



Définition d'un profil multidimensionnel de l'utilisateur : Vers une technique basée sur l'interaction entre dimensions

Lynda Tamine, Wahiba Bahsoun

► To cite this version:

Lynda Tamine, Wahiba Bahsoun. Définition d'un profil multidimensionnel de l'utilisateur : Vers une technique basée sur l'interaction entre dimensions. Actes de la Conférence francophone en Recherche d'Information et Applications (CORIA 2006), 2006, Lyon, France. pp.225-236, 2006. <hal-00359573>

HAL Id: hal-00359573

<https://hal.archives-ouvertes.fr/hal-00359573>

Submitted on 8 Feb 2009

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

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

Définition d'un profil multidimensionnel de l'utilisateur :

Vers une technique basée sur l'interaction entre dimensions

Lynda Tamine, Wahiba Bahsoun

*Laboratoire IRIT (Equipe SIG)
Université Paul Sabatier
118 Route de Narbonne
31400 Toulouse CEDEX 06
tamine, wbahsoun@irit.fr*

RÉSUMÉ. La personnalisation d'un processus d'accès à l'information a pour objectif de délivrer à l'utilisateur une information appropriée à ses préférences, ses centres d'intérêts ou plus globalement son profil. Ce papier présente une technique de construction du profil de l'utilisateur qui s'inscrit dans une approche statistique utilisant le comportement de l'utilisateur comme source permettant de prédire implicitement son modèle. Cette technique s'articule plus particulièrement sur l'interaction entre dimensions du profil représentées par l'historique des recherches et centres d'intérêt de l'utilisateur.

ABSTRACT. Personalized information access is to carry out retrieval for each user incorporating his preferences, interests or more globally his profile. This paper presents a technique to built and learn a user profile; it's based on a statistical approach which exploits the user's behaviour in order to infer implicitly the corresponding model. The technique focuses on the interaction of user's past search history and interests viewed as his profile dimensions.

MOTS-CLÉS : accès personnalisé à l'information, profil utilisateur, contexte

KEYWORDS: personalized information access, user profile, context

1. Introduction

L'essor du Web, conforté par le rapide développement des nouvelles technologies de l'information et de la communication, lance à la communauté en recherche d'information (RI) de nouveaux défis. Certes, les systèmes de recherche d'information (SRI) sont des outils qui ont permis, jusqu'à aujourd'hui, d'améliorer sans cesse la qualité des services d'accès à l'information, grâce à la capitalisation des théories issues de nombreux travaux de recherche ; cependant, la surabondance de l'information ainsi que sa large accessibilité a engendré une dégradation de leurs performances tant en efficacité qu'en efficience. Du point de vue de l'utilisateur, ceci se traduit par le volume et l'inintelligibilité des informations retournées [BUD 00, NUN 97], sa désorientation et démotivation face à des interfaces qui induisent une surcharge cognitive. C'est pourquoi les travaux s'orientent actuellement vers l'adaptation du cycle de vie d'un processus d'accès à l'information à un utilisateur spécifique en vue de lui délivrer une information pertinente relativement à ses besoins précis, son contexte et ses préférences. Ces travaux s'inscrivent dans le cadre général de la personnalisation de l'information. C'est une direction de recherches qui combine un ensemble de technologies et de connaissances portant sur la requête et le contexte de l'utilisateur, dans une même infrastructure et ce, dans le but de délivrer les réponses les mieux appropriées à son besoin en information [ALL 02]. Les applications sont diverses : systèmes de recommandation, systèmes de filtrage de messages et d'informations, systèmes d'apprentissage, moteurs de recherche d'information etc.

Indépendamment de l'objectif applicatif visé, on identifie trois principaux aspects à promouvoir dans les systèmes d'accès personnalisés : (a) capacité à identifier l'intention conceptuelle de l'utilisateur (b) flexibilité du processus de sélection de l'information en vue de s'adapter au contexte d'utilisation courant (c) intelligibilité des interactions utilisateur - système. C'est précisément le premier aspect que l'on aborde dans le présent article. Notre objectif, dans ce cadre, est en effet de proposer une structure descriptive de l'utilisateur, appelée communément profil, qui permet de traduire deux principaux facteurs qui constituent des raisons fondamentales qui plaident pour la personnalisation de l'accès à l'information, à savoir que [SU 03] :

- les utilisateurs ont des objectifs différents, des centres d'intérêts différents, en conséquence des perceptions différentes de la notion de pertinence,
- un même utilisateur peut avoir différents centres d'intérêts à différents instants.

A cet effet, on décrit l'utilisateur par un profil à deux dimensions [TAM 05]. La première représente l'historique de ses interactions avec le SRI. La seconde caractérise ses besoins récurrents en information ; elle est inférée et évolue à partir de la première dimension, et permet de définir différents centres d'intérêt. Ces derniers sont à intégrer, en perspective, dans une librairie exploitée dans un processus d'appariement requête - document - utilisateur.

L'organisation retenue pour cet article est la suivante : la section 2 résume la problématique de l'accès personnalisé à l'information sous l'angle de la modélisation de l'utilisateur, puis rapporte une synthèse des travaux du domaine. La section 3 présente

de manière détaillée notre approche pour la définition des dimensions descriptives du profil. La section 4 résume notre proposition et en présente les perspectives.

2. Accès personnalisé à l'information : quel modèle pour l'utilisateur ?

Les modèles classiques de recherche d'information sont fondés sur la supposition naïve que l'utilisateur est complètement représenté par sa requête ; ceci a pour corollaire que les résultats retournés pour une même requête sont identiques même si elle est exprimée par des utilisateurs différents. Les problèmes immédiats posés par une telle hypothèse sont notamment l'ambiguïté du sens des mots, l'impossibilité de sélectionner des sources opportunes et l'inintelligibilité des résultats [BUD 00]. En outre, ces problèmes sont d'autant plus accentués que les requêtes sont courtes (2.29 mots par requête) et que les sources d'information sont volumineuses et hétérogènes.

Les premières solutions apportées à ce type de problèmes et pouvant s'apparenter à la personnalisation sont les techniques de reformulation de requêtes par injection de pertinence [ROC 71, RUT 03a]. Cependant, vu le contexte actuel lié au volume d'informations, ces techniques sont peu viables [JAN 98, SPI 98]. Les travaux s'orientent actuellement vers une définition plus large de l'utilisateur, qui a donné naissance à la RI contextuelle. C'est un courant de recherche qui vise la mise en oeuvre de systèmes centrés utilisateurs, non dans le sens d'un utilisateur générique, mais d'un utilisateur spécifique et ce, en vue d'adapter son fonctionnement à son contexte précis. Cette section présente les notions ayant émergé de cette direction de travaux puis en rapporte les principales contributions portant sur la modélisation de l'utilisateur.

2.1. *L'utilisateur ou les notions de contexte, situation et profil*

Les notions de contexte et situation sont en amont de l'interaction utilisateur-système d'accès à l'information. Ces notions ont été initialement introduites, sans distinction de sens, par les travaux de Saracevic [SAR 97, BEL 96] et Ingerwersen [ING 96]. Le contexte (ou situation) y est défini comme l'ensemble des facteurs cognitifs et sociaux ainsi que les buts et intentions de l'utilisateur au cours d'une session de recherche. Une tentative de distinction entre ces notions a fait l'objet d'autres travaux [ALL 97, SON 99, COO 01] qui précisent que le contexte couvre des aspects larges tels que l'environnement cognitif, social et professionnel dans lesquels s'inscrivent des situations liées à des facteurs tels que le lieu, temps et l'application en cours. C'est le sens générique du contexte qui a été largement exploré cette dernière décennie en RI contextuelle [LAW 00, QUI 02, RUT 03b, ING 04, BOT 04]. Même si les auteurs ne convergent vers une même définition, on retrouve toutefois des dimensions descriptives communes telles que l'environnement cognitif, le besoin mental en information, et l'interaction liée à la recherche d'information [COO 02]. Par ailleurs, la notion de profil couvre, à notre connaissance ces mêmes aspects, à la seule différence que la définition initiale (réductrice de la définition actuelle du contexte) et l'usage de ce terme revient à la communauté en filtrage d'information.

En résumé, on peut définir globalement le contexte ou profil de l'utilisateur, dans le cadre d'une activité de recherche d'information, comme l'ensemble des dimensions qui permettent de décrire et/ou inférer ses intentions et perception de la pertinence. Les travaux en RI contextuelle abordent alors peu ou prou l'une ou l'autre de ces dimensions pour décrire l'utilisateur et l'intégrer à terme dans le processus d'accès à l'information.

2.2. Aperçu de l'état de l'art

Le but fondamental de la plupart des travaux actuels en RI contextuelle, portant particulièrement sur la modélisation du profil de l'utilisateur, est de représenter, construire puis faire évoluer ses besoins en information à court et moyen terme. C'est une question qui pose la double difficulté de traduire les centres d'intérêt de l'utilisateur d'une part et faire émerger leur diversité d'autre part. On synthétise dans ce qui suit, les principales approches de représentation, construction et évolution du profil.

1) *Représentation du profil*. Il existe principalement trois types de représentation :

- *ensembliste* : le profil y est généralement formalisé comme des vecteurs de termes pondérés [BUD 00, DUM 03] ou classes de vecteurs [MCG 03],

- *sémantique* : la représentation du profil met en évidence, dans ce cas, les relations sémantiques entre informations le contenant. La représentation est essentiellement basée sur l'utilisation d'ontologies [GAU 03, CHA 04, NAN 03, LIU 04] ou réseaux sémantiques probabilistes [LIN 05, WEN 04],

- *multidimensionnelle* : le profil est structuré selon un ensemble de dimensions, représentées selon divers formalismes [AMA 99, BOU 05].

2) *Construction du profil*. La construction du profil traduit un processus qui permet d'instancier sa représentation. Ce processus est généralement implicite basé sur un procédé d'inférence du contexte et préférences de l'utilisateur via son comportement lors de l'utilisation d'un système d'accès à l'information [KEL 04] ou de diverses autres applications [BUD 00, MCG 03, DUM 03]. Les informations exploitées sont généralement : le contenu des documents consultés et/ou imprimés et/ou sauvegardés, les liens explorés, les dernières pages visitées etc.

3) *Evolution du profil*. L'évolution des profils désigne leur adaptation à la variation des centres d'intérêt des utilisateurs qu'ils décrivent, et par conséquent, de leurs besoins en information au cours du temps . L'évolution y est d'avantage abordée comme un problème de représentation de la diversité des domaines d'intérêts de l'utilisateur en utilisant des techniques de classification [PAZ 96, MIZ 02] ou heuristiques liées la notion de cycle de vie artificielle d'un centre d'intérêt [CHE 98] .

Comparativement aux travaux les plus proches cités ci-dessus, notre approche de définition du profil s'en distingue par les principaux points suivants :

- le profil est formellement représenté par deux dimensions corrélées : historique des interactions et centres d'intérêt,

– la dimension *historique des interactions* est le résultat d’une agrégation des informations issues de l’évaluation implicite de la pertinence en relativisant leur importance par rapport au profil de l’utilisateur ;

– la dimension *centres d’intérêt* est dérivée à partir de la seule dimension *historique des interactions*, sans utilisation d’autres ressources telles que les ontologies ou classifieurs de concepts ;

3. Notre définition du profil

Selon notre approche, le profil de l’utilisateur est multidimensionnel, décrit plus précisément, par deux dimensions. La première représente l’historique de ses interactions avec le SRI ; elle est exploitée pour inférer la seconde dimension représentée par les divers centres d’intérêt de l’utilisateur. Les deux dimensions évoluent corrélativement au cours du temps. De manière sommaire, on définit le profil d’un utilisateur à l’instant s par $U = (H^s, I^s)$ où H^s représente l’historique des interactions de l’utilisateur jusqu’à l’instant s et I^s représente la librairie de ses centres d’intérêt inférés jusqu’à l’instant s . Plus précisément, notre procédé de modélisation du profil se décline en un cycle comportant deux principales étapes.

La première étape consiste à représenter l’historique des interactions de l’utilisateur avec le SRI par agrégation des informations collectées à partir de ses sessions de recherche successives. Une session de recherche est particulièrement décrite par l’association d’une requête et d’un ensemble de documents jugés explicitement ou implicitement par l’utilisateur. La seconde étape a pour but d’apprendre les centres d’intérêt de l’utilisateur en se basant sur les contextes associés aux sessions de recherche passées. Les contextes décrivent des besoins à court terme et sont exprimés à l’aide de listes de termes dominants pondérés, construites à partir de l’historique des interactions. L’apprentissage des centres d’intérêt est alors basé sur une mesure de corrélation qui évalue le degré de changement entre contextes de la recherche sur une période donnée. Les paragraphes qui suivent détaillent les principes de représentation de l’historique des interactions et dérivation des centres d’intérêt de l’utilisateur.

3.1. *Historique des interactions*

Soit q^s la requête soumise par un utilisateur U à la session de recherche se déroulant à l’instant s , soit S^s , et D^s l’ensemble des documents pertinents pour l’utilisateur durant cette session. Un document est considéré comme pertinent s’il a été ainsi jugé par l’utilisateur de manière explicite ou implicite. A cet effet, on se réfère à une catégorisation du comportement de l’utilisateur, traduisant des jugements implicites de pertinence, largement adoptée par les travaux du domaine [KEL 04]. Cette catégorisation construit une fonction *pertinence* qui associe à chaque action type (impression, lecture, sauvegarde etc.) un degré de pertinence. Pour notre part, on utilise une fonction constante qui traduit d’avantage un indicateur de pertinence sûre. On note

alors $R_u^s = \cup_{s_0..s} D^s$ l'ensemble des documents déjà *visités* et jugés pertinents par l'utilisateur lors des sessions de recherche passées depuis l'insatnt s_0 .

On propose dans ce qui suit l'utilisation de matrices pour la représentation d'une session de recherche et de l'historique des interactions.

3.1.1. Représentation d'une session de recherche

La session de recherche S^s est représentée par une matrice Document-Terme $D^s * T^s$ où T^s est l'ensemble des termes qui indexent les documents D^s (T^s est une partie de l'ensemble des termes représentatifs des documents préalablement jugés pertinents noté $T(R_u^s)$). Chaque ligne de la matrice S^s représente un document $d \in D^s$, chaque colonne représente un terme $t \in T^s$.

Dans le but d'affiner la représentation Document-Terme, on propose de décliner l'importance d'un terme relativement au profil de l'utilisateur dans le schéma de pondération terme-document. A cet effet, on calcule pour chaque terme t dans un document d à l'instant s , un coefficient de pertinence $CPT^s(t, d)$ qui traduit la pertinence relative d'un terme compte tenu des jugements de pertinence qu'il a émis et qui sont supposés êtres des indicateurs de son centre d'intérêt courant. L'expression de ce coefficient est fondée sur l'hypothèse qu'un terme est d'autant plus important pour l'utilisateur qu'il cooccure avec les termes qui lui sont *familiers* en ce sens qu'ils sont présents dans des documents déjà jugés. Les dépendances entre termes associés à des documents préalablement jugés sont vues comme des règles d'association [LIN 98].

Définition 1. Le coefficient de pertinence d'un terme t dans un document d à l'instant s noté $CPT^s(t, d)$ est défini comme suit :

$$CPT^s(t, d) = \frac{w_{td}}{dl} * \sum_{t' \neq t, t' \in T(R_u^s)} coc(t, t') \quad [1]$$

w_{td} est le poids du terme t dans le document d calculé selon le schéma classique Tf-Idf, dl est la longueur du document d , $coc(t, t')$ est le degré de confiance de la règle ($t \rightarrow t'$), $coc(t, t') = \frac{n_{tt'}}{n_t}$, $n_{tt'}$ est le nombre de documents contenus dans l'ensemble R_u^s et contenant les termes t et t' , n_t est le nombre de documents contenus dans l'ensemble D^s contenant le terme t .

$S^s(d, t)$ est alors ainsi construit :

$$S^s = (CPT^s)^T \quad [2]$$

Où T est l'opérateur transposée de matrice

3.1.2. De la session de recherche à l'historique des interactions

L'historique des interactions de l'utilisateur est représenté par une matrice notée H^s de dimensions $R_u^s * T(R_u^s)$. Cette matrice est construite de manière incrémentale en ce sens qu'elle est mise à jour à chaque session de recherche en y reportant, par

agrégation, les informations issues de la matrice S^s . A cet effet, on propose la définition d'un opérateur d'agrégation qui combine pour chaque terme son poids classique dans le document et ses poids atténués par les coefficients de pertinence calculés lors des sessions de recherche passées.

Définition 2. L'opérateur d'agrégation des sessions de recherche, noté \oplus , est défini comme suit :

$$H^0(d, t) = S^0(d, t)$$

$$H^{s+1}(d, t) = H^s(d, t) \oplus S^{s+1}(d, t) = \begin{cases} \alpha * w_{t,d} + \beta * S^{s+1}(d, t) & \text{si} \\ t_j \notin T(R_u^{(s-1)}) \\ \alpha * H^s(d, t) + \beta * S^{s+1}(d, t) & \text{si} \\ t_j \in T(R_u^{(s-1)}) \text{ et } d \in R_u^{(s-1)} \\ H^s(d, t) & \text{sinon} \end{cases} \quad [3]$$

$$(\alpha + \beta = 1), s > s_0$$

La définition de l'opérateur \oplus est fondée sur l'hypothèse que les termes associés aux centres d'intérêt de l'utilisateur sont récurrents. L'idée est alors d'affiner les descripteurs des documents déjà jugés par :

- expansion éventuelle avec des termes associés présents dans des documents pertinents,
- combinaison de l'importance classique de ses termes (relativement à la collection de documents) et de leur pertinence relative au profil, calculée à l'aide du coefficient $CPT(t, d)$ en cours des sessions de recherche passées.

3.2. Les centres d'intérêt

Les centres d'intérêt de l'utilisateur, contenus dans une librairie notée I^s , constituent la seconde dimension de son profil traduisant ses besoins récurrents en information ; cette dimension est précisément construite sur la base de la première, en l'occurrence, l'historique de ses interactions. On propose à cet effet une méthode cyclique qui procède en deux étapes. La première a pour objet d'extraire à partir de l'historique des interactions un centre d'intérêt candidat qualifié de *contexte d'usage*, qui traduit un besoin ponctuel en information. L'objectif de la seconde étape est alors d'intégrer le contexte ainsi découvert dans la librairie I^s en respectant l'hypothèse de diversité éventuelle des centres d'intérêt. Ceci traduit, à juste titre, la phase d'apprentissage des centres d'intérêt qui induit l'évolution du profil.

3.2.1. Extraction d'un contexte d'usage

A l'issue d'un cycle d'apprentissage représentant un nombre déterminé de sessions de recherche S^s agrégées dans l'historique H^s , est construit un contexte d'usage courant c^s défini comme suit :

Définition 3. Un contexte d'usage traduit un besoin en information ponctuel exprimé sur une courte période d'interaction avec le SRI ; il est représenté par un vecteur de termes pondérés, ordonnés par leur degré de représentativité du contexte ; pour chaque terme $t \in T(R_u^s)$, on calcule le poids associé comme suit :

$$c^s(t) = \sum_{d \in R_u^s} H^s(d, t) \quad [4]$$

$$c^s(t) \text{ est normalisé } c^s(t) = \frac{c^s(t)}{\sum_{t \in T(R_u^s)} c^s(t)}.$$

Un contexte d'usage est ainsi un vecteur extrait à partir de l'historique des interactions en sommant chaque colonne de la matrice associée.

3.2.2. Apprentissage des centres d'intérêt

On se base sur l'hypothèse qu'un utilisateur a divers centres d'intérêt et qu'il peut basculer d'un centre d'intérêt vers un autre en cours de sessions de recherche successives. Dans le but d'évaluer le degré de changement des contextes liés aux centres d'intérêt de l'utilisateur, on compare le contexte courant cc avec le contexte extrait à la période précédente pc en utilisant le coefficient de corrélation des rangs de Kendall comme suit :

$$\Delta I = \frac{\sum_t \sum_{t'} S_{tt'}(pc) * S_{tt'}(cc)}{\sqrt{(\sum_t \sum_{t'} S_{tt'}(pc))^2 (\sum_t * \sum_{t'} S_{tt'}(cc))^2}} \quad [5]$$

où $S_{tt'}(pc) = \text{Signe}(pc(t) - pc(t')) = \frac{pc(t) - pc(t')}{|pc(t) - pc(t')|}$, $S_{tt'}(cc) = \text{Signe}(cc(t) - cc(t'))$.

La valeur du coefficient ΔI est inscrite dans l'intervalle $[-1..1]$, où une valeur proche de -1 signifie que les contextes sont non similaires alors qu'une valeur proche de 1 signifie que les contextes sont proches sémantiquement. Le coefficient ΔI est approché par une loi Laplace Gauss [SAP 90] : $\Delta I \approx LG \left(0; \sqrt{\frac{2(2n+5)}{9n(n-1)}} \right)$ avec n , le nombre de termes traités, soit dans notre cas, $T(R_u^s)$. Le seuil critique de corrélation σ est alors donné par la table correspondante.

L'apprentissage des centres d'intérêt, qui a pour effet la mise à jour éventuelle de la librairie I^s , est alors déterminé par le résultat de comparaison de ΔI relativement au seuil de corrélation σ . Plus précisément, la stratégie est la suivante :

1) $\Delta I > \sigma$. Les sessions de recherche sont inscrites dans le même contexte : pas d'indication sur l'évolution des centres d'intérêt l'utilisateur ;

2) $\Delta I < \sigma$. Détection d'un changement de contexte ; deux configurations se présentent : découverte d'un nouveau centre d'intérêt ou évolution d'un autre préalablement découvert. On procède alors de la manière suivante :

- sélectionner $c^* = \operatorname{argmax}_{c \in I^s} (c \circ cc)$,
- si $cc \circ c^* > \sigma$ alors affiner le descripteur du centre d'intérêt c^* , mettre à jour la matrice H^s par élimination des lignes les moins récemment recalculées R_u^s ,
- si $cc \circ c^* < \sigma$ alors élargir la librairie des centres d'intérêt ie $I^{s+1} = I^s \cup c^*$, réinitialiser la matrice H^s de manière à privilégier l'apprentissage de ce nouveau centre d'intérêt, poser $s_0 = s$.

Les centres d'intérêt ainsi construits peuvent être réutilisés dans différentes étapes d'un processus d'accès personnalisé à l'information. La librairie I^s constitue alors une ressource pour la définition d'heuristiques et/ou techniques de reformulation de requête, adaptation de la fonction d'évaluation de la pertinence, réordonnancement des résultats de recherche etc.

4. Conclusion

Dans la perspective d'un accès personnalisé à l'information, nous avons défini dans cet article, un profil multidimensionnel pour modéliser l'utilisateur. Plus précisément, ce dernier est décrit au travers l'interaction de deux dimensions. La première traduit l'historique de ses interactions avec le SRI, représentée par une matrice issue de l'application d'un opérateur d'agrégation sur les informations collectées implicitement lors des sessions de recherche successives. La seconde dimension traduit les centres d'intérêt de l'utilisateur dérivés automatiquement à partir de l'historique des interactions. Ces centres d'intérêt évoluent selon une stratégie basée sur une mesure de corrélation permettant de scruter et traduire, en cours du temps, tant la variation de leur structure que leur diversité.

La technique de définition du profil ainsi présentée est essentiellement caractérisée par :

- 1) la définition et utilisation d'une mesure de pertinence relative des termes pour un profil : cette mesure considère l'information véhiculée par les documents jugés par l'utilisateur,
- 2) l'exploitation de la seule dimension historique des recherches de l'utilisateur pour la construction et évolution de ses centres d'intérêt : aucune autre ressource n'est requise,
- 3) l'utilisation d'une méthode statistique pour maintenir la diversité des centres d'intérêt : cette méthode produit, à terme, une librairie qui peut être utilisée comme ressource pour personnaliser l'accès à l'information.

Des points méritent cependant d'être affinés. Le plus important porte sans doute sur la définition des périodes d'apprentissage des centres d'intérêt. En effet comme la variation des centres d'intérêt de l'utilisateur, décelée à travers les requêtes qu'il a émises, ne présentent pas forcément des régularités prévisibles, la méthode statistique proposée serait confrontée à un risque d'erreur difficilement mesurable. Même si ce risque pourrait être amoindri en réduisant au mieux ces périodes, il n'en demeure pas

moins qu'une réflexion plus poussée sur un compromis entre les différents paramètres qui régulent l'évolution des centres d'intérêt d'un utilisateur, mérite d'être menée.

Par ailleurs, nous envisageons à court terme, de valider ces propositions dans le cadre d'une tâche d'évaluation inscrite dans les étapes du projet APMD . A cet effet, nous mettrons en oeuvre un cadre d'évaluation où seront définis des collections de test et mesures d'efficacité selon des protocoles proches de ceux qui sont définis dans le cadre de la campagne internationale d'évaluation en RI, en l'occurrence TREC .

Remerciements

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 (<http://apmd.prism.uvsq.fr/>).

5. Bibliographie

- [ALL 97] ALLEN B., « Information seeking in context : Proceedings of an International Conference on Research in needs, seeking and use in different context », 1997, p. 111-122.
- [ALL 02] ALLAN J., ALL, « Challenges in information retrieval and language modelling », *Workshop held at the center for intelligent information retrieval*, Septembre 2002.
- [AMA 99] AMATO G., STARACCIA U., « User profile modelling and applications to digital libraries », *Proceedings of the 3rd European Conference on Research and advanced technology for digital libraries*, 1999, p. 184-187.
- [BEL 96] BELKIN N., *Intelligent model of information retrieval : Whose intelligent ?*, In J. Krause and M. Hrfurth and J. Marx Editions, 1996.
- [BOT 04] BOTTRAUD J.C. B. G. B. M., « Expansion de requêtes par apprentissage automatique dans un assistant pour la recherche d'information », 2004, p. 89-105.
- [BOU 05] BOUZEGHOUB M., KOSTADINOV D., « Personnalisation de l'information : Aperçu de l'état de l'art et définition d'un modèle flexible de définition de profils », *Actes de la 2nde Conférence en Recherche d'Information et Applications CORIA*, 2005, p. 201-218.
- [BUD 00] BUDZIK J., HAMMOND K., « User interactions with every applications as context for just-in-time information access », *Proceedings of the 5th international conference on intelligent user interfaces*, Mars 2000, p. 44-51.
- [CHA 04] CHALLAM V., « Contextual information retrieval using ontology based user profiles », *Master of science in computer science*, Jawaharlal Nehru Technological University, 2004.
- [CHE 98] CHEN L. AND SYCARA K., « WebMate : A personal agent for browsing and searching », *Proceedings of the 2nd international conference on autonomous agents and multi agent systems*, Minneapolis, 1998, p. 10-13.
- [COO 01] COOL C., « The concept of situation in information science », *Annual review of information science and technology*, vol. 35, 2001, p. 5-42.

- [COO 02] COOL C., SPINK A., « Issues of context in information retrieval : an introduction to the special issue », *In Journal of Information Processing and Management (IPM)*, vol. 38, n° 55, 2002, p. 605-611.
- [DUM 03] DUMAIS S., CUTTREL E., CADIZ J., JANCKE G., SARIN R., ROBBINS D., « Stuff I've seen : A system for a personal information retrieval and re-use », *Proceedings of the 26th ACM SIGIR International Conference on Research and Development*, 2003, p. 72-79.
- [GAU 03] GAUCH S., CHAFFE J., PRETSCHNER P., « Ontology based user profiles for search and browsing », vol. Special issue on user modelling for Web and hypermedia information retrieval, 2003.
- [ING 96] INGERWERSEN P., « Cognitive perspectives of information interactions : elements of a cognitive information retrieval theory », *Journal of documentation*, vol. 52, n° 1, 1996, p. 3-50.
- [ING 04] INGERWERSEN P., JARVELIN K., « Information retrieval in context », *In Proceedings of the 27th ACM SIGIR Workshop on information retrieval in context*, July 2004, p. 6-8.
- [JAN 98] JANSEN B., SPINK A., BATEMAN J., « Searchers : the subjects they search and sufficiency : a study of large sample of EXCITE searches », *Proceedings of the 1998 Web-net World conference of the WWW, Internet and Intranet*, Faculty of science, University college Dublin, 1998, p. 7-12.
- [KEL 04] KELLY N. J., « Understanding implicit feedback and document preference : a naturalistic study », *PHD dissertation*, Rutgers University, New Jersey, January 2004.
- [LAW 00] LAWRENCE S., « Context in Web search, IEEE Data Engineering », *Bulletin*, vol. 23, n° 3, 2000, p. 25-32.
- [LIN 98] LIN S., SHIH C., CHEN M., HO J., KO M., HUANG Y. M., « Extracting classification knowledge of Internet documents with mining term-associations : A semantic approach », *In the 21th International SIGIR Conference on Research and Development in Information Retrieval*, 1998, p. 241-249.
- [LIN 05] LIN C., XUE G., ZENG H., YU Y., « Using probabilistic latent semantic analysis for personalised Web search », *Proceedings of the APWeb Conference*, 2005, p. 707-711.
- [LIU 04] LIU F., YU C., « Personalized Web search for improving retrieval effectiveness », *IEEE Transactions on knowledge Data Engineering*, vol. 16, 2004, p. 28-40.
- [MCG 03] MC GOWAN J., « A multiple model approach to personalised information access », *Master thesis in computer science*, Faculty of science, University college Dublin, February 2003.
- [MIZ 02] MIZARRO S., TASSO C., « Ephemeral and persistent personalisation in adaptive information access to scholarly publications on the Web », *Proceedings of the 2nd International Conference on adaptive hypermedia and adaptive Web-based systems*, 2002, p. 306-316.
- [NAN 03] NANAS N., UREN U., DEROECK A., « Building and applying a concept hierarchy representation of a user profile », *Proceedings of the 26th Annual ACM Conference on Research and Development in Information Retrieval SIGIR*, 2003, p. 154-204.
- [NUN 97] NUNBERG G., « As Google goes, so goes the nation », *New York times*, mai 1997.
- [PAZ 96] PAZANNI M., MURAMATSU J., BILLSUS D., « Syskill and Webert : Identifying interesting Web sites », *Proceedings of the 13th National Conference on Artificial intelligence*, 1996, p. 54-61.

- [QUI 02] QUIROGA L.M.AND MOSTAFA J., « An experiment in building profiles in information filtering : the role of context of user relevance feedback », *Journal of Information Processing and Management (IPM)*, vol. 38, 2002, p. 671-694.
- [ROC 71] ROCCHIO J., « Relevance feedback in Information retrieval », Englewood Cliffs, 1971, Prentice Hall.
- [RUT 03a] RUTHVEN I., LALMAS M., « A survey on the use of relevance feedback for information access systems », vol. 18, 2003, p. 95-145.
- [RUT 03b] RUTHVEN I., M. L., RIJSBERGEN V., « Incorporating user search behaviour into relevance feedback », *Journal of the American Society for Information Science and Technology*, vol. 54, n° 6, 2003, p. 529-549.
- [SAP 90] SAPORTA G., « Probabilités, analyse de données et statistique », *Eds Technip*, , 1990.
- [SAR 97] SARACEVIC T., « The stratified model of information retrieval interaction : extension and applications », *Proceedings of the 60th annual meeting of the American Society for Information Science. Medford, NJ*, 1997, p. 313-327.
- [SON 99] SONNENWALD D. H., « Evolving perspectives of human behaviors : contexts, situation, social networks and information horizons », *Exploring the contexts of information behaviour : Proceedings of the 2nd international conference on reserach in information needs, seeking and use in different contexts*, 1999, p. 176-190.
- [SPI 98] SPINK A., BATEMAN J., JANSEN B., « User's searching behaviour on the EXCITE Web search engine », *Proceedings of the 1998 World Conference of the WWW, Internet and Intranet*, 1998, p. 7-12.
- [SU 03] SU J., LEE M., « An exploration in personalized and context-sensitive search », *Proceedings of the 7th annual UK special interest group for computatonal linguists research colloquium*, 2003.
- [TAM 05] TAMINE L., BOUGHANEM M., ZEMIRLI N., « Learning the user's interests using the search history », *Proceedings of NIPS 2005 Workshop on Machine Learning for Implicit Feedback and User Modeling, to appear*, 2005.
- [WEN 04] WEN J., LAO N., MA W. Y., « Probabilistic model for contextual retrieval », *Proceedings of the 27th annual internationsl ACM SIGIR Conference on Research and development in Information retrieval*, August 2004, p. 57-63.