



Accès personnalisé à l'information : Vers un modèle basé sur les diagrammes d'influence

Lynda Tamine-Lechani, Mohand Boughanem, Claude Chrisment

► To cite this version:

Lynda Tamine-Lechani, Mohand Boughanem, Claude Chrisment. Accès personnalisé à l'information : Vers un modèle basé sur les diagrammes d'influence. Revue I3 - Information Interaction Intelligence, Cépaduès, 2006, pp.69-90. <hal-00359533>

HAL Id: hal-00359533

<https://hal.archives-ouvertes.fr/hal-00359533>

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.

Accès personnalisé à l'information : Vers un modèle basé sur les diagrammes d'influence

L. Tamine-Lechani, M. Boughanem, C. Chrisment

Institut de Recherche en Informatique de Toulouse
Equipe Systèmes d'Information Généralisés
118 Route de Narbonne, 31062 Toulouse CEDEX 9
tamine, bougha, chrisme@irit.fr

Résumé

L'accroissement sans précédent des masses d'informations via le Web, lance à la communauté en recherche d'information (RI) de nouveaux défis. En effet, compte tenu des exigences liées à l'efficacité, la personnalisation du processus d'accès à l'information devient incontournable. Son objectif fondamental est de présenter à l'utilisateur des informations pertinentes et intelligibles compte tenu de ses centres d'intérêts, ses préférences, son intention ou plus globalement son profil. Nous présentons dans cet article, les bases de construction d'un modèle d'accès à l'information basé sur les diagrammes d'influence, et mettant en jeu trois composantes : le document, la requête et l'utilisateur. Ce modèle vise à mettre l'utilisateur, décrit par ses centres d'intérêt, au centre du processus de sélection de l'information

Mots-clés : recherche d'information personnalisée, profil utilisateur, contexte, diagramme d'influence

Abstract

As the amount of information on the Web rapidly increases, it creates new challenges for information retrieval. Indeed, the guarantee of effective and efficient search suggests to personalize the process of information access. The main goal of the model is to present to the user suitable information according to his query, interests and goals or more generally his profile. We present in this paper an influence diagram based retrieval model which requires three components : the document, the query and the user. This model focuses on the user's interests in order to carry out the process of information access.

Key-words: personalized information retrieval, user profile, context, influence diagram

1 INTRODUCTION

Les récents progrès des technologies de l'information de manière générale, des réseaux de communication de manière particulière, ont redonné à l'information de nouveaux contours et d'avantage de valeur selon divers aspects : scientifique, technique, économique, d'usage etc. De surcroît, les progrès techniques de numérisation et de compression de l'information, ont encouragé sa production, circulation et exploitation. 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, en raison de la surabondance de l'information d'une part et de sa large accessibilité à travers le Web, d'autre part, leur mise en oeuvre est confrontée à de nouveaux problèmes. En effet la situation est actuellement paradoxale : la masse d'informations est telle que l'accès à une information pertinente, adaptée aux besoins spécifiques d'un utilisateur donné, devient à la fois difficile et nécessaire. En clair, le problème n'est pas tant la disponibilité de l'information mais sa pertinence relativement à un contexte d'utilisation particulier. C'est pourquoi, les travaux [1, 6, 10] s'orientent actuellement vers la révision de la chaîne d'accès à l'information dans la perspective d'intégrer l'utilisateur comme composante du modèle global de recherche et ce, dans le but de lui délivrer une information pertinente, adaptée à ses besoins précis, son contexte et ses préférences. Ces travaux s'inscrivent dans le cadre général de la personnalisation de l'information. Dès lors l'accès à l'information tend vers une nouvelle définition [1] : "*Combine search technologies and knowledge about query and user context into a single framework in order to provide the most appropriate answer for a user's information needs*".

Dans ce cadre, notre objectif consiste à proposer les bases de définition d'un modèle formel capable d'intégrer l'utilisateur dans le processus d'accès à l'information. Plus précisément, notre approche globale se focalise particulièrement sur deux aspects. Le premier aspect porte sur la représentation et la maintenance d'une dimension qui caractérise les besoins récurrents de l'utilisateur en informations. Cette dimension qui est déterminée et évolue à partir de ses interactions avec le SRI, permet de définir ses centres d'intérêt. Le second aspect concerne l'intégration de ces contextes dans un modèle décisionnel d'accès à l'information. A cet effet, on s'oriente vers l'utilisation des diagrammes d'influence [29, 17] comme support théorique nous permettant de formaliser l'utilité des décisions associées à la pertinence des documents compte tenu de la requête et du profil de l'utilisateur. C'est précisément ce second aspect qui est développé dans le présent article, organisé comme suit : la section 2 présente la problématique à l'origine de la personnalisation de l'accès à l'information puis rapporte une synthèse des travaux du domaine. La section 3 décrit globalement les bases du modèle d'accès personnalisé que nous proposons ainsi que les composants fonctionnels du SRI qui le

supporte. La section 4 aborde de manière sommaire les dimensions descriptives du profil utilisateur. La section 5 détaille le modèle formel d'évaluation de la pertinence des documents. Ce modèle étant basé sur les diagrammes d'influence, nous en décrivons alors la topologie ainsi que les composantes qualitative et quantitative. On y illustre également le principe d'évaluation de requête selon le modèle proposé. La section 6 résume notre contribution et présente les perspectives envisagées.

2 PROBLÉMATIQUE ET APERÇU DE L'ÉTAT DE L'ART

2.1 Problématique

L'objectif fondamental d'un SRI est de retourner, à partir d'une collection de documents, ceux qui sont pertinents à un besoin en information exprimé par l'utilisateur à travers une requête composée généralement de mots-clés. Le processus de recherche d'information est alors fondé sur un modèle théorique (booléen, vectoriel, probabiliste etc.) qui permet la mise en correspondance entre représentants de documents et la requête de l'utilisateur en vue d'estimer leur pertinence. Les documents sont présentés à l'utilisateur par ordre décroissant de la mesure de pertinence estimée.

La mise en oeuvre d'un modèle classique de recherche d'information suppose cependant que l'utilisateur est complètement représenté par sa requête et que les résultats retournés pour une même requête sont identiques même si elle est exprimée par des utilisateurs différents. Les problèmes immédiats posés par de telles hypothèses sont notamment l'ambiguïté du sens des mots, l'impossibilité de sélectionner des sources opportunes, et l'inintelligibilité des résultats. En outre, ces problèmes sont d'autant plus accentués que les requêtes sont courtes (2.29 mots en moyenne par requête) et que les sources d'information sont volumineuses et hétérogènes [6]. Outre les facteurs liés à l'information, d'autres sont étroitement liés à l'utilisateur en tant qu'entité participante dans le processus de recherche d'information. En effet deux raisons fondamentales plaident pour la personnalisation [33] : (a) les utilisateurs ont des objectifs différents, des contextes différents et perceptions différentes de la notion de pertinence, (b) un même utilisateur peut avoir différents besoins à différents instants.

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 [26, 27]. Cependant, vu le contexte actuel lié au volume d'informations, ces techniques sont peu viables [15, 32]. Sous un autre angle, un aspect de la personnalisation a été également initié en filtrage d'information [3]. Les systèmes de filtrage d'information sont fondés sur le principe de dissémination sélective d'information utilisant une structure descriptive de l'utilisateur appelée communément profil. Ce dernier

décrit généralement ses centres d'intérêt ainsi que ses préférences portant généralement sur le contenu des documents.

Les travaux s'orientent actuellement vers une définition plus large de l'utilisateur, qui a donné naissance à un courant de recherche : *la recherche d'information contextuelle* qui définit globalement le contexte et/ou profil de l'utilisateur comme l'ensemble des dimensions cognitives et sociales (environnement social, professionnel et culturel) qui le caractérisent ainsi que ses buts et intentions au cours d'une session de recherche. Les travaux en RI contextuelle abordent alors peu ou prou l'une ou l'autre de ces questions :

- Comment représenter l'utilisateur ?
- Comment intégrer l'utilisateur dans le cycle de vie de la requête : reformulation, optimisation et appariement ?

Le paragraphe suivant présente une synthèse des travaux donnant des éléments de réponse à ces questions fondamentales.

2.2 Aperçu de l'état de l'art

On rapporte dans ce qui suit une brève synthèse des travaux portant sur la personnalisation en RI. On focalisera notamment sur les principes de définition du profil et modèle d'accès à l'information.

2.2.1 Sur la modélisation du profil

Le profil de l'utilisateur est une structure d'informations hétérogènes qui couvre des aspects larges tels que l'environnement cognitif, social et professionnel de l'utilisateur ; ces informations sont généralement exploitées dans le but de préciser ses intentions au cours d'une session de recherche [10]. La plupart des travaux actuels en RI contextuelle, se focalisent à juste titre, sur l'exploitation d'une partie de l'environnement cognitif représentée par les centres d'intérêt de l'utilisateur. Les centres d'intérêt traduisent des domaines de connaissances ciblés régulièrement par l'utilisateur au cours de ses sessions de recherche. Dans cette perspective, la modélisation du profil de l'utilisateur a pour objectif fondamental de représenter 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. Il existe principalement trois types de représentation du profil :

- *ensembliste* : le profil y est généralement formalisé comme des vecteurs de termes pondérés [6, 12] ou classes de vecteurs [21],

- *sémantique* : la représentation du profil met en évidence, dans ce cas, les relations sémantiques entre concepts qui le décrivent. La représentation est essentiellement basée sur l'utilisation d'ontologies [14, 7, 23, 20] ou réseaux sémantiques probabilistes [19, 36],
- *multidimensionnelle* : le profil est structuré selon un ensemble de dimensions, représentées selon divers formalismes [2, 4].

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 des préférences de l'utilisateur via son comportement lors de l'utilisation d'un système d'accès à l'information [18] ou de diverses autres applications [6, 21, 12]. Les informations exploitées sont généralement : le contenu des documents consultés et/ou imprimés et/ou sauvegardés, les liens explorés, les dernières pages visitées etc.

L'évolution des profils désigne leur adaptation à la variation des centres d'intérêts des utilisateurs qu'ils décrivent, et par conséquent, de leurs besoins en information au cours du temps. L'évolution y est davantage abordée comme un problème de représentation de la diversité des domaines d'intérêts de l'utilisateur en utilisant des techniques de classification [24, 22] ou heuristiques liées la notion de cycle de vie artificielle d'un centre d'intérêt [9].

2.2.2 Sur l'accès personnalisé à l'information

L'accès personnalisé à l'information traduit l'utilisation du profil lors des opérations éventuelles de sélection de sources d'information, reformulation de requête et calcul d'un score de pertinence.

La sélection personnalisée de sources d'information est une pratique courante dans les méta-moteurs de recherche [8]. Leur principe est d'identifier, à travers le profil utilisateur, le type de requête (besoin général, actualité, scientifique etc) puis l'adresser à un moteur de recherche approprié afin d'augmenter la précision de résultats.

La reformulation de requête qui est une des premières techniques qui s'apparentent à la personnalisation, consiste dans le cadre de la RI contextuelle, à augmenter la requête avec des informations issues du profil de l'utilisateur, avant de lancer le processus d'appariement [28].

La personnalisation peut également porter sur la définition de la fonction de calcul de la pertinence. Dans ce sens, on trouve dans [13] une proposition pour l'adaptation des paramètres de la fonction de pertinence au contexte de l'utilisateur, en utilisant les techniques de programmation génétique. Jeh et Widom [16] ont proposé une variante personnalisée de l'algorithme PageRank en l'occurrence PPV (Personalized PageRank Vector). L'algorithme adapte le principe de calcul de l'*Authority* d'une page, donné par l'algorithme *PageRank*, en utilisant une distribution de probabilité tenant compte de la présence de liens entrants et liens sortant vers et/ou les pages favorites

de l'utilisateur. Plus précisément, pour un utilisateur u donné, l'équation posée par l'algorithme PPV est : $v = (1 - c)Av + cu$ où c est une constante avec $c \in (0, 1)$, A est la matrice de contingence du graphe et u est un vecteur de préférences de l'utilisateur où chaque composante traduit le poids de préférence de l'utilisateur pour chacune de ses pages favorites. Dans [7, 31, 35], une autre approche de personnalisation de la fonction de mesure de la pertinence est proposée. Celle ci consiste à réordonner les résultats fournis par le processus de sélection en combinant l'ordre produit par le processus de sélection et celui donné par le contexte de l'utilisateur, représenté par une liste de mots clés issus des documents sélectionnés, via un calcul de similarité. Dans [30], les auteurs proposent un modèle probabiliste d'accès à l'information utilisant des sources de croyance liées aux contextes. Plus précisément, ils exploitent, en premier lieu, l'historique des interactions de l'utilisateur avec le SRI pour étendre le modèle de langage de la requête ; en second lieu ils appliquent la méthode classique de Kulback-Leiber pour calculer le taux de divergence du modèle de langage des requêtes et documents, traduisant ainsi le score de pertinence.

Notre contribution s'inscrit précisément dans le cadre de l'accès personnalisé à l'information. Notre objectif est de proposer un modèle inférentiel pour la mise en oeuvre du raisonnement lié à la prise de décision quant à la pertinence des documents compte tenu du profil de l'utilisateur d'une part et de la requête d'autre part. Cette fonction est fondée sur l'estimation d'une valeur de l'utilité de la décision prise quant à la sélection de documents pertinents. Comparativement aux travaux du domaine, notre approche s'en distingue fondamentalement par :

- l'intégration du profil de l'utilisateur comme composante à part entière du modèle formel d'accès personnalisé et non comme source de définition d'heuristiques ou techniques permettant la réécriture de la requête et/ou la fonction d'appariement,
- la caractérisation de l'accès personnalisé à l'information comme un problème décisionnel, ce qui justifie notre recours à la modélisation basée sur les diagrammes d'influence.

3 LE MODÈLE GÉNÉRAL D'ACCÈS PERSONNALISÉ À L'INFORMATION

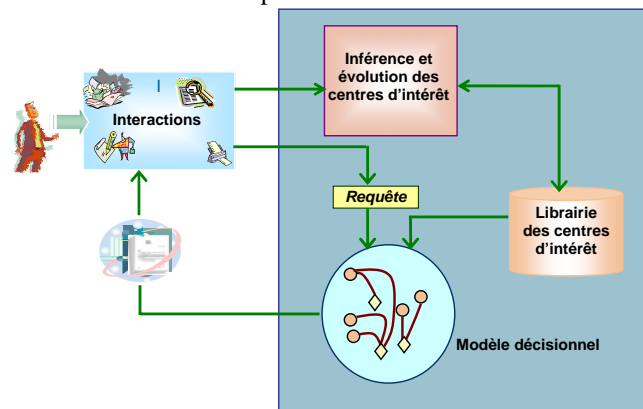
Le modèle d'accès à l'information que nous proposons peut être défini par un quadruplet $M = \langle I, U, E^t(U), P(I, U) \rangle$ où I est le modèle de représentation de l'information (documents et requêtes), U est le modèle de représentation du profil de l'utilisateur, $E^t(U)$ est la fonction d'évolution du profil de l'utilisateur en fonction du temps et $P(I, U)$ est la fonction d'évaluation de la pertinence d'une information relativement à un utilisateur décrit

par son profil. La figure 1 illustre les composants fonctionnels du SRI supporté par le modèle. L'accès personnalisé s'y traduit précisément par :

1. l'inférence des centres d'intérêt de l'utilisateur à partir de l'historique de ses interactions avec le SRI. Ces interactions correspondent à l'expression de(s) requête(s) et de(s) jugement(s) explicite(s) ou implicite(s) de l'utilisateur quant à la pertinence des documents présentés lors de(s) sessions de recherche ;
2. l'intégration de ces centres d'intérêt dans la phase d'accès à l'information. On s'oriente à cet effet vers la définition d'un modèle décisionnel pour l'évaluation de la pertinence d'une requête relativement à un utilisateur. Ce modèle est fondé sur une fonction cumulative permettant d'estimer l'utilité de présenter à l'utilisateur des documents et ce, compte tenu de sa requête et de ses centres d'intérêt spécifiques.

Dans ce papier, nous focalisons sur le modèle décisionnel d'évaluation de la pertinence de l'information (composante $P(I, U)$ du modèle). Nous présentons dans cette section, le principe général de représentation (composante U du modèle) et évolution du profil de l'utilisateur (composante $E^t(U)$ du modèle) puis détaillons la description du modèle décisionnel pour l'évaluation de la pertinence des documents.

FIG. 1 – Composants fonctionnels du SRI



4 LE PROFIL DE L'UTILISATEUR

4.1 Représentation

Le profil de l'utilisateur, noté U , est multidimensionnel et utilisé lors de la phase d'appariement requête-document. Plus précisément, le profil de l'utilisateur est représenté par deux dimensions : historique des interactions et

centres d'intérêt . De manière sommaire, on définit le profil d'un utilisateur à l'instant s par $U = (H^s, C_I^s)$ où H^s représente l'historique des interactions de l'utilisateur jusqu'à l'instant s et C_I^s représente la librairie de ses centres d'intérêt inférés jusqu'à l'instant s . L'historique des interactions H^s est une matrice termes-documents qui traduit la représentation vectorielle et agrégée des documents jugés pertinents par l'utilisateur en cours de ses sessions de recherche. Les centres d'intérêt sont des vecteurs de termes dominants ayant émergé de l'historique des interactions.

4.2 Evolution

Les deux dimensions H^s et C_I^s évoluent corrélativement au cours du temps. Plus précisément, la fonction d'évolution du profil de l'utilisateur, notée $E^t(U)$, se décline en un cycle comportant deux principales étapes : évolution de l'historique des interactions puis évolution des centres d'intérêt. Dans ce qui suit, nous décrivons globalement ces étapes. Une description détaillée en est présentée dans [34].

4.2.1 Evolution de l'historique des interactions

Cette 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, notée S^s , est particulièrement décrite par l'association d'une requête et d'un ensemble de documents jugés explicitement ou implicitement pertinents par l'utilisateur. Plus précisément l'opérateur d'agrégation des sessions de recherche, noté \oplus , est défini comme suit :

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

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

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

avec S^s une matrice Document-Terme $D^s * T^s$ où T^s est l'ensemble des termes qui indexent les documents supposés ou jugés explicitement pertinents 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 $t \in T^s$. Le facteur w_{td} traduit l'importance du terme t dans le document d .

4.2.2 Evolution des centres d'intérêt

L'évolution des centres d'intérêt est effectuée selon des périodes d'apprentissage définies par un nombre déterminé de sessions de recherche et basée sur la récurrence éventuelle des contextes d'usage associés aux sessions de recherche passées. Un contexte d'usage traduit un besoin en information ponctuel exprimé sur une courte période d'interaction avec le SRI. A chaque période d'apprentissage, est extrait un contexte d'usage représenté par un vecteur de termes pondérés, ordonnés par leur degré de représentativité du contexte.

Plus précisément, pour chaque terme $t \in T(R_u^s)$, on calcule le poids associé comme suit :

$$c^s(t) = \sum_{d \in R_u^s} H^s(d, t) \quad (2)$$

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

Le degré de récurrence des contextes d'usage est, à notre sens, un indicateur d'évolution des centres d'intérêt de l'utilisateur. Sur la base de cette hypothèse, on évalue le degré de changement des contextes d'usage de l'utilisateur en comparant le contexte courant cc avec le contexte extrait à la période précédente pc . On utilise, à cet effet, le coefficient de corrélation des rangs de Kendall comme suit :

$$\Delta C_I = \frac{\sum_t \sum_{t'} S_{tt'}(pc) * S_{tt'}(cc)}{\sqrt{(\sum_t \sum_{t'} S_{tt'}(pc))^2 (\sum_t * \sum_{t'} S_{tt'}(cc))^2}} \quad (3)$$

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 ΔC_I est inscrite dans l'intervalle $[-1..1]$, où une valeur proche de -1 signifie que les contextes d'usage sont non similaires alors qu'une valeur proche de 1 signifie que les contextes d'usage sont proches sémantiquement. Le coefficient ΔC_I est approché par une loi Laplace Gauss : $\Delta C_I \approx LG\left(0; \sqrt{\frac{2(2n+5)}{9n(n-1)}}\right)$ 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'évolution des centres d'intérêt déjà appris et/ou construction de nouveaux centres d'intérêt, sont alors déterminées par le résultat de comparaison de ΔC_I relativement au seuil de corrélation σ .

5 EVALUATION DE LA PERTINENCE DE L'INFORMATION

Dans le but de formaliser la fonction d'évaluation de la pertinence, notée $P(I, U)$, des documents relativement à la requête de l'utilisateur d'une part et de son profil d'autre part, on s'oriente vers l'utilisation d'une extension des réseaux bayésiens [25, 17] en l'occurrence les diagrammes d'influence [29]. L'idée de base est de substituer à la fonction de pertinence classique qui mesure le degré d'appariement requête-document $RSV(Q, D) = p(Q/D)$, une fonction indexée par l'utilisateur $RSV_u(Q, D) = p(D/Q, U)$ où $p(A/B)$ est la probabilité conditionnelle de l'événement A sachant l'événement B et U représente l'utilisateur décrit par ses centres d'intérêt.

Suite à la présentation des principes théoriques des réseaux bayésiens et des diagrammes d'influence, nous en illustrons l'application à notre problème précis : *mettre en oeuvre un appariement requête-document-utilisateur*.

5.1 Le cadre formel : les réseaux bayésiens

Les réseaux bayésiens (RB) sont des modèles graphiques capables de représenter et manipuler efficacement des distributions de probabilités multidimensionnelles. Un RB utilise deux composantes pour codifier une connaissance qualitative et quantitative. La composante qualitative est représentée par un graphe acyclique $G = (V, E)$ où V comprend des noeuds qui représentent des variables aléatoires X_1, X_2, \dots, X_n et E un ensemble d'arcs entre noeuds qui traduisent des relations de causalité. A chaque variable X_i est associée un ensemble de valeurs mutuellement exclusives définies dans $dom(X_i)$. La composante quantitative comprend des distributions de probabilités conditionnelles où pour chaque variable $X_i \in V$, est attachée une classe de probabilités $p(X_i/pa(X_i))$ avec $pa(X_i)$ une combinaison de valeurs associées à $PA(X_i)$, parents de X_i dans G . Un RB permet une représentation compacte de la loi jointe :

$$p(X) = \prod_{i=1}^n p(X_i/pa(X_i)) \quad (4)$$

Ainsi, un processus d'inférence permet d'attribuer des degrés de croyance à de nouveaux événements.

Les diagrammes d'influence (DI) constituent une extension des RB à un problème de décision. La composante qualitative comporte différents types de noeuds : des noeuds chance associés aux variables aléatoires observables, des noeuds décision associés aux décisions contrôlables et des noeuds utilité associés aux décisions ; on distingue en outre deux types d'arcs : les arcs d'information qui relient entre les noeuds chance et les arcs d'influence qui relient les noeuds chance et les noeuds utilité. Les noeuds chance et les noeuds décision ont un ensemble fini d'états mutuellement exclusifs. La

composante quantitative est décrite par des probabilités conditionnelles attachées à chaque noeud chance X_i ainsi que des valeurs d'utilité attachées aux noeuds utilité notées $u(X_i/pa(X_i))$.

Notre choix pour l'utilisation des DI est motivé par le besoin d'utiliser un cadre théorique pour la formalisation du problème décisionnel lié à la présentation d'un document à l'utilisateur compte tenu de son utilité liée l'influence mutuelle de son contenu, de la requête émise et des centres d'intérêt spécifiques de l'utilisateur.

5.2 Topologie du diagramme

La topologie du diagramme d'accès à l'information est présentée sur la figure 2. L'ensemble des variables V est composé de trois sous-ensembles : le premier sous-ensemble $V^{(info)}$ qui comprend les noeuds chance, le deuxième sous-ensemble $V^{(decision)}$ qui comprend les noeuds décision et le troisième sous-ensemble $V^{(utilité)}$ qui comprend les noeuds utilité. En outre, on distingue deux types d'arc : les arcs d'information et les arcs d'influence.

5.2.1 Description des noeuds

– *Noeuds chance.* Il existe quatre types de noeud chance $V^{(info)} = Q \cup D \cup T \cup C$. L'ensemble Q représente une unique requête.

L'ensemble $D = \{D_1, D_2, \dots, D_n\}$ représente l'ensemble des documents de la collection, $T = \{T_1, T_2, \dots, T_m\}$ représente l'ensemble des termes qui indexent ces documents, $C = \{C_1, C_2, \dots, C_u\}$ représente l'ensemble des centres d'intérêt associés à un utilisateur donné.

Il existe un noeud unique Q qui représente une variable aléatoire binaire définie dans l'ensemble $dom(Q) = \{q, \bar{q}\}$ où q désigne que *la requête Q est exprimée dans un contexte donné (lié au centre d'intérêt de l'utilisateur qui l'a émise) et introduit de l'information dans le réseau* et \bar{q} désigne qu' *aucune requête n'est exprimée* ; dans notre cas, on ne s'intéressera qu'à l'instanciation positive q .

Chaque noeud document D , représente une variable aléatoire binaire prenant des valeurs dans l'ensemble $dom(D) = \{d, \bar{d}\}$, où d représente *'le document D est pertinent pour une requête donnée'*, et \bar{d} représente *'le document D n'est pas pertinent pour la requête donnée'*.

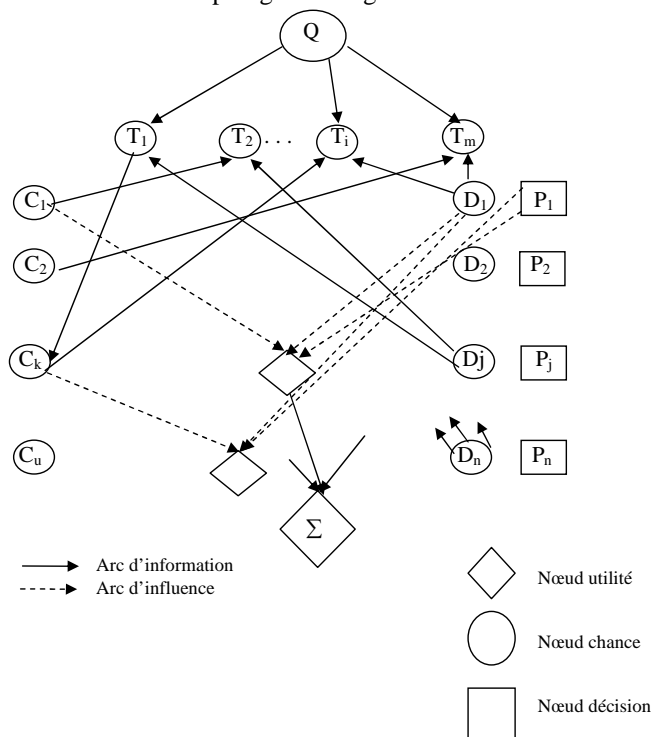
Chaque noeud terme T représente une variable aléatoire binaire prenant des valeurs dans l'ensemble $dom(T) = \{t, \bar{t}\}$, où t représente *'le terme T est pertinent pour la requête'* et \bar{t} représente *'le terme T n'est pas pertinent pour une requête'*. La pertinence d'un terme signifie sa présence éventuelle dans un document pertinent.

Chaque noeud centre d'intérêt C représente une variable aléatoire binaire prenant des valeurs dans l'ensemble $dom(C) = \{c, \bar{c}\}$, où c représente *'le contexte C est pertinent pour une requête donnée'* et \bar{c} représente *'le contexte C n'est pas pertinent pour une requête donnée'*. La

pertinence d'un contexte signifie son adéquation relativement au centre d'intérêt de l'utilisateur véhiculé par la requête en cours.

- *Noeuds décision.* Le modèle évalue l'utilité de la décision de présenter ou non un document à l'utilisateur, compte tenu de sa requête et de son profil décrit par ses centres d'intérêt. Donc à chaque noeud document D_i est associé un noeud décision P_i qui prend des valeurs dans $dom(P) = \{p, \bar{p}\}$ qui correspondent respectivement aux décisions de présenter le document D à l'utilisateur ou ne pas le présenter.
- *Noeuds utilité.* La décision de présenter un document doit dépendre non seulement du contenu du document mais également du profil de l'utilisateur qui l'a émise. Pour cela, nous distinguons deux types de noeud d'utilité. Le premier permet de quantifier une utilité partielle qui correspond à l'adéquation relative du document vis-à-vis de la requête, et au regard de chaque centre d'intérêt ; le second type de noeud (unique et noté Σ dans la figure 2) permet de quantifier une utilité globale qui correspond à l'agrégation des utilités partielles.

FIG. 2 – Topologie du diagramme d'influence



5.2.2 Description des arcs

On distingue deux types d'arcs :

- *Les arcs d'information.* Ces arcs relient chaque noeud chance n de type Q, D ou C aux noeuds termes qui les indexent notés $\tau(n)$. Plus précisément, des arcs relient chaque noeud terme $T_i \in \tau(D_j)$ (termes qui indexent le document D_j) et chaque noeud document $D_j \in D$. De même, des arcs relient chaque noeud terme $T_i \in \tau(C_k)$ (termes représentatifs du centre d'intérêt C_k) et chaque noeud $C_k \in C$. Il existe également des arcs d'information qui relient chaque terme $T_i \in \tau(Q)$ avec le noeud requête Q .
- *Les arcs d'influence.* Ces arcs traduisent le degré d'influence des variables associées à la décision prise. Dans le cas de notre modèle, des arcs d'influence relient les noeuds décision, centres d'intérêt et documents en utilisant un opérateur de sommation.

5.3 Distribution de probabilités

Pour chaque noeud n du diagramme, on spécifie une distribution de probabilités conditionnelles $p(n/pa(n))$ selon le modèle des RB. Dans la suite, on note $R(PA(n))$ l'ensemble des termes instanciés positivement parmi les termes parents du noeud n soit $R(PA(n)) = \{T_i \in \tau(n), t_i \in pa(n)\}$.

- *Noeud requête.* La probabilité $p(Q/pa(Q))$ traduit le degré de correspondance entre la configuration de termes donnée par $pa(Q)$ avec la configuration initiale de la requête. Le calcul de cette probabilité $p(Q/pa(Q))$ est effectuée sur la base d'une agrégation *Noisy-OR* proposée dans [25, 5]

$$p(Q/pa(Q)) = \left\{ \begin{array}{l} 0 \text{ si } (Pa(Q) \cap R(Pa(Q)) = \emptyset) \\ \frac{1 - \prod_{T_i \in R(Pa(Q))} Idf(T_i)}{1 - \prod_{T_i \in Pa(Q)} Idf(T_i)} \text{ sinon} \end{array} \right\} \quad (5)$$

$Idf(T_i)$ est la valeur de la fréquence documentaire inverse du terme T_i .

- *Noeud terme.* L'estimation de la probabilité $p(t_i/c_k, d_j)$ est basée sur la représentativité du terme t_i dans le centre d'intérêt c_k et le document d_j selon une variante du schéma $Tf - Idf$. En supposant que les documents et centres d'intérêt sont indépendants, on calcule :

$$\begin{aligned} p(t_i/c_k, d_j) &= p(t_i/c_k) * p(t_i/d_j), \quad p(t_i/c_k) = c_{ik} \\ p(t_i/d_j) &= ndf_i * nt f_{ij} \\ p(t_i/\overline{c_k}) &= 0, \quad p(t_i/\overline{d_j}) = 0 \end{aligned} \quad (6)$$

avec $nt f_{ij} = \frac{t f_{ij}}{\max_{t_r \in d_j} t f_{rj}}$, $ndf_i = \frac{\log(\frac{n}{n_i})}{\log n}$, n est le nombre total de documents dans la collection, n_i est le nombre de documents contenant le terme t_i , c_{ik} est le poids du terme t_i dans la représentation du centre d'intérêt C_k calculé précédemment (Cf paragraphe 4.2).

5.4 Valeurs d'utilité

Le degré d'utilité de la décision de présenter ou non un document dépend à la fois du contenu du document et des centres d'intérêt relativement à la requête en cours d'évaluation. La valeur d'utilité est cumulative pour un même centre d'intérêt. On assume que la décision la plus utile est de présenter un document pertinent dans un contexte de requête pertinent et que la décision la moins utile est de présenter un document non pertinent dans un contexte de requête non pertinent ; on pose alors :

$$u(p/c, d) = 1, u(p/\bar{c}, \bar{d}) = 0 \quad (7)$$

Par ailleurs, il est plus utile, indépendamment du contexte, de présenter un document pertinent que de ne pas présenter un document non pertinent. On peut alors définir l'ordre suivant des utilités :

$$u(p/c, d) \geq u(p/\bar{c}, \bar{d}) \text{ et } u(\bar{p}/c, \bar{d}) \geq u(\bar{p}/c, d) \quad (8)$$

$$u(p/\bar{c}, d) \geq u(p/\bar{c}, \bar{d}) \text{ et } u(\bar{p}/\bar{c}, \bar{d}) \geq u(\bar{p}/\bar{c}, d) \quad (9)$$

De plus, il est trivialement plus utile de ne pas présenter un document non pertinent que de le présenter, on a alors :

$$u(\bar{p}/c, \bar{d}) \geq u(p/c, \bar{d}) \text{ et } u(\bar{p}/\bar{c}, \bar{d}) \geq u(p/\bar{c}, \bar{d}) \quad (10)$$

Dans le cas d'un contexte pertinent, il est plus utile de présenter un document non pertinent que de ne pas présenter un document pertinent ; A l'inverse, lorsque le contexte est non pertinent, il est plus utile de ne pas présenter un document non pertinent que de présenter un document pertinent ; On a alors :

$$u(p/c, \bar{d}) \geq u(\bar{p}/c, d) \text{ et } u(\bar{p}/\bar{c}, \bar{d}) \geq u(p/\bar{c}, d) \quad (11)$$

on obtient ainsi, l'ordre suivant des utilités :

$$\begin{aligned} u(p/c, d) &\geq u(\bar{p}/c, \bar{d}) \geq u(p/c, \bar{d}) \geq u(\bar{p}/c, d) \\ &\geq u(p/\bar{c}, d) \geq u(\bar{p}/\bar{c}, \bar{d}) \geq u(\bar{p}/\bar{c}, d) \geq u(p/\bar{c}, \bar{d}) \end{aligned} \quad (12)$$

En privilégiant le rappel, on obtient l'ordre global suivant :

$$\begin{aligned} 1 = u(p/c, d) &\geq u(p/\bar{c}, d) \geq u(\bar{p}/c, \bar{d}) \geq u(\bar{p}/\bar{c}, \bar{d}) \\ &\geq u(\bar{p}/\bar{c}, d) \geq u(p/c, \bar{d}) \geq u(\bar{p}/c, d) \geq u(p/\bar{c}, \bar{d}) = 0 \end{aligned} \quad (13)$$

Une manière simple de faire respecter cet ordre est d'assigner des valeurs directes aux différentes utilités. Cependant, il existe d'autres opérateurs d'agrégation permettant une meilleure prise en compte de l'ordre des valeurs, on cite notamment les opérateurs OWA [37], Leximin et Discrimin [11]. Les valeurs d'utilité $u(p/c, d)$ ainsi définies sont dépendantes des valeurs d'instances d et c mais sont indépendantes des structures respectives des documents et centres d'intérêt associés D et C . Dans le but de décliner le degré de correspondance entre document et centre d'intérêt dans le calcul de la valeur d'utilité, on propose d'introduire un facteur de coût :

$$u'(p/c, d) = u(p/c, d) - \text{coût}(c, d) \quad (14)$$

$$\text{avec } \text{coût}(c, d) = \frac{\sum_{T_i \in \tau(D) - \tau(C)} \text{nid}f_i}{\sum_{T_i \in \tau(D)} \text{nid}f_i}$$

5.5 L'évaluation de requête vue comme un processus décisionnel dépendant du profil de l'utilisateur

L'évaluation de requête consiste en l'application d'un algorithme de propagation de nouvelles croyances dans le diagramme, comme dans le cas des RB, et ce dans le but d'optimiser une mesure d'utilité [17]. Dans le cadre de notre approche, cette mesure est basée sur la valeur cumulative d'utilité correspondant aux décisions prises quant à la pertinence des documents compte tenu de la requête en cours d'évaluation et profil de l'utilisateur qui l'a émise décrit par ses centres d'intérêt.

Plus précisément, soit une requête Q représentée par un ensemble de termes représentatifs ($\tau(Q) = \{T_1, T_2, \dots, T_r\}$), le processus d'évaluation est enclenché comme dans un problème décisionnel en maximisant la mesure d'utilité $\frac{EU(p/Q)}{EU(\bar{p}/Q)}$ équivalente à $RSV_u(Q, D)$ calculée comme suit :

$$EU(p_j/Q) = \sum_{C_k \in I^s} \sum_{c_k \in \{c, \bar{c}\}, d_j \in \{d, \bar{d}\}} u'(p/c_k, d_j) p(c_k, d_j/Q) \quad (15)$$

$$EU(\bar{p}_j/Q) = \sum_{C_k \in I^s} \sum_{c_k \in \{c, \bar{c}\}, d_j \in \{d, \bar{d}\}} u'(\bar{p}/c_k, d_j) p(c_k, d_j/Q) \quad (16)$$

En procédant à l'approximation basée sur la supposition d'indépendance des centres d'intérêt et documents, on obtient :

$$EU(p/Q) = \sum_{C_k \in I^s} \sum_{c_k \in \{c, \bar{c}\}, d_j \in \{d, \bar{d}\}} u'(p/c_k, d_j) p(c_k/Q) p(d_j/Q) \quad (17)$$

$$EU(\bar{p}/Q) = \sum_{C_k \in I^s} \sum_{c_k \in \{c, \bar{c}\}, d_j \in \{d, \bar{d}\}} u'(\bar{p}/c_k, d_j) p(c_k/Q) p(d_j/Q) \quad (18)$$

On a $p(d_j/Q) = p(Q/d_j) \frac{p(d_j)}{p(Q)}$ et $p(\bar{d}_j/Q) = 1 - p(d_j/Q)$. Le facteur $\frac{p(d_j)}{p(Q)}$ étant constant (on pose $p(d_j) = 1$ qui est une probabilité a priori sur l'observation du document d_j), on obtient alors $p(d_j/Q) \approx p(Q/d_j)$. En adoptant le même principe d'approximation, on obtient $p(c_k/Q) \approx p(Q/c_k)$. En utilisant la loi jointe d'un RB on calcule :

$$p(Q/d_j) = \sum_{c_k, \bar{c}_k} \sum_{\theta^k \in \theta} p(Q/\theta^k) \prod_{T_i \in (\tau(Q) \cap (\tau(D_j)))} p(\theta_i^k/d_j, c_k) \quad (19)$$

où θ représente l'ensemble des combinaisons d'instances possibles des termes d'indexation. En considérant l'agrégation de la requête basée sur les formules (5) et (6), on obtient :

$$p(Q/d_j) = \sum_{\theta^k \in \theta^q} \frac{1 - \prod_{T_i \in R(Pa(Q))} Idf(T_i)}{1 - \prod_{T_i \in Pa(Q)} Idf(T_i)} \prod_{T_i \in (\tau(Q) \cap (\tau(D_j)))} p(\theta_i^k/d_j) p(\theta_i^k/c_k) \quad (20)$$

où θ^q représente les combinaisons d'instances possibles des termes de la requête comprenant au moins une instance positive.

De manière analogue, on calcule :

$$p(Q/c_k) = \sum_{\theta^r \in \theta^q} \frac{1 - \prod_{T_i \in R(Pa(Q))} Idf(T_i)}{1 - \prod_{T_i \in Pa(Q)} Idf(T_i)} \prod_{T_i \in (\tau(Q) \cap \tau(C_k))} p(\theta_i^k/d_j) p(\theta_i^k/c_k) \quad (21)$$

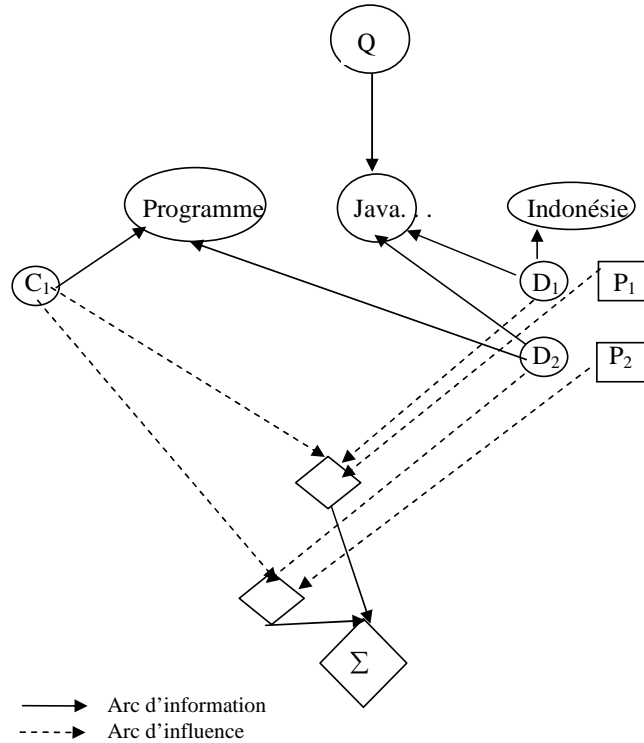
6 ILLUSTRATION

Soit un utilisateur s'intéressant à la *programmation JAVA* décrit préalablement par un centre d'intérêt $C = \{programme\}$ et qui émet la requête $Q = \{java\}$. On supposera, pour simplifier la présentation des calculs, que la collection ne comprend que de deux documents : $d_1 = \{java, Indonésie\}$ et $d_2 = \{java, programme\}$. Le document d_1 porte des informations sur *l'île de java en Indonésie* alors que d_2 traite de *la programmation Java*. La figure 3 illustre la partie du diagramme associée à l'exemple.

On développe à titre d'exemple :

$$\begin{aligned} EU(p_1/Q) &= u'(p_1/c, d) * p(c/Q) * p(d_1/Q) + \\ &+ u'(p_1/\bar{c}, d) * p(\bar{c}/Q) * p(d_1/Q) \\ &+ u'(p_1/c, \bar{d}_1) * p(c/Q) * p(\bar{d}_1/Q) \\ &+ u'(p_1/\bar{c}, \bar{d}_1) * p(\bar{c}/Q) * p(\bar{d}_1/Q) \end{aligned} \quad (22)$$

FIG. 3 – Diagramme d'influence associé à l'exemple



Dans le cas de cet exemple précis, le terme *java* est l'unique terme de la requête, lequel est de surcroît, présent dans tous les documents de la collection ce qui conduit à $Idf(java) = 0$ et par conséquent à $p(d_1/Q) = p(d_2/Q) = 0$; par ailleurs le terme *java* est absent de l'unique centre d'intérêt de l'utilisateur en l'occurrence $C = \{programme\}$, ce qui donne $p(c/Q) = 0$. Cependant le calcul de l'utilité permet de discerner l'intérêt du document D_1 relativement au document D_2 compte tenu de la requête Q et du centre d'intérêt C . En effet on calcule :

$$u'(p_1/c, d_1) = u(p_1/c, d_1) - \text{coût}(c, d_1)$$

$$u'(p_2/c, d_2) = u(p_2/c, d_2) - \text{coût}(c, d_2)$$

En assignant simplement les valeurs d'utilité graduelles suivantes, sur la base de l'ordre global préalablement défini :

$$1 = u(p/c, d) \quad u(p/\bar{c}, d) = 1/2 \quad u(\bar{p}/c, \bar{d}) = 1/3 \quad u(\bar{p}/\bar{c}, \bar{d}) = 1/4$$

$$u(\bar{p}/\bar{c}, d) = 1/5 \quad u(p/c, \bar{d}) \geq u(\bar{p}/c, d) = 1/6 \quad u(p/\bar{c}, \bar{d}) = 0 \quad (23)$$

On obtient :

$$EU(p_1/Q) = -1, EU(p_2/Q) = -1/2, EU(\overline{p}_1/Q) = 4/3, EU(\overline{p}_2/Q) = 5/6$$

Ce qui nous permet de calculer :

$$RSV_u(Q, D_1) = \frac{EU(p_1/Q)}{EU(\overline{p}_1/Q)} = -\frac{4}{3}, RSV_u(Q, D_2) = \frac{EU(p_2/Q)}{EU(\overline{p}_2/Q)} = -\frac{6}{10}$$

Ce qui nous amène à conclure que compte tenu de la requête (*java*) et du centre d'intérêt de l'utilisateur qui l'a émise défini par (*cgi*) le document D_2 (qui traite de la programmation *java*) est plus pertinent que le document D_1 (qui traite de l'île de *Java*)

7 CONCLUSION ET PERSPECTIVES

Nous avons proposé dans cet article les bases d'un modèle d'accès à l'information centré sur le profil spécifique d'un utilisateur U . Ce dernier est décrit par ses centres d'intérêt dérivés automatiquement à partir de l'historique de ses interactions avec le SRI. Ces centres d'intérêt traduisent les besoins pérennes en information de l'utilisateur et sont décrits par une liste de termes dominants pondérés par leur degré de représentativité ; ils sont intégrés dans un modèle décisionnel pour l'évaluation de la pertinence, basé sur une extension des RB en l'occurrence les DI. Du point de vue qualitatif, un DI traduit la structure des centres d'intérêt de l'utilisateur, des documents et index de la collection. Du point de vue quantitatif, le DI traduit la mutuelle influence existante entre un besoin en information exprimé par une requête et contexte associé, dans une situation liée à la prise de décision quant à la pertinence d'un document. L'évaluation d'une requête y est alors vue comme un processus de propagation d'inférence ayant pour objet de maximiser l'utilité cumulée des décisions parcellaires associées aux noeuds du diagramme.

Outre le cadre formel, l'utilisation d'un DI pour modéliser le problème de l'accès personnalisé à l'information offre des atouts majeurs :

- introduire le profil de l'utilisateur comme composante, à part entière, du modèle,
- considérer l'influence du contexte sur la pertinence relative d'un document pour un utilisateur,
- ordonnancer les documents sur la base d'une fonction de pertinence cumulative tenant compte de l'utilité des décisions associées à leur présentation compte tenu des centres d'intérêt de l'utilisateur qui a émis la requête

Inscrits dans le cadre du projet *APMD*, ces travaux seront suivis par une validation utilisant un prototype de gestion de profils d'une part et un environnement expérimental de RI d'autre part. A cet effet, nous envisageons à court terme de mettre en oeuvre un cadre d'évaluation où seront définis des collections de test et mesures d'efficacité adaptés à l'évaluation de scénarios d'accès personnalisé à l'information.

8 REMERCIEMENTS

Cette recherche a été partiellement soutenue par le Ministère délégué à la Recherche et aux Nouvelles Technologies, dans le programme ACI Masse de Données, projet MD-33 (<http://apmd.prism.uvsq.fr/>).

RÉFÉRENCES

- [1] J. Allan et al. Challenges in information retrieval and language modeling. In *Workshop held at the center for intelligent information retrieval*, Septembre 2002.
- [2] G. Amato et U. Staraccia. User profile modelling and applications to digital libraries. In *Proceedings of the 3rd European Conference on Research and advanced technology for digital libraries*, pages 184–187, 1999.
- [3] N.J Belkin et B. Croft. Information retrieval and filtering : two sides of the same coin ? In *ACM Transactions*, volume 35, pages 29–38, 1992.
- [4] M. Bouzeghoub et D. Kostadinov. Personnalisation de l'information : Aperçu de l'état de l'art et définition d'un modèle flexible de définition de profils. In *Actes de la 2nde Conférence en Recherche d'Information et Applications CORIA*, pages 201–218, Mars 2005.
- [5] A. Brini, M. Boughanem et D. Dubois. A model for information retrieval based on possibilistic networks. In *Proceedings of the symposium on String Processing and Information REtrieval (SPIRE)*, pages 271–282, 2005.
- [6] J. Budzik et K. J Hamond. User interactions with every applications as context for just-in-time information access. In *Proceedings of the 5th international conference on intelligent user interfaces*, pages 44–51, Mars 2000.
- [7] V.K.R Challam. Contextual information retrieval using ontology based user profiles. In *Master of science in computer science*. Jawaharlal Nehru Technological University, 2004.
- [8] M. Chau, D. Zeng et H. Chen. Personalised spiders for web search and analysis. In *Proceedings of ACM/IEEE Joint conference on digital libraries*, pages 79–87, 2001.
- [9] C. Chen, M. Chen et Y. Sun. A web document personalization user model and system. In *International retrieval and user modeling, Workshop shedule UM*, 2001.
- [10] C. Cool et A. Spink. Issues of context in information retrieval : an introduction to the special issue. In *Journal of Information Processing and Management (IPM)*, 38(55) :605–611, 2002.

- [11] D. Dubois, H. Prade et C. Testemale. Weighetd minimum and maximum operations in fuzzy set theory. information sciences. In *Information Sciences*, pages 39 : 205–210, 1986.
- [12] S. Dumais, E. Cadiz Cuttrel, J.J. Jancke, G. Sarin et D.C. Robbins. Stuff i've seen : A system for a personal information retrieval and reuse. In *Proceedings of the 26th ACM SIGIR International Conference on Research and Development*, pages 72–79, 2003.
- [13] W. Fan, M.D Gordon et P. Pathak. Discovery of context specific ranking functions for effective information retrieval using genetic programming. volume 16, pages 523–527, 2004.
- [14] S. Gauch, J. Chaffe et P. Pretschner. Ontology based user profiles for search and browsing. volume Special issue on user modelling for Web and hypermedia information retrieval, 2003.
- [15] B. Jansen, A. Spink et J. Bateman. Searchers : the subjects they search and sufficiency : a study of large sample of excite searches. In *Proceedings of the 1998 Web-net World conference of the WWW, Internet and Intranet*, pages 7–12. Faculty of science, University college Dublin, 1998.
- [16] G. Jeh et J. Widom. Scaling personalized web search. In *Proceedings of the 12th International World Wide Web Conference*, pages 271–279, 2003.
- [17] F.V Jensen. Bayesian networks and decision graphs. Springer, 2001.
- [18] N. J Kelly. Understanding implicit feedback and document preference : a naturalistic study. In *PHD dissertation*. Ritgers University, New Jersey, January 2004.
- [19] C. Lin, G.R Xue, H.J Zeng et Y. YU. Using probabilistic latent semantic analysis for personalised web search. In *Proceedings of the APWeb Conference*, pages 707–711, 2005.
- [20] F. Liu et C. Yu. Personalized web search for improving retrieval effectiveness. In *IEEE Transactions on knowledge Data Engineering*, volume 16, pages 28–40, 2004.
- [21] J.P Mc Gowan. A multiple model approach to personnalised information access. In *Master thesis in computer science*. Faculty of science, University college Dublin, February 2003.
- [22] S. Mizarro et C. Tasso. Ephemeral and persistent personalisation in adaptive information access to scholarly publications on the web. In *Proceedings of the 2nd International Conference on adaptive hypermedia and adaptive Web-based systems*, pages 306–316, 2002.
- [23] N. Nanas, U. Uren et A. Deroeck. Building and applying a concept hierarchy representation of a user profile. In *Proceedings of the 26th Annual ACM Conference on Research and Development in Information Retrieval SIGIR*, pages 154–204, 2003.

- [24] M. Pazanni, J. Muramatsu et D. Billsus. Syskill and weber : Identifying interesting web sites. In *Proceedings of the 13th National Conference on Artificial intelligence*, pages 54–61, 1996.
- [25] J. Pearl. Probabilistic reasoning in intelligent systems : Networks for plausible inference. In *Morgan Kaufman, Olivier, R.M and Smith J. Eds*, 1988.
- [26] J. Rocchio. Relevance feedback in information retrieval. Englewood Cliffs, 1971. Prentice Hall.
- [27] I. Ruthven et M. Lalmas. A survey on the use of relevance feedback for information access systems. volume 18, pages 95–145, 2003.
- [28] L. Scime, A. and Kershberg. Websifter : An ontology based personalizable search agent for the web. In *Proceedings of the 3rd International Conference on digital libraries : Research and Practice*, pages 439–446, 2000.
- [29] R.D Shachter. Probabilistic inference and influence diagrams. In *Operating Research*, volume 36, pages 589–604, 1988.
- [30] X. Shen, Tan B. et Zhai J. C. Context-sensitive information retrieval using implicit feedback. In *Proceedings of the 29th annual international ACM SIGIR Conference on Research and development in Information retrieval*, pages 43–50, August 2005.
- [31] M. Speretta et S. Gauch. Personalizing search based user search histories. In *Proceedings of the 13th International Conference on Information Knowledge and Management (CIKM)*, pages 238–239, 2004.
- [32] A. Spink, J. Bateman et B. Jansen. User’s searching behaviour on the excite web search engine. In *Proceedings of the 1998 World Conference of the WWW, Internet and Intranet*, pages 7–12, 1998.
- [33] J. Su et M. Lee. An exploration in personalized and context-sensitive search. In *Proceedings of the 7th annual UK special interest group for computational linguists research colloquium*, 2003.
- [34] L. Tamine et W. Bahsoun. Définition d’un profil multi-dimensionnel de l’utilisateur : Vers une technique basée sur l’interaction entre dimensions. In *Actes de la 3ème Conférence en Recherche d’Information et Application (CORIA)*, page A paraître, Lyon Mars 15-17 2006.
- [35] J. Teevan, S. Dumais et E. Horvitz. Personalizing search via automated analysis of interests and activities. In *Proceedings of the 28th Annual ACM Conference on Research and Development in Information Retrieval SIGIR*, pages 449–456, August 15-19 2005.
- [36] J.R Wen, N. Lao et Ma W. Y. Probabilistic model for contextual retrieval. In *Proceedings of the 27th annual international ACM SIGIR Conference on Research and development in Information retrieval*, pages 57–63, August 2004.

- [37] R. Yager. A note on weighted averaging aggregation operators in multicriteria decision making. In *IEEE Transactions on Systems, Man and Cybernetics*, pages 18(1) 183–190, 1988.