour ne pas biaiser les résultats avec un seul jeu de test. Nous considérons ici que les sessions de recherche sont définies préalablement par

l'ensemble de requêtes annotées des domaines de TREC. La validation croisée (Mitchell, 1997) ou la *k-fold cross validation* est une méthode d'évaluation qui consiste à diviser la collection de requêtes de test en  $k$  sous ensembles de tailles égales (approximativement), d'utiliser  $k - 1$  sous ensembles pour l'apprentissage du profil utilisateur, et le  $k^{ième}$  sous ensemble pour le test. Dans notre cas, on subdivise l'ensemble des  $n$  requêtes du domaine en un sous-ensemble d'apprentissage de  $n - 1$  requêtes pour apprendre le profil utilisateur et en un sous-ensemble de test contenant la  $n^{ième}$  requête à tester.

### 3.2. Le cadre d'évaluation issu de TREC HARD 2003

Ce cadre est défini dans le but d'évaluer l'efficacité un modèle de RI personnalisé sur des requêtes difficiles et en l'absence d'une connaissance préalable de corrélation entre ces requêtes. Ce cadre d'évaluation consiste à définir une stratégie de test permettant d'évaluer l'efficacité du modèle à travers une séquence de sessions de recherche simulées traitant de sujets différents.

#### 3.2.1. Collection de test

##### A. Requêtes

Le choix des requêtes de la collection *HARD* TREC 2003 a pour but d'augmenter la précision de recherche sur des requêtes difficiles. Vu qu'aucune information concernant la corrélation entre ces requêtes n'existe, nous procédons par la définition des sous-requêtes issues d'une même requête. La requête principale représente un sujet auquel les sous-requêtes générées sont rattachés définissant une session de recherche. Le processus de génération des sous-requêtes d'une même requête est détaillé comme suit :

- 1) Extraire le profil *pertinence* de la requête principale  $q$  en construisant l'ensemble des  $N$  vecteurs documents pertinents associés extraits du fichier de jugements de pertinence fourni par TREC, soit  $dp_q$ ,
- 2) Subdiviser ce profil en  $p$  sous-profils, notés  $sp_i$ ,  $sp_i \subset dp_q$ ,
- 3) Pour chaque sous-profil *pertinence*  $sp_i$ , créer un vecteur centroïde selon la formule :  $c_i(t) = \frac{1}{|sp_i|} \sum_{d \in sp_i} w_{td}$ ,  $w_{td}$  est le poids du terme  $t$  dans le document  $d$  calculé selon la fonction de pondération classique  $tf * idf$ ,
- 4) Représenter la sous-requête par les  $k$  termes les mieux pondérés du vecteur centroïde,
- 5) Eliminer les documents pertinents  $dp_q$  de la requête de la collection de test.

##### B. Collection de documents

Le corpus *HARD* comprend des documents comprenant des textes issus du *NewsWire 1999*, *AQUAINT corpus* et *U.S. government*.

### C. Le profil utilisateur

Le principe de construction du profil utilisateur est analogue à celui décrit dans *TREC adhoc*. Dans ce cadre précisément, (1) la notion de domaine, clairement identifié dans le cas de la collection *TREC ad-hoc* est remplacée par la notion de sujet de requête principal, non connu *a priori*, (2) les requêtes associées aux domaines, sont remplacées par les sous-requêtes associées à la requête principale en cours de traitement, (3) les requêtes servant à la construction du profil sont des sous-requêtes corrélées le long d'une séquence de sessions de recherche simulés.

#### 3.2.2. Stratégie d'évaluation

La stratégie de validation dans ce protocole consiste à diviser l'ensemble des requêtes en un ensemble de requêtes d'apprentissage permettant de paramétrer le système (définir le seuil du mécanisme de délimitation de sessions de recherche) et un ensemble de requêtes de tests permettant d'évaluer notre modèle.

#### A. Phase d'apprentissage

Cette phase est une étape préliminaire dont le but est de déterminer le seuil de corrélation optimal à partir d'une séquence des sessions d'apprentissage. Cette phase consiste à calculer les valeurs de corrélation requête-profil le long d'une séquence des sessions d'apprentissage. Une séquence des sessions d'apprentissage est définie par alignements successifs des sous-requêtes des requêtes d'apprentissage. Les valeurs de corrélations sont calculées entre chaque sous-requête traitée de la séquence et le profil utilisateur créé sur l'ensemble des sous-requêtes précédentes et liés à une même requête.

Pour chaque valeur de seuil de corrélation obtenu, on calcule la précision de détection des requêtes corrélées  $P_{intra}$  et celle de délimitation de sessions de recherche  $P_{inter}$  selon les formules suivantes :

$$P_{intra}(\sigma) = \frac{|CQ|}{|TCQ|}, P_{inter}(\sigma) = \frac{|FQ|}{|TFQ|} \quad [1]$$

où  $|CQ|$  est le nombre de sous-requêtes correctement classifiées comme corrélées,  $|TCQ|$  est le nombre total de sous-requêtes devant être identifiées comme corrélées sur la séquence,  $|FQ|$  est le nombre de sous-requêtes indiquant correctement des frontières de sessions de recherche et  $|TFQ|$  est le nombre total de frontières de sessions de la séquence.

Le seuil de corrélation optimal  $\sigma^*$  est ensuite identifié pour des valeurs de précisions maximales de ( $P_{intra}(\sigma)$  et  $P_{inter}(\sigma)$ ). En effet, le seuil optimal est calculé comme suit :

$$\sigma^* = \operatorname{argmax}_{\sigma} (P_{intra}(\sigma) * P_{inter}(\sigma)) \quad [2]$$

Ce seuil de corrélation est exploité dans la phase de test dans le but de classifier des sous-requêtes dans une même session le long d'une séquence des sous-requêtes de tests.

### **B. Phase de test**

La phase de test est basée sur l'évaluation du système de RI personnalisé le long d'une séquence de sessions issue d'un ensemble de requêtes de tests traitant de sujets différents. Cette évaluation consiste à comparer la performance de recherche classique (requête seule) à la performance de recherche personnalisée (requête et profil associé). Toute sous-requête de la séquence traitée ayant une valeur de corrélation plus grande que le seuil optimal est considérée corrélée au profil construit dans la session en cours de traitement. Par conséquent, le profil utilisateur de la session est utilisé dans le processus de RI personnalisé de cette sous-requête.

Notons que les documents pertinents ayant servi à la création des profils utilisateurs dans cette phase ne sont pas considérés pour l'évaluation des performances associées à ces sous-requêtes. Ceci permet en effet de ne pas biaiser les résultats dans le sens des documents pertinents déjà considérés dans la création du profil.

## **4. Mise en œuvre et résultats**

Nous avons mis en œuvre ces cadres d'évaluation dans le but de tester notre système d'accès personnalisé à l'information. Nous présentons dans la suite notre système de RI personnalisé ainsi que l'évaluation de son efficacité selon les deux cadres d'évaluation proposés.

### **4.1. Conception d'un système d'accès personnalisé à l'information à base de sessions**

Notre approche en RI personnalisée porte sur la définition d'un profil utilisateur selon un graphe de concepts sémantiquement liés et issus d'une ontologie de référence, l'ODP (Daoud *et al.*, 2009). Le profil utilisateur est construit sur une session de recherche définie comme étant une séquence de requêtes liées à un même besoin en information. Notre approche peut être décrite selon deux principales composantes : (1) La construction du profil utilisateur sur une session de recherche, (2) la personnalisation du processus de recherche.

#### *4.1.1. Construction du profil utilisateur selon un graphe issu d'une ontologie*

Le processus de construction du profil utilisateur consiste principalement à combiner les profils des requêtes de la même session. En effet, la méthode se résume par le suivant :

– Pour chaque requête de la session, on extrait la liste des vecteurs associés aux documents pertinents  $dpq$  de chaque requête  $q$ . Partant des vecteurs documents, un vecteur basé mots clés appelé contexte de la requête est construit puis projeté sur l'ontologie de l'ODP aboutissant à la construction du profil  $G_q^s$  de la requête  $q$  soumise à l'instant  $s$ .



– Le profil utilisateur est ainsi construit par combinaison des profils des requêtes d'une même session. Les requêtes sont groupées dans des sessions selon un mécanisme de délimitation des sessions de recherche. Celui-ci est basé sur la corrélation de rangs des concepts du profil courant avec ceux de la nouvelle requête soumise. Le profil utilisateur est alors représenté par un graphe de concepts sémantiquement liés via l'ontologie de l'ODP, noté  $G_u^s$ .

#### 4.1.2. La personnalisation du processus de recherche

Le profil utilisateur  $G_u^s$  construit sur la base d'une session de recherche est exploité dans le réordonnement des résultats de recherche d'une requête  $q^{s+1}$  de la même session. Notre fonction de réordonnement est basée sur la combinaison des scores d'appariement original et contextuel du document :

$$S_f(d_k) = \gamma * S_i(q, d_k) + (1 - \gamma) * S_c(d_k, G_u^s) \quad [3]$$

$$0 < \gamma < 1$$

Le score contextuel du document est calculé selon une mesure de similarité entre son vecteur représentatif  $d_k$  et le vecteur contextuel représentatif du profil adéquat  $G_u^s$ .

$$S_c(d_k, G_u^s) = \frac{1}{h} \cdot \sum_{j=1..h} score(c_j) * \cos(\vec{d}_k, \vec{c}_j) \quad [4]$$

Où  $c_j$  représente un concept du profil,  $score(c_j)$  est le poids de la catégorie  $c_j$  dans le vecteur contextuel et  $h$  est le nombre de concepts du profil utilisateur considérés dans le réordonnement des résultats.

## 4.2. Résultats expérimentaux

La mise en œuvre des cadres d'évaluation proposés a pour objectif d'évaluer l'efficacité de notre modèle d'accès personnalisé à l'information intégrant le profil utilisateur dans le processus de recherche. Cette évaluation consiste à comparer les résultats obtenus par notre modèle aux résultats obtenus par la recherche classique ignorant le profil utilisateur.

Nous avons mené nos expérimentations en utilisant le moteur de recherche "MERCURE" et suivant la stratégie de test associée à chacun des protocoles. Le modèle de recherche classique est basé sur la fonction d'appariement BM25 donnée dans la formule suivante :

$$w_{td} = tf_d \times \frac{\log\left(\frac{n-n_t+0.5}{n+0.5}\right)}{K_1 \times ((1-b) + b \times \frac{dl}{avgdl}) + tf} \quad [5]$$

où  $tf_d$  est la fréquence du terme  $t$  dans le document  $d$ ,  $n$  est le nombre total des documents de la collection de test et  $n_t$  est le nombre de documents contenant le terme  $t$ ,  $K_1 = 2$  and  $b = 0.75$ .

Le modèle de RI personnalisé est basé sur le réordonnement des résultats de recherche de la requête utilisant le profil avec  $\gamma = 0,3$  dans l'équation (3) et  $h = 3$  dans l'équation (4).

Concernant le protocole TREC adhoc, nous avons simulé six domaines présentées dans le tableau 1 dont les requêtes sont numérotées de 51 à 100. Quant au protocole TREC HARD, nous avons utilisé 30 requêtes de HARD TREC subdivisé en 15 requêtes d'apprentissage et 15 requêtes de tests définissant respectivement la séquence des sessions d'apprentissage et celle de test.

Le profil utilisateur est construit selon un processus automatique qui se charge de récupérer tout d'abord la liste de premiers documents pertinents fournies par TREC ( $\|dpq\| = 10$ ) pour chaque requête d'apprentissage dans le cas du protocole TREC adhoc ou la liste des documents du sous-profil pertinence ( $\|sp_i\| = 10$ ) de chaque sous-requête dans le cas du protocole HARD TREC. Partant de cette liste de documents, le profil de la requête est construit, puis le profil utilisateur est défini par combinaison des profils des requêtes d'apprentissage dans TREC ad hoc ou des profils des sous-requêtes corrélées selon le seuil optimal dans HARD TREC.

Les résultats obtenus sont présentés en termes de précision et rappel calculées à diffé-

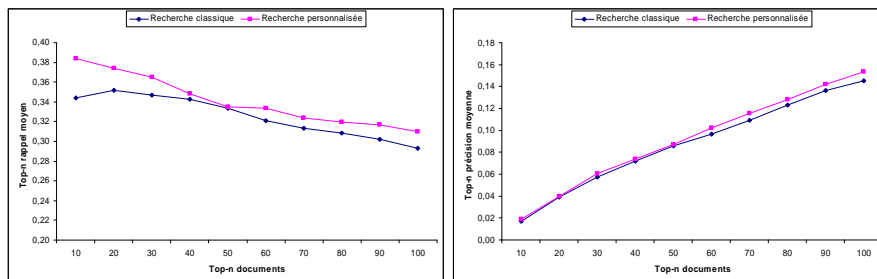
Domaines	Requêtes
Environment	59 77 78 83
Military	62 71 91 92
Law and Government	70 76 85 87
International Relations	64 67 69 79 100
US Economics	57 72 84
International Politics	61 74 80 93 99

**Tableau 1.** Domaines de TREC choisies pour la simulation des profils utilisateurs

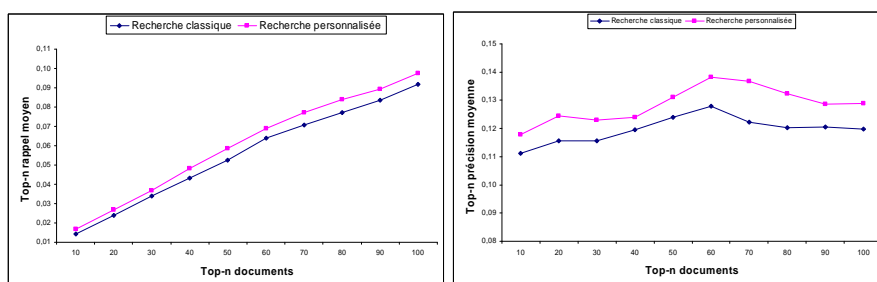
rents points (5, 10, ..., 100 premiers documents restitués). Les résultats présentés dans la figure 1 sont obtenus selon le protocole TREC adhoc et montrent un taux d'accroissement significatif de notre modèle sur l'ensemble des requêtes de tests. Plus précisément, les pourcentages d'amélioration sont de 10% et de 11.6% respectivement pour le rappel au Top-10 rappel et la précision au Top-10. La figure 2 montrent les résultats obtenus selon le protocole TREC HARD. Nous pouvons constater une amélioration significative pour notre modèle aussi bien selon la mesure du rappel que de la précision sur les  $n$  premiers documents restitués par le système. Plus précisément, les pourcentages d'amélioration sont de 23.6% et de 6% respectivement pour le rappel au Top-10 rappel et la précision au Top-10.

La différence du taux d'efficacité de la recherche personnalisé sur les deux collections TREC est du à plusieurs facteurs. Le premier concerne le degré de corrélation des requêtes d'un même domaine (TREC adhoc) ou des sous-requêtes d'une même requête (HARD TREC). Le deuxième facteur est lié à la précision du mécanisme de délimitation des sessions de recherche. En effet, ce mécanisme a un impact sur la précision de représentation du profil utilisateur qui est créé sur la base des requêtes identifiées comme corrélées selon le seuil optimal. Toutefois, les résultats montrent bien un pour-

centage d'amélioration significatif sur les deux collections ce qui montre la stabilité de performance de notre modèle de RI personnalisé.



**Figure 1.** Evaluation de performance du modèle de RI personnalisée en termes de Top-n précision moyenne et Top-n rappel moyen sur TREC adhoc



**Figure 2.** Evaluation de performance du modèle de RI personnalisée en termes de Top-n précision moyenne et Top-n rappel moyen sur HARD TREC

## 5. Bilan et perspectives

Les cadres d'évaluation que nous avons proposés dans ce papier suivent une méthodologie de simulation du profil de l'utilisateur par découpage thématique des sessions de recherche en utilisant des ressources TREC. Notre approche de personnalisation proposée intègre le profil utilisateur construit sur une session de recherche dans le processus de recherche d'information. L'évaluation de notre modèle selon les deux cadres montre bien la stabilité de performance des résultats sur des collections de test différentes.

Les perspectives de recherche ouvertes par ce travail portent sur l'amélioration des cadres d'évaluation proposés. Ces cadres peuvent être étendus dans le sens d'adapter une stratégie d'évaluation basée sur l'utilisation des données réelles issues du log des interactions utilisateurs. Plus généralement, nous envisageons de définir dans le futur

Sixième édition de la Conférence en Recherche d'Information et Applications (CORIA 2009)

un cadre d'évaluation standard des systèmes d'accès contextuel à l'information guidé par les centres d'intérêts de l'utilisateur.

## 6. Bibliographie

- Borlund P., « The IIR evaluation model : A framework for evaluation of interactive information retrieval systems », *Journal of Information Research*, vol. 8, n° 3, p. 152, 2003.
- Challam V., Gauch S., Chandramouli A., « Contextual Search Using Ontology-Based User Profiles », *Proceedings of RIAO 2007, Pittsburgh USA*, 2007.
- Cleverdon C., « The Cranfield test on index language devices », *Aslib*, p. 173-194, 1967.
- Daoud M., Tamine L., Boughanem M., « Using a concept-based user context for search personalization », *International Conference of Data Mining and Knowledge Engineering (ICDMKE), London, UK*, p. 57-64, 2008.
- Daoud M., Tamine L., Boughanem M., Chebaro B., « A Session Based Personalized Search Using An Ontological User Profile », *ACM Symposium on Applied Computing (SAC), Hawaii (USA)*, ACM, p. 1031-1035, march, 2009.
- Ingwersen P., Jarvelin K., *The turn : Integration of information seeking and retrieval in context*, Springer, 2005.
- Jarvelin K., Kekäläinen J., « Cumulated gain-based evaluation of IR techniques », *ACM Trans. Inf. Syst.*, vol. 20, n° 4, p. 422-446, 2002.
- Liu F., Yu C., Meng W., « Personalized Web Search For Improving Retrieval Effectiveness », *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, n° 1, p. 28-40, 2004.
- Mitchell T. M., « Machine Learning, McGraw-Hill Higher Education », 1997.
- Mizzaro S., « How many relevances in information retrieval ? », *Interacting with Computers*, vol. 10, n° 3, p. 303-320, 1998.
- Shen X., Tan B., Zhai C., « Context-sensitive information retrieval using implicit feedback », *Proceedings of the 28th annual international ACM SIGIR conference*, ACM, New York, NY, USA, p. 43-50, 2005.
- Sieg A., Mobasher B., Burke R., « Web search personalization with ontological user profiles », *CIKM'07 : Proceedings of the sixteenth ACM conference on information and knowledge management*, ACM, New York, NY, USA, p. 525-534, 2007.
- Tamine L., Boughanem M., Daoud M., « Evaluation of contextual information retrieval : overview of issues and research », *Knowledge and Information Systems (Kais)*, 2009.
- Tamine L., Boughanem M., Zemirli W., « Exploiting Multi-Evidence from Multiple User's Interests to Personalizing Information Retrieval », *IEEE International Conference on Digital Information Management(ICDIM 2007)*, 2007.
- Tamine L., Boughanem M., Zemirli W. N., « Personalized document ranking : Exploiting evidence from multiple user interests for profiling and retrieval », *Journal of Digital Information Management*, vol. 6, n° 5, p. 354-365, octobre, 2008.
- Voorhees E. M., « Overview of TREC 2001 », *TREC*, 2001.