Institut de Recherche en Informatique de Toulouse
CNRS
INPT
UPS
UT1

Rapport interne : Janvier 2005 Référence : IRIT/2005- -

Accès personnalisé à l'information : Approches et Techniques

Lynda Lechani Tamine, Mohand Boughanem
Equipe **SIG / RFI**
{lechani, bougha}@irit.fr

.

Table des Matières

I- 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 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 œuvre 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 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 spécifique.
C'est pourquoi les travaux 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 précis de la *personnalisation de l'information* qui est vue comme l'une des solutions pouvant maintenir le Web comme une ressource viable [Gow 2003].

I.1 Accès à l'information : préambule

L'objet de toute méthode d'accès à l'information peut être traduit très simplement par l'expression de Kuhlen [Kuh, 1991] : « *Retrieve that amount of knowledge wich a user needs in a specific situation for solving his/her current problem ».*
Les différentes théories, techniques et heuristiques qui ont supporté les stratégies d'accès à l'information ont abordé et permis de résoudre des problèmes liés principalement aux propriétés de l'information et à la finalité de l'accès. Les propriétés de l'information portent notamment sur :

- le média : texte, graphe, image, son, vidéo,
- la structure : non structuré, semi-structuré, structuré, hypertexte,
- l'hétérogénéité : langue, média, structure, service.

La finalité de l'accès caractérise l'objet poursuivi par l'utilisateur : recherche *adhoc,* filtrage, question-réponse, extraction, etc..

Les différentes méthodes d'accès proposées dans la littérature ne sont pas globales. En effet, elles sont généralement adaptées à une finalité bien précise, tenant compte d'un type d'informations particulier, ce qui accroît leurs performances dans leur cadre dédié d'utilisation. Outre ces paramètres, la grande question abordée cette dernière décennie, est l'adaptation de ces méthodes au contexte de l'application qui inclut l'utilisateur et son usage de l'information. Dès lors l'accès à l'information tend vers une nouvelle définition [Fuh 2000] : *« 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 ».* C'est précisément cette perspective qui est explorée dans la suite du rapport.

I.2 Motivations

L'essor du Web, conforté par le rapide développement des nouvelles technologies de l'information et de la communication, lance à la communauté scientifique en recherche d'information, de nouveaux défis. En effet, compte tenu des exigences liées à l'efficacité et efficience, les outils de recherche d'information doivent prendre en considération les facteurs liés à la quantité, structure, contenu des informations produites et leurs contextes d'usage via le Web. Parmi les facteurs dominants qui constituent des enjeux actuels dans le domaine de la recherche d'information, on retient : le volume, l'hétérogénéité et disparité des informations. Pour chacun de ces facteurs, on citera dans ce qui suit, l'apport des techniques de personnalisation de l'information.

- *Volume*

L'explosion de l'utilisation de la toile est à l'origine d'une croissance très significative des volumes d'informations accessibles. Ainsi, le volume d'informations ne se mesure plus actuellement en giga-octets mais en téra-octets voire en péta-octets et exa-octets. Malheureusement, tous les algorithmes en recherche d'information ne sont pas de complexité linéaire en fonction du volume des informations. Ceci fait émerger le vaste problème du «passage à l'échelle» qui engendre des dégradations des performances des processus de recherche tant en termes d'efficience que d'efficacité [CMB & al 2004]. Plus précisément, le passage à l'échelle est à l'origine de:

- l'accroissement des volumes de stockage,
- l'allongement des délais de réponse,
- l'augmentation des coûts d'indexation,
- la diminution de la précision de la recherche.

Pour pallier ces problèmes, de nombreux travaux sont menés dans diverses directions : factorisation de l'information [Dum 1993 ; BDJ 1999], techniques de compression [WB 1999] et personnalisation [PG 1999 ; Kim 2003 ; Gow 2003].
Dans le contexte du passage à l'échelle, la personnalisation permet de réduire virtuellement l'échelle de l'espace de recherche en considérant les caractéristiques de l'utilisateur et de son contexte. Ceci permet alors de :

- mieux identifier l'information pertinente dans le volume considéré et par conséquent améliorer la précision des résultats de recherche,
- rendre plus fiable l'interaction avec l'utilisateur en présentant des informations plus synthétiques et donc plus assimilables.

- *Hétérogénéité*

Le Web est caractérisé par une forte hétérogénéité des sources d'information. Cette hétérogénéité porte sur divers aspects : langue (plus de cent langues actuellement sur le Web), média (texte, image, vidéo), structure etc…A juste titre, le Web sémantique est une nouvelle infrastructure qui permet d'assister les utilisateurs afin d'accéder plus efficacement à diverses ressources du Web. L'intégration globale et automatique d'informations provenant de sources hétérogènes est une question largement débattue dans le domaine [Mcc 02].
Outre la définition de normes standards dans la description des ressources (RDF, Dublin Core) et structure des documents (XML), la personnalisation est un processus qui permet dans ce contexte, d'adapter la structure et contenu de l'information à présenter à l'utilisateur en fonction de ses préférences. La sélection des sources d'information à explorer étant basée sur un contexte d'utilisation spécifique, l'hétérogénéité est de fait diminuée.

- *Disparité*

La disparité est une caractéristique qui traduit l'occurrence disséminée de l'information dans de larges collections de documents, généralement interconnectés. Les outils de navigation hypertextes sont, à ce titre, destinés à matérialiser la proximité des informations autour d'un besoin particulier. Cependant, compte tenu du volume important d'informations disponibles, les utilisateurs sont vite submergés par le nombre considérable de liens proposés, ce qui engendre les phénomènes fort connus de désorientation de l'utilisateur et de surcharge informationnelle .
La personnalisation permet d'y pallier en considérant le profil utilisateur dans le processus de *recommandation* [Lie 1995; Han 1998 ; CS 1998] de liens et de pages, ce qui permet de recentrer virtuellement les informations disséminées autour d'un besoin spécifique.

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 [SL 2003]:

- les utilisateurs ont des objectifs différents, des contextes différents et perceptions différentes de la notion de pertinence,
- un même utilisateur peut avoir différents besoins à différents instants.

I.3 Organisation du rapport

L'organisation retenue pour le présent rapport est la suivante : la section 2 présente la problématique générale de la personnalisation en mettant en évidence l'évolution des besoins ayant conduit à l'émergence de la recherche d'information contextuelle. La section 3 traite de la théorie autour de la modélisation utilisateur. On y montre essentiellement que les techniques d'accès personnalisé à l'information en puisent en grande partie les bases théoriques. La section 4 décrit les systèmes d'accès personnalisé à l'information. Leur architecture est développée par présentation des principales fonctions liées à la gestion des profils utilisateurs et exploitation de ces profils dans le processus d'accès à l'information. La section 5 aborde le volet de l'évaluation des performances de ces systèmes. Enfin, un bilan ainsi que des questions ouvertes seront posés en conclusion.

II- Problématique

L'accès à une information pertinente, adaptée aux besoins et profil de l'utilisateur est un enjeu capital dans le contexte actuel caractérisé par une prolifération massive de ressources d'informations hétérogènes. Malgré les développements récents dans le domaine de la recherche d'information, force est de constater que les résultats produits par un moteur de recherche sont en deçà des attentes des *85%* d'utilisateurs exploitant un moteur de recherche lors de leurs activités quotidiennes [SJW & al 2002]. Les raisons évoquées portent essentiellement sur la grande masse, l'incompréhensibilité et ambiguïté des informations retournées à leurs requêtes. De récentes études [Gow 2003] montrent que l'origine de ces problèmes réside dans le caractère non personnalisé du processus d'accès à l'information.
Dans ce cadre, la personnalisation est une dimension qui permet la mise en œuvre 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 aborde de manière générale la problématique ayant conduit à l'émergence de la personnalisation ainsi que les verrous technologiques et scientifiques posés par l'introduction de cette dimension dans la mise en œuvre de systèmes de recherche d'information actuels.

II.1 Emergence du problème

De nombreux modèles théoriques sont à la base de la conception des systèmes de recherche d'information. Un modèle de recherche d'information est généralement décrit comme un quadruplet [D, Q, F, R(qi, di])] [YN 1999] où :

D : ensemble de documents,
Q : ensemble de requêtes,
F : schéma de représentation des documents et requêtes,
R(qi, dj) : fonction de pertinence.

La mise en œuvre naïve d'un tel modèle suppose 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**. Dans le but de montrer l'incidence d'une telle représentation, considérons à titre illustratif la requête : *information about cats* [BH 2000]. On peut supposer différents scénarios d'utilisation :
- ***scénario 1*** : étudiants vétérinaires qui écrivent un papier sur les cancers de chats,
- ***scénario 2*** : architectes qui travaillent sur un projet de construction,
- ***scénario 3*** : étudiants qui écrivent un papier sur l'Egypte.

Ces différents scénarios illustrent sans doute la différence d'interprétation d'une requête en fonction du contexte de l'utilisateur qui l'exprime. Les problèmes immédiats posés par la non considération du contexte en cours du processus de recherche d'information sont notamment :

- l'ambiguïté du sens des mots,
- l'impossibilité de sélectionner des sources opportunes,
- l'inintelligibilité des résultats.

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.
Ceci a pour corollaire la non pertinence des résultats de recherche et de fait, l'insatisfaction de l'utilisateur. 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 1971 ; SB 1990] et expansion de requêtes en utilisant des techniques de désambiguïsation [CW 1997 ; DH 1999].

- *Reformulation de requêtes par injection de pertinence (Relevance feedback)*
  C'est un processus évolutif et interactif, dirigé par l'utilisateur ayant pour objectif la génération d'une nouvelle requête plus adaptée que celle initialement exprimée par l'utilisateur. Son principe fondamental est d'utiliser la requête initiale pour amorcer la recherche puis modifier celle-ci à partir des jugements de pertinence et/ou de non pertinence de l'utilisateur. La nouvelle requête obtenue à chaque itération feedback, permet de corriger la direction de recherche dans le sens des documents pertinents au sens exprimé explicitement par l'utilisateur.

- *Désambiguïsation du sens des mots*
  Les techniques de désambiguïsation ont pour objectif d'identifier précisément le sens évoqué par les termes de la requête et cibler ainsi les documents contenant les mots cités dans le contexte défini par le sens correspondant. Ces techniques sont généralement basées sur une intervention explicite de l'utilisateur ou exploitation de ressources telles que les thesaurus et ontologies.

Cependant, vu le contexte actuel lié au volume d'informations, ces techniques sont peu viables [JSB 1998] [SBJ 1998]. En effet, la recherche devient performante après un nombre relativement élevé d'itérations feedback alors que celles-ci engendrent une surcharge cognitive de l'utilisateur et par conséquent, une démotivation de ce dernier et donc un manque de fiabilité voire absence de son intervention (jugement de pertinence de documents, choix des termes etc…). De plus, la stratégie de recherche n'est pas adaptée à un contexte d'utilisation dans différents domaines d'intérêts ; le contexte de l'utilisateur demeure peu connu et par conséquent de moindre impact sur le processus de recherche. C'est pourquoi les travaux de recherche se sont orientés vers la *recherche d'information contextuelle* [BH 2000 ; SL 2003].

II.1.1 Notion de contexte

Le contexte de l'utilisateur peut être assimilé à l'ensemble des facteurs qui permettent de décrire ses intentions et perceptions de ce qui l'entoure. Ces facteurs peuvent couvrir divers aspects : psychologiques, sociaux, culturels, professionnels etc…
Vu sous l'angle de la recherche d'information, le contexte possède, d'après N. Fuhr [Fuh 2000], trois principales dimensions illustrées sur la figure 1: social, application et temps. La dimension sociale définit la composante d'appartenance de l'utilisateur : individuel, groupe ou communauté. La dimension application définit le contexte applicatif du besoin exprimé : recherche *ad-hoc*, résolution de problème ou *workflow*. La dimension temps permet de décrire la circonscription temporelle du besoin exprimé : temps passé (*Batch)*, instant courant ou à court terme (*interactive*), intention ou long terme (*personnalisation*). Sous l'angle de la dimension temps, on distingue deux types de contextes avec des démarches de personnalisations appropriées. Le contexte courant ou à court terme décrit les besoins et préférences de l'utilisateur lors d'une session de recherche. Le contexte persistant décrit les besoins à long terme de l'utilisateur sur diverses sessions de recherche. La personnalisation à long terme introduit dès lors des mécanismes d'adaptation du contexte de l'utilisateur en fonction de la variation de ses besoins inscrits sur une longue période.

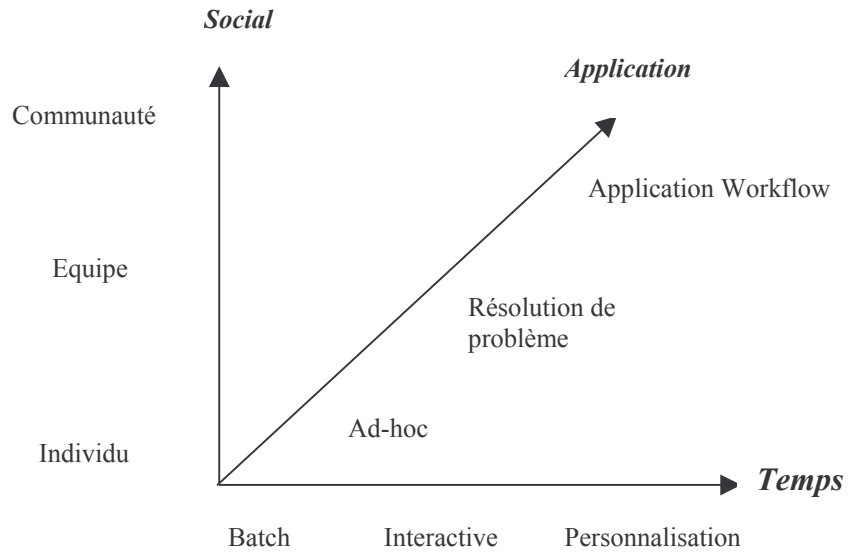

Fig 1. : *Dimensions d'un contexte [Fuh 2000]*

II.1.2 Recherche d'information contextuelle

La recherche d'information contextuelle combine un ensemble de technologies et de connaissances portant sur la requête et le contexte utilisateur, dans une même infrastructure et ce, dans le but de délivrer les réponses les mieux appropriées à son besoin en informations [Fuh 2000]. La recherche d'information contextuelle est une activité qui fait intervenir en grande partie l'utilisateur, assimilé en pratique, non plus à la requête, mais à un ensemble de facettes qui le décrivent. Ces facettes décrivent une structure qui est appelée communément *profil* et qui couvre, de manière non exhaustive, ses caractéristiques socio - professionnelles, ses centres d'intérêt et ses préférences.

Ce courant de recherches est assez récent (début des années 1990) et a suscité d'emblée des travaux émanant de plusieurs communautés : apprentissage automatique, modélisation utilisateur, recherche d'information, Web *learning,* analyse du langage naturel etc…les travaux visent principalement trois objectifs : accès personnalisé à l'information, assistance personnalisée à la navigation et présentation personnalisée des résultats de recherche [BM 2002].

Indépendamment de l'objectif applicatif visé, on identifie trois aspects à promouvoir dans les systèmes de recherche d'information contextuelle :

- capacité à identifier l'intention conceptuelle de l'utilisateur,
- possibilité d'exprimer des informations liées au contexte d'utilisation courant,
- convivialité des interactions utilisateurs – système.

II.2 Verrous scientifiques et technologiques

En tant que thématique de recherche, la personnalisation de l'information est confrontée à des verrous que l'on peut projeter sur deux niveaux. Le premier niveau est lié à l'introduction de la dimension utilisateur lors de la mise en oeuvre d'un processus d'accès à l'information. Le second niveau se rapporte à l'exploitation des systèmes d'accès personnalisé.

- *Mise en œuvre*

L'introduction de la dimension utilisateur dans un processus d'accès à l'information, mérite voire nécessite des réflexions sur la modélisation de l'entité *utilisateur* de manière intrinsèque puis sous l'angle de son rapport avec une activité de recherche d'information circonscrite dans un court et/ou long terme [PG 1999 ; Li 2000; Gow 2003] . Dans ce sens, les questions fondamentales posées par la conception de systèmes d'accès personnalisé sont de type Quoi, Comment et Quand :

- Quoi ?
  - Quelles propriétés décrivent un utilisateur ?
  - Quelle représentation ou quel modèle de l'utilisateur ?
  - Quel contexte d'utilisation ?

- Comment ?
  - Comment construire le modèle de l'utilisateur ?
  - Comment découvrir son intention courante ?
  - Comment exploiter le modèle utilisateur lors du processus de recherche ?
  - Comment évaluer l'impact de la personnalisation sur le processus de recherche d'information ?

- Quand ?
  - Quand faut il mettre à jour le modèle de l'utilisateur ?

L'ensemble des approches de personnalisation de l'information proposées dans la littérature tentent de répondre peu ou prou à ces questions.

- *Exploitation*

L'exploitation de systèmes d'accès personnalisé pose un problème fondamental, d'ordre technologique, qui porte sur la protection de la vie privée. En effet, la définition, utilisation et dissémination des profils constituent à la fois un atout et une contrainte pour assurer la portée des systèmes qui les supportent. C'est tout d'abord un atout dans le sens où les différents profils sont maintenus et sont accessibles, pouvant donc contribuer à mettre en œuvre une recherche d'information collective et dynamique par l'introduction de techniques d'apprentissage. Cependant, c'est ensuite une contrainte dans le sens où elle doit être impérativement soutenue par une réflexion sur les droits des personnes pour assurer la sécurité globale des profils. La personnalisation de l'information doit donc s'accompagner d'une réflexion sur les architectures des systèmes de traitement d'informations (couches fonctionnelles et protocoles de communication), d'un effort de standardisation (protocoles de communication et d'authentification) et de la définition de techniques défensives (supports de stockage ou de calcul inviolable comme la carte à puce, méthode de cryptographie, modèles d'octroi ou révocation de droits).

III- Modélisation de l'utilisateur : clé de la personnalisation

La modélisation de l'utilisateur est une discipline de recherche datant des années 70 et évoquant en premier lieu les travaux d'Allen, Cohen et Perrault [All 1979 ; CP 1979]. La préoccupation majeure de cette discipline est d'améliorer la qualité des interactions homme-machine par inférence et prédiction des buts, préférences et contexte des utilisateurs à partir de faits observés. Les premières applications étaient les systèmes de reconnaissance de plans. Par la suite, les méthodes issues de la modélisation de l'utilisateur ont investi et continuent à investir de nombreux domaines portant sur la mise en œuvre de systèmes intelligents tels que les systèmes ayant recours à l'analyse du langage naturel [Joh 2002], système d'aide à l'apprentissage, systèmes hypermédia adaptatifs [BAH & al, 2000] et tous les systèmes personnalisés de manière générale [PG 1999].
Indépendamment du domaine d'application, tout système mettant en œuvre des méthodes de modélisation de l'utilisateur inclut en partie les paquets d'informations suivants :

- des informations personnelles associées à l'utilisateur telles que l'âge, le pays, la langue,
- les préférences : peuvent être de différents niveaux telles que préférences de forme (style de la page, longueur d'un document) et préférences de domaine permettant de cibler le centre d'intérêts de l'utilisateur,
- historique de l'utilisateur : les interactions passées de l'utilisateur représentent une source pour prédire ses intentions et lui recommander des objets.

Les approches et techniques de la modélisation utilisateur peuvent être basées sur des modèles simples ou complexes dépendant de l'objectif final ou domaine d'application du système; un effort de standardisation pour la généralisation de tels systèmes afin de produire des *Shell* a toutefois été mené et semble donner une meilleure portée au devenir des méthodes de modélisation de l'utilisateur [Kob 2001].

En outre, ces méthodes peuvent être interactives ou implicites, peuvent avoir une portée d'adaptation sur une session d'utilisation du système avec des informations très générales sur l'utilisateur ou sur ou plusieurs sessions d'utilisation du système avec un modèle plus élaboré. Cette section tente de donner un aperçu des différentes approches et techniques puis met en exergue leur utilisation dans les systèmes d'accès à l'information.

III.1 Approches et techniques

III.1.1 Approches

On distingue trois principales approches de modélisation de l'utilisateur [Gow 2003] : approche canonique, approche explicite et approche automatique.

- *Modèles canoniques*

Cette approche préconise l'intégration de modèles d'utilisateurs typiques lors de la conception du système. Les interactions permettent de *cataloguer* l'utilisateur courant par rapport à un modèle prédéfini dans le système. Cette approche a été peu performante, notamment en raison de l'inadéquation des langages des concepteurs et des utilisateurs pour décrire les situations permettant d'apparier l'utilisateur à un modèle canonique [Gre 1984].

- *Modèles explicites*

Dans le cas de cette approche, le système maintient un *panel* de modèles canoniques caractérisés par une partie flexible qui est contrôlée par l'utilisateur lors de ses interactions avec le système. Cette approche remédie aux inconvénients de l'approche canonique en réduisant l'erreur de catalogage due à une description incertaine des situations. Cependant, elle induit une surcharge cognitive pour l'utilisateur et une complexité dans la conception du système.

- *Modèles automatiques*

Dans le but de pallier au problème de surcharge cognitive et d'incertitude engendré par les approches précédentes, l'approche automatique préconise d'inférer le modèle de l'utilisateur, non pas à partir de ses interactions explicites avec le système, mais à partir d'informations collectées implicitement lors de ses sessions d'utilisation du système. En clair, le comportement de l'utilisateur est la source permettant de prédire implicitement son modèle. C'est le modèle le plus répandu actuellement. Deux principales classes de techniques sont issues de cette approche et peuvent être combinées : les techniques collaboratives et techniques statistiques, développées dans le paragraphe suivant.

III.1.2 Techniques

III.1.2.1 *Techniques collaboratives*

Les techniques collaboratives sont basées sur l'idée de prédire le modèle individuel d'un utilisateur courant sur la base d'un comportement assimilable à celui d'un groupe d'utilisateurs. Les utilisateurs du système participent ainsi collectivement à alimenter des stéréotypes qui sont affectés à des groupes d'intérêts communs puis utilisés pour prédire les préférences inconnues de nouveaux utilisateurs. Les systèmes de recommandation sont généralement basés sur de telles techniques. Cependant leur efficacité dépend fortement du degré de corrélation du groupe [PCG 2003]. De plus, des problèmes liés à la taille et composition des groupes sont posés. En effet, l'approche reste peu performante pour un nouvel utilisateur (avec peu d'informations collectées à partir du groupe) et de nouveaux centres d'intérêts [ZA 2001].

III.1.2.2 *Techniques statistiques*

Ces techniques sont basées essentiellement sur des modèles théoriques issus de la statistique, soutenus par des heuristiques et algorithmes appropriés. Les principaux modèles sont : le modèle linéaire, le modèle markovien, les réseaux de neurones, la classification, les règles d'induction et les réseaux bayésiens.

- *Modèle linéaire*
  Le modèle linéaire a une structure simple. L'hypothèse de base est que la valeur de prédiction présumée et inconnue d'un objet cible du système (article à recommander [RBW 1997], degré d'intrusion par prédiction de l'intervalle entre deux accès sécurisés [Orw 1995], etc…) est une combinaison linéaire des valeurs calculées à partir d'un comportement passé de l'utilisateur. Le modèle linéaire peut être aisément combiné avec des techniques collaboratives où les valeurs connues sont issues de l'appréciation des membres du groupe associé à l'utilisateur courant [RIW 1994].

- *Modèle Markovien*
  Ce modèle est essentiellement basé sur l'hypothèse markovienne qui permet de représenter une séquence d'événements ultérieurs sur la base d'un nombre fixe d'événements antérieurs. Etant donnée la distribution de probabilités d'occurrences des événements passés, la théorie markovienne offre alors des éléments pour calculer la probabilité d'occurrence des événements futurs.

- *Réseaux de neurones*
  Les réseaux de neurones sont des structures basées sur l'interconnexion de nœuds et un principe d'activation par propagation de signaux à travers les connexions depuis les entrées jusqu'aux sorties. La signification effective des nœuds, des connexions et valeurs d'activation, dépend du problème pour lequel est dédié le réseau. De manière générale, les réseaux de neurones sont destinés à résoudre des problèmes de décision non linéaires. Dans le cas précis du vaste domaine d'application de la modélisation utilisateur, l'entrée représente une situation ou faits observables à partir de l'utilisateur, les sorties représentent des objets cibles du système avec des valeurs d'activation qui traduisent le degré de prédiction.

- *Classification*
  Les méthodes de classification permettent de partitionner un espace d'objets en classes de manière à réduire sa dimension. Les objets d'une même classe ont des propriétés partageables, dont le degré de corrélation est calculé à l'aide métriques basées sur la similarité. De nombreuses stratégies de classification sont proposées dans la littérature. Du point de vue de la modélisation utilisateur, les méthodes de classification permettent généralement d'identifier la classe de caractéristiques de l'utilisateur courant à partir d'informations dérivées de son comportement.

- *Induction de règles*
  L'induction de règles consiste en l'apprentissage de règles de prédiction à partir d'un ensemble de faits observés. Contrairement aux méthodes de classification, les techniques d'induction de règles requièrent, durant l'apprentissage du système, l'identité de la classe associée à chaque observation. Ces techniques produisent un ensemble de règles proprement dit ou des arbres de décision.

- *Réseaux bayésiens*
  Les réseaux bayésiens [Per 1988] ont largement investi, ces dernières années, les travaux sur la modélisation utilisateur [Jam 1995]. Les réseaux bayésiens sont des graphes acycliques orientés où les nœuds correspondent à des variables aléatoires. Les nœuds sont interconnectés à l'aide de liens orientés qui représentent des liens de causalité entre nœuds parents et nœuds fils. A chaque nœud est associée une distribution de probabilités conditionnelles qui permet d'assigner au nœud, une valeur de probabilité dépendante de la combinaison des valeurs de probabilités possibles des nœuds parents. Les réseaux bayésiens sont plus flexibles que les techniques précédentes, en ce sens qu'ils permettent de représenter explicitement les relations de causalité entre faits et d'émettre des prédictions sur de nombreux paramètres du système [ZA 2001].

La modélisation utilisateur pose globalement de nombreux enjeux se rapportant principalement à [Kim 2003] :

- la complexité des calculs lors de la phase d'identification du modèle d'utilisateur courant,
- la dérive de concept du au caractère très dynamique de l'utilisateur,
- la nécessité de disposer d'un volume important de données d'apprentissage.

III.2 Modélisation de l'utilisateur pour la personnalisation de l'accès à l'information

Les approches et techniques de modélisation de l'utilisateur sont au cœur de la mise en œuvre de processus d'accès personnalisé à l'information. On distingue principalement trois approches de la personnalisation : approche interactive, approche collaborative et approche basée sur l'apprentissage. L'approche interactive évoque en partie les modèles d'utilisateur explicites où un ensemble d'informations est explicitement exprimé par l'utilisateur pour guider le processus de personnalisation. La reformulation de requête par injection de pertinence est une technique qui s'apparente à cette approche mais s'avère peu efficace voire non viable dans le contexte du Web [CCL 2001]. L'approche collaborative est similaire à son homonyme décrite dans le cadre de la modélisation utilisateur. Cette approche a été particulièrement utilisée pour réaliser des systèmes de recommandation basés sur le filtrage collaboratif [RIW 1994 ; GSK & al, 1999]. Ces systèmes ont toutefois présenté de nombreux problèmes [SC 2003] dus principalement:

- *au passage à l'échelle* : la complexité des calculs dus à l'exécution des algorithmes du plus proche voisin (algorithmes permettant d'associer un utilisateur à un groupe) en termes de temps croît linéairement avec le nombre d'utilisateurs. En conséquence, le système de recommandation est peu performant pour des applications à grande échelle,

- *à la diversité* : en raison de a diversité des articles à recommander d'une part, et du nombre considérable d'utilisateurs d'autre part, les matrices de profils sont généralement creuses. Par conséquent, la précision et même nombre de recommandations diminue.

L'approche basée sur l'apprentissage évoque l'utilisation des techniques statistiques pour la prédiction du modèle utilisateur. On cite à titre d'exemples les travaux de:

- Bestavros [Bes 1996], Horvitz [HBH & al, 1998] et Zukerman et al [ZAN, 1999] ont utilisé le modèle Markovien pour la recommandation de pages web à partir des dernières pages vistées par l'utilisateur, liens explorés, taille et contenu de ces pages. Pitkow et Pirolli [PP 1999] ont montré le dilemme entre complexité du modèle et précision de la prédiction. Deshapande et Karypis [DK 2001] ont utilisé des techniques d'élagage pour réduire la complexité du modèle Markovien et maintenir sa précision.
- Jennings et Higuchi [JH 1993] ont utilisé les réseaux de neurones pour représenter les préférences des utilisateurs pour les nouveaux articles. Pour chaque utilisateur, l'approche proposée, consiste à apprendre un réseau où les nœuds représentent les mots qui apparaissent dans les articles préférés de l'utilisateur et les liens représentent des associations entre termes apparaissant dans ces mêmes articles,
- Perkowitz zt Etzioni [PE 2000] ont utilisé les techniques de classification dans le but de créer un index sur des pages WWW reliées et fréquemment visitées par l'utilisateur durant une même session. J.P. Gowan [Gow 2003] ont identifié les différents centres d'intérêts d'un utilisateur à l'aide d'un algorithme de classification qui opère sur le contenu des documents consultés via des applications.
- Billsus et Pazanni [BP 1999] ont appliqué des techniques d'induction de règles et modèles linéaires pour la recommandation de nouveaux articles. SurfLen[FBH 2000] est un système de recommandation basé sur la génération de règles d'association à partir des contenus des pages visitées par l'utilisateur.
- Horvitz et al [HBH & al, 1998] ont utilisé un réseau bayésien pour modéliser des requêtes sur le Web et prédire la prochaine requête de l'utilisateur en termes de spécialisation ou généralisation.

IV- Les systèmes d'accès personnalisé à l'information

Nous définirons dans ce qui suit les systèmes d'accès personnalisé à l'information puis développerons leurs objectifs ainsi que leur architecture générique.

IV.1 Définition

Un système d'accès personnalisé à l'information est un système qui intègre l'utilisateur, en tant que structure informationnelle, tout au long de la chaîne d'accès à l'information.
Un tel système inclut alors :

- des structures permettant de représenter l'utilisateur. Ces structures traduisent essentiellement les centres d'intérêts, les préférences et contexte de l'utilisateur,
- des techniques pour collecter les informations descriptives de l'utilisateur et instancier les structures associées,
- un processus d'accès à l'information intégrant les structures descriptives des utilisateurs,
- un mécanisme d'évolution de ces structures.

IV.2 Objectifs

L'objectif d'un système d'accès personnalisé à l'information est de délivrer une information pertinente en fonction des caractères spécifiques de l'utilisateur. La personnalisation de l'information peut être alors vue comme un processus de définition,

construction, exploitation et évolution des profils d'un utilisateur ou groupe d'utilisateurs en vue de répondre de façon adaptée à un besoin en informations exprimé par un type de requête. Ces systèmes peuvent être abordés selon deux points de vue orthogonaux [Bou 2004] : les domaines d'application et les technologies de base.

- les domaines d'application qui ont recours à la personnalisation de l'information sont nombreux: le commerce électronique (*e-commerce*), la dissémination sélective d'informations, l'assistance à la navigation, l'apprentissage assisté par ordinateur (*e-learning*), l'accès aux bibliothèques électroniques (*digital librairies*), les systèmes d'information mobiles (téléphonie mobile, agendas personnels, systèmes embarqués), la (re)configuration de logiciels (réseaux, composants), etc. Selon les domaines, la personnalisation consistera en l'une ou plusieurs des tâches suivantes : filtrer un flux d'informations entrant pour éliminer le bruit, guider la navigation dans un espace d'informations trop vaste, recommander un ensemble d'informations à l'utilisateur de manière plus intrusive (nouvelles offres par exemple), ajuster le résultat d'une requête selon une interface (ordre de présentation des résultats par exemple), adapter l'interaction à la situation de l'utilisateur (matérielle, géographique), etc.

- les technologies qui permettent de supporter ces applications. On distingue entre autres les systèmes de bases de données, les moteurs de recherche d'informations, les interfaces homme-machine et les intergiciels (ou *middlewares*). Chacune de ces technologies a une offre différente en personnalisation : introduction de préférences dans les langages de requêtes, utilisation du *feedback* des utilisateurs, pondération et ordonnancement des résultats des recherches, exploitation d'ontologies, etc

IV.3 Architecture

Nous tentons dans ce paragraphe de dégager une architecture standard pour un système d'accès personnalisé à l'information. Notre volonté est de mettre en évidence l'ensemble des fonctionnalités de tels systèmes même si elles ne sont pas toujours présentes collectivement dans tout système. Cette architecture est centrée autour de l'utilisateur (Fig 2.). En ce sens que l'ensemble des modules du système fait intervenir en partie les informations descriptives du profil utilisateur. Sur la base de cette architecture, nous dégageons principalement deux fonctions fondamentales qui sont la gestion des profils et sélection de l'information.

IV.3.1 Gestion des profils

La représentation de l'utilisateur à travers la notion de profil permet de mieux comprendre ses mécanismes cognitifs, notamment ceux permettant de percevoir le concept subjectif de la pertinence et au-delà, cibler ses besoins spécifiques dans le but d'améliorer les performances de recherche. [Dan 1986] définit deux classes de modèles de profils utilisateurs :

- les modèles quantitatifs et empiriques : leur but est de modéliser le comportement externe de l'utilisateur,
- les modèles analytiques et cognitifs : leur but est de comprendre le comportement interne de l'utilisateur : connaissance, raisonnement etc…

Ces deux aspects sont généralement combinés pour représenter, construire et faire évoluer les profils.

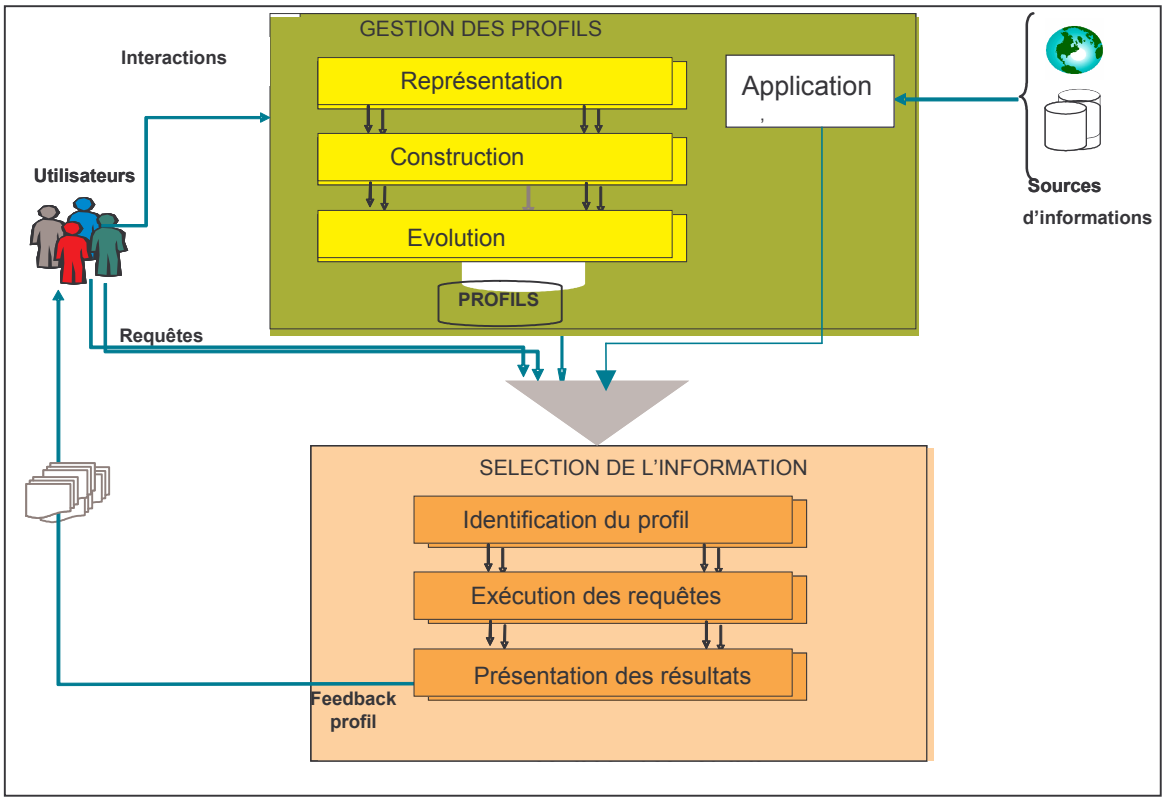

Fig 2. : *Architecture fonctionnelle d'un système d'accès personnalisé à l'information*

IV.3.1.1 Représentation

Le profil de l'utilisateur n'a pas forcément de structure explicite qui le représente. Il peut être constitué de paquets divers d'informations qui traduisent une connaissance éparse sur l'utilisateur. Dans ce sens, la représentation des profils rejoint en grande partie la représentation de l'information dans le contexte de la recherche d'information. Il n'y a pas à notre connaissance de modèle spécifique, dédié à la représentation du profil de l'utilisateur. Les modèles proposés puisent largement de ceux proposés en recherche d'information. On en cite principalement quatre types de représentation : vectorielle, sémantique, connexionniste et multidimensionnelle.

- *Représentation vectorielle*

Ce type de représentation s'appuie généralement sur le modèle vectoriel [Sal 1971]. Le profil est représenté par un ou plusieurs vecteurs définis dans un espace de termes obtenus implicitement ou explicitement à partir de plusieurs sources d'information. Les coordonnées des vecteurs correspondent aux poids des termes dans le profil. L'utilisation de plusieurs vecteurs permet de prendre en compte la diversité des domaines d'intérêts ou évolution dans le temps.

Ce type de représentation offre l'avantage indéniable de la simplicité de mise en œuvre. Cependant les modèles proposés ne mettent pas en évidence ni la dimension liée au temps marquant l'évolution des profils, ni à l'organisation des informations pour hiérarchiser les centres d'intérêt.

- *Représentation sémantique*

La représentation sémantique met d'avantage en relief les relations de sens entre unités d'informations représentant le profil en apportant des solutions aux problèmes de dissémination et synonymie. La direction proposée dans ce contexte, est la construction hiérarchique de concepts plutôt qu'une liste de structures indépendantes, à partir d'informations issues des fichiers *logs*. La hiérarchie peut rendre compte des niveaux de préférences de l'utilisateur et des associations latentes entre concepts et donc un raisonnement sémantique pour la dérivation du profil s'y prête aisément. Ce type de représentation utilise généralement des ontologies [PG 1999 ; SKW 2000].

- *Représentation connexionniste*

C'est un type de représentation basé sur l'interconnexion de nœuds représentant les termes, préférences [JH 1993] ou documents. Il offre le double avantage de la structuration et de la représentation associative permettant de considérer l'ensemble des aspects représentatifs du profil.

- *Représentation multidimensionnelle*

C'est un type de représentation qui se veut global dans le sens où il permet de capturer puis catégoriser l'ensemble des informations caractérisant le profil de l'utilisateur. Dans cette direction, les propositions de standards P3P pour la sécurisation des profils ont défini des classes distinguant les attributs démographiques des utilisateurs, les attributs professionnels et les attributs de comportement. D'autres travaux ont adopté cette structuration [Kos 2003]

IV.3.1.2 Construction

La construction du profil traduit un processus qui permet d'instancier sa représentation. Ce processus peut être explicite ou implicite. La construction explicite est basée sur une collecte d'informations directement fournies par l'utilisateur via l'interface du système. La construction implicite, largement motivée par les travaux actuels dans le domaine, repose sur un procédé d'inférence du contexte et préférences de l'utilisateur via son comportement lors de l'utilisation du système ou d'autres applications quotidiennes. Les informations exploitées pour la construction sont généralement issues :

- directement de l'utilisateur :
  - jugement explicite sur la pertinence des termes, documents,
  - définition de différents attributs : domaine d'intérêts, niveau, langue, etc…
  - sélection de thèmes, sites favoris

- indirectement de l'application :
  - contenu des documents crées, consultés,
  - liens explorés,
  - durée de lecture des documents,
  - dernières pages visitées,
  - type d'application

IV.3.1.3 Evolution

L'évolution désigne l'adaptation des structures représentatives des profils utilisateurs à la variation des besoins en informations de ces derniers.
Les principaux travaux mettant en exergue le processus d'évolution des profils utilisateurs ont porté notamment sur les systèmes de filtrage d'information [All 1990 ; Cro 1993] où la dimension temps est dominante pour un routage permanent des informations aux profils correspondants.

A notre connaissance, peu de travaux ont abordé le problème de l'évolution du modèle de l'utilisateur dans les systèmes personnalisés d'accès à l'information. Dans le cas de ces systèmes, l'évolution est d'avantage abordée comme un problème de représentation de la diversité des domaines d'intérêts de l'utilisateur [PMB & al, 1996 ; Gow 2003]. Cette représentation est généralement basée sur un processus de classification dynamique qui tient compte des contextes courants issus de chaque session d'utilisation du système ou des applications courantes de l'utilisateur.

IV.3.2 Sélection de l'information

Cette phase consiste à intégrer le profil ou modèle utilisateur préalablement construit dans le processus de recherche d'information proprement dit. En ce sens, les informations contenues dans le profil courant sont exploitées pour identifier éventuellement le profil parmi ceux qui sont en cours de construction, réécrire puis exécuter la requête, enfin présenter les résultats de la recherche.

IV.3.2.1 Identification du profil

C'est une opération existante dans le cas de systèmes qui maintiennent un panel de profils canoniques ou dynamiques. Elle consiste à apparier la structure instanciée du profil avec ceux définis ou construits préalablement dans le système. L'appariement est généralement basé sur le calcul d'un score de probabilité de prédiction de profil [HBH & al, 1988] ou similarité avec une classe de profils [PMB & al, 1996 ; Gow 2003].

IV.3.2.2 Exécution des requêtes

L'exécution d'une requête traduit la succession éventuelle des opérations de sélection de sources d'information, reformulation et calcul d'un score de pertinence. La sélection personnalisée de sources d'information est une pratique courante dans les méta-moteurs de recherche [CZC 2001 ; GLG &al, 1999]. Leur principe est d'identifier, à travers le profil utilisateur, le type de requête (besoin général, actualité, scientifique…) puis l'adresser à un moteur de recherche approprié afin d'augmenter la précision de résultats. Le système *Inquirus* présenté ci-après est un exemple type de tels systèmes. La reformulation de requête qui est une des premières techniques qui s'apparentent à la personnalisation, consiste à augmenter la requête avec des informations issues du profil avant de lancer le processus d'appariement [LYM 2002 ; SKW 2000]. La personnalisation peut également porter sur la définition de la fonction de calcul de la pertinence. Dans ce sens, [FGP 2004] ont proposé l'adaptation des paramètres de la fonction de pertinence au contexte de l'utilisateur, en utilisant les techniques de programmation génétique. Jeh et Widom [JW 2003] ont proposé une variante *personnalisée* de l'algorithme *PageRank* en l'occurrence PPV (*Personalized PageRank Vector)*. Son principe fondamental est de privilégier les pages reliées aux pages préférées de l'utilisateur ou pages citées par ces dernières durant le processus de calcul des scores de sélection.

IV.3.2.3 Présentation des résultats

La présentation des résultats est la phase ultime du processus d'accès à l'information. Cette phase peut également considérer le profil de l'utilisateur en réordonnant les résultats fournis par le processus de sélection. En ce sens que l'ordre final des documents à présenter à l'utilisateur est une combinaison de l'ordre produit par le processus de sélection et celui donné par le contexte de l'utilisateur via un calcul de similarité [Vis 2004] ou jugements explicites de la pertinence [Gow 2003].

IV.4 Prototypes

Les systèmes d'accès personnalisé à l'information peuvent être catégorisés selon différents critères : technologie de base, objectif, architecture etc.. Pour notre part, nous présentons quelques systèmes, reconnus comme des références dans le domaine, selon une classification retenue dans [SC 20003] qui est en l'occurrence : systèmes de recommandation, systèmes d'accès contextuel, méta-moteurs de recherche personnalisée. Pour chaque classe de systèmes, nous présentons un ou plusieurs prototypes en décrivant notamment les principes de représentation des profils, les sources d'informations utilisées pour leur construction et la stratégie de sélection de l'information.

IV.4.1 Les systèmes de recommandation

Les systèmes de recommandation sont des assistants à la navigation qui aident et s'adaptent à l'utilisateur soit en lui sélectionnant les liens pertinents d'une page, soit en lui soumettant des documents en relation avec le document en cours de lecture via une exploration automatique de liens. Cette approche est notamment illustrée dans Syskill & Webert [PMB & al, 1996], Letizia [Lie 1995], WAIR [SZ 2000], WEBACE [BGG &al, 1998] et WEBWATCHER [AFJ & al, 1995]. Ci-dessous, une description des systèmes Letizia et Syskill & Webert.

- ***Letizia [Lie 1995]***

  Letizia est un assistant à la recherche qui recommande des liens à explorer en fonction de la page en cours d'être visitée par l'utilisateur.

  - *Représentation du profil*

  Aucune information explicite n'est fournie sur le modèle de représentation du profil. Cependant, comme, les documents sont représentés selon le modèle vectoriel (vecteur de termes), on suppose qu'il en est de même des profils.

  - *Construction du profil*

  Le profil est construit de façon implicite à partir du contenu des pages explorées par l'utilisateur. En supposant que la lecture est effectuée de haut en bas et de gauche vers la droite, les liens non explorés sont supposés non pertinents et participent à la définition du profil

  - *Sélection*

  On suppose que la sélection des pages à recommander est basée sur le calcul d'un score d'appariement vectoriel entre contenu de la page et profil ainsi construit.

- ***Syskill & Webert [PMB & al, 1996]***
  Syskill & Webert est un système de recommandations de pages qui apprend des profils à plusieurs centres d'intérêts.

  - *Représentation du profil*

  Le profil est représenté sous forme de classes qui ne présentent aucune relation hiérarchique. Chaque classe représentant un centre d'intérêts, est modélisée à l'aide d'un vecteur booléen de mots-clés.

  - *Construction du profil*

  Le profil est construit de manière explicite, à partir du jugement de pertinence d'un index des pages recommandées.

  - *Sélection*

  La sélection des pages à recommander est basée sur un calcul de probabilités de prédiction issues d'un classifieur bayésien qui donne de meilleurs résultats relativement à une classification basée sur les mesures TF-IDF. Les mesures calculées constituent des probabilités d'appartenance de pages candidates aux différentes classes du profil. Le système construit également des requêtes comme une combinaison des termes les plus discriminants (valeur de IDF élevée) et les plus fréquents (valeur de TF élevée) issus du profils. Les résultats issus de l'exécution de la requête constituent des résultats présentés à l'utilisateur.

IV.4.2 Les systèmes d'accès contextuel

Relativement aux systèmes précédents, ces systèmes utilisent des informations contextuelles, non liées à la recherche en cours (autres qu'un *feedback* implicite ou explicite sur les documents en cours), pour effectuer la recherche ou désambiguïser la requête. Dans cette catégorie, on citera les systèmes Watson [BH 2000], Web Personae [Gow 2003] et SIS (Stuff I've Seen) [DCC & al, 2003].

- ***Watson [BH 2000]***
  Watson est un assistant à la recherche qui est exécuté en arrière plan d'applications courantes. Son principe est de scruter le comportement de l'utilisateur pendant l'exécution de tâches courantes (ex : traitement de texte) dans le but d'inférer son contexte pour anticiper des besoins en informations ultérieurs. Les informations portant sur le contexte de l'utilisateur sont exploitées pour construire implicitement des requêtes et lui recommander des documents ou reformuler des requêtes exprimées de manière explicite. La figure 3 illustre l'architecture conceptuelle du système Watson.

  - *Représentation du profil*

  Pas de structure explicite pour le profil. Ce dernier peut être assimilé à une requête construite implicitement à partir des documents manipulés lors d'applications courantes. Le document est représenté comme une liste pondérée de termes puis transformé sous forme d'une requête réécrite pour être adressée à différentes sources d'information.

  - *Construction du profil*

  Le profil-requête est construit implicitement à partir du contenu de documents en cours de création ou consultation par l'utilisateur. La construction de la requête est basée sur l'utilisation de différentes heuristiques : les mots fréquents sont importants, les mots figurant dans en-têtes, titres sont mieux pondérés que les autres, les termes apparaissant

avec des petites fontes sont moins importants, les termes apparaissant en marge des documents sont à ignorer (ex : barre de navigation). Le profil- requête est constitué des 20 termes les plus pondérées et dans l'ordre de leurs poids.

- *Sélection*

Le profil requête est adressé à différentes sources d'information. La sélection des meilleurs documents est basée sur la classification des résultats de recherche issus des différentes sources dans le but d'éliminer les redondances et réduire le volume d'informations présenté à l'utilisateur. La représentativité des classes de documents est exprimée à l'aide d'heuristiques liées à la similarité des titres et URL.

- ***WEB personae [Gow 2003]***

Web Personae est un système qui interagit en mode *off line* avec des applications Web dans le but de fournir un accès personnalisé. Il est principalement composé de deux parties : le constructeur et l'identificateur de profil. Le constructeur construit de manière évolutive, le modèle de l'utilisateur comme une liste de profils correspondant à divers centre d'intérêts inférés à partir de ses interactions avec des applications Web. L'identificateur permet de découvrir le profil courant de l'utilisateur lié à une session d'utilisation courante. WEB Personae combine ainsi les deux types de personnalisation à court et long terme. La figure 4 illustre l'architecture conceptuelle du système.

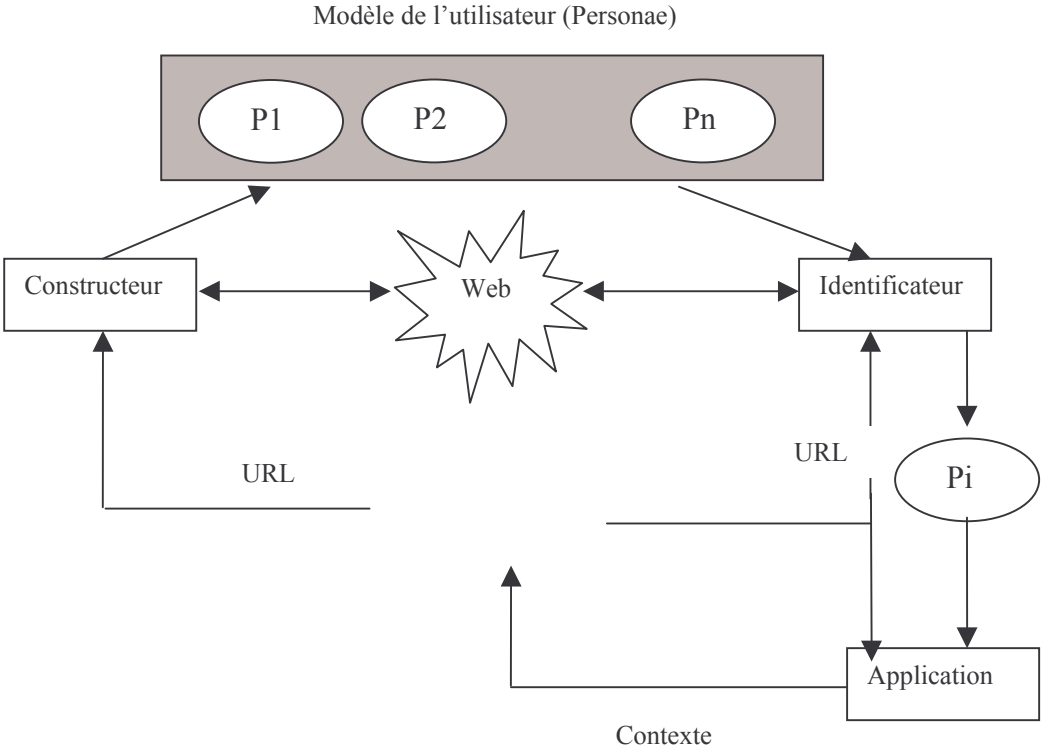

Fig 3. *Architecture fonctionnelle du système WebPersonae*

- *Représentation du profil*

Les profils de l'utilisateur sont représentés par des classes hiérarchiques de descripteurs de documents décrits selon le modèle vectoriel (vecteur de termes avec des poids TF-IDF). Chaque classe décrit un domaine d'intérêts de l'utilisateur. Le processus de classification tend à maximiser la grandeur :

$$R = \frac{I}{E}$$

Où :

I : similarité moyenne intra-classes, calculée comme suit :

$$I = \sum_{r=1}^{n} n_r \left(\frac{1}{n_r^2} \sum_{di,dj \in Sr} \cos(d_i, d_j) \right)$$

Avec S_r : classe, +

n_r : taille de la classe S_r,
d_i, d_j : documents,
n : nombre total de classes

E : similarité moyenne inter-classes, calculée selon la formule :

$$E = \sum_{r=1}^{k} n_r \cos(C_r, C)$$

Avec C: centroïde de la collection,
C_r : centrïde la classe C_r

- *Construction du profil*

Les profils sont construits implicitement à partir :

* du contenu des documents associés à une liste d'URL fournies initialement par l'utilisateur (*bookmark, favoris…),*

* de l'historique du comportement de l'utilisateur (fichier log construit par observation de l'utilisateur : pages visitées, liens explorés, …)

- *Sélection*

La sélection se résume réellement à une identification de profil courant qui est exploité ultérieurement pour effectuer l'accès personnalisé proprement dit. Le principe est simple : un vecteur de termes pondéré selon la formulation TF est construit à partir d'une combinaison des n dernières pages visitées par l'utilisateur. Le système calcule ensuite un score de similarité entre ce vecteur contexte et chacun des vecteurs centroïdes des classes construites pour la représentation des différents profils. Le profil qui est à l'origine du score le plus élevé constitue le profil courant utilisé pour effectuer la personnalisation.

- ***SIS [DCC &al 2003]***

Stuff I've Seen est un système qui offre un accès personnalisé à des informations préalablement traitées ou consultées par l'utilisateur via diverses formes : e-mail, page Web, document en intranet, etc…Ce système a pour objectif essentiel de remédier à la difficulté d'accès à des informations organisées et gérées par des applications indépendantes (fichier de

messages, Web, agenda etc…) qui offrent de surcroît des moyens limités pour l'accès et réutilisation de l'information. A cet effet, le système opère en deux phases. La première phase consiste à créer un index unifié pour toutes les informations consultées indépendamment de la source et de la forme. Cet index unique constitue réellement le contexte de l'utilisateur. La seconde phase exploite ce contexte, à travers une interface adaptée, pour effectuer un accès personnalisé à des sources d'information déjà utilisées.

- *Représentation du profil*

Le profil est représenté par un index unifié des informations collectées par les différentes sources. Pas d'informations explicites sur le modèle de représentation utilisé.

- *Construction du profil*

Le profil est construit à partir d'un fichier log qui décrit les diverses informations manipulées par les différentes applications ainsi que les requêtes adressées au système.

- *Sélection*

La sélection est basée sur la définition de critères par l'utilisateur tels que date, auteur et score d'appariement. Les documents associés à la requête de l'utilisateur sont présentés à l'utilisateur selon un ordre qui respecte ce critère ainsi que le score d'appariement requête-informations calculé selon l'algorithme d'Okapi.

IV.4.3 Les méta-moteurs de recherche personnalisée

Il a été montré que le degré de couverture des moteurs de recherche diminue remarquablement avec l'accroissement de la taille du Web [LG 1999]. Par conséquent l'utilisation d'un seul moteur de recherche peut sembler infructueuse pour répondre à un besoin en informations. Ceci a suscité l'apparition de méta-moteurs de recherche, tels que MetaCrawler[1], et DogPile[2], qui accroissent la couverture de recherche en combinant les résultats issus de différents moteurs de recherche (Voir Fig 5). Cependant, en raison du volume important d'informations accédé, l'utilisateur est vite submergé par la quantité considérable de résultats retournés. C'est pourquoi, les méta-moteurs personnalisés sont apparus. Leur objectif est d'améliorer l'efficience de la recherche combinée en intégrant le contexte de l'utilisateur dans les différentes phases du processus de recherche. Les techniques de personnalisation utilisent généralement les informations liées au contexte lors de la phase de formulation de la requête [CZC 2001], de sélection du moteur de recherche [GFL & al 1999] et/ou phase de fusion des résultats [ZDC & al 2001]. Le méta-moteur *Inquirus*, ci-dessous présenté, combine les techniques de personnalisation pour la reformulation de requêtes et sélection des sources d'informations.

- ***INQUIRUS [GFL & al 1999]***

  Le système *Inquirus* (voir Fig 6.) est un méta-moteur de recherche qui exploite la requête explicitement exprimée par l'utilisateur ainsi qu'un ensemble de préférences pour réécrire la requête et identifier le moteur de recherche à exploiter. La stratégie de sélection de l'information varie en fonction de la source d'information accédée, de la nature des modifications opérées sur la requête ainsi que de la fonction d'appariement utilisée.

[1] http://www.metacrawler.com/
[2] http://www.dogpile.com/

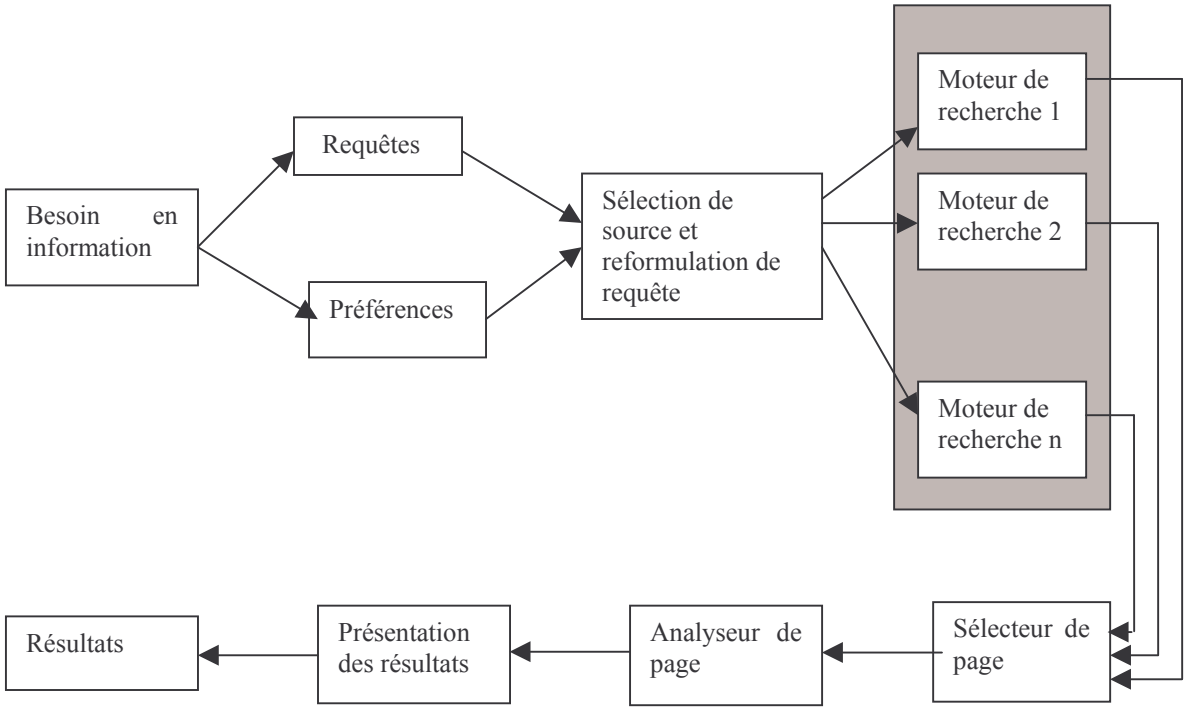

Fig 6 : *Arcitecture fonctionnelle du système Inquirus* [GFL & al 1999]

- Représentation du profil

Le profil n'est pas explicitement structuré. Il est constitué de la requête et d'une catégorie d'intérêts explicitement choisie par l'utilisateur. Cette catégorie permet au système d'associer un moteur adéquat, à même de retourner des résultats adéquats. Le tableau 1 illustre les associations catégorie - moteur de recherche.

- Construction du profil

Le profil correspond à une requête reformulée, adaptée à la catégorie de recherche. A titre d'exemple, on ajoute les termes :

* *in the last two weeks* pour les requêtes de la catégorie *recent events,*
* *abstract, keywords, introduction* pour les requêtes de la catégorie *research papers,*
* *what is X* pour les requêtes de la catégorie *general introductory* comprenant le mot *X.*

Nom	**Description**	**Moteur de recherche**
Papiers de recherche (références)	Pages détaillées, de préférence articles actuels	Google, Alta Vista, Snap, Yahoo, HotBot, Northern Light
Home pages individuelles	Home Pages associées à des requêtes	Snap, Google, HotBot, Yahoo
Evénements actuels, nouvelles récentes	Articles récents	ABCNews, News.com, Snap, AltaVista, Yahoo, HotBot
Introduction générale à …	Pour aborder, réponses à qu'est ce que ?	Google, AltaVista, Snap, Yahoo

Tableau 1. *Catégories de besoins en informations et moteurs de recherche associés [GFL & al, 1999]*

- *Sélection*

  La sélection est abordée comme un problème de décision en utilisant la théorie de l'utilité [RKH 1976]. A chaque catégorie de besoins est associée une liste d'attributs caractéristiques (voir Tableau 2) qui sont utilisés pour évaluer l'utilité d'un document selon la forme :

$$U(d_j) = \sum_k w_k v_k(x_{jk})$$

  Où :
  w_k : poids du kième attribut qui décrit son importance,
  v_k : valeur du kième attribut,
  x_{jk} : niveau du kième attribut pour le document d_j

  Ces attributs peuvent être prédéfinis ou définis par l'utilisateur en fonction de ses préférences. Les résultats de recherche issus des moteurs sélectionnés sont fusionnés puis ordonnées selon la valeur d'utilité ainsi calculée

Nom	**Description**
Wordcount	Nombre de termes par page
Anchorcount	Nombre de liens dans la page
Imagecount	Nombre d'images dans la page
Sectioncount	Nombre de sections dans la page
Patlenght	Profondeur de la page à partir du niveau domaine
Topicalrelevance	Attribut qui dépend de la requête, décrit le degré de pertinence de la page relativement à la page en cours.
Latex	Attribut binaire : vrai si la page est générée par LaTex2HTML, faux sinon

Tableau 2. *Attributs caractéristiques pour le calcul d'utilité [GFL & al, 1999]*

V- Evaluation

L'évaluation orientée vers l'utilisateur est une composante nécessaire dans la méthodologie d'évaluation d'un système d'accès à l'information, à plus forte raison dans le cas d'un modèle d'accès personnalisé ; cependant, elle semble encore peu formalisée pour être appliquée dans des campagnes d'évaluation effectives. En effet, s'il est indéniable que la meilleure façon d'évaluer un système interactif est de le confronter à un panel d'utilisateurs, la difficulté de réaliser un scénario d'évaluation objective reste un obstacle. Cette section aborde les principes de l'évaluation empirique, largement adoptée dans le domaine de la recherche d'information, puis met en évidence les obstacles posés par ce type d'évaluation dans le cas précis de systèmes d'accès personnalisé.

V.1 Evaluation empirique : principes [Chi, 2001]

L'évaluation empirique d'une théorie, procédé ou système désigne l'évaluation par l'observation expérimentale. La clé de la fiabilité d'une évaluation empirique est la rationalité de la méthodologie de conception et réalisation des expérimentations associées. Chaque expérimentation couvre généralement plusieurs scénarios d'évaluation qui ont pour objectif

de mesurer l'effet de l'introduction d'un facteur associé à ce qui est en cours d'évaluation. Ces facteurs, qui représentent des variables contrôlées par celui qui expérimente, peuvent être dépendantes ou indépendantes. Les variables dépendantes sont des variables qui dépendent d'autres variables. Dans une expérimentation idéale, seules les variables indépendantes sont à faire varier dans un domaine de valeurs précises. Les variables dépendantes sont maintenues fixes de manière à ce que toute variation dans les variables dépendantes soit directement attribuée à la variation des variables indépendantes. Ces expérimentations induisent deux types de validation : validation interne et validation externe. La validation interne « résume » les expérimentations en ce sens qu'elle permet d'évaluer les effets des variables sur les performances globales de ce qui est évalué et détecter éventuellement des relations de cause à effet. La validation externe a trait à la représentativité des résultats obtenus lors de la validation interne. En ce sens qu'une tentative de généralisation de ces résultats, selon des aspects à préciser, est effectuée de manière à élargir leur portée.

V.2 Evaluation des systèmes d'accès personnalisé à l'information

Les méthodes d'évaluation largement adoptées en recherche d'information, sont empiriques. Elles sont souvent basées sur une évaluation d'avantage quantitative que qualitative. En effet, les résultats obtenus sont issus de la comparaison de mesures et de métriques en termes de rappel[3] et précision[4], sur les réponses fournies par le système relativement à celles issues des réponses attendues qui constituent le référentiel. Ce type d'évaluation, est orienté vers une approche comparative de plusieurs systèmes reposant sur le principe d'évaluation des collections de test. Bien qu'adopté par des campagnes d'évaluation de référence en recherche d'information, tels que *TREC*[5], il est cependant contesté [BLP & al, 1996] notamment en raison de la non considération ni du contexte dans lequel se fait la recherche, ni de la perception de la pertinence des utilisateurs dans ce même contexte. L'introduction de la dimension utilisateur dans le processus d'accès à l'information, accentue d'avantage la difficulté d'évaluation. En effet, en raison du caractère subjectif des utilisateurs d'une part et du caractère dynamique ou adaptatif du système, d'autre part, il est difficile de fournir des valeurs absolues aux métriques. On cite dans ce qui suit les principaux problèmes liés à l'évaluation des systèmes personnalisés puis présentons quelques recommandations qui augmentent la fiabilité des résultats qui en sont issus.

V.2.1 Problèmes de l'évaluation

La rationalité des résultats issus d'un scénario d'évaluation d'un système personnalisé est compromise pour les principales raisons suivantes [Chi, 2001]:

- Si l'évaluation des variables indépendantes est effectuée par différents utilisateurs, alors des différences personnelles exogènes à l'expérimentation, telles que l'intelligence, la capacité de raisonnement, l'expérience etc.., ont un impact sur les valeurs des autres variables.

- Si le même utilisateur est impliqué dans des différents scénarios d'évaluation, alors son expérience passée avec le système peut influer sur sa perception de la pertinence.

[3] Le rappel désigne la proportion de documents pertinents sélectionnés par le système par rapport à l'ensemble des documents pertinents contenus dans la collection

[4] La précision désigne la proportion de documents pertinents sélectionnés par le système par rapport à l'ensemble des documents sélectionnés

[5] Text REtrieval Conference

- Les conditions de déroulement des expérimentations ont un impact non négligeable sur les résultats : configuration des machines (processus lent conduit à une lassitude de l'utilisateur), interfaces peu ergonomes, lieu inapproprié (bruit, exiguité, présence de caméras) etc…

- La validation interne est difficile à mettre en œuvre. En effet, en raison des problèmes évoqués ci-dessus, il n'est pas aisé de séparer l'effet intrinsèque des variables pertinentes du système indépendamment des facteurs exogènes. La validation externe est dès lors difficile également. Elle l'est d'autant plus que la généralisation des résultats obtenus lors de la validation interne doit considérer une combinaison des divers types d'utilisateurs, des variables d'expérimentation et des situations d'utilisation.

V.2.2 Recommandations

Une attention particulière aux protocoles d'expérimentation est accordée lors de l'évaluation de systèmes personnalisés de manière générale [Chi, 2001] et ceux dédiés à l'accès à l'information de manière particulière [BH 2000 ; BBB 2004 ; DCC & al, 2003] . Dans le but d'éviter des erreurs de mesures dus à l'aspect subjectif de l'évaluation, on cite dans ce qui suit quelques recommandations [Chi 2001] :

- définir un nombre suffisant de groupes d'utilisateurs avec des effectifs adéquats,
- isoler au mieux les utilisateurs,
- s'assurer de l'ergonomie des applications,
- préparer un canevas unique qui décrit le protocole d'expérimentation et l'adresser à l'ensemble des utilisateurs,
- effectuer des expérimentations pilotes avant de passer aux expérimentations effectives,
- les utilisateurs ne doivent pas être informés des facteurs à évaluer dans le système (ex avec et sans modèle utilisateur) ; leur appréciation doit porter sur des aspects perceptibles, liés à son fonctionnement,
- les variables exogènes (âge, expérience, aptitudes …) doivent être identifiées explicitement et leur influence mesurée pour être prise en compte dans le processus global d'évaluation

Hormis ces recommandations, force est de reconnaître que les méthodes d'évaluation des systèmes personnalisées restent peu formelles, laissant de nombreuses questions ouvertes sur la méthodologie de construction des collections de test et définition de scénarios d'évaluation objective. Ceci explique finalement l'impossibilité de comparer actuellement les performances des systèmes d'accès personnalisé [PRE 1999].

VI- Conclusion

Au terme de ce rapport, nous dressons un bilan des acquis dans le domaine de la personnalisation de l'accès à l'information puis dégageons quelques questions pendantes qui peuvent être abordées lors de nos recherches futures.

VI.1 Bilan

Notre synthèse sur l'accès personnalisé à l'information fait ressortir d'emblée que les approches et techniques associées puisent largement des acquis reconnus dans les domaines de la recherche d'information d'une part et modélisation de l'utilisateur d'autre part. La théorie autour de la recherche d'information permet d'aborder les problèmes liés à la représentativité des unités d'informations, de leur organisation et de leur accès. La théorie autour de la modélisation utilisateur permet d'aborder cette même problématique d'accès sous l'angle d'un utilisateur spécifique décrit par un ensemble de facettes qui représentent son contexte, ses préférences, son besoin, en clair, son rapport avec une activité de recherche d'information. Cette théorie propose alors des solutions pour mieux décrire l'utilisateur selon des aspects cognitifs et comportementaux. L'association de ces acquis est motivée principalement par le volume important d'informations disséminées et hétérogènes sur le Web. Ceci a en effet, engendré une dégradation des performances des systèmes de recherche d'informations en termes d'efficience et d'efficacité. Du point de vue de l'utilisateur, ceci se traduit par le volume et l'inintelligibilité des informations retournées, sa désorientation et démotivation face à des interfaces qui induisent une surcharge cognitive. La personnalisation est dès lors perçue comme une solution à ces problèmes. C'est une direction de recherches qui a pour objectif dominant d'adapter le cycle de vie d'un processus d'accès à l'information, aux caractères spécifiques d'un utilisateur en vue de lui délivrer une information appropriée. Les applications sont diverses : systèmes de recommandation, systèmes de filtrage de messages et d'informations, systèmes d'apprentissage, moteurs de recherche d'information etc…Les approches proposées se scindent principalement en deux : interactives ou automatiques. Les approches interactives ne passent pas l'échelle. Les approches automatiques, largement en vogue actuellement, sont basées sur des modèles statistiques qui tentent d'inférer implicitement une connaissance sur l'utilisateur. Les questions fondamentales abordées sont alors les suivantes :

- *Quelles informations pour représenter l'utilisateur ?*
  La prise en compte des dimensions liées au contexte de l'utilisateur intégrant l'application en cours, fait ressortir de nombreuses sources d'informations : l'utilisateur lui-même, les documents qu'il consulte, les liens qu'il explore, les pages qu'il sélectionne favorites, les applications qu'il utilise, les pages visitées etc…

- *Quel modèle pour l'utilisateur ?*
  Une question préliminaire à celle-ci est sans doute: est ce que l'utilisateur est explicitement modélisé ou assimilé aux informations qui le décrivent ? c'est plutôt la seconde alternative qui est largement adoptée actuellement. Peu voire pas de modèles globaux spécifiques à l'utilisateur. La modélisation se résume à la structuration des paquets d'information qui décrivent l'utilisateur : classes de termes, réseau connexionniste ou bayésien de préférences, ontologies, vecteurs de documents préférés ou jugés pertinents. La description d'un modèle unificateur de l'utilisateur n'est pas le souci fondamental tant les informations descriptives sont de sources, natures et degrés de fiabilité diverses.

- *Comment adapter le cycle de vie du processus d'accès ?*
  L'adaptation est effectuée principalement à l'un des différents niveaux : sélection des sources d'information, reformulation de requête, sélection de l'information et filtrage des résultats. La sélection personnalisée des sources d'informations est opérée par les méta-moteurs de recherche. La reformulation de requête a pour objectif d'introduire dans la structure de la requête les termes issus du profil de l'utilisateur ; c'est la technique la plus largement répandue. La sélection adaptée de l'information, promue par de récents travaux, évoque la contextualisation de la fonction de pertinence en définissant des paramètres issus du profil de l'utilisateur. Enfin, le filtrage des résultats traduit la prise en compte des préférences de l'utilisateur à l'étape précédant la présentation des résultats et suivant leur sélection. L'adaptation consiste généralement à réordonnancer les résultats en tenant compte de critères descriptifs de l'utilisateur.

La diversité et quantité des travaux actuels autour de la personnalisation, traduisant un engouement dans le domaine, montrent, selon notre point de vue, deux aspects. Le premier, aspect de forme, porte sur la longue portée de la personnalisation quant à la viabilité et devenir du Web en tant que source d'informations privilégiée. Le second, aspect de fond, porte sur la difficulté liée au dilemme de formaliser au mieux le problème de l'accès à l'information en introduisant de nombreux paramètres subjectifs liés à l'utilisateur. Le dilemme est d'autant plus important que la connaissance utile se rapportant à ce dernier est peu cernée.

VI.2 Questions ouvertes

L'état actuel des travaux dans le domaine de l'accès personnalisé à l'information laisse, malgré une avancée notable des réflexions pour une problématique aussi récente (début des années 1990), quelques questions pendantes. La première question a trait incontestablement à l'évaluation. Il n'y a en effet aucune méthodologie d'évaluation reconnue et partagée dans la communauté. Les évaluations sont effectuées à l'aide de collections locales, de surcroît, en l'absence d'un protocole précis. Ceci diminue la fiabilité des systèmes qui sont proposés et ne permet pas d'envisager des campagnes d'évaluation effectives pour la comparaison entre systèmes. La seconde question porte sur la modélisation proprement dit de l'utilisateur. Les systèmes développés actuellement focalisent sur des aspects parcellaires de l'utilisateur qui sont exploités, selon leur nature, dans une ou plusieurs phases du processus. Aucun modèle n'est proposé pour traduire une connaissance globale de l'utilisateur ; la description formelle du modèle de recherche est par conséquent inchangée, hormis l'introduction de techniques qui traduisent des procédés d'inférence d'informations. La troisième question concerne l'adaptation à long terme de ces systèmes. La plupart des travaux ont focalisé en effet sur la personnalisation contextuelle relativement à une session d'utilisation du système. La personnalisation dite persistante est une direction qui fera surgir d'autres problèmes, tels que la dérive des concepts, abordé notamment en filtrage automatique d'informations. La combinaison des deux types personnalisation dans une même infrastructure élargira sans doute le champ de la problématique de la personnalisation.

REFERENCES

[AFJ & al, 1995] Armstrong R., Freitag D., Joachims T., Mitchell T., WebWatcher: A learning apprentie for the World Wide Web, AAAI Spring symposium on Information Gathering from Heterogeneous, distributed environments, Standford, 1995

[All 1990] R. Allen, User models, theory, methods and practice. INT. J. Man Machine Stud, 1990.

[All, 1979] Allen J.F, A plan based approcah to speech act recognition. Technical report, 131/79, Dept. of computer science, University of Toronto, Canada, 1979

[Ama 1999] G. Amato, U. Straccia, User profile modeling and applications to digital libraries, In Proceedings of the 3rd European Conference on Research and advanced technology for Digita Libraries, ECDL, 1999

[BAH & al, 2000] P. de Bra, A. Aerts, G. Houben, H. WU. Making Genral Purpose adaptive hypermedia work. WebNet 2000

[BBB 2004] J.C Bottraud, G. Bisson, M.F. Bruandet, Expansion de requêtes par apprentissage automatique dans un assistant pour la recherche d'information, Actes de la première conférence francophone en recherche d'information et applications, 89 :108, Toulouse mars 2004

[BDJ, 1999] M. Berry, Z. Darmac, E. Jessup, Matrices, vector spaces and information retrieval, SIAM, 41(2), 335:362, 2002

[Bes 1996] Bestarovs A., Speculative data dissemination and service to reduce server load, network traffic and service time in distributed information systems, In Proceedings of the 1996 International Conference on data Engineering,

[BGG & al, 1998] Boley D., Gini M., Gross R., Han E.H., Hastings K., Karypis G., Kumar V., Mobasher B., Moore J., Document categorization and query generation on the Wordld Wide Web using WebAce, 1998

[BH, 2000] J. Budzik, K.J Hammond, Users interactions with everyday applications as context for just-in-time information access, Proceedings of the 5th international conference on intelligent user interfaces, 44-51, 2000

[BLP & al, 1996] Balpe J.P, Lelu A, Papy F, Saleh, Techniques avancées pour l'hypertexte. Hermes (Eds) 1996

[BM, 2002] Brusilovsky, P., Marbury, M.T, From adaptive hypermedia to adaptive web. In P. Brusilovsky and M. T Marbury (Eds), Communication of the ACM 45(5), Special issue on the adaptive Web, 31:33, 2002

[Bou 2004] M. Bouzeghoub, Personnalisation de l'information, AS98 / RTP9, 2004

[BP 1999] Billsus, D., Pazzani, M., A hybrid user model for news stories classification, In UM99, Proceedings of the seventh International Conference on User Modelins, Banff, Canada, 99:108, 1999

[CCL, 2001] W. B. Croft, S. Cronen Towsend, V. Lavrenko, Relevance feedback and personalization: a language modeling perspective, Workshop Personnalization and recommander systems in digital librairies, 2001

[Chi, 2001] David N. Chin, Empirical evaluation of user models and user adapted systems, User modeling and user-adapted interaction 11 181:194, 2001, Kluwer Academic Publishers, 2001

[CMB &al, 2004] J.P Chevallet, J. Martinez, M. Boughanem, L. Lechani-Tamine, S. Calabretto, Rapport final de l'AS Passage à l'échelle dans la taille des corpus, janvier 2004

[CP, 1979] Cohen P.R, Perrault, C.R : Elements of a plan based theory of speech acts, Cognitive science 3, 177:212, 1979

[Cro 1993] W.B. Croft, Knowledge_based and statistical approaches to text retrieval. IEEE Expert, 8(2), 1993
[CS, 1998] L. Chen, K. Sycara, WebMate, a personal agent for searching and browsing, Proceedings of the 2nd International Conference on autonomous agents, 1998
[CW, 1997] Cheng, I., Wilensky R., An experiment in enchaancing information access by natural language processinf, Technical report CSD-97-963, Computer science division, University of California, Berkeley, 1997
[CZC 2001] Chau M., Zeng, D., Chen, H. , Personalised spiders for Web search and analysis, In Proceedings of ACM/IEEE Joint Conference on Digital Librairies, 2001
citesser.nj.nec.com/article/croft01relevance.html
[Dan, 1986] J.P Daniels, Cognitive models in information retrieval- An evaluation review, Journal of documentation, 42(4), 272:304, 1986
[DCC &al, 2003] S. Dumais, Edward Cuttrell, J.J Cadiz, G. Jancke, Raman Sarin, Daniel C. Robbins, Stuff I've seen : a system for a personal information retrieval and re-use, In Proceedings of the 26th ACM SIGIR , pp 72-79, Toronto July 2003
[DHF, 1999] Dean J., Henzinger, M.R, Finding related pages in the World Wide Web, In Proceedings of WWW-8, the Eight International World Wide Web Conference, Fortec Seminars, 1999
[DK 2001] Deshpande M., Karypis G., Selective Markov models for predicting Web-page accesses, In 1st SIAM International Conference on Data Mining, 2001
[Dum, 1993] S. Dumais, LSI meets TREC: A status report, In Proceedings of the 1st Text Retrieval Conference (TREC-1), 137:152, NIST Special publication, March 1993
[FBH 2000] Fu, X., Budzik, J., Hammond K.J., Mining navigation history for recommenadation, In Proceedings of the 2000 Conference on Intelligent User Interfaces, 2000
[FGP 2004] W. Fan, M.D. Gordon, P. Pathak, Discovery of context specific ranking functions for effective information retrieval using genetic programming, IEEE Transactions on knoledge and data engineering, Volume 16, issue 4, pp 523-527, 2004
[Fuh, 2000] N. Fuhr, Information retrieval : introduction and survey, Post-Graduate course on information retrieval, University of Duisburg-Essen, Germany, 2000
[GCP 2003] S. Gauch, J. Chaffe, A. Pretschner, Ontology based user profiles for search and browing, In User modeling and user adapted systems, Special issue on user modeling for Web and Hypermedia information retrieval, 2003
[GFL & al, 1999] Glover E., Flake, G.W., Lawrence, S., Birmingham, W.P., Kruger, A., Giles, C.L., Pennock, D.M., Improving category specific web search by learning query modifications, In Proceedings of Symposium on Applications and the Internet, 1999
[GLG & al, 1999 a] Glover E.J., Lawrence, S., Gordon, M.D., Birmingham, William, P., Birmingham, C., Lee Giles, C., Recommending Web documents based on user preferences, In proceedings of the Workshop SIGIR on recommender systems, August 1999
[Gow, 2003] J.P Mc Gowan : A multiple model approach to personalised information access, Thesis of Master in computer science, Faculty of science, university College Dublin, Fenrurary 2003
[Gre, 1984] S. Greenberg, User modeling in interactive computer systems. M.Sc thesis, Dpt of computer science, University of Calgary, Calgary, Report 85/193/6, 1984
[GSK & al, 1999] Good, N, Schafer, J., Konstan, J., Borchers, J., Sarwar, B. Herlocker, J. Riedl, J., Combining collaborative filtering with personal agents for better recommandations, In Proceedings of the 1999 Conference of the American Association of Artificial Intelligence, 439:446, 1999
[HAN, 2000] Han, J., Kamber, M., Data Mining: concepts et techniques, Morgan Kauffmann Publishers,2000

[Han, 1998] E.H. Han, WebAce, A Web agent for document categorization and exploration, Proceedings of the 2nd International Conference on Autonomous agents, 1998

[HBH & al, 1998] Horvitz E., Breese, J., Heckerman, D. Hovel D., Rommelse, K., The lumiere project: Bayesian user modeling for inferring the goals and needs of software users, In Proceedings of the Fourteenth conference on uncertainty in artificial intelligence, Madison, Wisconsin, 256:265, 1998
[Jam, 1995] Jameson, A., Numerical uncertainty management in user and student modeling: an overview of systems and issues, user modeling and user adapted interaction 5(3-4), 193:251, 1995
[JH 1993] Jenningd, A., Higuchi, H., A user model neural network for a personal news service. User modeling and user adapted interaction, 3(1), 1:25, 1993
[Joh, 2002] P. Johansson, User modelling in dialogue systems. St Anna Report SAR 02-2, 2002
[JSB, 1998] Jansen, B. Spink, A. Bateman J. , Searchers, the subjects they serch and sufficiency: astudy of a large sample of EXCITE serches. In Proceedings of Web-Net 1998, World conference of the WWW, Internet and Intranet. AACE Press, 1998
[JW 2003] Jeh, G., Widom, J. : Scaling personalized Web search, In Proceedings of the 12th International World Wide Web Conference (2003)
[Kim, 2003] Hyonug-rae Kim, Web personnalization,
[Kob, 2001] A. Kobsa, Generic user modeling systems, User modeling and user adapted interaction, 11 49:63, Kluwer Academic Publishers, 2001
[Kuh, 1991] *Hypertext*. Ein nicht-lineares Medium zwischen Buch und Wissensbank. Berlin: Springer.
[Kos 2003] Kostadinov D., La Personnalisation de l'information. Définition de modèle de profil utilisateur, Rapport de DEA, Université de Versailles, France, 2003
[LG 1999] Lawrence, S., Giles, C.L. : Accessibility of Information on the Web. Nature Vol 400 (1999), 107:109, 1999
[Li, 2000] X. Li, Adaptive personalized information retrieval, a thought analysis and utilization of user preferences in IR, 2000
[Lie, 1995] H. Lieberman, Letizia, Un agent that assists web browsing, Proceedings of the Internation Joint Conference on Artificial Intelligence (IJCAI'95), 924:929, Montreal, August 1995
[LYM 2002] Liu, F., YU, C.T., Meng, W. : Personalized Web search by mapping user queries to categories. In Proceedings of CIKM 2002
[Mcc 02] C. McCathieneville : Web sémantique : philosophie et standards, 2002. disponible sur http://www.w3.org/2002/talks/1213-lp/
[Orw, 1995] Orwan J., Heterogenous learning in the Doppelgänger user modeling system, User modeling and user adapted interaction, 2), 107:130, 1995
[PCG, 2003] D. Poo, B. Cheng, J.M. Goh, A hybrid approach for use profiling, 36th annual Hawaï International Conference on System Sciences (HICSS'03)- Track 4, January 2003
[Per, 1988] Peral J., Probabilistic reasoning in intelligent systems, San Mateo, California: Morgan Kaufmann Publishers
[PE 2000] Perkowitz, M., Etzioni, O., Towards adaptive Web sites: conceptual framework and case study. Artificial intelligence 118(1-2), 245:275, 2000
[PG, 1999] A. Pretschner, S. Gauch, Personalization on the Web, Technical report ITTC-FY2000-TR-13591-01, Information and telecommunication technology center, Deprtement of electrical engineering and computer science, University of Kansas, December 1999

[PMB & al, 1996] Pazzani, M., Muramatsu J., Billsus, D., Syskill & Webert : Identifying interesting Web sites, In Proceedings of the Fourteenth national conference on Artificial intelligence, AAAI Press, 1996
[PP 1999] Pitkow J., Pirolli, P., Mining longest repeating subsequences to predict world wide web surfing, In Proceedings of the 2nd USENIX Symposium on Internet Technologies and Systems, 1999
[RBW, 1997]Raskutti, B., Beitz A. and Ward B., A feature based approach to recommending selections based on past preferences. User Modeling and User Adapted Interactions 7(3), 179:218, 1997
[RIW, 1994] Resnick, P., Iacovou, N. And Ward, B., An open architecture for collaborative filtering of netnews, In CSCW'94 Proceedings of the Conference on Computer Supported Collaborative Work, 175:186, 1994
[RK, 1976] Ralph L., Keeney, Howard Raiffa, Decisions with multiple objectives, John Wiley and sons, New York, 1976
[Roc, 1971] J. Rocchio , Relevance feedback in information retrieval al, In G. Salton editor, The SMART retrieval system – experiments in automated document processing. Prentice-Hall, Englewood Cliffs, NJ, 1971
[Sal 1971] G. Salton, The SMART retrieval system: experiments in automatic document processing, Prentice Hall Inc, NJ., 1971
[SB, 1990] Salton G., Buckley C., Improving retrieval performance by relevance feedback. In Sparck Jones and Willet, (Eds) Readings in information retrieval, San Francisco, CA, Morgan Kauffman, 1990
[SBJ, 1998] Spink, A., Bateman, J., Jansen, B., User's searching behavior on the EXCITE web search engine. In Proceedings of Web-Net 1998, World conference of the WWW, Internet and Intranet. AACE Press, 1998
[SC 2003] C. Shahabi, Y.S Chen, Web information personalization: Challenges and approaches, Databases in networked information systems, 3rd International Workshop, 2822 5:15, Japan September 2003
[SJW &al, 2002] A. Spink, B. Jansen, D. Wolfram, T. Saracevic. From E-sex to E-commerce : Web serche changes. IEEE Computer 35(3): 107-111, 2002
[SKW 2000] Scime, A., Kershberg, L., WebSifter: An Ontology-based personalizable search agent for the Web, In Proceedings of International Conference on Digital Librairies: Research and Practice (2000)
[SL, 2003] J. Su, M. Lee, An exploration in personalized and context-sensitive serach, In Proceedings of SIGIR Conference, In proceedings of the 7th annual CLUK (the UK special interest group for Computational Linguists) research colloquium, 2003
[SZ 2000] Seo Y. W., Zhag B.T., A reinforcement learning agent for personalized information filtering, In Proceedings of the 2000 International Conference on Intelligent User Interfaces, New-Orleans, USA, January 9-12, 248:251, 2000
[Vis 2004] Vishnu Kanth Reddy Challam, Contextual information retrieval using ontology based user profiles, Master's thesis, January 2004
[WM, 1999] T. B. I. Witten, A. Moffat, Managing gigabytes, compressing and indexing documents and images, Morgan Kaufmann publishers, second edition, Butterworths, 1999
[YN, 1999] R. B Yates R. Neto, Modern information retrieval. ACM Press, Addison Wesley, 1999
[ZA, 2001] I. Zuckerman, D.W. Albrecht, Predictive statistical models for user modeling, User modeling and user adapted interaction, 11 5:18, 2001
[ZAN 1999] I. Zukerman, D.W. Albrecht, A.E. Nicholson, Predicting users' requests on the WWW, In UM99 Proceedings of the seventh International conference on user modeling, Banff, Canada, 275:284, 1999

[ZDC & al 2001] Zhu, S., Deng, X., Chen, K., Zheng, W., Using online relevance feedback to build effective personalised metasearch engine, In Proceedings of 2nd International Conference on Web Information Systems Engineering, 2001