

وزارة التعليم العالي و البحث العلمي

BADJI MOKHTAR-ANNABA UNIVERSITY
UNIVERSITE BADJI MOKHTAR-ANNABA



Faculté des sciences de l'ingénieur

Année : 2012

Département d'informatique

MEMOIRE

Présenté en vue de l'obtention du diplôme de **MAGISTER**

**Découverte des préférences des apprenants dans
un environnement de e-learning**

Option

TIC & Ingénierie du document

Par

Amor BOULEDROUA

DIRECTEUR DE MEMOIRE : Hassina SERIDI Docteur Université de Annaba

DEVANT LE JURY

PRESIDENT: Pr Khadir Mohamed Tarek Pr Université Badji Mokhtar Annaba

EXAMINATEURS: Dr Bahi Halima MCA Université Badji Mokhtar Annaba

Dr Yacine Lafifi MCA Université de Guelma

يهدف إ تسهيل وصول المستخدمين إلى البيانات

المعلومات مفهومة و

ين التفضيلات الفردية التي يمثلها

(القيم) هذه الأنماط من التركيز على

، ونوعية البيانات التي تريد أو كيفية تقديم البيانات. يتم تمثيل كل هذه

الهدف من هذه الأطروحة هو اقتراح

، باستخدام هذه

تطبيق التحديث إلى ملف تعريف المستخدم

كلمات مفاتيح: تحليل الوثائق، البحث عن الوثائق،

ABSTRACT

The goal of information retrieval is to facilitate accessing data by users in huge information databases. The information relevance can be defined by a set of features and preferences specific to each user or community of users. These features should describe the user's interest domains, the quality of the retrieved data and the modalities of data presentation. The data describing the users is often presented in the form of profiles. We use to propose in this thesis a question answering system to help learners answering their questions, based on these questions the system updates automatically the user profile information.

Keywords: documents analysis, QA System, information retrieval, user profile, preferences.

RESUME :

La recherche d'information a pour objectif de faciliter la disponibilité des données pour l'utilisateur et de rendre l'information sélectionnée facilement accessible à l'utilisateur et exploitable. Chaque utilisateur, ou communauté d'utilisateur, peut être défini, par un ensemble de préférences individuelles représentées sous forme de couples (attribut, valeur), par des ordonnancements de critères ou par des règles sémantiques spécifiques à chaque utilisateur. Ces modes de spécification servent à décrire le centre d'intérêt de l'utilisateur, le niveau de qualité des données qu'il désire ou des modalités de présentation de ces données. L'ensemble de ces informations est représenté dans un modèle d'utilisateur appelé souvent profil. L'objectif de cette thèse est de proposer un système de question réponse qui permet de répondre aux questions posées par l'apprenant. Tout en assurant la mise à jour du profil utilisateur à partir de ces questions appliquées une mise à jour au profil de l'utilisateur.

Mots clés : analyse de documents, système QA, recherche d'information, profil utilisateur, préférence.

REMERCIEMENT

Je souhaite remercier très particulièrement mon directeur de thèse Madame Hassina Seridi. Je la remercie de m'avoir fait confiance et de m'avoir fourni les meilleures conditions de travail possibles. Les discussions et le temps qu'elle m'a accordés tout au long de ma thèse m'ont permis d'avancer dans mon travail et ont contribué à améliorer mes connaissances. Ses remarques ont toujours été claires et précises ce qui a facilité le processus d'apprentissage du métier. Ses conseils aussi bien sur le plan professionnel que personnel m'ont toujours été d'une grande aide. Pour tout ce que vous avez fait pour moi.

Je voudrais également remercier mon frère Ahmed Boulemden avec qui nous avons beaucoup collaboré durant ma thèse. Les discussions que nous avons eues m'ont permis d'avancer plus rapidement dans mon travail.

merci du fond du cœur

Je souhaite exprimer toute ma gratitude envers ma famille, mon père, ma mère et ma sœur ainsi que mes frères.

Et merci pour tous qui m'a aidé de près ou de loin.

LISTE DES TABLEAUX

Tableau 1 : Comparaison entre le E-learning et l'apprentissage traditionnel	9
Tableau 2 : La catégorisation du comportement utilisateur selon	42
Tableau 3 : Synthèse des techniques d'acquisition implicite	45
Tableau 4 :Quelques EIAH qui implémentent les styles d'apprentissage.....	59
Tableau 5 : Résumé des études menées sur la relation entre les styles dépendants et indépendants du domaine et le comportement des utilisateurs avec des applications Web.....	62
Tableau 6 :Tableau de probabilité des positions des réponses en sein des documents.	81

LISTE DES FIGURES

Figure 1 La plateforme e-learning.....	15
Figure 2 Un exemple de profil représenté par des mots clés.....	34
Figure 3 Un extrait d'un profil utilisateur sémantique.....	36
Figure 4 Représentation multidimensionnelle du profil utilisateur	38
Figure 5 Schéma général du Système Q/A Proposé.....	67
Figure 6 Schéma général de la partie de mise à jour du profil Utilisateur	78
Figure 7 Exemple de Structuration d'un cours	79
Figure 8 Exemple indexé de Structuration d'un cours.....	80
Figure 9 déroulement classique d'une requête vers le serveur d'applications Tomcat...86	
Figure 10 Page d'accueil de SAKAI.....	88
Figure 11 Interface principale de l'apprenant.....	91

TABLE DES MATIERES

.....	i
Abstract.....	ii
Résumé	iii
Remerciement	iv
Liste des tableaux.....	v
Liste des figures.....	vi
Table des matières	viii
Introduction générale	1
Contexte et problématique.....	2
Contribution.....	3
Organisation du mémoire.....	3
Chapitre 01 : E-learning et le système de question réponse.....	5
Introduction	6
1. E-learning.....	6
1.1 histoire et évolution du e-learning.....	7
1.2 définition du e-learning.....	8
1.3 acteurs du e-learning.....	9
1.4 développement rapide de e-learning.....	11
1.5 les principaux bénéfices de e-learning.....	11
1.6 inconvénient du e-learning.....	12
1.7 les plateformes d'apprentissage.....	13
1.8 les enjeux de la plateforme e-learning.....	14
1.9 caractéristiques des plateformes e-learning.....	15
1.10 rôle des principaux utilisateurs	15
2. Système QA	16
2.1 histoire des systèmes de Q/A.....	17
2.2 fonctionnement des systèmes Q/A.....	19

3. Conclusion	22
Chapitre 02 : Le profil utilisateur et les style d'apprentissage.....	23
Introduction	24
1. Profil utilisateur.....	25
1.1 définition de la notion du « profil utilisateur ».....	25
1.2 Les préférences.....	26
1.3 Contenu du profil utilisateur	27
1.4 La modélisation de l'utilisateur.....	30
1.5 Classification du contenu des profil utilisateur.....	32
1.6 Approches de représentation du profil utilisateur.....	33
1.7 Approche de construction du profil utilisateur.....	38
1.8 Synthèse des techniques implicites.....	44
1.9 Indicateurs pour la prédiction des centres d'intérêts.....	45
1.10 Gestion du profil.....	46
1.11 Les inconvénients de quelques système de gestion du profil	49
2. Style d'apprentissage.....	50
2.1 Définition.....	50
2.2 Quelques modelé de style d'apprentissage	50
2.3 Style d'apprentissage enseignement.....	56
2.4 Style d'apprentissage et EIAH.....	58
2.5 Relation entre style d'apprentissage et comportement sur le web.....	61
Chapitre 03 : Conception.....	66
Introduction	68
1. Architecture du système proposé.....	68
2. Phase recherche (Système QA)	69
Indexation	71
Test FAQ.....	75
Analyse et reformulation.....	76
Appariement.....	77
Scénario Pédagogique	78
2 Phase mise a jours profil	81
Profil Utilisateur.....	81
Procédure de mise a jour	83
Capture du type de question	83
3 Conclusion	84

Chapitre 04 : Implémentation.....	85
Introduction	86
1. Outils de developpement	87
Java	87
JSP et Servlet.....	88
Caractéristiques des java server pages.....	88
Eclipse.....	89
Tomcat.....	89
Mysql	90
2. Installation de Sakai.....	91
Configuration de la machine pour Sakai.....	91
Configuration de Sakai.....	92
Lancement de Sakai.....	94
3. Environnement d'exécution.....	91
Bibliographie.....	101

INTRODUCTION GENERALE

1. Contexte et problématique

Les systèmes d'information actuels donnent accès à un grand nombre de sources hétérogènes et distribuées. Au fur et à mesure que les sources se multiplient et que le volume de données disponibles s'accroît, l'utilisateur se voit confronté à une surcharge informationnelle dans laquelle il est difficile de distinguer l'information pertinente de l'informations secondaire et même du bruit. En outre, l'évaluation d'une requête se fait généralement sans tenir compte du contexte et/ou des besoins spécifiques de l'utilisateur qui l'a émise. La même requête, faite par deux utilisateurs différents, produit les mêmes résultats même si ces utilisateurs n'ont pas les mêmes attentes.

Les systèmes de recherche d'information sont des outils qui ont permis, jusqu'à aujourd'hui d'améliorer la qualité des services d'accès à l'information. Au fur et à mesure que le volume de données s'accroît et les informations se diversifient, ces systèmes (système QA, moteur web, SGBD...) délivrent des résultats massifs, générant ainsi une surcharge informationnelle en réponse aux requêtes des utilisateurs dans laquelle il est difficile de distinguer l'information pertinente d'une information secondaire. Donc le problème n'est pas dans la disponibilité de l'information mais dans sa pertinence relative aux besoins précis de l'utilisateur, c'est pourquoi les travaux s'orientent actuellement vers la révision de cycle de vie d'une requête dans la perspective d'intégrer l'utilisateur comme composante du modèle globale de recherche et ce dans le but de lui délivrer une information pertinente adaptée à ses besoins et préférences précis.

2. Contribution

Dans ce travail, nous nous intéressons à la recherche d'information ainsi que la gestion du profil utilisateur. Nous nous intéressons principalement dans ce projet au fait de mise à jour du profil utilisateur. Il existe des multiples de façon et de critères qu'ils peuvent agir sur le profil utilisateurs. Nous proposons un nouvel indicateur au sein de cette thèse qui est les questions posées par l'apprenant ainsi que nous montrerons la façon d'implémenter un système de Question Réponse propre à nous au sein d'une plateforme E-learning (SAKAI).

3. Organisation du mémoire

Ce mémoire est organisé autour de quatre chapitres, dans ce qui suit nous donnons une brève description de leurs contenus respectifs.

Chapitre 1 : E-learning et le système Q/A

Dans la première partie de ce chapitre, nous allons définir la notion du E-Learning, citer les différents acteurs et intervenants et les principales plateformes d'apprentissage existantes.

Dans la deuxième partie, nous aborderons quelques notions sur les systèmes de question réponse (Q/A), en montrerons leurs acheminements de développements ainsi que le fonctionnement d'un système de Q/A générale.

Chapitre 2 : Profil Utilisateur

Le but de ce chapitre est de fournir un état de l'art sur le profil d'utilisateur, la gestion de profil et l'exploitation du profil pour l'extraction des données. La section 2 est divisée en plusieurs parties, on commence par donner une définition du profil utilisateur et des préférences, puis on montre quelques modèles utilisateurs. Après on passe à la gestion du profil partant de la construction et allant vers la mise à jour du profil. On finit la section en citant quelques inconvénients des systèmes de gestion de profil. Enfin, la section 3 conclut le chapitre.

Chapitre 3 : Conception

Afin d'améliorer la performance de notre Système Q/A, Nous avons injecté un module LSA qui permet d'indexer et faire l'appariement entre la requête et la table d'indexe générer à partir de la base de documents.

Afin de réaliser notre système nous procédons en deux phases : une phase de recherche et détection de la réponse de la question posé par l'apprenant ; et une phase de mise ajours des préférences du profil de l'utilisateur (Apprenant).

Chapitre 4 : Implémentation

Ce chapitre est organisé de la manière suivante : dans la section 2 nous illustrons l'environnement expérimental de notre système et le contexte des expérimentations, ensuite dans la section 3 nous présentons la base de documents utilisées pour évaluer la partie de recherche. Dans la section 4, nous décrivons les différentes expérimentations, les résultats obtenus des expérimentations, avant de conclure.

Chapitre

1

E-Learning

Et

Le Système Q/A



Introduction

Les modalités d'apprentissage évoluent. Cette évolution correspond à un besoin induit par la transformation de nos sociétés et à une attente nouvelle de nos concitoyens. Elle s'appuie, par ailleurs, sur de nouveaux outils d'apprentissage. Depuis plusieurs années, les technologies de l'information et de la communication (TIC) constituent non seulement un nouvel outil, un nouveau média, mais aussi un moyen d'ouverture sur des ressources du monde entier.

L'application des Technologies de l'Information et de la Communication pour l'Enseignement (TICE) a donné naissance à une nouvelle forme d'apprentissage appelée E-learning. Le E-Learning est la solution qui a permis actuellement l'admission des TIC dans la formation à distance. Il s'agit d'une transformation rapide des technologies pour l'apprentissage, rendue possible par le développement de l'Internet. L'introduction de ces TICE vise à améliorer la qualité de la formation en facilitant l'accès aux ressources et aux services du web d'une part; et la collaboration à distance d'autre part.

Dans la première partie de ce chapitre, nous allons définir la notion du E-Learning, citer les différents acteurs et intervenants et les principales plateformes d'apprentissage existantes.

Dans la deuxième partie, nous aborderons quelques notions sur les systèmes de question réponse (Q/A), en montrerons leurs acheminements de développements ainsi que le fonctionnement d'un système de Q/A générale.

1 Le e-Learning

Avec l'avènement des TIC au domaine de la formation, une nouvelle réalité a été créée: il s'agit du l'E-Learning. Ainsi des changements remarquables sont apparus. On retrouve de plus en plus de cours destinés à un enseignement ouvert et en ligne, autrement dit, enseignement à distance (EAD) ou encore le E-Learning.

Le E-Learning n'est pas uniquement de l'enseignement à distance par Internet, c'est un mode de formation en ligne qui accompagne souvent une formation avec formateur. Son objectif est surtout de créer un environnement plus attractif et plus interactif.

1.1 Histoire et évolution du e-learning :

Une longue histoire a précédé ce qu'on regroupe aujourd'hui sous la notion du « E-learning ». Il s'agit de la dernière forme de l'Enseignement à Distance (EAD). L'enseignement à distance ou l'apprentissage à distance se compose de techniques et méthodes permettant l'accès aux programmes éducatifs pour les étudiants qui sont séparés par le temps et l'espace. Les systèmes de e-Learning souffrent du manque de la relation élève-enseignant (une à une).

Il existe plusieurs moyens pour assurer l'enseignement à distance: la correspondance sur papier, des cassettes vidéo éducatives, éducation par ordinateur (enseignement multimédia, utilisation d'Internet pour l'éducation sur le Web), etc. Où trois phases de développement sont distinguées (qui sont tous des formes de développement de l'enseignement traditionnel):

1) Première phase : Connue sous le nom d'enseignement par correspondance; l'enseignement dans cette phase est basé sur la diffusion du contenu de cours élaboré par l'enseignant suivant sa propre logique qui s'impose à tous les apprenants. Le rythme de l'enseignement et lui aussi imposé par l'organisation de la formation.

2) Seconde phase : Caractérisé par le développement de l'enseignement assisté par ordinateur (EAO), basée sur l'approche « behavioriste » ; qui a cherché à se dégager de cette programmation linéaire et uniforme en proposant des parcours différents pour les élèves en fonction des résultats de tests.

3) Troisième phase : Fondé sur l'approche constructiviste (le savoir ou la connaissance est construite par l'apprenant). Cette phase se caractérise par l'introduction d'une plus grande variété de technologies dont notamment le développement d'Internet. C'est une mixture entre l'enseignement à distance et l'enseignement présentiel (traditionnelle) connue sous le nom « blended learning ». Il essaye de cumuler les avantages des deux formules.

Avec le e-learning c'est différent, il y'a une personnalisation des parcours pédagogiques en fonction des résultats de tests et des conseils du tuteur.

1.2 Définition du e-learning :

Aujourd'hui, il y'a beaucoup de termes utilisés pour désigner l'éducation basée sur le Web comme l'E-learning, E-formation, l'enseignement en ligne, enseignement basé web (web-based learning), apprentissage basé web (web-based training),...etc.

A l'heure actuelle, la notion de E-Learning renferme des concepts et des technologies qui ne sont pas figés, sa définition reste alors assez approximative. Nous retenons la définition proposée par le Conseil Européen (2001), qui considère que le e-learning est : « un ensemble de concepts, de méthodes, et d'outils utilisant les nouvelles technologies multimédias et de l'Internet, pour améliorer la qualité de l'apprentissage en favorisant l'accès à des ressources et des services, ainsi que les échanges et la collaboration éventuellement à distance ». [CON 01]

Le e-learning est donc née pour permettre aux apprenants de se former sans se déplacer dans un lieu de formation, et sans s'inquiéter du temps de début ou de fin de formation puisque le formateur ne sera pas présent physiquement.

Pendant, dans ce nouveau mode de transmission et d'acquisition du savoir, concernant aussi bien des formations académiques que professionnelles, la relation directe enseignant/apprenant (face à face) est remplacée par une relation médiatisée par un support et un ensemble de techniques.

Remarque

Plusieurs termes sont utilisés pour traduire le terme **e-learning**. La traduction la plus fidèle est **apprentissage en ligne**. On parle aussi parfois de **e-formation**, le « e » comme dans e-learning étant une référence explicite aux technologies de l'information.

Le principe consiste alors à remplacer les anciennes méthodes « temps/place/contenu » de l'apprentissage par des processus d'apprentissage rapides/ouverts/personnalisés. Voici quelques points de différences entre l'apprentissage traditionnelle dans les classes et se nouveau mode e-learning illustrés dans le **Tableau 1**.

Dimensions	Apprentissage traditionnel	E-learning
Centré	Enseignant	Apprenant
Rôle de l'apprenant	Passif	Actif
Personnalisation	Enseignement de masse avec un contenu qui doit satisfaire les besoins de plusieurs apprenants	Personnalisation avec un contenu adapté au besoin de chacun
Processus d'apprentissage	Statique, fondé sur un contenu prédéfini	Dynamique, fondé sur les interactions entre les apprenants
Flexibilité	Peu de liberté dans l'organisation du travail	Autonomie importante de l'apprenant pour organiser son apprentissage
Espace	Espace circonscrit : salle de classe	Distribué, ouvert
Technologies utilisées	Tableau, écran mural, etc.	Technologie de l'information et de la communication (TIC)

Tableau 1 Comparaison entre le E-learning et l'apprentissage traditionnel [BEN 05]

Cependant, le domaine du e-learning est vu comme un processus d'apprentissage, où la pratique pédagogique se focalise davantage sur l'apprenant, en mettant à sa disposition des dispositifs de formation en ligne et interactifs. La formation est adaptée aux besoins et au niveau de l'apprenant, en lui proposant un environnement où il peut progresser à son rythme et bénéficier d'un suivi personnalisé.

1.3 Acteur du E-learning

Les différents acteurs d'une formation en ligne peuvent être classés par catégories suivant les rôles qu'ils seront amenés à jouer. Toutefois, trois catégories principales sont recensées: les apprenants, les enseignants (ou tuteur) et les administrateurs.

L'apprenant : est une personne engagée et active dans un processus d'acquisition ou de perfectionnement des connaissances et de leur mise en œuvre. L'apprenant est l'acteur central pour lequel la formation est conçue. Il peut, consulter en ligne ou télécharger les contenus pédagogiques, participer à des activités d'apprentissage en ligne (activités individuelles ou collaboratives), échanger des données, effectuer des exercices, s'auto évaluer et transmettre des travaux à son tuteur pour les corriger ou bien poser des question qui est la tâche qui nous intéresse le plus parmi les activité qu'il fait.

Afin que l'apprenant accède à ses cours, il faudra d'abord qu'il s'inscrive, il fournira ensuite les informations concernant son profil. De cette manière, il pourra travailler en collaboration avec les membres de son groupe ainsi que son tuteur (formateur).

L'apprenant peut être un étudiant désirant suivre un certain cours, un employé d'entreprise ayant besoin d'une formation dans un certain domaine, une personne désirant perfectionner ses connaissances dans une branche quelconque, etc.

Le tuteur : aide à faire progresser les apprenants en mettant davantage au premier plan les fonctions d'évaluation, de suivi « Tracking » et d'accompagnement, pour atteindre un objectif d'apprentissage, plutôt que la capacité à transférer une expertise. Il communique et interagit avec eux, en jouant un rôle d'administrateur dans le cadre d'une activité collective. Ses rôles consistent alors à gérer les communications. Ainsi, il facilite l'apprentissage et gère les apprenants et les environnements. Donc son rôle est de suivre l'évolution du travail de l'apprenant et de l'assister.

La qualité du suivi d'un tuteur permet d'assurer au mieux l'encadrement d'un apprenant et ainsi maintenir sa motivation afin de réduire les risques d'abandon au cours de la formation.

L'administrateur : chargé d'entretenir l'environnement technique, c'est à dire assure l'installation et la maintenance du système, gère les droits d'accès, crée des liens vers d'autres systèmes et ressources externes (dossiers administratifs, catalogues, ressources pédagogiques, etc.).

1.4 Développement rapide de e-learning

Le e-learning s'est bien développé ces dernières années et s'est imposé dans le domaine de l'enseignement public ou privée grâce à plusieurs facteurs :

a. **Facteurs technologiques** : L'Internet est en plein développement ; il permet en effet aux utilisateurs d'avoir accès à l'information, en grande quantité, de traiter l'information, de l'organiser, de communiquer, d'apprendre, d'agir et de réagir...Plusieurs outils sont intégrés à l'ordinateur et à l'Internet : des outils de communication et d'échange (la messagerie électronique, les forums...), de gestion (agenda...), de production, de recherche...

b. **Facteurs psychopédagogiques** : Il s'agit des modèles et des théories d'enseignement et d'apprentissage comme le constructivisme, le cognitivisme en passant par les aspects sociaux de l'apprentissage...

c. **acteurs sociocognitifs** : C'est un besoin qui se fait sentir de plus en plus, surtout avec l'évolution dans le domaine des TIC : les compétences, le marché du travail doivent suivre l'évolution ; d'où la nécessité d'une mise à jour pour assurer une certaine compatibilité entre les outils, ses fonctions, les nouveaux besoins, l'usage en termes de compétences.

d. **Facteurs socio économiques** : L'augmentation massive des besoins de formation, liée à la nécessité du développement de nouvelles compétences sur le marché du travail est à l'origine de l'apparition de plusieurs instituts, organismes qui se spécialisent dans le domaine et proposent leurs nouveaux services.

1.5 Les principaux bénéfices de e-learning

1) Permet la formation «juste à temps» :

La formation virtuelle permet aux apprenants de se former et de se rafraîchir la mémoire en tout temps sur une foule de connaissances sans avoir à attendre la diffusion d'un cours donné à un moment bien précis.

2) Offre une formation en toute souplesse :

Les modes de diffusion de l'apprentissage virtuel offrent une de large possibilités pour recevoir la formation: dans une salle de classe traditionnelle aménagée en conséquence, dans un bureau réservé spécifiquement à l'apprentissage individuel ou à des stations de travail aménagées un peu partout dans l'entreprise pour les employés n'ayant pas accès à un ordinateur, et même à la maison.

3) Contribue à réduire les coûts de formation :

L'apprentissage virtuel requiert un investissement initial important, mais son utilisation peut souvent engendrer des économies substantielles. Elles se réalisent principalement par la réduction des frais de déplacement des apprenants, des honoraires des formateurs, des pertes de temps de travail grâce à la diminution du temps requis pour l'apprentissage, des pertes de productivité en rendant accessible la formation sur les lieux de travail.

Chaque projet de formation virtuelle aura une rentabilité différente d'après les critères de calculs énoncés précédemment.

4) Uniformise et personnalise l'apprentissage :

L'apprentissage virtuel assure à la fois l'uniformité de la formation et rend accessible aux employés des cours et des parties de cours selon les besoins exprimés par chacun.

5) Permet de diffuser l'apprentissage de manière simultanée à un large auditoire :

Que vous lanciez sur le marché un nouveau produit ou un nouveau service et que vous deviez former une multitude de personnes simultanément, à un moment précis, l'apprentissage virtuel est la stratégie toute désignée. En effet, vous pouvez joindre, au moment où vous choisissez, tous ceux qui ont accès à un ordinateur branché sur Internet ou sur votre intranet. Vous pouvez dès lors vous assurer que tous auront la même formation au même moment donné.

1.6 Inconvénients du e-learning :

1) La diffusion de cours e-learning nécessite des équipements multimédia :

L'équipement de l'entreprise doit permettre la diffusion des contenus des cours. Cela concerne notamment les postes utilisateurs (PC récents, logiciels installés compatibles, éventuellement réseau avec suffisamment de bande passante). Cet inconvénient tend à perdre de son importance au fil des années, étant donné l'évolution des technologies.

2) La mise en place de l'infrastructure technique et la création des contenus sont coûteuses :

Un investissement est nécessaire au lancement d'une politique du e-learning. Cet investissement (coût fixe et risque important) remplace le paiement à la prestation de formation (coût variable et risque faible). Par ailleurs, le marché n'est pas encore consolidé et l'offre en matière de cours spécialisés reste limitée suivant les domaines.

3) L'accès à l'outil informatique est nécessaire :

L'utilisation de l'outil informatique limite la diffusion du e-learning auprès d'une partie des collaborateurs. Ceci est un obstacle par exemple dans l'industrie.

4) L'e-learning limite les interactions entre les individus :

Certains mécanismes de communication ne peuvent pas être reproduits (langage du corps par exemple), alors qu'ils jouent un rôle important dans la diffusion du savoir.

1.7 Les plateformes d'apprentissage :

Une plateforme de formation ouverte et à distance est un logiciel qui assiste la conduite de ces formations. Ce type de logiciel regroupe les outils nécessaires aux trois principaux utilisateurs: enseignant, apprenant et administrateur. C'est un dispositif qui a pour premières finalités la consultation à distance de contenus pédagogiques, l'individualisation de l'apprentissage et le télé-tutorat.

Ce système logiciel dit LMS ou LCMS est développé pour accompagner les enseignants dans leur gestion des cours d'éducation en ligne pour leurs étudiants. Les services offerts incluent généralement un contrôle d'accès, des outils de communication (synchrones et/ou asynchrones) et l'administration des groupes d'utilisateurs.

a) LMS: Learning Management System

LMS (Learning Management System) ou MLE (Managed Learning Environment) ou VLE (Virtual Learning Environment) ou CMS (Course Management System) ou LSS (Learning Support System) sont des termes équivalents pour désigner la plateforme de e-learning. Cette plateforme est un système d'information conçu pour optimiser, sur un réseau Internet ou Intranet, la gestion de l'ensemble des activités de formation, l'hébergement du contenu pédagogique, l'inscription des participants, la distribution des ressources, l'organisation de parcours individualisés, le suivi par le tuteur et du tutorat (gestion intégrée des interactions apprenants formateur).[UNI 06].

b) LCMS: Learning Content Management System

Il s'agit d'applications logicielles multiutilisateurs. Une solution de LCMS est un environnement permettant aux concepteurs de cours de créer, stocker, réutiliser, gérer et distribuer des contenus pédagogiques à partir d'un référentiel unique [PCB 03]. Ce référentiel stocke des objets pédagogiques et la plateforme LCMS permet de les associer et de les ordonner afin de construire un cours cohérent.

c) La relation LMS/LCMS

Les LMS et LCMS sont non seulement distincts les uns des autres, ils se complètent également bien. Une fois étroitement intégrés, l'information des deux systèmes peut être échangée, ayant pour résultat, une expérience d'étude plus riche pour l'utilisateur et un outil plus complet pour l'administrateur du système d'apprentissage. Un LMS peut contrôler les communautés des utilisateurs, permettant à chacun d'eux de lancer les objets appropriés stockés et contrôlés par le LCMS. En fournissant le contenu, le LCMS suit également le progrès de l'étudiant individuel, enregistre les points d'essai et les passe de nouveau au LMS pour les gérer. [GRE 02]

1.8 Les enjeux de la plateforme e-learning

- Rendre plus efficaces, plus solides, plus adaptés les processus d'apprentissage et l'accès à la connaissance
- Gestion des contraintes de lieu et de temps (atteindre rapidement une population géographiquement dispersée).
- Compression du temps de formation.
- Parcours pédagogique ciblé et personnalisé qui permet d'apprendre à son rythme.
- Apprendre à mieux maîtriser l'outil informatique et Internet.
- Accès à de nombreux contenus pédagogiques.
- Mise à jour régulière des contenus pédagogiques et des informations.
- Accès à des informations supplémentaires et complémentaires aux sessions de formation en face à face pédagogique.
- Optimisation des coûts.
- Suivi et évaluation des apprenants performants

1.9 Caractéristiques des plateformes e-learning

Des études sur les différentes plates-formes [PCB 03] et [ORA 00] démontrent que la majeure partie d'entre elles peuvent fonctionner sans exiger sur le poste de l'étudiant un logiciel autre que des navigateurs universels (Netscape et Internet Explorer). Cependant, plusieurs plates-formes proposent des extensions sous forme de plug-ins, des modules externes, des outils associés, etc., et exploitent alors des fonctionnalités particulières.

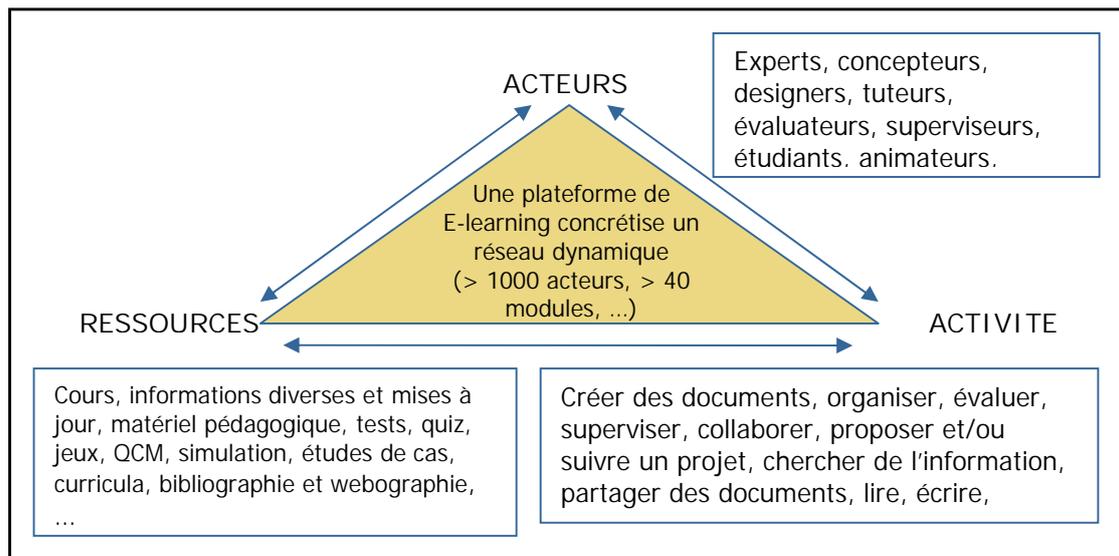


Figure 1 : La plateforme e-learning

1.10 Rôle des principaux utilisateurs

Les dernières versions de plate-forme peuvent être décrites selon trois espaces de travail correspondant aux trois catégories d'acteurs identifiés : enseignant, administrateur et apprenant.

a) Espace enseignant

Dans la partie création, les enseignants auteurs peuvent créer leurs applications multimédias grâce au système auteur de la plateforme qui permet notamment de mettre en place le graphe des concepts enseignés et les différentes relations entre les concepts et les modules.

Dans la partie gestion, accessible à tous les enseignants inscrits sur le serveur, suivant les droits attribués à chaque enseignant : tuteur, auteur ou correcteur, La plateforme met à la disposition de la personne connectée un ensemble de fonctionnalités qui lui permet :

- D'intégrer les modules réalisés par l'outil ou d'autres modules externes.
- De gérer les apprenants, les groupes d'apprenants ou les plans de formation.
- De créer les cursus personnalisés.
- De communiquer avec les autres utilisateurs de la plate-forme.

b) Espace administrateur

Cet espace est partagé en trois parties :

La première permet de gérer toutes les informations récupérées par la plateforme une fois que l'apprenant a fini l'exécution de son cours ou ses évaluations. Les enseignants obtiennent ainsi des informations fines sur le déroulement de chaque session d'apprentissage.

La deuxième permet de gérer la formation en ligne (gestion des groupes, des Intervenants, des plannings, gère les accès et les droits des uns et des autres etc.) via un poste connecté au réseau Internet.

La troisième permet d'installer et assure la maintenance de la plate-forme, crée des liens avec les systèmes d'information externes (scolarité, catalogues, ressources pédagogiques, etc.).

c) Espace apprenant

Cet espace est accessible à tous les apprenants inscrits sur le serveur. La plateforme met à leur disposition des fonctionnalités pour suivre la formation, par exemple consultation du livret élève, exécution des modules de formation en ligne.

2 Système de Question Réponse (Q/A)

Développés pour répondre aux besoins croissants de recherche d'information dans des corpus toujours plus fournis tels que le Web, les systèmes de Question Réponse correspondent à une double problématique : permettre à l'utilisateur recherchant une information précise de poser sa question en langage naturel, et extraire la réponse à cette question d'un grand nombre de documents également en langage naturel.

Les systèmes de question réponse Q/A prennent de plus en plus d'importance dans le domaine de la recherche d'information et d'enseignement, c'est une forme de recherche d'information qui traite des questions en langage naturel en utilisant généralement plusieurs types d'analyses et de ressources.

L'objectif principal est de rechercher des réponses explicites aux questions plutôt que les documents entiers. Cette tâche a été déjà abordée par de divers systèmes suivant des approches linguistiques et statistiques.

Dans les systèmes Q/A de domaine fermé, les réponses correctes ne peuvent normalement être trouvées que dans très peu de documents. Le système n'a donc pas beaucoup de chance, comme dans le cas du Q/A de domaine ouvert, où il y a plus de choix dans un ensemble plus large de réponses candidates. De plus, si le système vise à répondre aux questions d'un utilisateur, il doit être capable de répondre à des questions complexes, et de formes et de styles librement variés.

Cette partie du chapitre présente un court historique de tels systèmes et présente quelques méthodes possibles de résoudre cette tâche. Le reste de cette partie est divisé en deux sections. La première section décrit les systèmes de Q/A actuels en présentant quelques exemples ainsi que l'historique des systèmes Q/A. Dans la deuxième section nous présentons une vue d'ensemble des modules importants composant un système Q/A.

2.1 Histoire des systèmes de Q/A

De très nombreuses équipes développent à l'heure actuelle des systèmes de question réponse du fait notamment des conférences TREC. La tâche QR (en anglais Q/A Track) a été créée lors de TREC-8 et s'est rapidement imposée comme la tâche la plus populaire de TREC, devant les tâches traditionnelles de recherche d'information.

De nombreuses recherches antérieures ont permis d'explorer la tâche de Q/A, indépendamment de TREC. Les années 1960 et 1970 ont ainsi vu la naissance de systèmes d'interrogation de bases de données en langage naturel. Le système BASEBALL [GRE 61] permettait l'interrogation de résultats de baseball, le système LUNAR l'analyse de roche lunaire [WOO 72] et le système LIFER [HEN 77] des statistiques sur des employés. Les systèmes devaient être capables de répondre à des questions telles que :

- *Who did the Red Sox lose on July 5? (BASEBALL)*
- *On how many days in July did eight teams play?*
- *What is the average concentration of aluminium in high alkali rocks? (LUNAR)*
- *How many Brescias contain Olivine?*
- *What is the average salary of math department secretaries? (LIFER)*
- *How many professors are there in the CompSci department?*

L'approche utilisée consistait généralement à analyser la phrase et à la traduire, directement ou non, en requête de base de données (grosso modo une variante de SQL). Ces systèmes souffraient de limites bien connues : ils étaient dédiés à un domaine et il était souvent difficile de les étendre ou de les adapter à d'autres domaines.

D'un autre côté, l'analyse syntaxique et sémantique avait souvent des visées génériques et reposait sur un ensemble de connaissances de domaine préalablement enregistrées.

Ces systèmes pouvaient donc mettre en œuvre des techniques sophistiquées pour la résolution des anaphores ou des ellipses, ce qui n'est pas les cas des systèmes de Q/A actuels, qui sont le plus souvent hors domaine mais qui, en contrepartie, emploient un jeu de connaissances beaucoup moins riche. En conséquence, les systèmes actuels mettent rarement en œuvre des techniques de dialogue pour résoudre des questions ambiguës contrairement aux systèmes décrits ci-dessus.

D'autres systèmes ont été développés sur la base de théories issues de l'intelligence artificielle. Le système SHRDLU, qui permet une interaction entre un humain un robot placé dans un univers composé de cubes et de formes diverses a aussi conduit à de nombreuses recherches impliquant la compréhension de questions et l'activation de commandes en fonction des questions et des ordres donnés au robot [WIN 72]. Ainsi le système QUALM [LEH 77,88] applique la théorie des scripts et des plans au domaine de la réservation au restaurant [SCH 77]. Le système UNIX CONSULTANT [WIL 82] propose une approche similaire pour répondre à des questions concernant le système Unix.

L'accroissement gigantesque des masses de documents et d'informations disponibles sous forme électronique, en particulier avec l'avènement du Web, a considérablement renouvelé les approches. Comme le soulignent [LIN 03], les systèmes développés avant l'apparition du web avait une audience limitée et toute l'information qu'ils contenaient devait être codée manuellement.

Le web offre à l'inverse la possibilité à des millions de personnes d'accéder aux services offerts par les systèmes de Q/A, ce qui implique une diversité de questions et de domaines jamais abordés auparavant. Enfin, les systèmes peuvent eux-mêmes utiliser la redondance de l'information sur le web pour valider les réponses et créer leurs propres bases de connaissances.

L'inconvénient de tous les systèmes de Q/A décrits jusqu'ici est le besoin de domaine spécifique.

Normalement, ceci inclut également le besoin de bases de connaissances ou de bases de données ouvrées par main. C'est non seulement cher, mais le rend également presque impossible d'utiliser ces systèmes dans une autre application que l'original. Tous ce sont des raisons pour lesquelles ces systèmes n'ont jamais joué un rôle décisif pour les systèmes commerciaux dans ce domaine de recherche. Ainsi, la conclusion logique était d'accumuler des systèmes plus flexibles de Q/A qui ne se fondent pas sur la connaissance spécifique de domaine.

Comme déjà mentionné, un avantage de tels systèmes ouverts de Q/A de domaine est qu'il n'y a aucun besoin de bases de connaissances. Mais ceci signifie qu'ils doivent trouver leurs informations dans d'autres sources, dans ce cas-ci utilisant des documents de texte en clair. Ces collections de document peuvent contenir des articles de journaux, livres, les histoires, encyclopédies aussi bien que des documents de Web, tels que des forums ou des blogs. C'est pourquoi de tels systèmes s'appellent également les systèmes de Q/A basés par texte.

2.2 Fonctionnement des systèmes Q/A

Développés pour répondre aux besoins croissants de recherche d'information dans des corpus toujours plus fournis tels que le Web, les systèmes de Question Réponse correspondent à une double problématique : permettre à l'utilisateur recherchant une information précise de poser sa question en langage naturel, et extraire la réponse à cette question d'un grand nombre de documents également en langage naturel.

Les systèmes de question/réponse partagent de nombreux traits communs avec les systèmes de recherche d'information. On peut en effet dessiner à gros traits le processus suivant, nécessaire pour un système de question/réponse :

1. Analyse de la question
2. Transformation de la question en requête
3. Recherche de documents pertinents par interrogation d'un moteur de recherche
4. Extraction, au sein de l'ensemble de documents pertinents, de la séquence linguistique qui constitue la réponse à la question posée.

Les utilisateurs peuvent poser des questions de langage naturel et le système recherchera la réponse dans la collection donnée des textes. Pour faire ainsi, le système doit analyser la question, alors choisit les documents qui sont dans un certain sens semblable à la question et essaye d'extraire l'information spécifique hors du texte. Ceci signifie qu'un système de Q/A contient des éléments de divers domaines de la recherche, comme l'extraction de traitement du langage naturel, de recherche documentaire et d'information.

Nous aurons l'occasion de détailler certains de ces éléments mais on peut constater que les étapes 2 et 3 décrivent un processus de recherche d'information relativement classique.

Les apports du Traitement Automatique des Langues (TAL) à la Recherche d'Information (RI) ont fait l'objet de nombreuses publications. Les deux publications suivantes donnent un très bon aperçu des techniques de TAL à l'œuvre pour la RI:

- Traitement Automatique des Langues pour la Recherche d'information.
- Traitement automatique des langues et recherche d'information. [GAU 03]

Nous détaillerons, par la suite, les éléments concernant l'analyse de la question et la recherche de réponses précises.

1. L'analyse syntaxique

Les phrases à analyser proviennent de textes authentiques : nous partons du principe qu'ils ne contiennent aucune erreur de grammaire ou d'orthographe. Nous avons choisi d'analyser des textes techniques, sans effet de style, mais constitués néanmoins de phrases complexes avec des structures subordonnées, des coordinations et des structures emboîtées.

L'analyseur syntaxique applique une grammaire ascendante, afin d'obtenir la meilleure couverture syntaxique, selon des critères linguistiques. Cette grammaire, qui offre une bonne couverture de la langue anglaise, a été généralisée à partir du formalisme des grammaires logiques et permet de tenir compte du contexte. C'est une grammaire logique contexte sensitive ; elle résulte d'une extension des grammaires logiques.

Pour une phrase donnée, le parseur génère dans la plupart des cas un arbre d'analyse complet. Lorsque ce n'est pas possible, le système produit un ensemble d'arbres contigus qui correspondent à toutes les parties de la phrase.

Les structures syntaxiques produites par l'analyseur syntaxique peuvent être "ouvertes". En effet, lorsque la syntaxe ne permet pas de résoudre une ambiguïté, certains attachements des structures syntaxiques sont laissés libres et marqués en vue d'un traitement ultérieur par l'analyseur sémantique. Ceci évite de créer un trop grand nombre d'arbres d'analyse.

Avant d'appliquer la grammaire, un analyseur spécial localise les expressions idiomatiques et les locutions. Ceci est nécessaire car elles sont souvent agrammaticales ; en outre, leur sémantique ne résulte pas de la combinaison de la sémantique des mots qui les constituent. Ce traitement reconnaît les locutions sécables et déclinables ; les formes non déclinées sont réunies dans un dictionnaire de locutions propres au domaine.

En ce qui concerne la reconnaissance des mots isolés, le système accède à un dictionnaire; il utilise également un analyseur morphologique de l'Anglais/Français. Quant aux mots outils de l'Anglais (articles, prépositions, etc.), ils sont regroupés dans un autre dictionnaire.

2. Traitement de question

La première étape dans un système de Q/A est le traitement de la question. Ceci est fait, parce que le système doit « comprendre » la signification des besoins d'informations qu'un utilisateur a, c.-à-d. quelle information devrait être renvoyée par le système pour répondre à la question. Le traitement peut être fait de plusieurs manières, selon l'approche théorique du système.

Mais normalement, la classe spécifique d'une question (type de question) est déterminée et une requête est produite pour les modules postérieurs de recherche, aussi selon le langage d'interrogation que les modules de recherche fournissent. Les étapes de transformation plus ultérieure sont facultatives, mais nous décrivons quelques approches que nous avons utilisées dans notre système Alyssa de Q/A.

3. Typage de Question

Cette étape consiste à extraire le type de la question posé par l'utilisateur. Il existe plusieurs approches parmi lesquelles, l'approche qui utilise la taxonomie de classification proposée par [LI 02]. Elle se compose de six classes granuleuses brutes et 50 fines. Par exemple, toutes les valeurs numériques, comme des dates, des prix ou des comptes, sont récapitulées en dedans sur la classe brute NUMÉRIQUE.

Conclusion

Nous pouvons découper les fonctionnalités de base d'un système E-Learning en un certain nombre de fonctions, qui peuvent alors être implantées séparément sous forme d'applications autonomes ou sous forme d'e-Services en utilisant des services web. La mise à disposition de ces services permet la réutilisation du contenu et des fonctionnalités dans une plate-forme E-Learning.

La fonction principale des plates-formes E-Learning est de fournir à l'utilisateur les bonnes activités avec les bons outils au bon moment en fonction de ses besoins. Pour ce faire, il devient de plus en plus nécessaire de prendre en compte le profil de l'utilisateur ou seront stockées les informations utiles à la personnalisation de son parcours pédagogique.

Ce parcours pédagogique au sein d'un système E-Learning est une collection d'activités ou de processus qui ont un effet d'une part sur les apprenants et d'autre part sur le contenu pédagogique convenablement choisi sous forme d'objets pédagogiques.

Chapitre

2

Le Profil Utilisateur Et Les Styles d'Apprentissage



Introduction

La notion du profil utilisateur est apparue vers les années 80 avec les assistants et les agents d'interface. Cette apparition est due principalement au besoin de créer des applications personnalisées capables de s'adapter avec l'utilisateur.

Personnaliser une application pour un utilisateur particulier nécessite de disposer d'informations sur ce dernier, ce qui permettra d'évaluer la pertinence des objets disponibles ou d'aider le système à faire des choix. Le modèle utilisateur est « une source de connaissances ou une base de données sur un utilisateur ». [MCT 93]

Une nouvelle génération de systèmes avancés d'apprentissage doit intégrer des nouvelles approches pédagogiques donnant à l'apprenant un rôle actif pour apprendre et construire ses connaissances. Ces systèmes doivent être plus interactifs, mais surtout intégrer une vision plus centrée sur l'utilisateur. Les systèmes hypermédias adaptatifs dans le domaine de l'enseignement à distance (E-Learning) proposent des solutions à ces problèmes, [BRU 05], [DEB 04]. L'objectif de ces systèmes est d'adapter la présentation de la connaissance et aider l'apprenant à naviguer à travers le graphe composé par l'ensemble des pages et des liens. Pour cela, un profil d'utilisateur doit être défini pour chaque utilisateur afin qu'il soit utilisé par la suite au cours de la personnalisation.

Le concept de profil utilisateur a été introduit pour l'accès à l'information en premier dans les travaux de filtrage d'information, pour décrire une structure représentative de l'utilisateur, en l'occurrence ses centres d'intérêts. Cette notion a ensuite été ré exploitée en RI personnalisée pour former les composantes du contexte directement dépendantes de l'utilisateur : centres d'intérêts, préférences, domaines professionnels, expertise, etc.

Le but de ce chapitre est de fournir un état de l'art sur le profil d'utilisateur, la gestion de profil et l'exploitation du profil pour l'extraction des données. La section 2 est divisée en plusieurs parties, on commence par donner une définition du profil utilisateur et des préférences, puis on montre quelques modèles utilisateurs. Après on passe à la gestion du profil partant de la construction et allant vers la mise à jours du profil. On finit la section en citant quelques inconvénients des systèmes de gestion de profil. La deuxième partie entame le concept de style d'apprentissage, puis nous allons présenter quelques modèles de style d'apprentissage. Avant de conclure ce chapitre.

1 Profil utilisateur

1.1 Définition de la notion du « Profil de l'utilisateur »

Il n'y a pas de consensus dans la définition des profils utilisateurs, il y a autant de définitions que de domaines d'application ou de technologies de recherche d'information.

Un profil (ou souvent appelé « modèle d'utilisateur ») regroupe l'ensemble des connaissances nécessaire à une évaluation efficace des requêtes et à une production d'une information pertinente adaptée à chaque utilisateur. [BOU 05]

Il représente un ensemble de préférences individuelles représentées par des couples (attribut, valeur), par des ordonnancements de critères ou par des règles sémantiques spécifiques à chaque utilisateur ou communauté d'utilisateurs. Ces modes de spécification servent à décrire le centre d'intérêt de l'utilisateur, le niveau de qualité des données qu'il désire ou des modalités de présentation de ces données. [BOU 04]

Depuis le début des années 70, les recherches menées dans ce domaine se sont principalement focalisées sur la possibilité de définir des approches de modélisation de l'utilisateur dans le contexte de différentes applications [POH 97]. L'objectif de ces approches est d'améliorer les interactions homme-machine (IHM) par inférence et prédiction des buts, préférences et contextes des utilisateurs à partir de faits observés. La notion de profil utilisateur a été largement abordée dans le domaine du « user modelling ».

On appelle profil utilisateur toute structure qui permet de modéliser et de stocker les données caractérisant l'utilisateur. Ces données représentent les centres d'intérêts, les préférences et les besoins en informations de l'utilisateur ou un groupe d'utilisateurs.

Il convient de distinguer la notion de profil de la notion de requête. Un profil peut être défini comme une mise en équation du centre d'intérêt et des préférences de l'utilisateur, alors qu'une requête est l'expression d'un besoin circonstancié que l'utilisateur souhaite voir satisfait en tenant compte de son profil. Un profil a un caractère plus invariant que les requêtes même si le centre d'intérêt et les préférences de l'utilisateur peuvent légitimement évoluer.

1.2 Les Préférences

1) Définitions:

Une préférence utilisateur est un ensemble de descriptions englobant: ce qu'un utilisateur envisage d'accomplir dans le système (les activités) et la manière de le faire (séquentielle, concurrente, conditionnelle), le type et l'ordre des résultats de ces activités (le contenu), et la manière dont il aimerait que l'information soit affichée.

Les préférences de l'utilisateur constituent un aspect très important pour la définition du profil de l'utilisateur : elles décrivent notamment les attentes de l'utilisateur par rapport au contenu et à la présentation de l'information dans son dispositif d'accès lors de son interaction avec le système. Dans la section suivante, nous présentons différents travaux sur la représentation des préférences de l'utilisateur.

Angela Carrillo-Ramos et al. [CAR 06] s'appuient aussi sur les travaux de Tamine et al. [TAM 06] pour la définition du Profil de l'utilisateur, ils adoptent à leurs tours une approche multidimensionnelle de ce profil.

Ils définissent le profil contextuel de l'utilisateur comme un ensemble de descriptions, appelées préférences de l'utilisateur. Qui sont un ensemble de descriptions englobant : ce qu'un utilisateur envisage d'accomplir dans le système (les activités) et la manière de le faire (séquentielle, concurrente, conditionnelle), le type et l'ordre des résultats de ces activités (le contenu), et la manière dont il aimerait que l'information soit affichée sur son DM (affichage).

Selon Angela Carrillo-Ramos et al [CAR 06] le profil de l'utilisateur est composé des préférences d'activité, de résultat et d'affichage qui s'appliquent compte tenu du contexte d'utilisation d'une session (c'est-à-dire, compte tenu des caractéristiques de la session en cours, de celles du dispositif, etc.). Pour Carrillo les préférences se divise en deux types, préférences définies indépendamment du contexte d'utilisation (préférences de type général) et celles définies pour la session en question (préférences de type spécifique).

2) Type de préférence:

Selon [CAR 06], nous distinguons trois types de préférences :

- **Les Préférences d'Activité:** Concernent les activités qu'un utilisateur souhaite et peut accomplir dans le système.
- **Les Préférences de Résultat:** Concernent le contenu et le format préféré des résultats des fonctionnalités (par exemple, de la vidéo, du texte, ou des images).
- **Les Préférences d'Affichage:** Concernent la manière dont l'utilisateur souhaite que l'information soit affichée.

3) Expression des préférences:

Il existe deux approches majeures pour l'expression des préférences :

- **L'approche quantitative:** Dans cette approche la comparaison 'mieux que' est exprimée en assignant des scores numériques définissant des degrés d'intérêt d'un utilisateur à ces objets. Soit P la fonction attribuant des scores numériques aux objets 1 et 2 On dira que 1 est préféré que 2 si $P(1) > P(2)$.
- **L'approche qualitative:** Dans cette approche la préférence est spécifiée directement par des relations binaires de préférence. Par exemple, les formules logiques et les constructeurs de préférences spécifiques, cette préférence est comme un ordre partiel strict sur le domaine des valeurs associées à l'ensemble des attributs : c'est une relation binaire (notée $>_p$) entre les tuples d'une même relation. On dira que le tuple X est « mieux que » le tuple Y dans le sens d'une préférence (P) Si $X >_p Y$.

1.3 Contenu du profil utilisateur

Les traits qui caractérisent le profil utilisateurs sont : ses références ou acquis (background), ses connaissances du sujet, ses objectifs (goals), ses préférences et son expérience. Peter Brusilovsky a cité deux autres caractéristiques qui sont : ses centres d'intérêts (interests) et ses traits individuels.

Centres d'intérêts : l'utilisateur peut les formuler à travers l'introduction d'un ensemble de mots clés, pour décrire ses centres d'intérêts. Sur la base de ces mots clés, le système pourra lui recommander des informations qui répondent au mieux à ses attentes.

Traits individuels : ce sont des caractéristiques stables, les traits individuels englobent les facteurs de personnalité. Ces traits ne sont pas tirés d'un simple entretien mais après des tests psychologiques.

Nous distinguons principalement six dimensions dans la définition d'un profil de [KOS 03] : Données personnelles, Centre d'intérêt, Qualité attendue, Préférences de livraison, Sécurité et Historique des interactions de l'utilisateur.

➤ Les données personnelles

Les données personnelles sont la partie statique du profil. Elles comprennent l'identité de l'utilisateur (nom, prénom, numéro de sécurité sociale, etc.), des données démographiques (âge, genre, adresse, situation familiale, nombre d'enfants, etc.), les contacts personnels et professionnels et d'autres informations comme le numéro de la carte bancaire ou de la carte Vitale.

Les données personnelles sont relativement stables dans le temps et ne demandent pas de mise à jour automatique par le gestionnaire de profils. Dans plusieurs approches ces données ne jouent pas un rôle dans le processus de recherche d'information, mais servent souvent comme monnaie d'échange contre les services de personnalisation ou les services d'accès.

➤ Le centre d'intérêt

Le centre d'intérêt exprime le domaine d'expertise de l'utilisateur ou son périmètre d'exploration. Il peut être défini par un ensemble de mots clés (concepts) ou un ensemble d'expressions logiques (requêtes). Dans de nombreuses approches, l'importance de chaque concept est définie par une pondération des mots clés du centre d'intérêt.

L'ontologie du domaine complète la définition du centre d'intérêt en explicitant la sémantique de certains termes. Le centre d'intérêt peut être vu comme une présélection virtuelle qui réduit la masse d'informations à prendre en compte.

➤ La qualité attendue

La qualité est un des facteurs clés de la personnalisation ; elle permet d'exprimer des préférences extrinsèques sur l'origine de l'information, sa précision, sa fraîcheur, sa durée de validité, le temps nécessaire pour la produire ou la crédibilité de sa source.

Les attributs de cette dimension expriment la qualité attendue (ou espérée) par l'utilisateur ; elle sera confrontée à la qualité effective produite par le système de recherche d'informations.

La qualité attendue peut donc être définie par des expressions de préférences sur un facteur de qualité donné. Par exemple l'utilisateur peut exprimer sa préférence de la source S1 devant la source S2 en terme de fraîcheur. Ceci se traduit par l'ajout dans son profil d'une sous-dimension 'ExpressionQualité' où le type d'expression est préférence de source, l'attribut de qualité est fraîcheur et l'expression de préférence est S1»S2.

➤ Les préférences de livraison

Les préférences de livraison concernent d'abord tout ce qui est lié aux modalités de présentation des résultats en fonction de la plateforme (média de livraison), de la nature et du volume des informations délivrées, des préférences esthétiques ou visuelles de l'utilisateur.

A ces modalités de présentation, on peut ajouter les modalités d'exécution, décrivant le moment d'exécution d'une requête (mode pull ou push), la manière de notifier les résultats (différé, immédiat par exemple) et la quantité de résultats que l'on souhaite recevoir (les Top k, les premiers calculés, l'ensemble des résultats mais à la fin du calcul, etc.).

➤ La sécurité

La sécurité est une dimension fondamentale du profil. Elle peut concerner les données que l'on interroge ou modifie, les informations que l'on calcule, les requêtes utilisateurs elles-mêmes ou les autres dimensions du profil.

La sécurité du processus exprime la volonté de l'utilisateur de cacher un traitement qu'il effectue. Ceci peut être fait en définissant le degré de visibilité de certaines opérations.

➤ L'historique des interactions de l'utilisateur

Cette dimension contient entre autre ce qu'on appelle communément le "feedback" de l'utilisateur. Elle regroupe l'ensemble des informations collectées sur le comportement de l'utilisateur, que ces informations soient directement fournies par lui (feedback explicite) ou qu'elles soient récupérées ou dérivées à son parcours (feedback implicite). Les informations de l'historique des requêtes de l'utilisateur ne sont pas utilisées directement, mais sont analysées afin qu'elles soient utilisées pour la mise à jour des valeurs de profil.

1.4 La modélisation de l'utilisateur

Dans le domaine des IHM, la notion de profil est souvent appelé modèle de l'utilisateur, le profil contient, par exemple, des informations permettant au système d'adapter l'affichage des résultats selon les préférences de l'utilisateur. Un exemple simple d'un tel profil est celui utilisé par les fournisseurs de services Web.

La modélisation de l'utilisateur consiste à décrire les caractéristiques informationnelles des utilisateurs à travers un modèle de profil. On trouve dans la littérature plusieurs définitions de la modélisation utilisateur, dont nous citons quelques-unes ci-dessous :

"The process of gathering information about the users of computer systems and of making this information available to systems which exploit it to adapt their behaviour or the information they provide to the specific requirements of individual users has been termed as user modelling." [POH 97]

La modélisation des utilisateurs est souvent utilisée pour adapter un système à un utilisateur. L'adaptation permet d'améliorer la convivialité du système. L'interaction entre le système et l'utilisateur peut aussi être améliorées, par exemple en adaptant le dialogue au niveau d'expertise de l'utilisateur ou encore en orientant le discours sur ses centres d'intérêts.

Les utilisateurs peuvent avoir des préférences générales, récurrentes et stables. Cependant, l'ensemble des informations contenues dans le profil ne sont pas forcément appropriées à toutes les situations de recherche. Le plus souvent, les systèmes n'utilisent seulement qu'un sous-ensemble de ces informations, qu'ils supposent pertinents pour la recherche en cours.

Dès lors, le choix du profil adéquat constitue la principale réflexion lors de la mise en œuvre du SRIP. De ce fait, les questions fondamentales posées pour modéliser le profil utilisateur sont le «Quoi », le « Comment » et le « Quand » [MTC 93] :

- Quoi ?
 - ❖ Quelles propriétés informationnelles caractérisent l'utilisateur ?
 - ❖ Quelle structure informationnelle utiliser pour représenter l'utilisateur ?
- Comment ?
 - ❖ Comment collecter les informations du profil ?
 - ❖ Comment construire le profil de l'utilisateur ?
 - ❖ Comment détecter le contexte, le but de la recherche et les besoins à court/long terme de l'utilisateur ?
 - ❖ Comment adapter le profil à l'évolution de l'utilisateur lui-même ?
 - ❖ Comment assurer la sécurité et la confidentialité des informations du profil ?
- Quand ?
 - ❖ Quand faut-il faire évoluer le profil de l'utilisateur ?

1) Les types de modèles

Les modèles présentés ci-dessous peuvent être combinés afin de réaliser le modèle utilisateur qui conviendra à l'application développée :

- Le modèle individuel : ce modèle regroupe les caractéristiques propres à un individu. Il s'agit d'informations pouvant être soit renseignées par l'utilisateur, soit déduites par le système, soit acquises lors de l'utilisation de l'application.
- Le stéréotype : il s'agit d'un modèle générique qui correspond à un condensé des caractéristiques les plus représentatives d'un groupe – ou classe – d'individus, que nous pouvons qualifier des valeurs par défaut. Le stéréotype peut être utilisé tel quel, ou dans une optique d'individualisation de modèle [PET 99].

- Le modèle de recouvrement : il ne peut exister que si le système possède un modèle de domaine car le modèle de recouvrement associe pour un utilisateur particulier, une valeur à chaque concept du modèle de domaine. Chaque valeur correspond à une estimation du niveau de connaissance de l'utilisateur pour le concept. Le modèle de recouvrement est facile à mettre à jour mais souvent difficile à initialiser surtout au niveau de la mesure du niveau de connaissance.
- Le modèle partagé : il s'agit d'un modèle réutilisable dans plusieurs applications. En effet, de nombreuses caractéristiques d'un utilisateur sont utilisées systématiquement par toutes les applications adaptatives. L'idée du modèle partagé est de disposer d'une partie commune et de parties spécifiques à chacune des applications où tâches à réaliser. Cette approche permet d'une part de partager des informations entre applications et d'autre part d'obtenir plus facilement une participation des utilisateurs qui n'ont pas à redéfinir un modèle pour chacune des applications. Kosba et Wahlster [POH 96] proposent de disposer d'un modèle central et de plusieurs sous modèle disjoints.

1.5 Classification du Contenu des profils Utilisateur

Chaque système de personnalisation propose un modèle de profil adapté au service qu'il offre. De façon formelle les données stockées dans les profils peuvent être classées selon quatre catégories :

1) Des exemples d'objets de l'espace de recherche

Ces modèles de profils sont utilisés en général par les systèmes de recommandation [BEH 09]. L'idée principale est de stocker des exemples d'objets que l'utilisateur a jugés pertinents ou inintéressants pour ensuite comparer le contenu des profils des différents utilisateurs entre eux [POH 97, BRU 05]. S'il y a des similarités entre deux profils le système va recommander à l'utilisateur des objets que d'autres clients de profils semblables ont appréciés.

2) Des caractéristiques extraites des objets de l'espace de recherche

L'idée de cette approche est de trouver une forme normale permettant la représentation des profils et des objets de l'espace de recherche. Ceci est fait le plus souvent en analysant les objets pour extraire les caractéristiques pertinentes pour l'utilisateur [DEB 04, BRA 00, DAI 02, FER 01, GAU 99, CRO 01]. Pour la plupart des approches existantes ces caractéristiques sont des mots clés. Dans certains cas les

mots clés sont combinés avec d'autres informations portant sur l'importance des mots pour l'utilisateur (fréquences d'apparition, poids, votes etc.).

3) Des attributs décrivant les préférences externes de l'utilisateur

Les informations des deux catégories précédentes sont directement liées aux objets de l'espace de recherche et à leur contenu, mais ne portent aucune information extrinsèque sur la qualité des informations ou sur l'utilisateur. Dans cette catégorie d'informations on retrouve les préférences de l'utilisateur liées aux sources des données, la qualité des informations, le type et la structure des éléments recherchés [SOL 98, SHE 01]. Les données personnelles font également partie de cette classe de données.

4) Des règles et des formules de préférences

Dans certaines approches [JUN, FIN 00, AMA 99] le profil de l'utilisateur est remplacé par des formules de préférence qui permettent de définir un ordre entre les éléments selon la pertinence de ces éléments pour l'utilisateur. Dans d'autres systèmes comme par exemple Personnalisation server [SOL 98] les préférences des utilisateurs sont capturées par des règles actives qui tentent de découvrir des généralités en regroupant les utilisateurs selon les caractéristiques comme l'âge, le sexe, les données démographiques, le nom du domaine, le type du browser, le système d'exploitation, la bande passante disponible, les activités personnelles ou la situation familiale.

1.6 Approches de représentation du profil utilisateur

La représentation de l'utilisateur à travers la notion de profil permet de mieux comprendre certains 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. Un modèle de représentation permet d'organiser les informations du profil afin de faciliter leur exploitation par la suite. On cite principalement quatre principales approches de représentation : ensembliste, connexionniste, conceptuelle et multidimensionnelle.

1) Représentation ensembliste :

On L'approche ensembliste consiste à représenter le profil de l'utilisateur par des paquets de termes pondérés obtenus implicitement ou explicitement à partir de plusieurs sources d'information. D'un point de vue RI, on parle plutôt d'une représentation vectorielle par analogie au modèle vectoriel de Salton [SAL 71] sur laquelle elle se base. Ces paquets de termes, traduisant les centres d'intérêts de l'utilisateur, peuvent être regroupés différemment selon l'approche suivie pour considérer le profil de l'utilisateur. L'utilisation de plusieurs vecteurs permet de prendre en compte la diversité des domaines d'intérêts ou évolution dans le temps.

On distingue dans la littérature trois grandes approches de représentation du profil utilisateur basées sur ce modèle :

- Par une liste de mots clés, où chaque mot correspond à un centre d'intérêt spécifique.
- Par un vecteur de termes pondérés pour chaque centre d'intérêt.
- Par un ensemble de vecteurs de termes pondérés (ou non) indépendants, pour prendre en compte des centres d'intérêt multiples où chaque vecteur correspond à un domaine d'intérêt. [PAZ 96]

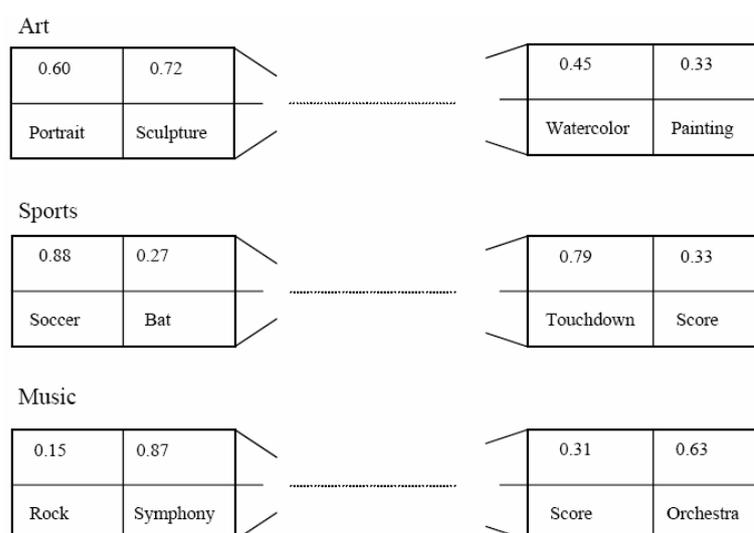


Figure 2 Un exemple de profil représenté par des mots clés [PAZ 96]

Plusieurs systèmes d'accès personnalisé à l'information utilisent ce type de représentation.

Notamment, dans Anatagonomy, un système personnalisé de consultation de nouvelles et de journaux en ligne, Fab est un système de recommandation de page web, Letizia [LIE 95], un système d'aide à la navigation, et Syskill & Webert un système de recommandation. Tous ces systèmes proposent des profils utilisateur représentés par une liste de mots clés.

La représentation par liste de mots clés et/ou par classe de vecteurs de termes apporte l'avantage de la simplicité de mise en œuvre. Néanmoins, même si ces systèmes prennent en considération des centres d'intérêts multiples en utilisant plusieurs vecteurs, cette représentation manque de structuration. Elle ne facilite ni l'interprétation ni la prise en compte des différents niveaux de généralité caractérisant l'utilisateur [BOT 04]. En effet, la plupart des utilisateurs ont des intérêts divers et multiples, leur généralisation dans un vecteur simple n'est pas clairement représentative de la réalité.

2) Représentation connexionniste :

Afin d'adresser le problème de polysémie des termes inhérents à la représentation ensembliste, une première solution consiste à représenter le profil par un réseau de nœuds pondérés dans lequel chaque nœud représente un concept traduisant un centre d'intérêt utilisateur. Ce type de représentation offre le double avantage de la structuration et de la représentation associative permettant de considérer l'ensemble des aspects représentatifs du profil.

Les centres d'intérêts sont souvent représentés par des relations de paires de nœuds dans laquelle chaque nœud contient un terme issu des documents du corpus de recherche. Les arcs reliant deux nœuds, sont créés sur la base des cooccurrences entre ces termes. Le profil utilisateur peut être augmenté par l'inclusion d'un ensemble de paires d'attribut-valeur correspondant à la partie structurée des documents [MIN 96] (par exemple, la taille, nombre d'images, etc.) qui ont précédemment intéressé l'utilisateur.

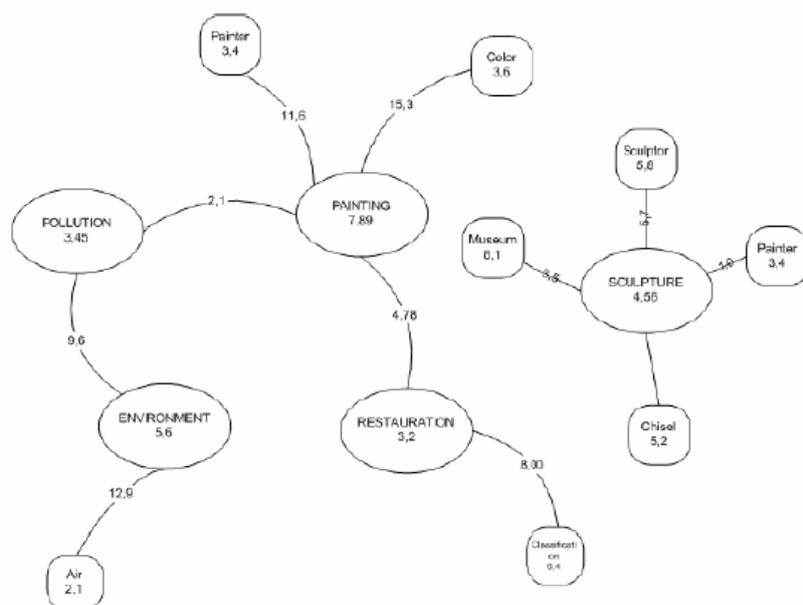


Figure 3 Un extrait d'un profil utilisateur connexionniste

3) Représentation sémantique :

La représentation du profil met en évidence, dans cette approche, les relations sémantiques entre les informations le contenant. Suivant une direction proposée dans un contexte plus général par Huhn [HUN 99], cette représentation offre une alternative intéressante à l'approche précédente.

Les travaux actuels tendent à représenter le profil sous forme d'une ontologie de concepts personnels en se basant sur les connaissances contenues dans les ontologies plutôt que de construire les profils d'utilisateur seulement à partir des documents collectés de son interaction [CHA 99]. La représentation est essentiellement basée sur l'utilisation d'ontologies [NAN 03, GAU 07, CHA 04, BAZ 04, LIU 04, HER 07] ou des réseaux probabilistes [LIN 05; WEN 04]. Dans ce type d'approche les liens entre les concepts sont explicitement induits de l'ontologie et le profil résultant inclura des relations informationnelles plus diverses et spécifiques.

Dans un contexte aussi dynamique et volumineux que le web, la représentation du profil basée sur les ontologies engendre certains problèmes liés à l'hétérogénéité et la diversité des centres d'intérêts de l'utilisateur. En dépit du fait que ces profils peuvent contenir un nombre considérablement élevé de concepts. D'ailleurs, les utilisateurs peuvent avoir de différentes perceptions pour un même concept, ce qui engendre des représentations imprécises de l'utilisateur [GOD 05].

4) Représentation multidimensionnelle :

La représentation multidimensionnelle du profil s'inscrit dans une réflexion globale sur la personnalisation de l'information, en partant du principe que la modélisation de l'utilisateur doit pouvoir capturer toutes les dimensions qui représentent l'utilisateur.

Différents travaux ont abordé cet aspect sans le couvrir dans son ensemble. Ainsi, les propositions de standards P3P [W3C 05] pour la sécurisation des profils ont défini des classes distinguant les attributs démographiques des utilisateurs (identité, données personnelles), les attributs professionnels (employeur, adresse, type) et les attributs de comportement (trace de navigation).

Dans le cadre du développement d'un service avancé de librairie digitale « Système EUROgatherer », Amato [AMA 99] a fait une autre proposition qui consiste à représenter le contenu du profil utilisateur par un modèle structuré de catégories (ou dimensions) prédéfinies : catégorie de données personnelles, catégorie de données de la source, catégorie de données de livraison, catégorie de données de comportement et catégorie de données de sécurité.

Kostadinov [KOS 03] distingue principalement huit dimensions pouvant contenir la plupart des informations susceptibles de caractériser l'utilisateur tels que les données personnelles, les centres d'intérêt, l'ontologie de domaine, la qualité attendue des résultats délivrés, la customisation, la sécurité et la confidentialité, le retour de préférences et les informations diverses.

Zemirli et al. [ZEM 05], en étudiant le travail de Kostadinov [KOS 03], ont proposé un profil à trois dimensions (cf. Figure 4) :

- i. Les données de préférences concernant les préférences de recherche et les centres d'intérêts de l'utilisateur ;
- ii. Les données personnelles (son identité et sa profession) ;
- iii. Les données sur l'environnement de travail de l'utilisateur (l'emplacement géographique, la configuration logicielle et matérielle). Les préférences de recherche représentent les préférences de l'utilisateur portant sur le processus de recherche (temps de réponse), sur les documents recherchés (contenu et contenant) et sur la personnalisation des résultats (livraison et mise en page).

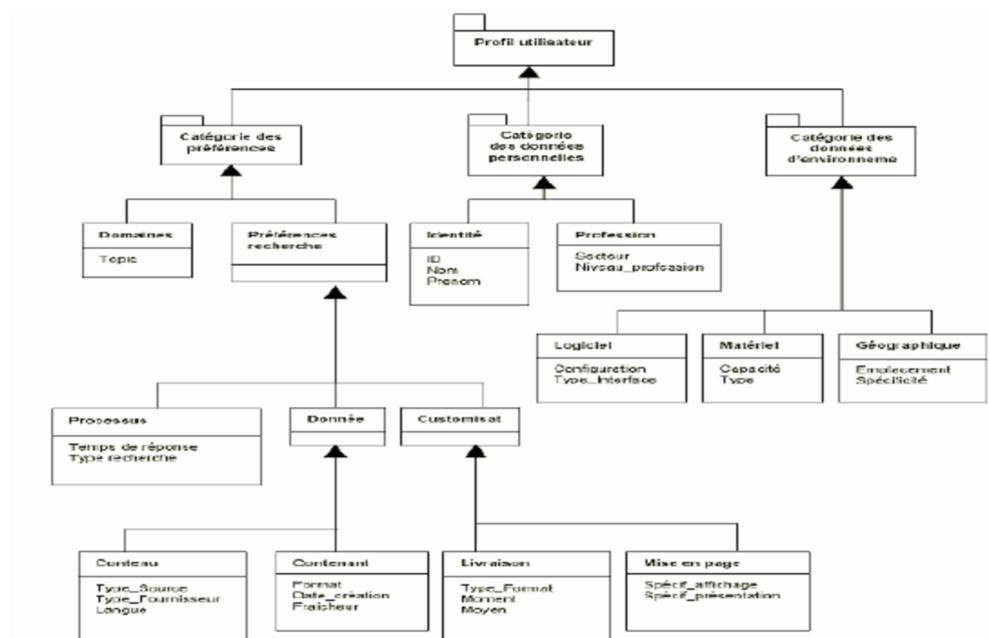


Figure 4 Représentation multidimensionnelle du profil utilisateur
(D'après Zemirli et al. [ZEM 05]).

1.7 Approches de construction du profil utilisateur

La construction du profil traduit un processus qui permet d'instancier sa représentation. L'approche de construction dépend fortement de la représentation choisie pour le profil utilisateur : les techniques utilisées par les systèmes différents selon qu'ils représentent le profil par un (des) vecteur(s) de termes ou par des classes (hiérarchiques ou pas). Cependant la démarche de construction commune à tous les systèmes est la suivante :

1. Acquisition et collection des informations sur l'utilisateur à partir de sources d'informations diverses,
2. Apprentissage de ces informations en appliquant des techniques et des algorithmes

On dispose de deux manières afin d'obtenir les informations sur les utilisateurs,

1. **L'approche explicite** : consiste à obtenir les informations directement de l'utilisateur en utilisant une source externe pour créer et / ou compléter le modèle.
2. **L'approche implicite** : largement motivée par les travaux actuels dans le domaine, implique l'exploitation des données de comportement de l'utilisateur pour inférer son profil. (Acquisition incrémentale).

Nous détaillons ces deux approches dans ce qui suit, puis nous dresserons un bilan comparatif et nous donnerons une synthèse des principaux travaux dans ce domaine

1) L'acquisition explicite

Cette technique constitue une approche simple pour obtenir des informations sur l'utilisateur. On interroge directement l'utilisateur ou on lui demande par exemple de remplir des formulaires pour collecter les données personnelles et démographiques tels que sa date de naissance, son statut marital, son activité professionnelle et ses centres d'intérêts.

Dans le cadre de l'accès personnalisé à l'information, l'approche explicite est assimilable au feedback explicite, largement utilisé dans les systèmes de filtrage et de reformulation de requête. En effet, l'utilisateur émet directement son jugement d'intérêt en donnant une valeur de pertinence sur une échelle graduée allant du moins intéressant au plus intéressant.

L'acquisition explicite a été largement utilisée dans les systèmes de E-commerce pour personnaliser l'interface des sites web en fonction des préférences des utilisateurs

Ce procédé d'acquisition direct des données est sûr et souvent utilisé pour déclencher la collecte implicite d'informations. Cependant cette méthode induit un désintéressement et l'abandon de l'utilisateur, ce qui en résulte une détérioration de l'efficacité du système de recherche. Ceci est dû à la surcharge cognitive de l'utilisateur engendré par l'obtention répétitive et fréquente du jugement de pertinence. De plus la construction du profil dépend fortement du degré d'implication de l'utilisateur : si l'utilisateur ne fournit pas volontairement les informations, aucun profil ne sera construit.

2) L'acquisition implicite

Certains systèmes sont misent sur un processus de découverte implicite des centres d'intérêts de l'utilisateur de manière interactive et incrémentale. Le fonctionnement de base de cette approche est réalisé par l'établissement d'un dialogue entre le système et l'utilisateur ou mieux encore en observant son comportement à travers ses différentes interactions avec le système pour récolter discrètement l'information nécessaire sur lui.

Le principal avantage de cette approche est qu'elle ne nécessite aucune implication directe de l'utilisateur, ni de temps passé à émettre des jugements, ni un effort d'attention particulier lors de sa recherche. En effet, toute interaction de l'utilisateur avec le système est considérée comme une estimation de son jugement d'intérêts, tel que l'explique [OUA 01] qui stipulent que :

"Implicit feedback (method) may bear only an indirect relationship to the user's assessment of the usefulness of any individual document"

Basé sur la durée de l'interaction avec le système, le procédé d'acquisition implicite du profil d'utilisateur peut être classifié comme suit :

- Un modèle peu profond basé sur l'observation du comportement d'interaction relativement à court terme avec un système ; il ne tient pas compte des interactions de l'utilisateur avec le système durant les sessions précédentes.
- Un modèle profond qui observe le comportement de l'utilisateur durant son interaction à long terme avec un système. Il tient compte de l'historique entier des interactions de l'utilisateur avec le système.

La difficulté de ce type de techniques est la définition d'un processus d'interprétation du comportement observé dans un contexte d'application spécifique. Le processus doit être capable d'inférer les prétentions (intention, centres d'intérêts, préférences) de l'utilisateur. Le plus souvent ces prétentions sont incertaines et doivent être validées par l'utilisateur ou par des observations du comportement à plus long terme.

Nous exposons dans ce qui suit les principaux comportements observables de l'utilisateur lors de son interaction avec les SRIP. Puis, nous abordons les différentes sources d'information pour la construction du profil utilisateur.

Par la suite, nous allons citer quelques méthodes couramment utilisées pour l'extraction des informations sur les utilisateurs:

- **L'observation directe** : il s'agit de la méthode la plus précise. Elle permet d'identifier des classes d'utilisateurs ainsi que les tâches de ces derniers. De plus, on peut identifier des facteurs critiques, comme la pression sociale, qui ont des effets néfastes sur l'utilisation du système. Malheureusement, il s'agit d'une méthode très coûteuse qui nécessite des personnes qualifiées derrière chacun des individus observés.
- **Les interviews** : cette technique permet d'obtenir un autre type d'information, l'expérience, les opinions, les motivations comportementales mais surtout les avis sur les outils existants. Ils sont plus courts et moins coûteux que la technique d'observation, néanmoins, ils nécessitent aussi du personnel qualifié.
- **Les questionnaires** : ou « comment obtenir à moindre coût un maximum de données » ? [BOU 05] les résultats obtenus permettent des études statistiques et des généralisations plus fortes que les interviews. Les questionnaires peuvent être collectés par des personnes non expérimentées. Ils permettent d'avoir à la fois un aperçu de la situation et des points d'information plus précis. La situation idéale, qui est d'utiliser à la fois ces trois méthodes a déjà été pratiquée – [BRA 99] – mais est rarement justifiée, le plus souvent une seule de ces techniques suffit à rassembler les informations nécessaires pour le modèle utilisateur d'une application.
- **L'apprentissage** : il est possible d'obtenir de l'information par l'intermédiaire d'un outil d'apprentissage. En effet, lorsqu'il existe déjà un système, il est intéressant, et très utile s'y ajouter un algorithme d'apprentissage pour obtenir des informations essentiellement comportementales sur l'utilisateur [BRA 96].

L'unité sur laquelle se porte l'observation

		Segment	objet	classe
Catégories des comportements	Examiner	Regarder Ecouter Défiler Trouver Soumettre une requête	Sélectionner	Naviguer
	Retenir	Imprimer	Marquer Sauvegarder Supprimer Envoyer un mail	
	Référencer	Copier - Coller	Répondre Ajouter un lien Citer	
	Annoter	Marquer	Juger Publier	Organiser
	Créer	Taper, Editer	Autre	

Tableau 2 La catégorisation du comportement utilisateur selon []

2.1) Sources d'information pour l'acquisition implicite

Différentes sources d'information sont utilisées par les systèmes d'accès pour construire le profil de l'utilisateur. Les informations sont principalement issues des interactions utilisateur-système lors de l'accès de ce dernier. On distingue principalement deux types de sources d'information :

L'historique des interactions (ou de navigation) : est obtenu généralement par le système en scrutant les actions effectuées sur la page web tel que les annotations, le référencement et sauvegarde. Il peut contenir les URLs des pages web consultées par l'utilisateur, ainsi que la date et le temps de chaque consultation dans le but d'extraire des données statistiques sur le comportement et préférences de l'utilisateur (tel que le nombre de consultation et le temps de lecture).

La technique la plus communément utilisée par la majorité des systèmes se base sur l'utilisation d'un serveur proxy pour observer les interactions de l'utilisateur avec le système. Le principal avantage est qu'elle ne nécessite qu'une implication minimale de l'utilisateur et aucune installation de logiciel. Cependant, peu de profils peuvent être construits car l'accès s'effectue souvent de manière anonyme par différents utilisateurs sur différents emplacements géographiques. Une alternative possible est de s'enregistrer sur le système. De ce fait, le profil sera alimenté par son historique à chaque connexion. Peu importe son emplacement géographique le système lui fournira un accès personnalisé en fonction de son profil [TRA 04, BAR 97, PRE 99].

D'autre part, les approches de personnalisation peuvent employer des agents (directement installés sur la machine de l'utilisateur, intégrés de manière autonome dans le système d'accès ou dans les navigateurs Web standards) pour capturer l'ensemble des activités effectuées par l'utilisateur parallèlement à sa recherche. Letizia [LIE 95] est l'un des premiers systèmes de personnalisation ayant exploité les feedback implicites à l'aide d'agents.

Basé sur les pages précédemment visitées et annotées par l'utilisateur, il suggère des liens intéressant l'utilisateur sur sa page courante. Plusieurs autres systèmes d'aide à la navigation ont suivi tels que Web-Mate [CHE 98], Vistabar [MAR 97], et Personal WebWatcher [MLA 98]. D'autres approches de personnalisation mettent en place des toolbars pour aider l'utilisateur à organiser l'information contenue sur son ordinateur. Parmi ces systèmes on trouve Seruku Toolbar et SurfSaver.

L'historique de recherche constitue une source d'information pour la construction du profil de l'utilisateur dans le cadre d'une activité de RI. Il contient des informations issues des requêtes de l'utilisateur (requêtes, documents résultants, dates et durée des requêtes). Le SRIP peut également exploiter les URLs et les résumés ou les documents retrouvés par le système en réponse à la requête, ayant été sélectionnés par l'utilisateur. Le principal avantage de cette ressource est qu'elle ne nécessite aucune installation logicielle.

En effet, le SRIP collecte directement l'historique de recherche de l'utilisateur à partir de son interaction avec les résultats de recherche. Si le système nécessite l'enregistrement de l'utilisateur, son profil sera alimenté à chaque connexion, et un même profil sera exploité à chaque nouvelle recherche. Cependant, le risque potentiel est que les informations collectées de l'utilisateur seront issues uniquement de l'historique des résultats de recherche retournées par le SRIP lui-même.

Ce qui constitue une source d'information moins générale que dans les approches précédentes. Néanmoins, plusieurs projets [LIU 04; SPE 04] ont obtenu de très bonnes performances de recherche personnalisée basées sur des profils d'utilisateur construits à partir de l'historique de recherche.

Dans le cas où l'accès n'est pas limité au web, le système inclut des sources d'information plus générales telles que les bases de données utilisées par l'utilisateur, ses documents personnels disponibles dans son environnement de travail. Cette approche est mise en application par des outils tels que Google Desktop et SIS (Stuff I've Seen) [DUM 03]. En effet, le profil de l'utilisateur est alimenté implicitement à partir de bases de données et de ses fichiers personnels.

Dans la même perspective le projet Haystack [ADA 99, CHI 00] fournit une infrastructure pour créer un environnement personnalisé : une base de données pour stocker tous les documents de l'utilisateur, un système de gestion de base de données, et un module d'apprentissage du profil utilisateur.

1.8 Synthèse des techniques implicites

Le tableau 3 résume les différentes techniques d'acquisition implicites. Il présente en particulier le type de source d'information collectée, l'avantage et l'inconvénient de la technique et son applicabilité dans le système.

Technique	Source	Applicabilité	Avantage-Inconvénient	Système
Historique du navigateur	Historique de recherche	Sites <i>web</i>	(+) Aucune installation (-) sauf mise à jour	OBIWAN [138]
Serveur proxy	Historique de recherche	Sites <i>web</i>	(+) Utilisation de navigateur standard et (-) d'un serveur proxy	OBIWAN [138] [181] [9]
Agent de recherche	Historique de recherche	Accès personnalisé	(+) L'agent collecte toute activité utilisateur (-) Installation de l'agent et utilisation de l'application spécifique lors de la recherche	Letizia [110] WebMate [39] Vistabar [118] Personal WebWatcher [125]
Agent Desktop	Toute activité utilisateur	Accès personnalisé	(+) Tous les fichiers et activité de l'utilisateur (-) Installation de l'agent	Seruku [156] Surfsaver [170] Haystack [1] [40] Google Desktop [55] Stuff I've Seen [57]
Les logs de navigation (Cookies)	Toute activité utilisateur	Le site <i>web</i>	(+) Information collaborative de plusieurs utilisateurs (-) Information minimale issue d'un seul site <i>web</i>	Mobasher [126]
Les logs des requêtes (Cookies)	Résultats de recherche	Moteur de recherche	(+) Information issue d'un même moteur (-) Information minimale issue d'un seul site <i>web</i> (-) Activation des Cookies à chaque connexion	Misearch [160] [114]

Tableau 3 Synthèse des techniques d'acquisition implicite

1.9 Indicateurs pour la prédiction des centres d'intérêts

La détection implicite des préférences et des centres d'intérêts se fait à travers les comportements observables collectés par le système lors de l'interaction de l'utilisateur avec l'environnement (session de travail, consultation d'un document clé ...etc.).

Dans ce contexte plusieurs comportements peuvent être examinés, tels que :

- Clic de la souris ;
- Défilement en utilisant la souris ou le clavier ;
- Temps passé sur une page Web ;
- Temps de lecture d'un document (au de la d'un seuil 20s) ;
- Impression du document ;
- Sauvegarde d'un document ;
- Copier/coller tout ou une partie d'un document ;
- Annotation de document ou de quelque uns de ses passages ;
- Suivi des liens hypertextes pour arriver à une page Web ou à un document ;
- Visualisation et consultation d'un document.
- Enregistrement des traces d'interaction

Les traces des interactions de l'utilisateur avec le système sont construites à partir de données qui sont produites par les fichiers logs (journal). Ces fichiers enregistrent discrètement les actions de l'utilisateur durant l'utilisation des logiciels ou les traces d'interactions de celui-ci avec l'environnement de travail. Ces fichiers logs ont plusieurs caractéristiques qui font d'eux un moyen idéal pour le domaine de la recherche en vue de concevoir par exemple des interfaces homme/machine, adaptées. Ils permettent la collection des données sur de nombreux utilisateurs à travers le temps et à chaque fois que le logiciel est utilisé.

Pour chaque évènement un ensemble de descripteurs est sauvegardé, la cible, le contenu, le contexte de l'évènement et les valeurs descriptives des composants ayant contribué dans l'évènement, tels que le nom du bouton ayant été appuyé, son texte descriptif,...etc.

Dans le cadre de l'environnement d'un navigateur Web, les évènements enregistrés seraient le mouvement de la souris, le clic sur un bouton (avec son nom « File », « Back », « Forward », « Reload »), l'usage de la barre de défilement (« Up », « Down »), le clic sur un emplacement du texte ou dans un endroit particulier de la page Web, le clavier avec un ensemble de touches utilisées, le clic sur des liens ou la fermeture du navigateur ou du document.

1.10 Gestion du profil

a) Représentation du profil

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'existe pas un 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.

b) Construction du profil

La construction des profils se décompose en 2 étapes :

➤ *La modélisation de profil :*

La modélisation est l'une des premières fonctionnalités que doit offrir un système de personnalisation considérant le profil comme une forme de connaissance sur les habitudes et préférences de l'utilisateur. L'existence d'une méta base facilitera l'élaboration de ces connaissances en offrant une typologie de connaissance et les mécanismes de spécialisation et d'instanciation qui permettent d'adapter totalement ou partiellement la méta base au besoin de personnalisation de l'utilisateur et/ou des applications.

➤ *Instanciation du profil :*

C'est de prendre le Meta modèle, et de lui affecter des valeurs précises en fonction de la connaissance actuelle que le système a de l'utilisateur. Ces valeurs peuvent donc s'affiner au fur et à mesure que cette connaissance augmente. Ceci peut être fait par le module de génération automatique, par apprentissage ou datamining par exemple, ou manuellement par l'utilisateur via l'interface graphique [BOU 99].

c) Evolution du profil

➤ *Mise à jour des profils*

Le gestionnaire de profils offre des fonctionnalités de mise à jour classique qui sont : l'insertion, la suppression et la modification de la structure et des valeurs des profils. La seule règle à respecter est que le modèle de profil générique doit inclure l'union des structures de toutes les instances de profils.

Pour maintenir le profil générique le plus complet possible, chaque insertion d'une nouvelle dimension, sous dimension ou attribut dans une instance de profil se fait par le modèle générique. Par exemple si un utilisateur veut ajouter un attribut à ces données personnelles, il est obligé de créer cet attribut dans le modèle générique de profil utilisateur pour ensuite le recopier dans son profil. De cette manière la typologie des données est préservée (par exemple le nom d'un utilisateur est une chaîne de caractères dans l'ensemble des instances de profils utilisateurs) [BOU 99].

➤ *Archivage des profils*

L'archivage de profils permet de sauvegarder le contenu d'une instance de profil. Lors de cette sauvegarde, la date de création de l'archive est sauvegardée avec l'ensemble des valeurs du profil. Plusieurs archives peuvent être créées pour le même profil ce qui permet de suivre l'évolution des préférences d'un utilisateur ou d'analyser son comportement.

Les archives peuvent être visualisées et il est possible de remplacer le contenu actuel du profil par une de ses archives. Ceci peut être très utile si on veut invalider une mise à jour non pertinente du profil utilisateur et revenir à une version précédente [BOU 99].

1.11 Les inconvénients de quelques systèmes gestion du profil

a) Le Munich Model

Le modèle Munich [BAL 97] est composé d'une classe gestionnaire d'utilisateurs, qui est composée d'un ou de plusieurs utilisateurs. Un utilisateur possède des attributs nom d'utilisateur et email. Il contient également d'autres attributs, potentiellement quelconques. Les attributs sont divisés en deux, ceux qui dépendent du domaine, et ceux qui n'en dépendent pas. Un attribut est constitué d'un nom et d'une valeur.

L'identifiant est unique pour un utilisateur dans toute l'application. Les attributs dépendant du domaine sont liés aux composants du domaine. Ce modèle a une seule catégorie d'attributs indépendants du domaine. Dans notre cas nous en avons plusieurs, afin de pouvoir raisonner différemment sur différentes sortes d'attributs. Ce modèle ne permet pas de conserver la trace des documents parcourus : il n'y a pas d'historique, dans notre cas si.

b) Modèle AHAM

Le modèle de l'utilisateur d'AHAM est relativement peu formalisé [BAL 97]. Il reste très général. Dans AHAM, le modèle de l'utilisateur est constitué d'attributs, qui correspondent le plus souvent à la connaissance de concepts. Pour chaque entité, un certain nombre de paires attribut-valeur sont stockées, comme, par exemple, le fait qu'un concept ait été lu ou le niveau de connaissance d'un concept. Les données relatives aux utilisateurs sont stockées dans une table.

Cette table contient l'état des connaissances que l'utilisateur a du domaine, ainsi que ses préférences si nécessaire. La façon dont les caractéristiques des utilisateurs sont mises en relation avec les autres éléments du modèle est implicite. Il n'y a pas d'historique. Il n'y a pas d'attributs prédéfinis.

2 Les Styles d'Apprentissage

2.1 Définition

Confrontés à une même situation, les apprenants présentent des réactions diversifiées, parfois très contrastées. Pour tenir compte de ces différences individuelles, les recherches se sont orientées vers la notion de « style d'apprentissage » relative à la manière préférée d'apprendre pour un individu (individu au sens large, sans se limiter aux apprenants).

Cependant, définir les styles d'apprentissage n'est pas simple. Un survol de la littérature met rapidement en évidence la pluralité et la diversité des définitions [LAM 04].

Riding, R., et Rayner dans [RID 98] nous donne une définition un peu générale: « Le terme style d'apprentissage renvoie à un ensemble individuel de différences qui incluent non seulement une préférence personnelle exprimée concernant l'enseignement ou une association avec une forme particulière d'activité d'apprentissage, mais aussi à des différences individuelles que l'on retrouve en psychologie de l'intelligence ou de la personnalité »

Le style d'apprentissage est donc relié à la fois à un ensemble de conduites - stratégies dans la manière de gérer et d'organiser l'information ainsi qu'à la manière de mettre en œuvre ces conduites et stratégies.

2.1 Quelques modèles de styles d'apprentissage

Diverses théories de styles d'apprentissage ont été élaborées avec une fréquence accrue au cours des dernières décennies. [COF 04] recensent 71 modèles, parmi lesquels 13 sont considérés les plus importants en raison de leurs apports, leurs larges usages et leurs influences sur d'autres modèles. Parmi ces derniers, cette section présente certains modèles que nous avons sélectionnés, vu leur implémentation dans des EIAH.

1. Modèle de Kolb

Le modèle de style d'apprentissage de [KOL 84] est basé sur sa théorie du processus d'apprentissage expérientiel en quatre étapes [KOL 74]: expérience concrète, observation réfléchie, conceptualisation abstraite et expérimentation active. L'expérience concrète est la base des observations et des réflexions. Ces observations sont utilisées pour former des concepts abstraits et des généralisations, ce qui agit en tant que base pour l'expérimentation de ces nouveaux concepts dans des situations nouvelles. La mise en œuvre des expérimentations résulte en l'expérience concrète, qui clôt le cycle d'apprentissage. Sur la base de ce processus, [KOL 74] suggère l'existence de deux dimensions : concret/abstrait, actif/réfléchi, qui, combinés deux à deux, forment quatre styles d'apprentissage [CHE 00] :

- le style *divergent* (concret-réfléchi) se caractérise par l'interprétation de situations concrètes de différents points de vue,
- le style *assimilateur* (réfléchi-abstrait) s'explique par l'appropriation d'une gamme étendue d'informations et leur intégration concise et logique,
- le style *convergent* (abstrait-actif) se caractérise par la recherche d'applications pratiques aux concepts et aux théories,
- le style *accommodateur* (concret-actif) présente de fortes capacités de mise en œuvre d'expériences pratiques et l'implication personnelle dans de nouvelles expériences comportant un défi.

Pour mesurer ces styles, Kolb propose le *Learning Style Inventory*, un questionnaire qu'il perfectionne depuis sa proposition en 1976, jusqu'à sa dernière version en 2005 [KOL 05].

Ce questionnaire est destiné aux adultes et adolescents. Il comporte 12 questions relatives aux préférences d'apprentissage, chacune ayant 4 réponses possibles. Enfin, ce modèle a inspiré d'autres modèles de styles d'apprentissage à travers sa théorie d'apprentissage expérientiel en quatre étapes tels que les modèles de [GRE 79], [MCC 81] et [HON 86]. Ce dernier est présenté dans la section suivante.

2. Modèle de Honey et Mumford

S'inspirant de Kolb, [HON 86,92] développent leur modèle de styles d'apprentissage qui s'appuie aussi sur chacune des étapes du processus d'apprentissage expérientiel mais sans supposer l'existence de dimensions bipolaires. Ils définissent ainsi quatre styles d'apprentissage relatifs aux quatre étapes : actif, réfléchi, théoricien et pragmatique.

- Le style *actif* se caractérise par un engagement dans l'expérience du moment présent et une préférence pour apprendre à partir d'expériences nouvelles et de situations problèmes. Il se caractérise également par l'enthousiasme, la flexibilité mais les *activistes* prennent parfois des risques inutiles, à se lancer dans l'action sans suffisamment de préparation.
- Le style *réfléchi* se caractérise par un recul face aux situations, un désir de les examiner selon différents points de vue et une préférence pour apprendre à partir d'activités exigeant de réfléchir, d'analyser et de pondérer une quantité d'informations. Les *réfléchis* sont méthodiques, mais ont parfois tendance à être trop prudents et de ne pas prendre assez de risques, et peuvent être lents à se faire une opinion et de parvenir à une décision.
- Le style *théoricien* se caractérise par un besoin de situer et d'intégrer les informations dans un cadre conceptuel, une structure, un modèle ou une théorie. Les *théoriciens* sont très logiques, rationnels et objectifs. Ils ont une préférence d'apprendre à partir d'activités où des modèles sont présentés et où il est possible d'en construire. Ils ne tolèrent pas l'incertitude, le désordre ou l'ambiguïté.
- le style *pragmatique* se caractérise par l'application pratique d'idées, de théories et de procédures et par une préférence pour apprendre des activités où il y a des liens entre les connaissances et la vie réelle et où il y a possibilité de mettre en pratique ces connaissances. Les apprenants *pragmatiques* sont plus axés sur les tâches et ont tendance à refuser des théories sans une application évidente.

Pour mesurer ces quatre styles, Honey et Mumford proposent le Learning Styles Questionnaire (LSQ), révisé en plusieurs versions [HON 06].

Ce questionnaire existe maintenant en deux versions : une à 40 questions et une à 80 questions. Pour chaque question deux choix sont possibles : vrai ou faux. Il a été adapté en français par LSQ-F [FOR 97].

3. Modèle de Dunn et Dunn

Le modèle de style d'apprentissage de Dunn et Dunn a été proposé initialement en 1974, puis affiné au fil des années [DUN 78] [DUN 03]. Le modèle adopte une perspective de construction d'un profil d'apprentissage de l'apprenant et comprend cinq variables qui peuvent être sources de différences individuelles, chacune est composée de plusieurs facteurs [CHE 00] :

1. les facteurs *environnementaux* (son, lumière, température et design),
2. les facteurs *affectifs* (motivation, persistance, responsabilité et structure),
3. les facteurs *sociologiques* (apprendre mieux seul, avec un autre, en équipe, avec un adulte ou de manière variée),
4. les facteurs *physiologiques* (modalités perceptives efficaces visuelles, auditives, tactiles/kinesthésiques, fluctuation du niveau d'énergie selon le moment de la journée, besoin de nourriture et de mobilité pendant l'apprentissage)
5. et enfin les facteurs *psychologiques* (traitement global vs analytique, degré de spécificité hémisphérique et fonctionnement réfléchi vs impulsif).

Pour mesurer ces variables, Dunn et Dunn ont élaboré le LSI (Learning Styles Inventory) qui existe en plusieurs versions : trois versions pour les enfants (de la maternelle à la 2^{ème} année, 3^{ème} et 4^{ème}, les classes de 5 à 12) avec 104 questions qui emploient 3 ou 5 choix [DUN 96] ; et le Building Excellence Inventory [RUN 00] pour les adultes avec 118 questions [GRA 07].

4. Modèle de Witkin

Parmi les modèles de styles d'apprentissage ayant une grande notoriété dans la communauté, nous pouvons citer celui de [WIT 71]. Ce modèle a été largement étudié vu qu'il reflète la capacité des apprenants à structurer l'information en se basant sur leur perception de l'environnement et du domaine [COF 04]. Dans ce sens, ce modèle différencie les apprenants selon qu'ils soient dépendants ou indépendants à l'égard du champ (ou du domaine) [CHE 08]:

- Les apprenants *indépendants du champ* ont tendance à présenter des comportements plus individualistes, car ils n'ont pas besoin d'aide externe pour le traitement de l'information. Ils sont plus susceptibles de développer leurs propres références et à restructurer leurs connaissances.

Par conséquent, ils ne sont pas facilement influencés par les autres, et sont plus analytiques dans leur approche d'apprentissage.

- Les apprenants *dépendants du champ* ont une orientation plus sociale que les apprenants indépendants, car ils sont plus susceptibles de solliciter des références et de l'aide externes. Ils sont plus facilement influencés par les opinions des autres, et ont tendance à percevoir l'image globale dans leur approche d'apprentissage.

Pour mesurer ces deux styles, plusieurs outils ont été développés. Parmi ces outils, le test de figures imbriquées GEFT (Group Embedded Figures Test) proposé par [WIT 71], est un livret de trois parties comportant des figures géométriques de plus en plus complexes dans lesquels il est demandé de retrouver des formes simples. Le score du sujet est le nombre de figures simples correctement identifiées dans les formes complexes. Le score total maximum est de 18. Plus le score est élevé, plus le sujet est considéré comme indépendant du domaine. Ce test est reconnu comme étant fortement corrélé aux performances académiques. En effet les étudiants en cursus scientifiques ont souvent de meilleurs scores que des étudiants en cursus littéraires [TIN 97 ; CLA 00 ; PIT 02] dans [TYN 05]. Nous pouvons également citer CSA (Cognitive Styles Analysis) proposé par [RID 91] qui est composé de deux parties. La première consiste à juger si des formes complexes sont semblables ou différentes, la seconde partie contient le même principe que le GEFT [CHE 08].

5. Modèle de Felder et Silverman

Parmi les modèles de styles d'apprentissage les plus utilisés, celui proposé par [FEL 88]. Ce modèle est défini en répondant aux cinq questions suivantes :

1. Quel type d'information l'apprenant préfère-t-il percevoir : *sensoriel* (externe), des sons, des sensations physiques ; ou *intuitif* (interne), des possibilités, des idées, des intuitions ?
2. A travers quelle modalité sensorielle l'information externe est-elle efficacement perçue : *visuelle*, des images, des diagrammes, des graphiques, des démonstrations ; ou *auditive*, des mots, des sons ?
3. Avec quelle organisation de l'information l'apprenant est-il le plus à l'aise : *inductive*, à partir des faits et des observations les principes sont déduits, ou *déductive*, allant des principes les conséquences et les applications sont déduites ?

4. Comment l'apprenant préfère-t-il traiter l'information : *activement* par l'engagement dans des activités physiques ou des discussions ; ou de manière *réfléchie*, grâce à l'introspection ?
5. Comment l'apprenant progresse-t-il pour comprendre : de façon *séquentielle* avec des petits pas dans un ordre linéaire ; ou de manière *globale* avec de grands sauts dans un ordre aléatoire, jusqu'à ce qu'il saisisse l'ensemble du concept ?

Ainsi, dans cette première version du FSLSM (Felder and Silverman Learning Style Model), le modèle comprenait 5 dimensions : perception (sensoriel/intuitif) tirée de [MYE 62] et aussi analogue à la dimension concret/abstrait du modèle de Kolb [KOL 84]; réception (visuel/auditif), organisation (inductif/déductif), traitement (actif/réfléchi) définie aussi dans le modèle de [KOL 84]; et compréhension (séquentiel/global) fondée sur le modèle de style d'apprentissage de [PAS 76].

Ce modèle a été réduit en quatre dimensions dans [FEL 95], par la suppression de la dimension organisation (inductif/déductif), et le changement de nom de la dimension réception (visuel/auditif) vers visuel/verbal pour renvoyer aux représentations textuelles, indépendamment du fait qu'elles soient écrites ou parlées. Les raisons de ces changements sont détaillées dans [FEL 02].

Bien que ces dimensions ne soient pas nouvelles, la façon de décrire les styles d'apprentissage est innovante. En effet, ce modèle utilise des échelles de -11 à +11 pour chaque dimension des styles, en ne considérant que les valeurs impaires. Ces échelles facilitent la description des préférences de style d'apprentissage plus en détail, comparée à la construction de types d'apprenants, comme dans les autres modèles.

De plus, Felder et Silverman considèrent ces préférences comme des tendances, ce qui signifie que même un apprenant avec une forte préférence pour un style d'apprentissage peut parfois agir différemment. Par ailleurs, [FEL 88] proposent pour chacune des dimensions des styles d'apprentissage les styles d'enseignement qui correspondent le mieux, afin d'être utilisés comme des recommandations pour la personnalisation des enseignements aux préférences des apprenants, ce qui explique sa forte popularité.

Enfin, pour identifier les styles d'apprentissage selon FSLSM, [FEL 96] ont développé ILS (Index of Learning Styles), un questionnaire de 44 questions (11 questions pour chaque dimension).

Les préférences sont exprimées avec des valeurs impaires comprises entre -11 et +11 par dimension, pour exprimer une faible, moyenne, ou forte préférence.

6. Synthèse

A travers l'étude des modèles de styles d'apprentissage existants dans la littérature et à partir de ceux présentés dans cette section, nous constatons que pour chacun, la définition des styles d'apprentissage revient à tenir compte des facteurs qui ont à la fois un rôle à jouer dans le processus d'apprentissage de l'apprenant et engendrent des différences individuelles importantes et pertinentes. Pour mesurer ces préférences, chacun propose un questionnaire, avec des questions à choix multiple ou unique, ou la reconnaissance de formes, dont le nombre est souvent important.

Notons également que les auteurs des modèles présentés ci-dessus, sauf pour Witkin et al. (1971), considèrent que le style d'apprentissage n'est pas stable, mais plutôt des préférences qui changent en fonction des situations, ou des périodes, mais qui peut être relativement stable pour une période donnée [KOL 00]. Nous soutenons cette supposition de non stabilité des styles, que nous utilisons plus loin dans cette thèse.

Enfin, vu la multitude des modèles proposés dans la littérature, plusieurs classifications ont été proposées, point que nous détaillons dans la section suivante.

2.3 Styles d'apprentissage et enseignement

Plusieurs recherches en psychologie et en science de l'éducation affirment l'impact du style d'apprentissage sur le processus d'apprentissage et encouragent sa prise en compte dans les stratégies d'enseignement afin de faciliter la tâche aux apprenants et améliorer leurs résultats.

Dans ce sens, chaque modèle de style d'apprentissage suggère un ensemble de recommandations relatives aux stratégies et aux types de ressources pédagogiques qui doivent être utilisés avec les apprenants, se rapportant à leurs préférences en termes de styles d'apprentissage. FSLSM par exemple, exprime clairement les styles d'enseignement nécessaires pour chaque dimension des styles d'apprentissage [FEL 88].

En outre, Felder signale que l'inadéquation des styles d'enseignement aux styles d'apprentissage peut avoir de graves conséquences [FEL 93] : « les apprenants ont tendance à obtenir des résultats inférieurs que ceux dont les styles d'apprentissage sont mieux adaptés au style d'enseignement et sont moins susceptibles de développer un intérêt au cours » [POP 08].

[DUN 03] suggèrent également d'adapter l'environnement d'apprentissage pour permettre aux apprenants de travailler avec leurs fortes préférences. D'autres psychologues soutiennent un point de vue opposé : utiliser une variété de styles d'enseignement et fournir des ressources et un environnement d'apprentissage qui ne sont pas forcément adaptés aux styles d'apprentissage peuvent aider les apprenants à développer de nouvelles stratégies d'apprentissage, à recouvrir leurs faiblesses et à développer des compétences, qui ne font pas partie de leurs préférences afin de se préparer à de futures situations auxquelles ils sont tenus à s'adapter [GRA 84][KOL 84][APT 01].

Par exemple, en fournissant aux apprenants ayant des préférences verbales que des ressources visuelles, cela les oblige à développer et à utiliser les compétences visuelles. Pour [GRA 84] cette approche est pertinente pour rendre l'apprentissage intéressant et stimulant pour les apprenants. Pour [KOL 84] les objectifs éducatifs de cette approche sont relatifs au développement personnel et à la créativité.

Pour résumer, nous pouvons conclure que l'intégration des styles d'apprentissage au processus de formation est bénéfique. Elle permet d'accroître l'efficacité de l'apprentissage, que ce soit dans une approche d'adéquation ou d'inadéquation des styles d'apprentissage aux enseignements.

L'approche d'adéquation peut être utilisée pour l'adaptation et la personnalisation de la formation rendant la tâche d'apprentissage plus facile aux apprenants. Tandis que l'approche d'inadéquation peut être par exemple utilisée pour déstabiliser ponctuellement les apprenants afin de les inciter à développer leurs capacités.

Une première étape serait donc de rendre les apprenants conscients de leurs styles d'apprentissage afin d'acquérir un savoir métacognitif sur leurs différences individuelles.

En effet, plusieurs études affirment que généralement les étudiants ignorent leurs styles d'apprentissage, et ont du mal à l'exprimer [STA 06] [MER 02]. Ce savoir métacognitif permet donc à l'apprenant d'avoir une réflexion sur ses méthodes et préférences d'apprentissage pour mieux comprendre ses difficultés, autoréguler son apprentissage, prendre conscience de l'existence d'autres manières d'apprendre et de s'exercer à les apprivoiser.

Dans le même ordre d'idées, force est également de constater que le plus souvent « l'enseignant façonne son enseignement selon sa manière d'apprendre » [KIN 95]. C'est pourquoi, le fait de décrire ces différences peut amener l'enseignant à les reconnaître, à les comprendre, et à être plus en mesure de les prendre en compte.

2.4 Styles d'apprentissage et EIAH

L'utilisation de l'informatique en éducation apporte chaque jour de nouvelles perspectives. Les recherches dans le domaine des Environnements Informatiques d'apprentissage Humain (EIAH) portent sur les principes de conception, de développement et d'évaluation de systèmes informatiques qui permettent à des êtres humains d'apprendre.

L'objectif de tels systèmes est de venir en complément, et/ou de répondre aux problèmes rencontrés dans une situation classique de formation en termes de distance, assistance, adaptation, personnalisation, individualisation et suivi. Pour cela, l'identification des caractéristiques relatives à la connaissance, aux intérêts, aux objectifs, aux prérequis et aux traits individuels s'avère indispensable [BRU 07].

Parmi les traits individuels considérés, plusieurs recherches en EIAH s'intéressent ces dernières années à la prise en compte du style d'apprentissage comme facteur clé dans le processus de formation.

L'intégration de cette théorie dans un environnement informatique permet de bénéficier des avantages cités dans la section précédente et apporte également la possibilité de considérer le style d'apprentissage de chaque apprenant individuellement par l'adaptation des contenus, en terme de forme, de structure, d'ordre de présentation des activités d'apprentissage et de choix de ces activités ; tâche difficile, voire impossible pour l'enseignant dans une situation traditionnelle de formation avec un groupe ou une classe d'apprenants.

A travers notre étude, et celles des travaux de [BRO 05] [GRA 07] [POP 08] [WOL 07], le tableau 4 établie par [BOU 11], qui présente une vue synthétique de quelques EIAH qui implémentent explicitement la théorie de styles d'apprentissage et les approches sur lesquelles ils fondent leurs préférences d'apprentissage. D'autres EIAH intégrant implicitement des préférences liées aux styles d'apprentissage tel que « iMANIC » [STE 01] ne sont pas intégrés à ce tableau.

Les systèmes présentés dans le tableau 4 sont des EIAH adaptatifs. Ces recherches s'appuient sur l'approche d'adéquation entre les styles d'apprentissage et les styles d'enseignement afin de fournir une formation personnalisée. D'autres systèmes utilisent les styles d'apprentissage également pour le suivi et pour apporter une assistance aux apprenants [SCH 08]. Par ailleurs, ces systèmes utilisent généralement l'une des dimensions ou l'un des modèles de styles d'apprentissage proposés dans la littérature.

EIAH	Modèle de styles d'apprentissage utilisé	Outil de mesure des styles
Arthur (Gilbert et Han, 1999; 2002)	(auditif, visuel et tactile) de (Sarasin, 1998)	Le style d'apprentissage est attribué aléatoirement en premier, ensuite les apprenants peuvent le modifier manuellement
SACS (Wang et al., 2008)	VARC (Flemming, 1995)	Le questionnaire VARK ou Mentionné volontairement par l'apprenant
AES-CS (Triantafillou et al., 2002)	Dépendant ou indépendant du domaine de Witkin (Witkin et al. 1971)	Le test (GEFT) (Witkin et al., 1971)
iWeaver (Wolf, 2002)	(Dunn et Dunn, 1978)	Building Excellence Inventory (Rundle and Dunn, 2000)
INSPIRE (Grigoriadou et al., 2001) (Papanikolaou et al., 2003)	(Honey et Mumford, 1992)	Questionnaire de (Honey et Mumford, 1992)
IDEAL (Shang et al., 2001)	Déterminer par l'enseignant	Questionnaire du style utilisé
AHA! (De Bra et al., 2003) (Stash et al., 2004) (Stash, 2007)	Déterminer par le concepteur du cours. Testé avec les styles divergeant / convergeant de (Kolb, 1984)	Les apprenants peuvent exprimer manuellement leurs préférences des styles d'apprentissage
CS383 (Carver et al., 1999)	Différentes dimensions du FSLSM (Felder et Silverman, 1988)	ILS (Felder et Soloman 1996)
CAMELEON (Laroussi et Benahmed, 1998)		
MASPLANG (Peña, 2004; Peña, et al., 2002)		
ILASH (Bajraktarevic et al., 2003.a)		
WHURLE (Brown et Brailsford, 2004)		
LSAS (Bajraktarevic et al., 2003.b)		
TANGOW (Paredes et Rodriguez, 2004)		

Tableau 4 Quelques EIAH qui implémentent les styles d'apprentissage [BOU 11].

Afin de faciliter la lecture de ce tableau, nous allons en premier présenter les systèmes qui utilisent le modèle de style d'apprentissage de Felder et Silverman, suivi par ceux utilisant les modalités de perception, par la suite les EIAH utilisant d'autres modèles de styles d'apprentissage, et enfin les systèmes proposant une architecture indépendante d'un modèle spécifique.

Pour chacun, nous précisons les préférences de styles d'apprentissage implémentées et l'approche utilisée pour les identifier. Enfin, nous finissons cette partie par une réflexion sur l'impact de l'usage des styles d'apprentissage en EIAH et les nouvelles tendances pour les modéliser et les identifier.

2.5 Relation entre style d'apprentissage et comportement sur le Web

Vu que la théorie des styles d'apprentissage a été conçue pour des situations traditionnelles d'apprentissage, et comme il n'existe pas de modèles spécifiques pour l'étude des documents numériques, plusieurs études ont été menées afin de montrer que même dans un tel environnement, les comportements des apprenants avec des hypermédias sont toujours influencés par leurs styles d'apprentissage.

En effet, au-delà des résultats positifs de la prise en compte des styles d'apprentissage dans les EIAH vus dans la section précédente, plusieurs études en psychologie, en science de l'éducation et aussi en informatique ont montré une forte relation entre les styles d'apprentissage des utilisateurs et leurs comportements de recherche et de navigation au sein des applications Web. Ces travaux suggèrent la prise en compte des différents styles d'apprentissage lors de la conception de ces environnements. De plus, certaines de ces études ont conclu qu'au-delà d'expliquer les comportements à travers les styles d'apprentissage, il est aussi possible de déduire les styles à partir de l'analyse de ces comportements [CHE 08].

Dans ce contexte, le modèle de style d'apprentissage de [WIT 71] (dépendance ou indépendance du domaine) a été le plus étudié, notamment en termes de format, d'accessibilité, de structure et de performance.

En ce qui concerne le format, plusieurs études suggèrent que les apprenants ayant un style indépendant du domaine pourraient en particulier bénéficier de la maîtrise des choix des médias (texte, animation, voix).

En ce qui concerne l'accessibilité, [FOR 00] ont mené une étude pour examiner la façon dont les styles d'apprentissage influencent les comportements de navigation des utilisateurs sur le Web.

En termes de structure, [DUF 97] [LEE 05] ont examiné les performances des utilisateurs dans des structures d'information linéaires et non linéaires. Ils ont constaté que les apprenants dépendants du domaine qui ont utilisé le système avec la structure non linéaire ont passé plus de temps à remplir le test que ceux qui ont utilisé le système avec une structure linéaire.

[BOU 11] a résumé les résultats de ces études dans le tableau 5. Les études présentées dans ce tableau confirment la forte influence des styles d'apprentissage sur les comportements.

Aspect	Utilisateurs Dépendants du domaine	Utilisateurs Indépendant du domaine
Format	Les options de média ne les affectent pas (Chuang, 1999).	Préfèrent plusieurs options de média. (Chuang, 1999)
	Les supports audio leurs sont bénéfiques (Lee, 1994)	Travaillent mieux avec les traitements de contrôle visuel (Chan-Lin, 1998)
	Les outils multimédias qui demandent aux apprenants d'extraire les indices eux-mêmes les embarrassent (Ghinea et Chen, 2003).	Préfèrent des ressources pédagogiques détaillées (Ghinea et Chen, 2003)
Accessibilité	Utilisent les outils qui les aident à trouver une information spécifique afin d'avoir une vision globale du contexte (ex. carte de site), préfèrent l'apprentissage par l'exemple, visitent plusieurs pages (Chen et Ford, 1998) (Ford et Chen, 2000) (Chen et al., 2004.a), (Chen et Liu, 2008), font plus de pas et prennent plus de temps que les apprenants indépendants (Reed et al., 2000) et ont tendance à suivre les liens prévus par la page Web, en particulier pour les novices (Palmquist et Kim, 2000).	Utilisent les outils qui les aident dans la navigation pour trouver l'information visée directement (ex. index, ordre alphabétique, les moteurs de recherche, la fonction rechercher, les URLs pour atteindre les sites Web, etc.) (Chen et Ford, 1998) [FOR 00]) (Palmquist et Kim, 2000), préfèrent les descriptions détaillées et visitent moins de pages. (Chen et al., 2004.a) (Chen et Liu, 2008)
Structure	Préfèrent un parcours linéaire [DUF 97] (Reed et Oughton, 1997) (Lee et al. 2005)	Ont tendance à suivre un parcours non linéaire [DUF 97] (Reed et Oughton, 1997) [LEE 05]
	Travaillent mieux avec la navigation en largeur (Ford et Chen 2001)	Travaillent mieux avec la navigation en profondeur (Ford et Chen, 2001)
	Dans une recherche, préfèrent que les sous catégories soient présentées en premier ensuite les résultats. (Chen et al., 2004.a) (Chen et Liu, 2008)	Dans une recherche, préfère avoir les résultats en premier suivis par les sous catégories. (Chen et al., 2004.a) (Chen et Liu, 2008)
	Nécessitent une structure claire, planifiée et du renforcement dans l'usage des répertoires Web (Chen et Macredie, 2004)	Peuvent bien travailler dans un environnement déstructuré (Chen et Macredie, 2004)
	Présentent de meilleurs résultats avec les pages courtes (Graff, 2003)	Présentent de meilleurs résultats avec des pages longues (Graff, 2003)
Performance	Utilisent avec un plus grand pourcentage les notes d'enseignement et les ressources de classes que les apprenants indépendants (Lu et al., 2003)	Dépassent ceux dépendants du domaine dans l'apprentissage avec des hypermédias (Cacciamani, 2002) (Korthauer et Koubek, 1994) (Lu et al., 2003) (DeTure, 2004)

Tableau 5. Résumé des études menées sur la relation entre les styles dépendants et indépendants du domaine et le comportement des utilisateurs avec des applications Web [BOU 11].

Conclusion

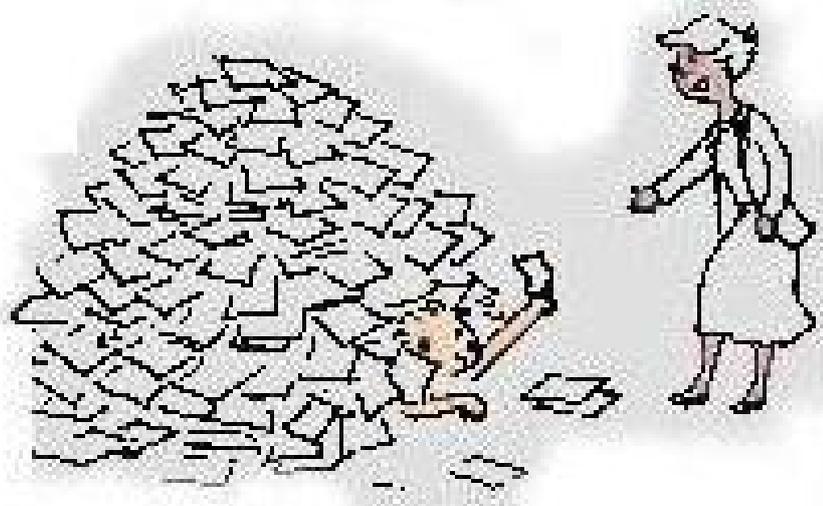
Nous avons présenté dans ce chapitre un état de l'art sur le profil utilisateur, en définition cette notion ainsi que la notion de préférence. Les données contenues dans le profil doivent être correctes et justes car elles ont un impact direct sur les résultats qui vont être obtenus. Le fait que l'utilisateur a souvent des idées floues sur ses préférences complique l'extraction des caractéristiques pertinentes et nécessite un processus de découverte de ses préférences. La mise à jour du profil est un des problèmes qui entre dans le cadre de nos objectifs.

Aussi nous avons présenté des études qui ont montrés que le domaine des styles d'apprentissage est complexe. Malgré le nombre important de travaux menés, plusieurs questions restent ouvertes, problématiques et discutables.

Chapitre

3

Conception



Introduction

La réussite d'un système E-Learning est liée, à la fois, au contenu de la formation et à la manière dont l'apprenant est mis en contact avec ce contenu. Par contenu de la formation, on entend : l'organisation et la modélisation du domaine de la formation (la structuration des modules suivis par l'apprenant en chapitres et les chapitres en notions, etc.).

Quant à l'apprenant, il doit disposer de moyens d'agir, car parmi les fondements du E-Learning : offrir un rôle actif à l'apprenant durant sa formation. Pour ce faire, il est indispensable de prendre en compte certaines informations concernant l'utilisateur telles que : buts, préférences, informations personnelles, ...etc.

Dans ce travail, nous envisageons d'aider l'apprenant à accomplir son parcours d'apprentissage par la conception d'un système de Question Réponse intelligent et qui sera par la suite doté d'un module de mise à jour automatiquement du profil de l'utilisateur.

Le présent chapitre est consacré à la présentation de l'approche proposée. La suite de ce chapitre est organisée de la manière suivante : tout d'abord, nous décrivons Le système Q/A proposé et ces fonctionnalités ainsi que la méthode LSA injecté. Puis, nous passons à la partie de la mise à jour où nous présentons notre contribution.

1 Architecture du système proposé :

L'objectif que nous nous sommes assignés s'articule autour du développement d'une technique de mise à jour automatique du profil Utilisateur basé sur un système de Question Réponse (Q/A) dans un environnement d'apprentissage.

Notre idée de base consiste à détecter les préférences de l'apprenant en se basant sur plusieurs paramètres parmi lesquelles : les questions posés, la ou les réponses retournées, type de question, emplacement de la question dans le cursus, ...etc.

Afin d'améliorer la performance de notre Système Q/A, Nous avons injecté un module LSA qui permet d'indexer et faire l'appariement entre la requête et la table d'indexation générée à partir de la base de documents.

Afin de réaliser notre système nous procédons en deux phases : une phase de recherche et détection de la réponse de la question posé par l'apprenant (Système Q/A) ; et une deuxième phase consacrée à la mise à jours des préférences du profil de l'utilisateur (Apprenant).

2.1 Première phase : Le système Q/A

Le but de notre contribution est d'assister l'apprenant dans son parcours d'apprentissage en lui proposant un système de question réponse basé sur les spécifications d'un auteur/tuteur et sur son assimilation antérieure afin d'atteindre l'objectif choisi. Cette phase a pour objectif de répondre aux questions posées par l'apprenant pendant son cursus d'apprentissage.

Nous avons amélioré notre système de Q/A en injectant une nouvelle base (Faq) qui stock en temps réel les questions fréquemment posée par les apprenants.

Notre système QA est un module à part entière, son but principal est d'améliorer le processus de tutorat dans une plate-forme en ajoutant une fonctionnalité qui va augmenter l'interaction entre le tuteur et apprenant.

Avant toute initiative, et dès que l'apprenant pose une question, le système Q/A consulte la base de question fréquemment posée (FAQ) afin qu'il affiche la réponse respective s'il existe, sinon la question de l'apprenant (en langage naturel) sera analyser en utilisant des outils de Traitement de Langage Naturel (TALN) puis

reformuler afin que la procédure LSA soit en mesure d'appliquer l'appariement (Document-Mots Clés).

Les TOP réponses seront extraits des documents pertinents et le système sélectionne la réponse la plus fiable pour l'afficher finalement à l'apprenant. Les différentes étapes de cette phase est illustrés dans la figure 5.

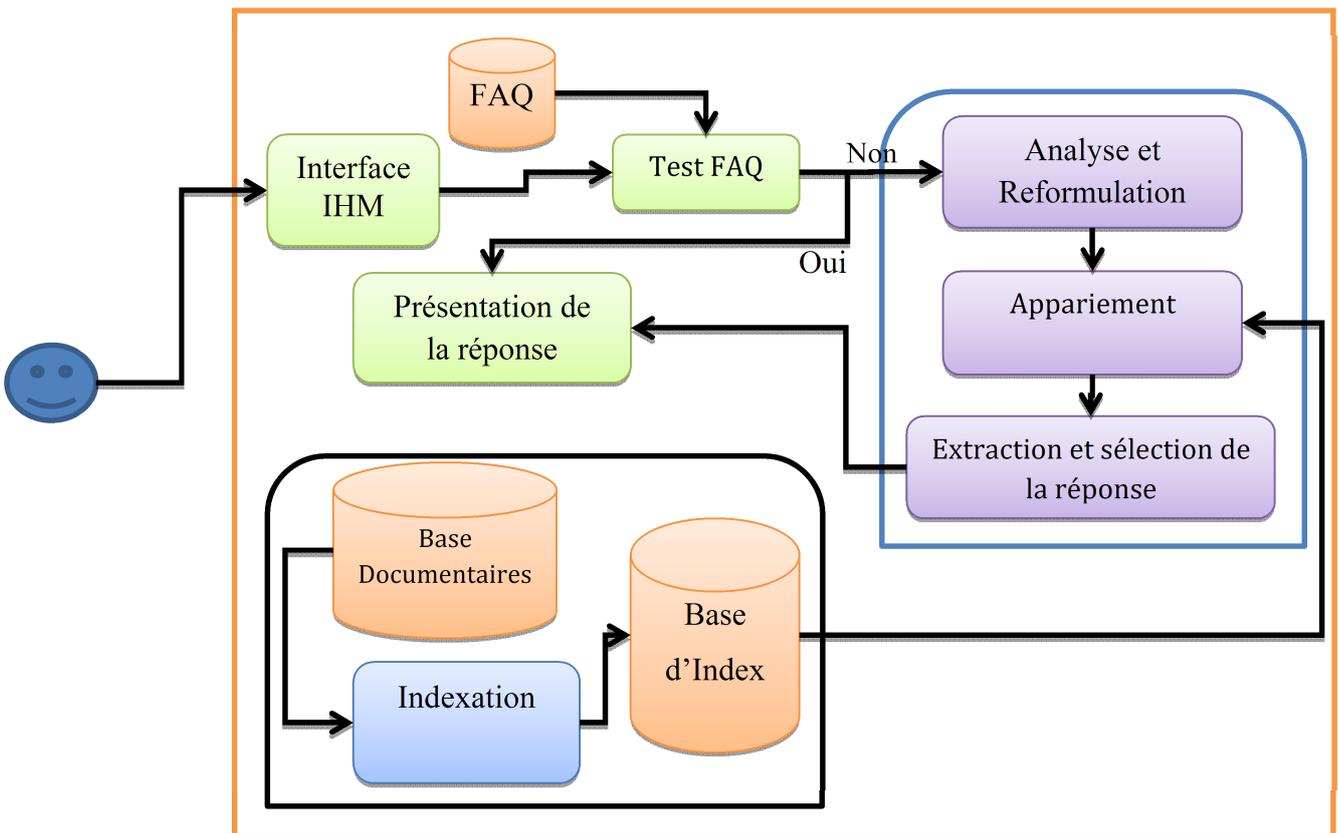


Figure 5 Schéma général du Système Q/A Proposé

Ce système peut automatiquement répondre aux questions qui sont posées par un apprenant en utilisant le langage naturel. L'architecture fondamentale de ce système est montrée sur la figure 5, et se compose de l'interface utilisateur (IHM), Le Test FAQ, d'Analyseur de question, la base de données des documents, Base de FAQ (frequently asked question)

L'interface utilisateur est l'interface du dialogue de l'apprenant avec le système en langage naturel. C'est la fenêtre qui permet au apprenant de poser leurs questions et au système de répondre à ces questions.

Dans le module d'analyse de question, le traitement de la question demandée en utilisant le langage naturel suit les méthodes de transformation incluant l'analyse morphologique, la classification de question, traitement des mots vides, l'extraction des mots-clés et l'expansion des mots-clés. Nous allons par la suite présenter les composantes de notre système une après l'autre.

Pour le traitement appliquée au document existant au niveau de la base documentaire, nous avons optez pour la méthode LSA pour l'indexation des documents et pour faire l'appariement entre les documents et les requêtes poser par les apprenants. Par la suite nous éclaircissons ces traitements.

1) Indexation :

L'indexation consiste à extraire des informations réduites et pertinente à partir des documents de la base, ces information vont être exploité pour caractériser ces document durant la phase de recherche.

Pour un texte, ce qui notre cas, un index très simple à établir automatiquement est la liste ordonnée de tous les mots apparaissant dans les documents avec la localisation exacte de chacune de leurs occurrences ; mais un tel index est volumineux et surtout peu exploitable.

L'indexation automatique tend donc plutôt à rechercher les mots qui correspondent au mieux au contenu informationnel d'un document. On admet généralement qu'un mot qui apparaît souvent dans un texte représente un concept important. Ainsi, la première approche consiste à déterminer les mots représentatifs par leur fréquence. Cependant, on s'aperçoit que les mots les plus fréquents sont des mots fonctionnels (ou mots outils, mots vides). En français, les mots « de », « un », « les », etc. sont les plus fréquents. En anglais, ce sont « of », « the », etc.

Il est évident que l'on ne peut pas garder ces mots à haute fréquence mais peu porteur de sens en terme. C'est pourquoi on introduit dans les moteurs de recherche des opérations de filtrage de ces mots. Ces listes de mots sont appelées anti-lexiques ou plus fréquemment stoplist1.

Chaque unité documentaire (chaque document ou chaque passage de document) peut alors faire l'objet d'une représentation vectorielle : les coordonnées représentent les fréquences des mots non vides. Lorsque l'on effectue cette opération pour un corpus de documents ou de pages web on obtient une matrice dont les colonnes représentent un document et les coordonnées la fréquence des termes.

Pour notre cas, nous avons optés pour la méthode LSA (Latent Semantic Analysis) que nous allons présenter par la suite.

- LSA : un modèle par analyse de la sémantique latente

Utilisé pour l'analyse de grand corpus de texte, LSA [DEE 90] pour « Latent Semantic Analysis » ou « analyse de la sémantique latente (ASL) », est une technique statistique automatique pour extraire et inférer des relations entre mots à partir de leur contexte.

L'utilisation de LSA dans le domaine de l'analyse de documents est pertinente, car des travaux en sciences cognitives montrent que la représentation et l'acquisition de connaissances à partir de textes, la compréhension et l'évaluation de textes, l'extension à des connaissances non issues de textes, à l'aide du modèle LSA, sont comparables à celles des sujets humains lors de tests standardisés [LAN 97, LAN 98, LEM 03]. LSA est à la fois vu comme un modèle d'acquisition et de représentation des connaissances.

- Principe de LSA :

Le sens d'un mot peut être défini statistiquement à partir de l'ensemble des contextes (phrases, paragraphes, textes) dans lesquels ce mot apparaît. Par exemple, le mot autobus sera souvent conjointement associé à démarrer, route, gare routière, et rarement à fleur, barbecue. Cependant, le contexte du mot n'est pas suffisant pour en définir le sens, car il ne dit rien sur les relations avec les mots qui n'apparaissent jamais ensemble.

Par exemple, si les mots autobus et autocar n'apparaissent jamais ensemble, nous n'avons aucune information sur les liens sémantiques entre ces mots. Or autocar doit être considéré comme proche de autobus car tous les deux sont cooccurrents avec les mêmes mots. Ce sont donc des enchaînements de liens de cooccurrences à plusieurs niveaux qui permettent une représentation correcte du sens des mots.

Pour résoudre cette difficulté, LSA construit une matrice de cooccurrences, constituée du nombre d'apparitions de chaque mot dans chaque contexte, sans tenir compte de leur ordre. Cette matrice est ensuite réduite à l'aide d'une décomposition en valeurs singulières (SVD) (généralisation de l'analyse factorielle) afin de capturer dans une certaine mesure les relations entre les mots et les documents et en espérant que les mots ayant un sens voisin (en particulier les synonymes) auront la même direction dans le nouveau sous-espace.

En résumé, le fonctionnement de LSA est basé sur deux principes :

- (1) le sens d'un mot peut-être défini statistiquement à partir de l'ensemble des contextes
- (2) deux mots sont similaires s'ils apparaissent dans des contextes similaires.

Il résout dans une certaine mesure les problèmes suivants :

- **synonymie** : deux mots ayant le même contexte seront proches dans l'espace latent,
 - **ambiguïté** : plusieurs documents possédant le même mot, mais dans des sens différents, n'auront pas le même contexte et seront donc éloignés dans l'espace latent,
 - **absence** : même si un mot est absent d'un document sur le même thème, les deux seront proches dans l'espace latent car ils ont le même contexte.
- L'Indexation avec LSA

Le modèle d'indexation associé à LSA est LSI (Latent Semantic Indexing) [[DEE 90]]. Comme dans le modèle vectoriel, chaque document est représenté sous la forme d'un vecteur (représentation «sac de mots»). On peut ainsi construire la matrice terme-document prenant en ligne les mots du lexique et en colonne les documents. Chaque case représente la fréquence d'un mot dans un document. Un problème du modèle vectoriel est que les mots utilisés dans la requête ne sont pas forcément les mêmes que les mots utilisés dans les documents pertinents.

En effet, des mots similaires peuvent avoir différents sens (polysémie) et différents mots peuvent avoir le même sens (synonymie). Pour résoudre ce problème, LSA utilise une SVD afin de prendre en compte le contexte des mots et de réduire les dimensions de l'espace.

Soit \mathbf{A} la matrice terme-document et r son rang, par la méthode algébrique de décomposition en valeurs singulières, elle peut être écrite sous la forme du produit de trois matrices telles que :

$$A = W\Sigma D^t \quad \begin{cases} A \in \mathbb{R}^{n_W \times n_D} \\ W \in \mathbb{R}^{n_W \times r} \\ \Sigma \in \mathbb{R}^{r \times r} \\ D \in \mathbb{R}^{n_D \times r} \end{cases} \quad (\text{EQ1})$$

Où W et D sont des matrices orthonormales contenant les vecteurs singuliers gauche et droit de A, et Σ est une matrice diagonale contenant les valeurs singulières de A :

$$\begin{cases} \Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r) & \text{avec } \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \\ WW^t = DD^t = I_r. \\ \text{rang}(\Sigma) = r \leq \min(n_W, n_D). \end{cases} \quad (\text{EQ2})$$

Pour diminuer le nombre de dimensions de l'espace, et si l'on suppose que les valeurs singulières de la matrice diagonale Σ sont ordonnées, alors on peut trouver une bonne approximation \tilde{A} de A en mettant à zéro les petites valeurs singulières de Σ afin d'obtenir un espace réduit de dimension \tilde{r} choisie :

$$A \cong \tilde{A} = \tilde{W}\tilde{\Sigma}\tilde{D}^t \quad \begin{cases} \tilde{A} \in \mathbb{R}^{n_W \times n_D} \\ \tilde{W} \in \mathbb{R}^{n_W \times \tilde{r}} \\ \tilde{\Sigma} \in \mathbb{R}^{\tilde{r} \times \tilde{r}} \\ \tilde{D} \in \mathbb{R}^{n_D \times \tilde{r}} \end{cases} \quad (\text{EQ3})$$

où

$$\begin{cases} \tilde{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_{\tilde{r}}) \\ \tilde{W}\tilde{W}^t = \tilde{D}\tilde{D}^t = I_{\tilde{r}}. \\ \text{rang}(\tilde{\Sigma}) = \tilde{r} \leq \text{rang}(\Sigma) = r \leq \min(n_W, n_D). \end{cases} \quad \text{avec } \begin{cases} \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\tilde{r}} \\ \text{et } \sigma_{\tilde{r}+1} = \dots = \sigma_r = 0 \end{cases} \quad (\text{EQ4})$$

Cette opération réalise la projection de l'espace original vers un espace réduit à \bar{r} dimensions. Il a été démontré que, sous certaines conditions, l'espace réduit capture dans une certaine mesure les relations sémantiques entre les mots du corpus [PAP 98]. Le nombre de dimensions optimal de l'espace réduit pour la langue anglaise a été estimé empiriquement à 300 dimensions [LAN 97].

La SVD effectue un changement de base pour se placer suivant les axes de plus grande variation de la matrice A . De manière intuitive, on peut se représenter un mot comme un point dans un espace dont la dimension est le nombre de documents n_D .

La matrice A donne les coordonnées des n_W mots. Ce nuage de points a des axes d'inertie qui sont précisément les axes de plus grande variation de A . En tronquant aux \bar{r} premières valeurs singulières, on conserve les axes d'inertie suivant lesquels s'alignent le mieux les points du nuage. Ainsi on capture la structure la plus significative de la matrice. Il faut voir la décomposition en valeurs singulières comme une méthode qui réduit la dimension du problème et, surtout, qui permet de représenter mots et documents dans un même espace de dimension \bar{r} .

L'espace de dimension \bar{r} s'interprète comme un espace de concepts. On ne peut pas vraiment espérer mettre un nom sur ces concepts. Mais ce n'est pas gênant : tout ce dont on a besoin est de savoir dans quelle mesure les différents concepts (abstraites) sont présents dans tel mot et tel document, de manière à comparer ceux-ci. Mathématiquement, puisque le mot et le document sont représentables dans un même espace, un simple calcul de la distance entre leurs représentants fournit une quantification de leur proximité. Au final, les documents renvoyés peuvent ne contenir aucun mot de la requête mais être pertinents.

2) Test FAQ :

Le rôle du teste de la FAQ est de retourner une réponse à l'apprenant plus rapidement sans aucun besoin de la base de documents de la plateforme. Pour cela, Une base de données est créer ; afin de stocker les questions fréquemment posées avec leurs réponses respectives.

Si la question posée par l'apprenant existe au niveau de la Faq, Le système Q/A retourne immédiatement la réponse correspondante, Sinon il passe au traitement de la question et l'extraction de la réponse à partir de la base de documents.

Algorithme de Recherche dans la FAQ ;

Entrée : La question Q

Sortie : La Réponse R

DEBUT

Pour tous les lignes i de la base FAQ Faire

 Si tous mots-clés existent dans une ligne i de la FAQ Alors $R \leftarrow \text{Reponse}(i)$;

Fin Pour

FIN.

Remarque :

Reponse(i) : est la réponse correspondante de la question situé dans la ligne i de la base FAQ.

3) Analyse et Reformulation:

Au niveau de cette étape, le système Q/A analyse la question posé par l'apprenant pour extraire les mots clés ainsi que le type de la question et d'autres informations à propos de la question. **Ces mots clés seront filtrés pour augmenter la probabilité de retourner la bonne réponse au sein de la base documentaire existante.**

- Analyse de la question :

Une des étapes les plus importantes dans le processus des systèmes de question/réponse est l'analyse des questions. En effet, les questions sont exprimées en langage naturel, ce qui diffère des tâches classiques en recherche d'information (RI) dans lesquelles les requêtes sont exprimées sous forme de mots clés.

Une étape essentielle consiste donc à extraire les éléments clés d'une question afin de constituer une (ou plusieurs) requête(s). L'autre fonction est la détermination du type de la réponse attendue. Ce type est généralement formalisé sous la forme d'une entité nommée (exp : PERSONNE, LIEU,...). Toutefois la notion d'entité nommée, communément utilisée, a été élargie pour prendre en compte d'autres types comme les titres d'œuvre ou les adresses,...etc. Lors de l'analyse d'une question, des mots clés sont extraits et sont par la suite utilisés pour former une requête qui permettra de ramener des documents susceptibles de contenir la réponse à la question posé.

L'analyse de question, basée sur des outils d'extraction, consiste en une cascade de 520 transducteurs combinés à des lexiques spécifiques. Cette analyse se place sur trois niveaux.

Tout d'abord, les entités nommées sont détectées. Puis une série de transducteurs est appliquée pour la détection du type de question et donc du type de réponse attendue, ainsi que l'identification des termes clés. Finalement, une analyse syntaxique est utilisée pour déterminer les liens entre les éléments de la question. La détection du type de question peut se faire à un niveau très fin (« WHAT IS THE BIRTH DATE OF PROPER_NAME ? » on sait qu'on cherche une date et la signification de cette date) ou à un niveau plus grossier si l'analyse fine a échoué (« WHO IS ... ? » on sait qu'on cherche une personne qui est liée à ce qui suit mais l'analyse n'a pas pu structurer tous les éléments).

Ces derniers automates de reconnaissances effectuent donc une analyse incomplète de la question, ce qui permet au moins de déterminer l'entité recherchée.

Cette étape est une phase très importante dans le cycle du système de question réponse. Car, tant que le traitement de la question donne de bons résultats, la performance du système va augmenter.

Tout d'abord le texte est traité par un analyseur syntaxique. Les phrases à analyser proviennent de textes authentiques : nous partons du principe qu'ils ne contiennent aucune erreur de grammaire ou d'orthographe.

L'analyseur syntaxique applique une grammaire ascendante, afin d'obtenir la meilleure couverture syntaxique, selon des critères linguistiques.

Pour une phrase donnée, le parseur génère dans la plupart des cas un arbre d'analyse complet. Lorsque ce n'est pas possible, le système produit un ensemble d'arbres contigus qui correspondent à toutes les parties de la phrase.

4) Appariement:

Afin d'extraire le document qui contient la réponse désirer. Nous avons utilisé la méthode LSA pour faire une recherche dans la base d'indexe généré précédemment afin d'extraire les documents pertinents.

• Rechercher avec LSA

Pour comparer les documents dans l'espace réduit à un vecteur requête d_q , nous transformons tout d'abord le vecteur d_q en un pseudo document \tilde{d}_q dans l'espace réduit.

Nous avons $X = \tilde{W}\tilde{\Sigma}\tilde{D}^t$ que nous pouvons dériver en $\tilde{D} = X^t\tilde{W}\tilde{\Sigma}^{-1}$.

La vecteur ligne \tilde{d}_q dans l'espace réduit peut donc être obtenu par :

$$\tilde{d}_q = d_q^t \tilde{W} \tilde{\Sigma}^{-1}. \quad (\text{EQ5})$$

Il contient le contexte associé au document. Nous pouvons alors utiliser une mesure de similarité classique comme le cosinus pour calculer la distance entre d_q et chacun des documents dans l'espace réduit. De même, pour un mot w_q , nous pouvons obtenir sa représentation \tilde{w}_q dans l'espace réduit par :

$$\tilde{w}_q = \tilde{\Sigma}^{-1} \tilde{D}^t w_q^t. \quad (\text{EQ6})$$

Pour mesurer la similarité entre le document i et le document j , il suffit de réaliser le produit scalaire entre les vecteurs lignes i et j de la matrice $\tilde{D}\tilde{\Sigma}$. Il est aussi possible de calculer les p mots les plus pertinents pour un document.

Des expériences réalisées sur des corpus de texte adaptés à la recherche d'informations montrent que LSI donne des résultats similaires ou légèrement meilleurs que les modèles classiques en RI [DEE 90].

2.1. Deuxième phase : Détection et mise à jours du profil

Dans cette phase, nous aborderons la partie de gestion de profil de l'utilisateur. Au niveau de notre approche, nous allons utiliser le contenu du profil défini par [Kostadinov 2003], avec quelques modifications. Le profil contient les informations suivantes : Données personnelles, Centre d'intérêt, Qualité attendue, Préférences de livraison, Sécurité et Historique des interactions de l'utilisateur et Style d'apprentissage.

Nous exploitons les résultats obtenus par la première phase notamment le type de question, la réponse, les mots clés de la question,...etc.

Le système a pour rôle la détection du profil utilisateur de l'apprenant en se basant sur les différentes caractéristiques. Nous avons proposé une théorie pour la mise à jour du profil de l'utilisateur basé sur un système de Q/A.

Avant d'entamer le fonctionnement et le parcours nécessaire pour arriver à mettre à jours les préférences du profil utilisateur, nous allons définir le profil utilisateur qu'on a utilisée tout au long de cette thèse.

Pour cela, nous allons en premier lieu présenter une définition des informations contenant le profil utilisateur utilisée par notre système. Dans un second lieu, nous

allons présenter un scénario pédagogique pour faire éclaircir la théorie présenté. Et enfin, nous décrivons la méthodologie développée pour faire la mise jours du profil.

- Profil utilisateur :

En se basant sur le profil proposé par [Kostadinov 2003] (Voir Chapitre 2), Nous avons utilisé un profil utilisateur qui contient 04 informations qui sont présenté par la suite :

Données personnel : Qui contient les informations de base sur l'utilisateur, à savoir le Nom, Prénom, l'Age, l'adresse, Numéro Tel,...etc.

Le centre d'intérêt : qui exprime le domaine d'expertise de l'utilisateur. Nous l'avons représenté en deux vecteurs. Le premier vecteur, exprime les termes du centre d'intérêt de l'utilisateur et leurs pondérations. Et pour le second vecteur, on propose un vecteur à deux dimensions qu'on appelle « un vecteur de localisation » afin de capturer la localisation temporelle de chaque type de question au sein du cours. Pour chaque centre d'intérêt (module), on attribue un vecteur de localisation.

La première partie, nous avons initialisé les pondérations à une valeur égale pour tous les termes d'un centre d'intérêts. Puis, au fur et à mesure que l'apprenant est en apprentissage, le système mis à jour ces pondérations.

La seconde partie, est construite par l'intermédiaire du système de Q/A proposé. L'utilisation de ce vecteurs donne de nombreuses avantages parmi-les :

- Augmenter la performance du système de Q/A, en réduisant le temps d'accès à la réponse.
- Aider par la suite à faire une mise jours sur le style d'apprentissage de l'utilisateur
- ... etc.

Historique des interactions : Soit q_s la requête soumise par un utilisateur U à la session de recherche se déroulant à l'instant s , soit S_s , et D_s l'ensemble des documents pertinents pour l'utilisateur durant cette session. Un document est considéré comme pertinent s'il a été ainsi jugé par l'utilisateur de manière explicite ou implicite. A cet effet, on se réfère à une catégorisation du comportement de l'utilisateur, traduisant des jugements implicites de pertinence, largement adoptée par les travaux du domaine [KEL 04]. Cette catégorisation construit une fonction

pertinence qui associe à chaque action type (impression, lecture, sauvegarde etc.) un degré de pertinence. Pour notre part, on utilise une fonction constante qui traduit d'avantage un indicateur de pertinence sûre. On propose dans ce qui suit l'utilisation de matrices pour la représentation d'une session de recherche et de l'historique des interactions.

Le Style d'Apprentissage : Nous avons adopté pour le modèle du style d'apprentissage proposé dans [FEL 95] qui définit le style d'apprentissage en répondant aux quatre questions (Voir Chapitre 2 Section 2.1)

Notre système est conçu afin de faire la mise à jour des informations du profil. Par conséquent l'architecture générale de la deuxième phase qui est donnée par le schéma de la figure 8:

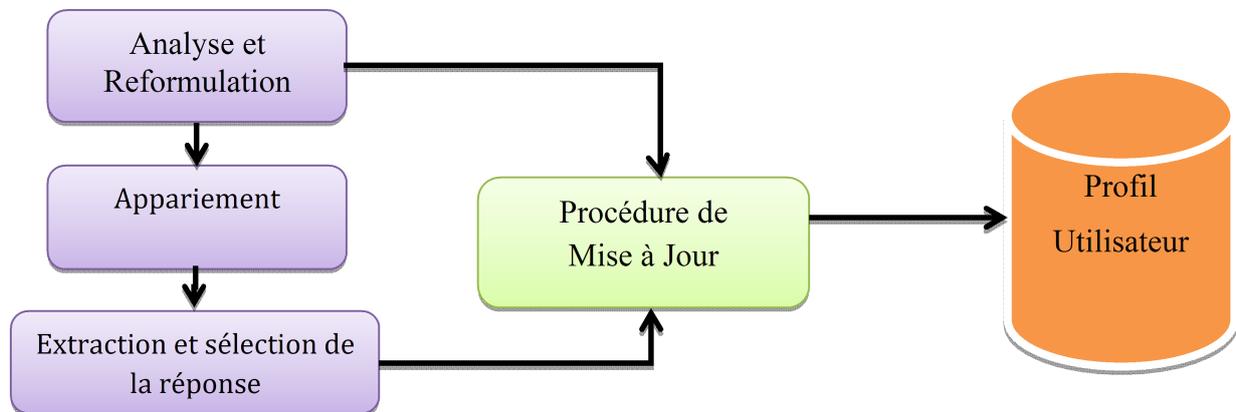


Figure 8 Schéma général de la partie de mise à jour du profil Utilisateur

Par la suite, nous allons détailler comment le système serai capable de faire une mise jours du profil à partir des questions posé par l'apprenant pendant une session de cours donnée et quelle sont les préférences spécifique qu'elles seraient mis à jours.

- Procédure de mise à jour:

Au sein de cette procédure, Notre système va devoir être capable de faire la mise à jour du profil utilisateurs. Comme l'indique la figure 8, La procédure de mise à jour a deux entrées provenant du système de question de Réponse.

Cette procédure reçoit la question ainsi que d'autres informations de la part du système de Q/A, certains d'entre elles sont extraite explicitement et d'autre d'une façon implicite. D'après ces informations, le profil soit mis à jour en calculant des paramètres et affectant des préférences bien précises.

- Capture des types de question :

Pour cela on a proposé une nouvelle caractéristique dans le centre d'intérêt afin de capturer les positions de la question posée par l'apprenant durant son cycle d'apprentissage.

Scénario

Afin de bien mener les étapes de construction de notre système, on propose que le cours à suivre pendant l'apprentissage de l'apprenant est découper d'une façon a distingué les multiples rôles pédagogique offerte au moment donné.

A titre d'exemple : L'apprenant X qui a un style d'apprentissage Y suit le cours qui est organisé de la façon suivante :

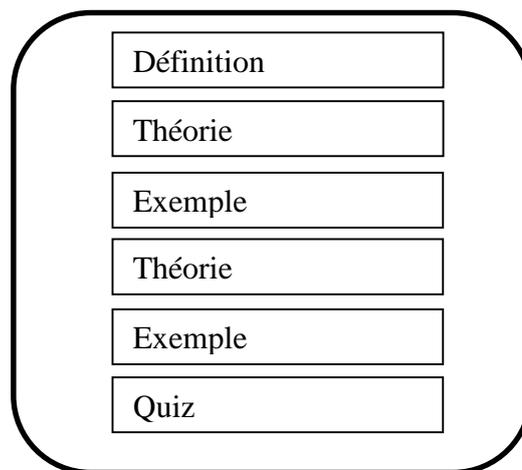


Figure 7 Exemple de Structuration d'un cours

L'organisation du cours suit le style d'apprentissage défini de chaque étudiant.

Pour faciliter le travail, nous avons fait une indexation ou plutôt une abréviation, pour chaque partie du cours (rôle pédagogique) afin de simplifier la sauvegarde et la capture du moment où l'apprenant à essayer de poser une question au système de Q/A.

Voici l'exemple après indexation :

Définition	A1
Théorie	A2
Exemple	A3
Théorie	A4
Exemple	A5
Quiz	A6

Figure 8 Exemple indexé de Structuration d'un cours

A chaque fois que l'apprenant pose une question, le système de QA analyse la question pour extraire le type de la question posé. Voici le groupe de question prise en charge par notre système : [Who, What, When, Where, Which, How, Yes/No].

Pour le type de question [YEs/No] elle inclut les questions :

- Did ?
- Do ?
- Have ?
- Had ... ?
- Can ;?

Chaque fois que l'apprenant pose une question au système de question réponse , qui est disponible tout au long de son cursus, le système capture automatique la position de l'apprenant dans son cours actuel dans une table qui sera sauvegarder au niveau du centre d'intérêt de l'utilisateur qui montrer dans le « Tableau 6 » .

	WHAT	WHO	WHEN	WHERE	WHICH	HOW	YES/NO
A1							
A2							
A3							
A4							
A5							
A6							

Tableau 6 Tableau de probabilité des positions des réponses en sein des documents

En analysant les informations de ce tableau, on peut retirer la fréquence d'apparition d'un type précis de question dans un endroit temporelle bien précis du cours. Aussi, il nous permet par la suite de diminuer la surface de recherche pour la réponse d'une question posé plus tard par l'apprenant.

Avant l'utilisation du tableau 6, Au sein de l'indexation des documents par la méthode LSA, la LSI fais la relation entre les documents et les mots existants de façons à ce que les documents ayants le plus de mots semblable soit proche dans l'espace de recherche. Pour rechercher une réponse, la LSA retourne un ou plusieurs documents qui peuvent avoir la réponse voulu.

Maintenant en utilisant le Tableau 6, La LSA va maintenant fais l'indexation non pas des documents par rapports aux mots existants mais plutôt les parties des documents par rapport aux mots existants.

En résultant, nous allons avoir un espace construit, non pas des documents, mais des rôles pédagogiques. Et lorsqu'un apprenant pose une question, le système détecte le type de cette question. Puis une recherche serait faite sur la partie du cours où la grande fréquence d'apparence du type de la question.

Conclusion

Dans ce chapitre, nous avons présenté un système qui comporte deux parties ; la première permet de répondre aux questions posées par l'apprenant pendant une séance d'apprentissage au sein d'une plateforme d'enseignement. Pour cela, Nous avons proposé d'ajouter une base FAQ afin de diminuer le temps de réponse du système. Afin de répondre à la question de la base de documents, nous avons utilisé la méthode LSA pour indexer la base ainsi pour faire l'appariement entre la base d'indexe et les termes des questions.

L'idée de base de la deuxième partie est que le système essaye de mettre à jours le profil de l'utilisateur en injectant une nouvelle caractéristique qui localise les positions des types de question posée, ce que nous aides par la suite à faire un recherche a la réponse plus rapide du fait que le système reconnais la partie du cours ou probablement la réponse soit trouvé. L'évaluation du système proposé, les résultats obtenus, et les discussions accompagnées seront l'objet du prochain chapitre.

Chapitre

4

Implémentation



Introduction

Le but de ce chapitre est d'évaluer notre système de détection et de mise à jour de profil utilisateur à base d'un système de question réponse.

L'évaluation est effectuée à travers un ensemble d'expérimentations et tests qui concernent à la fois la partie analyse de documents et la partie recherche.

Les tests et les évaluations sont faits sur une base de documents de 100 documents.

La suite de ce chapitre est organisée de la manière suivante : dans la section 2 nous illustrons l'environnement expérimental de notre système et le contexte des expérimentations, ensuite dans la section 3 nous présentons la base de documents utilisées pour évaluer la partie de recherche. Dans la section 4, nous décrivons les différentes expérimentations, les résultats obtenus des expérimentations, avant de conclure.

1 Outils de développement :

Nous allons expérimenter notre système au sein de SAKAI qui est un environnement numérique d'apprentissage et de collaboration en code source libre, développé en général en utilisant des outils libres conformes aux normes et standards internationaux. En se basant sur ces critères et aux recommandations trouvées sur **Sakaipédia**, l'encyclopédie de la communauté SAKAI nous avons choisi comme outils de développement ceux présentés ci-dessous :

1.1 JAVA

Java est à la fois un langage de programmation et un environnement d'exécution. Le langage Java a la particularité principale d'être portable sur plusieurs systèmes d'exploitation tels que Unix, Microsoft Windows, Mac OS ou Linux... C'est la plateforme qui garantit la portabilité des applications développées en Java.

Ces concepteurs ont privilégié l'approche orientée objet de sorte qu'en Java, tout est objet à l'exception des types primitifs (nombres entiers, nombres à virgule flottante, etc.).

1.2 JSP et Servlet :

Le **JavaServer Pages** ou **JSP** est une technologie permettant de développer des applications Web interactives, c'est-à-dire dont le contenu est dynamique, basées sur Java. La technologie permet au code Java et à certaines actions prédéfinies d'être ajoutés dans un contenu statique tel que les pages HTML.

Les JSP sont compilées par un compilateur JSP pour devenir des Servlets Java. Un compilateur JSP peut générer un Servlet Java en code source Java qui peut à son tour être compilé par le compilateur Java, ou peut générer le pseudo-code Java interprétable directement.

1.3 Eclipse :

Eclipse IDE est un environnement de développement intégré extensible, universel et polyvalent, permettant potentiellement de créer des projets de développement mettant en œuvre n'importe quel langage de programmation en particulier le langage JAVA. Eclipse IDE est écrit en Java (à l'aide de la bibliothèque graphique SWT, d'IBM), et ce langage, grâce à des bibliothèques spécifiques, est également utilisé pour écrire des extensions.

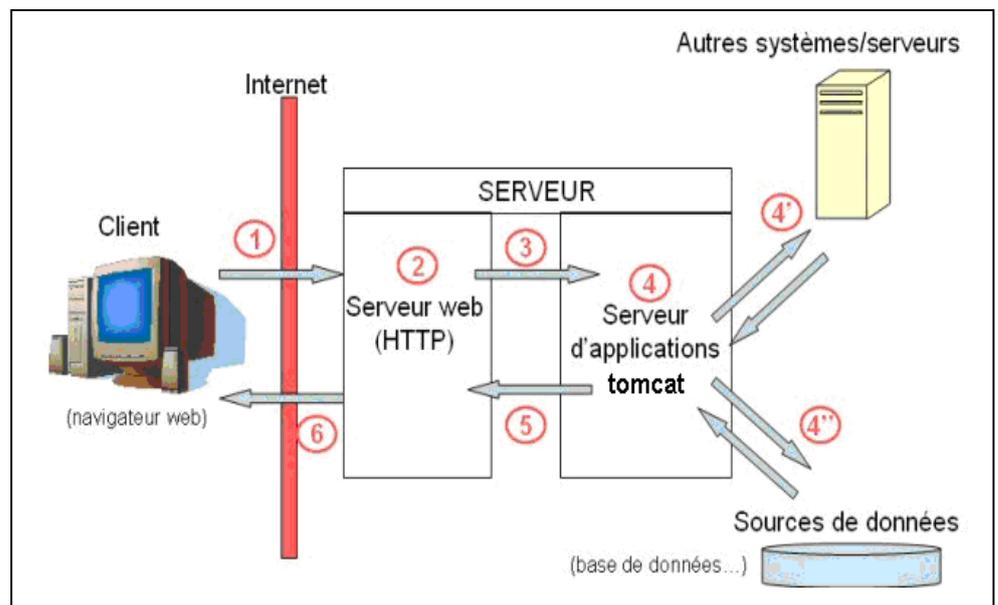
La spécificité d'Eclipse IDE vient du fait de son architecture totalement développée autour de la notion de plug-in (en conformité avec la norme OSGi) : toutes les fonctionnalités de cet atelier logiciel sont développées en tant que plug-in.

1.4 Tomcat

Tomcat est un serveur d'applications web. Issu du projet Jakarta, Tomcat est désormais un projet principal de la fondation Apache. Tomcat implémente les spécifications des Servlets et des JSP de Sun Microsystems. Il inclut des outils pour la configuration et la gestion, mais peut également être configuré en éditant des fichiers de configuration XML .Elle permet de générer une réponse HTML à une requête après avoir effectué un certain nombre d'opérations (connexion à une base de données, à un annuaire LDAP...).

Le schéma suivant montre le déroulement classique d'une requête vers le serveur d'applications Tomcat :

Figure 9 : déroulement classique d'une requête vers le serveur d'applications Tomcat



1.5 MYSQL

MySQL est Système de gestion de base de données relationnelles SQL développé dans un souci de performances élevées. C'est un logiciel libre développé sous double licence en fonction de l'utilisation qui en est faite : dans un produit libre (open-source) ou dans un produit propriétaire. Dans ce dernier cas, la licence est payante, sinon elle est libre.

1.5.1 Caractéristiques de MYSQL :

- multi-utilisateurs.
- Il est multi-thread,
- Rapide et flexible

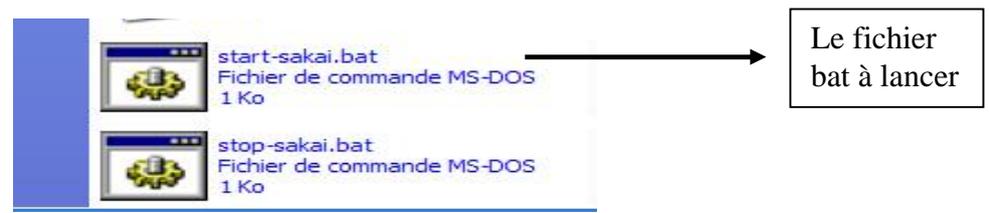
Remarque

- La totalité de ces outils sont libres d'utilisation (Open source).
- Ils sont recommandés par les développeurs et la communauté SAKAI.

Tous les tests effectués sont exécutés sur un PC disposant d'un processeur Intel Core 2 Duo cadencé à 3.0 GHz, de 2 Go de mémoire centrale et d'un disque local de 250 Go.

2.4 Le lancement de SAKAI

Pour démarrer SAKAI, le fichier star-sakai.bat est lance. Si TOMCAT n'affiche aucune erreur alors l'installation est correcte. On peut tester la configuration et sur le browser on lance la requête.



Les deux fichiers de commande pour lancer et stopper SAKAI sont :

- http://localhost :8080/.**
- http://127.0.0.1 :8080/.**

Dans le navigateur, nous pouvons voir la page de démarrage de SAKAI, cette page offre des services publics qui ne nécessitent pas d'avoir un compte, et offre aussi la possibilité d'ouvrir une session sous SAKAI selon le compte utilisateur simple ou administrateur.

Par défaut SAKAI est configurée avec un nom d'administrateur **ADMIN** et des mots de passe **ADMIN**, il faut modifier ce compte pour bien sécuriser la plateforme dans un cadre professionnel.

Le lancement du serveur TOMCAT

Démarrage du serveur après 116032 MS sur le port 8080

```
INFO: Jk running ID=0 time=0/1062 config=null
17 déc. 2006 18:28:08 org.apache.catalina.storeconfig.StoreLoader load
INFO: Find registry server-registry.xml at classpath resource
17 déc. 2006 18:28:08 org.apache.catalina.startup.Catalina start
INFO: Server startup in 116032 ms
```

La figure ci-dessous montre le Portal principal de démarrage de SAKAI :

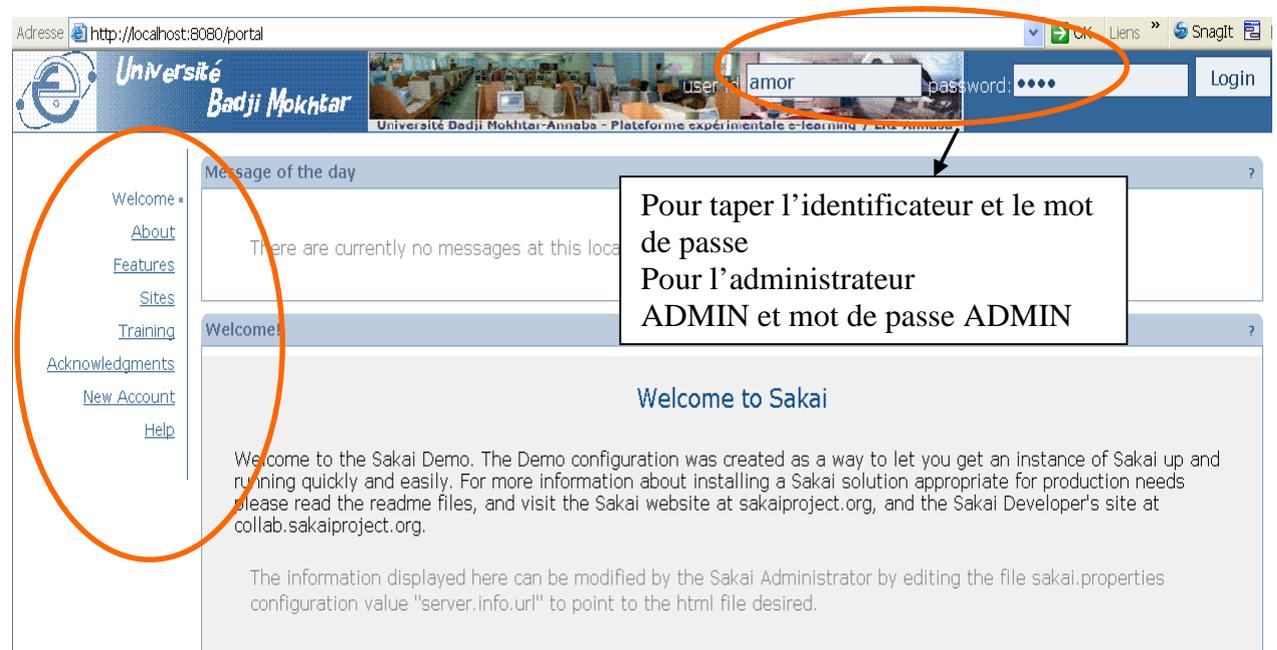
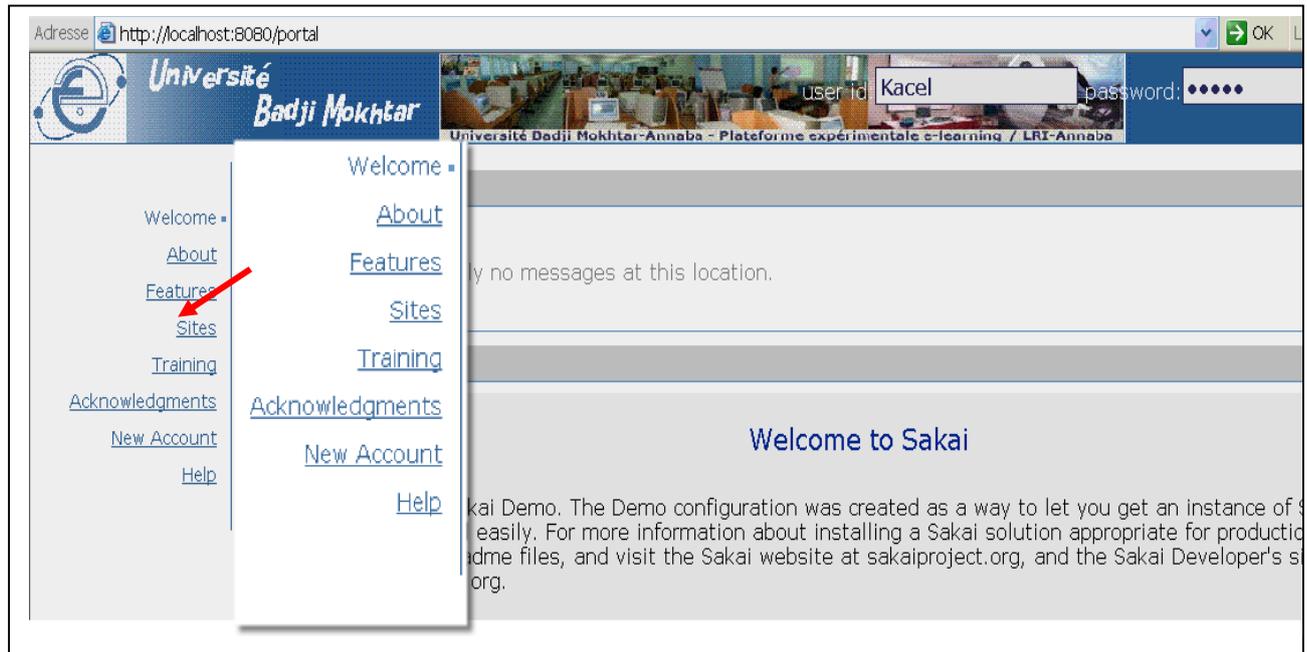


Figure 10 Page d'accueil de SAKAI

3 Environnement d'exécution :

Notre **QASystem** interagit avec deux utilisateurs principaux, le tuteur et l'apprenant. Ces deux derniers opèrent différemment avec notre système.

De ce fait chaque utilisateur doit posséder un compte utilisateur sur la plateforme en tant que tuteur ou apprenant.



Nous remarquons l'absence de notre outil dans le menu principal de sakai.

► Pour l'ajouter nous procédons comme suit :

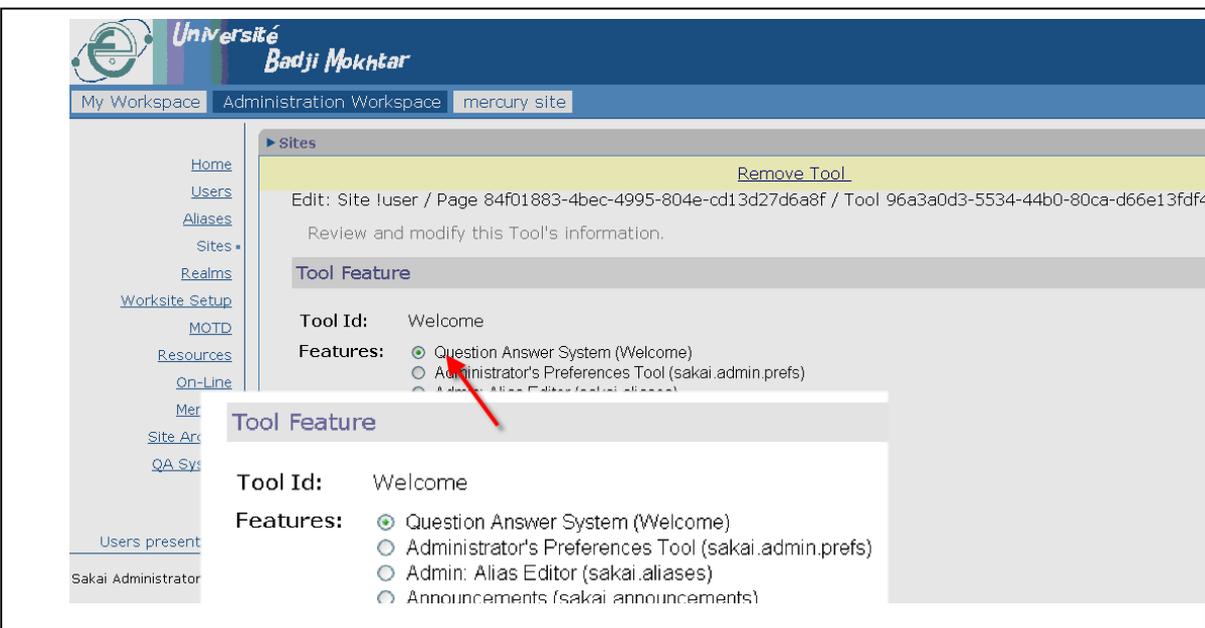
1) Si le site est déjà créé les étapes à suivre sont :

1. login like admin
2. Click on the "sites" link on the left.
3. Select site id in sites.
4. Add page in the low of site
5. Click the "Edit Tools" button.
6. Check the box of the tool you wish to add.
7. Complete any customization pages presented to you.
8. Confirm the new choice by pressing "Finish".

2) Si c'est un nouveau site il suffit de l'ajouter pendant l'étape de la sélection des outils qui composeront le site :

1. Login to Sakai
2. Create a new site or edit an existing site.
3. In the Tool selection menu a new tool should appear labelled " **QASystem** "
4. Add the **QASystem** Tool to your site.

- 5. Access the site. The **QASystem** tool should appear as an option in the Site's left hand menu.
- 6. Click on the **QASystem** menu option.



La figure montre l'apparition de notre module au sein de la plateforme Sakai.

3.1 Interface tuteur :



3.2 Interface apprenant:

Comme nous avons déjà fait nous continuons de détailler l'interface de l'apprenant :

3.2.1 Accès à QASystem :

Comme nous avons dit tout utilisateur de SAKAI peut personnaliser son espace de travail à son choix ce qui nous expose à deux cas différent pour accéder a notre outil :

- 1) Au moment de création du site l'apprenant ajoute l'outil QASystem a son menu personnel.
- 2) Si le site est déjà créé l'apprenant peut le mettre à jour insérant l'outil QASystem aux autres outils déjà dans le menu.

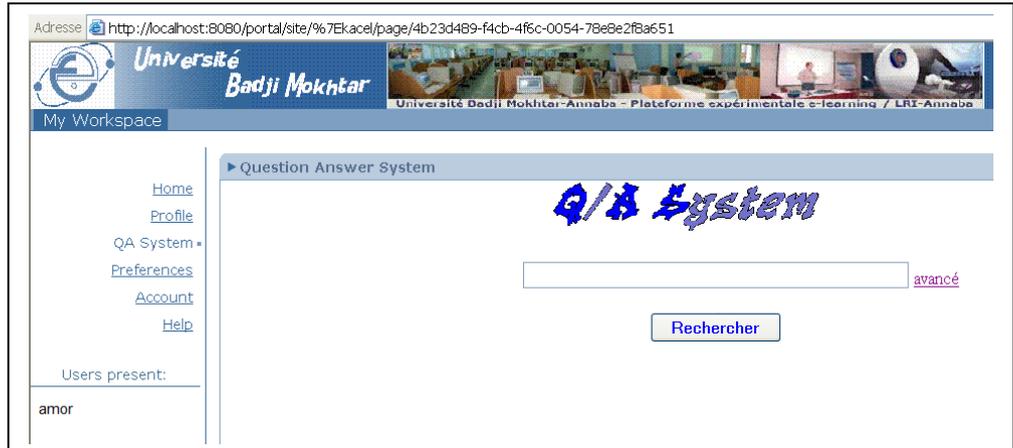


Figure 11 Interface principale de l'apprenant

3.2.2 Base de Données

Pour estimer les performances de notre système, nous avons besoin d'une base de données. Une base locale de 100 Documents a été construite. Les documents composant cette base ont été collectés à partir du web, et donc elles ne possèdent pas les mêmes caractéristiques.

Les documents sont structurés manuellement sous les rôles pédagogiques montrées dans le chapitre 3 : Définition – Théorie – Exemple – Théorie – Exemple – Quiz. En utilisant des balise dans marquant le début et la fin du document.

1.2 Portée des contraintes de domaine

Les contraintes de domaine sont des contraintes qui restreignent les valeurs qu'une donnée peut prendre. Pour mieux en mesurer le sens, il nous faut comprendre ce qu'est un domaine au sens SQL. Par exemple, un domaine POURCENT se verra doté d'une contrainte qui ne lui permet que des valeurs comprises entre 0 et 100. Nous découvrirons les domaines et leur utilisation un peu plus tard dans cet article.

1.3 Portée des assertions

Au sens commun du terme, une assertion n'est autre qu'une proposition devant toujours être vérifiée. Il en va de même en SQL. Une assertion est donc une contrainte. Une assertion porte sur différents objets de la base (en général des tables ou des vues) et on l'exprime sous la forme d'un prédicat¹ devant toujours se vérifier. La portée d'une assertion est donc la base de données dans son ensemble. Par exemple notre base pourrait être composée d'une table des prospects et d'une table des clients, chacune des tables ayant pour clef primaire un nombre entier. Nous aimerions qu'une clef

2.3 Application à un système de Questions-Réponses

Aujourd'hui, de nombreuses applications autour du Traitement Automatique des Langues se fondent sur des connaissances syntaxiques pour résoudre la tâche à laquelle elles sont vouées. Cependant, il est parfois difficile, face à la diversité des analyseurs syntaxiques, de sélectionner au premier abord, pour une application donnée, un analyseur plutôt qu'un autre. De ce fait, ceux-ci sont parfois inappropriés pour l'application envisagée et surtout ont une mauvaise couverture, voire une performance moindre pour les phénomènes importants pour celle-ci.

Ainsi, dans le cadre du développement du système de question-réponse QALC du LIMSI [FER 02], nous avons pu constater que les analyseurs syntaxiques n'étaient pas toujours appropriés pour analyser des questions. Or, pour ce système, l'analyse syntaxique de la question est essentielle puisqu'elle sert de base pour l'extraction d'informations pertinentes (comme le type attendu de la réponse, l'objet de la question, etc.) afin de cerner au mieux où se trouve la réponse possible. C'est le travail de la thèse de Laura Monceaux [MON 02], dont certains résultats sont repris ici.

Plus les informations syntaxiques retournées seront fiables, plus les informations extraites de l'analyse de la question le seront aussi. Or, même si l'on dispose d'un outil d'évaluation des analyseurs syntaxiques permettant de déterminer celui qui est le plus approprié pour ce type d'application, cela ne garantira aucunement la fiabilité de

Quelques Documents de notre base avant structuration

Par la suite Nous présentons quelque exemples des documents de la base de données en injectant les balise de contrôle du cours.

1.2 Portée des contraintes de domaine

Les contraintes de domaine sont des contraintes qui restreignent les valeurs qu'une donnée peut prendre. Pour mieux en mesurer le sens, il nous faut comprendre ce qu'est un domaine au sens SQL. Par exemple, un domaine POURCENT se verra doté d'une contrainte qui ne lui permet que des valeurs comprises entre 0 et 100. Nous découvrirons les domaines et leur utilisation un peu plus tard dans cet article.

[fin_theorie]

1.3 Portée des assertions

[debut_theorie]

Au sens commun du terme, une assertion n'est autre qu'une proposition devant toujours être vérifiée. Il en va de même en SQL. Une assertion est donc une contrainte. Une assertion porte sur différents objets de la base (en général des tables ou des vues) et on l'exprime sous la forme d'un prédicat¹ devant toujours se vérifier. La portée d'une assertion est donc la base de données dans son ensemble. Par exemple notre base pourrait être composée d'une table des prospects et d'une table des clients, chacune des tables ayant pour clef primaire un nombre entier. Nous aimerions qu'une clef

[fin_theorie]

2.3 Application à un système de Questions-Réponses

[debut_theorie]

Aujourd'hui, de nombreuses applications autour du Traitement Automatique des Langues se fondent sur des connaissances syntaxiques pour résoudre la tâche à laquelle elles sont vouées. Cependant, il est parfois difficile, face à la diversité des analyseurs syntaxiques, de sélectionner au premier abord, pour une application donnée, un analyseur plutôt qu'un autre. De ce fait, ceux-ci sont parfois inappropriés pour l'application envisagée et surtout ont une mauvaise couverture, voire une performance moindre pour les phénomènes importants pour celle-ci.

Ainsi, dans le cadre du développement du système de question-réponse QALC du LIMSI [FER 02], nous avons pu constater que les analyseurs syntaxiques n'étaient pas toujours appropriés pour analyser des questions. Or, pour ce système, l'analyse syntaxique de la question est essentielle puisqu'elle sert de base pour l'extraction d'informations pertinentes (comme le type attendu de la réponse, l'objet de la question, etc.) afin de cerner au mieux où se trouve la réponse possible. C'est le travail de la thèse de Laura Monceaux [MON 02], dont certains résultats sont repris ici.

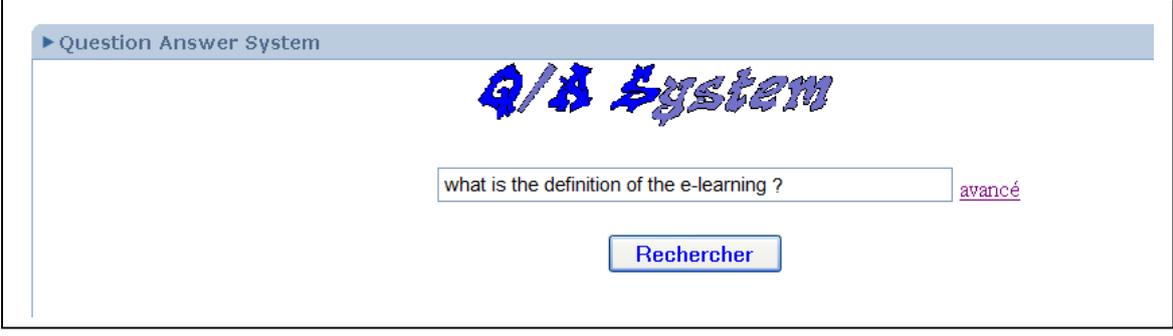
Plus les informations syntaxiques retournées seront fiables, plus les informations extraites de l'analyse de la question le seront aussi. Or, même si l'on dispose d'un outil d'évaluation des analyseurs syntaxiques permettant de déterminer celui qui est le plus approprié pour ce type d'application, cela ne garantira aucunement la fiabilité de

[fin_theorie]

Quelques Documents de notre base après structuration

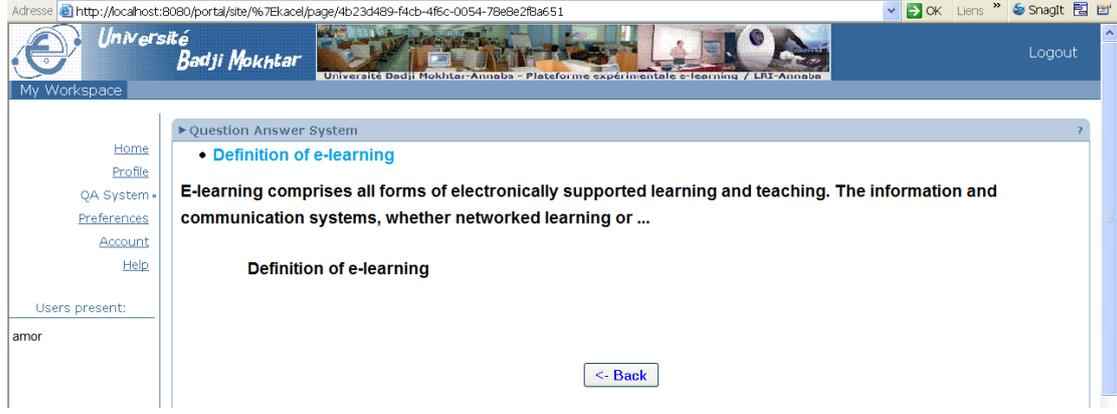
Nous avons utilisé dans nos tests une partie de la base de données composée de 20 Documents choisi au hasard. Dans les environs de 20 étudiants du département d'informatique, on a assisté la phase de test du module

a) Saisie de la question :



c) Affichage du résultat de la recherche :

En remplissant le champ de la question et en cliquant sur le bouton **Rechercher**, **QASystem** affichera le résultat à l'apprenant comme suit :



Interface pour l'affichage des résultats.

3.2.3 Initialisation de la table de classification des centres d'intérêt

Afin d'utiliser la table de classification proposé, nous devons remplir la table avec les probabilités des réponses de chaque question dans un fragment de cours. Pour cela, nous avons lancé une petite expérimentation en utilisant un ensemble de requêtes sur un groupe de 20 documents choisi au hasard parmi le corpus existant afin d'évaluer les probabilités et les affectés dans la table. Le tableau suivant démontre le nombre de question posé pour chaque type de question

TYPE DE QUESTION	NOMBRE DE QUESTION	NOMBRE DE REPONSE CORRECTE RETOURNE
WHAT	15	9
WHO	15	7
WHEN	15	8
WHERE	15	6
WHICH	15	4
HOW	15	2
YES/NO	15	5

Après que le système met à jours la table de classification, les résultats sont montrés dans le tableau suivant :

	WHAT	WHO	WHEN	WHERE	WHICH	HOW	YES/NO
A1	0.333	0.142	0.5	0.333	0	0	0
A2	0.333	0.571	0.25	0	0.25	0.5	0.2
A3	0.111	0	0	0	0.5	0	0.4
A4	0.222	0.285	0.125	0.333	0	0.5	0
A5	0	0	0	0.166	0	0	0.2
A6	0	0	0.125	0.166	0.25	0	0.2

3.2.4 TEST DES RESULTATS

Pour le test des résultats, nous avons utilisé les mesures standards d'évaluation des systèmes d'information qui sont le RAPPEL et la PRECISION.

Le tableau suivant montre les résultats obtenu sur une base de 60 documents.

TYPE DE QUESTION	RAPPEL	PRECISION
WHAT	0.68	0.35
WHO	0.62	0.45
WHEN	0.50	0.64
WHERE	0.39	0.51
WHICH	0.35	0.41
HOW	0.39	0.42
YES/NO	0.42	0.62
MOYENNE	0,47	0.48

TYPE DE QUESTION	RAPPEL	PRECISION
WHAT	0.68	0.41
WHO	0.64	0.50
WHEN	0.53	0.61
WHERE	0.46	0.62
WHICH	0.43	0.52
HOW	0.39	0.49
YES/NO	0.46	0.68
MOYENNE	0.51	0.54

DISCUSSION

Nous remarquons qu'à partir des résultats que l'application de notre contribution à améliorer les rappels et les précisions au niveau de quelques types de questions comme : WHO, WHEN.

Tandis que pour d'autres types de questions les rappels et les précisions ont été diminués mais en général, on peut remarquer une amélioration pour la valeur moyenne du rappel et les précisions.

CONCLUSION GENERALE ET PERSPECTIVES

Dans le cadre de cette thèse, nous avons étudié principalement deux domaines (Système Q/A, profil utilisateurs) afin de soutenir une nouvelle démarche de détection de profil utilisateurs dans les environnements de formation sur le Web en se basant sur les questions posées par l'apprenant.

Comme perspective, nous avons le but de faire une mise à jour du style d'apprentissage de l'utilisateur en se basant sur les questions. En plus faire une augmentation des performances du système de Q/A.

Référence



Bibliographique

- [ADA 99] E. Adar and D. Karger. Haystack : Per-user information environments. In Proceedings of the 8th International Conference on Information Knowledge Management, pages 413-422, Kansas City, Missouri, November 2-6 1999
- [AMA 99] 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 digital libraries, pages 184-187, 1999.
- [APT 01] Apter, M. J. (2001). Motivational Styles in Everyday Life: a Guide to Reversal Theory. Washington DC: American Psychological Association
- [BAL 97] M. Balabanovic and Y. Shoham. Fab : Content-based collaborative recommendation. Communications of the ACM, 3(40) :66–72, March 1997.
- [BAR 97] R. Barrett, P. Maglio, and D. Kellem. How to personalize the web. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 75–82, New York, NY, USA, 1997. ACM.
- [BAZ 04] M. Baziz. Towards a semantic representation of documents by ontology-document mapping. In The Eleventh International Conference on Artificial Intelligence, pages 33-43. Christoph Bussler, Diete Fensel (Eds.), LNCS/LNAI 3192, Springer, Springer-Verlag Berlin, Heidelberg, Germany, 2-4 septembre 2004.
- [BEH 09] A. Behaz, M. Djoudi, « Approche de modélisation d'un apprenant à base d'ontologie pour un hypermédia adaptatif pédagogique », CHIA, 2009.
- [BEN 05] E. Ben Romdhane, H. Skik, « E-learning: éléments de réflexions autour d'une expérience en 'Blended Learning' développée dans le milieu universitaire », First International E-business Conference, 2005.
- [BRA 99] C. Bradford and I. Marshall. Analysing users www search behaviour. Lost in the Web- Navigation on the Internet, IEE Colloquium, 6(169):1–4, 1999.
- [BRA 96] G. Brajnik, S. Mizzaro, C. Tasso. Evaluating user interfaces to information retrieval systems: a case study on user support. In Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 128–136, Zurich, 1996.
- [BRU 05] Brusilovsky, P. Nejd, « Adaptive hypermedia and Adaptive Web », In: M. P. Singh (ed.) Practical Handbook of Internet Computing. Baton Rouge: Chapman Hall & CRC Press, pp. 1.1-1.14, 2005
- [BOT 04] J. Bottraud, G. Bisson, and M. Bruandet. Expansion de requêtes par apprentissage automatique dans un assistant pour la recherche d'information. In Actes du congrès CORIA, pages 89–105, Mars 2004.

- [BOU 99] M. Boughanem, C. Chrismont, and C. Soulé-Dupuy. Query modification based on relevance back-propagation in an ad hoc environment. *Information processing & management*, 35(2) :121-139, 1999.
- [BOU 05] Bouzeghoub M., Kostadinov D. « Personnalisation de l'information : aperçu de l'état de l'art et définition d'un modèle flexible de profils ». In *proc of CORIA 2005*, Grenoble, France, 9-11 Mars 2005, pp.201-218 (2005)
- [BOU 04] Bouzeghoub M., Kostadinov D., « Une approche multidimensionnelle pour la personnalisation de l'information », *Rapport PRiSM*, Versailles, France, 2004
- [BOU 11] Bousbia Nabila « Analyse des traces de navigation des apprenants dans un environnement de formation dans une perspective de détection automatique des styles d'apprentissage » thèse doctorat, université pierre été marrie curie, janvier 2011
- [BRA 00] K. Bradley, Rafter, B. Smyth, « Case-Based User Profiling for Content Personalisation », *Lecture Notes in Computer Science*, 2000
- [BRU 07] Brusilovsky, P., et Millán, E. (2007). User models for adaptive hypermedia and adaptive educational systems. *The Adaptive Web: Methods and Strategies of Web Personalization*. A. K. a. W. N. Brusilovsky. Berlin Heidelberg, New York, 4321: 3-53, Springer-Verlag.
- [BRO 05] Brown, E., Cristea, A., Stewart, C., et Brailsford, T.J. (2005). Patterns in Authoring of Adaptive Educational Hypermedia: A Taxonomy of Learning Styles. *Journal of Educational Technology and Society*, 8(3): 77-90.
- [CAR 06] Angela Carrillo-Ramos, Marlène Villanova-Oliver, Jérôme Gensel et Hervé Martin. « Gestion des préférences utilisateurs pour les Systèmes d'Information ubiquitaires » *Laboratoire LSR-IMAG, équipe SIGMA*. 681, 2006.
- [CHA 04] V. Challam. Contextual information retrieval using ontology based user profiles. Master's thesis, Jawaharlal Nehru Technological University, 2004
- [CHA 99] P. Chan. A non-invasive learning approach to building web user profiles. In *Workshop on Web usage analysis and user profiling*, Fifth International Conference on Knowledge Discovery and Data Mining, pages 7-12, San Diego, August 1999
- [CHE 00] Chevrier, J., Fortin, G., Théberge, M., et Le Blanc, R. (2000). Le style d'apprentissage : une perspective historique. *Le style d'apprentissage*, ACELF, XXVIII (1).

- [CHE 08] Chen, S. Y., et Liu, X. (2008). An integrated approach for modeling learning patterns of students in Web-based instruction: A cognitive style perspective. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 15(1): 28.
- [CHE 98] L. Chen and K. Sycara. Webmate : A personal agent for browsing and searching. In *Proceedings of the 2nd international conference on autonomous agents and multi agent systems*, Minneapolis, pages 10–13, 1998
- [CHI 00] W. Chien. Learning query behavior in the haystack system. Master's thesis, MIT, USA, June 2000.
- [CLA 00] Clark, S., Seat, E., et Weber, F. (2000). « The performance of engineering students on the group embedded figures test ». 30th ASEE/IEEE Frontiers in Education.
- [COF 04] Coffield, F., Moseley, D., Hall, E., et Ecclestone, K. (2004). Learning styles and pedagogy in post -16 learning. A systematic and critical review. London: Learning and Skills Research Centre.
- [CON 01] Conseil de l'Union Européenne, « Résolution du conseil sur le e-learning », *Journal Officiel des Communautés européennes*, Vol. 20, n°7, 2001.
- [CRO 01] W. B. Croft, St. Cronen-Townsend, V. Lavrenko, Relevance Feedback and Personalization : A Language Modeling Perspective, , *Computer Science*, DELOS Workshop, 2001.
- [DAI 02] H. Dai, T. Luo, M. Nakagawa, Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization, B. Mobasher, , *Data Mining and Knowledge Discovery*, 6(1):61-82, 2002.
- [DEB 04] De Bra P., Nejd W., « Adaptive Hypermedia and Adaptive Web-Based Systems », *Third International Conference*, AH 2004, Eindhoven, The Netherlands, August 23 - 26, 2004.
- [DEE 90] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. 1990. Indexing by Latent Semanti Analysis. *Journal of the American Society of Information Science*, 41(6), 391–407.
- [DUF 97] Dufresne, A., et Turcotte, S. (1997). Cognitive style and its implications for navigation strategies. *Artificial intelligence in education knowledge and media learning system*. B. Boulay, and Mizoguchi, R. Kobe (Japan), IOS Press, Amsterdam, pp. 287-293.
- [DUM 03] S. Dumais, E. Cuttrel, J. Cadiz, G. Jancke, R. Sarin, and D. Robbins. Stuff i've seen : a system for a personal information retrieval and re-use. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and*

Development, pages 72-79, Toronto, Canada, 2003. ACM Press

- [DUN 78] Dunn, R., et Dunn, K (1978). Teaching Students Through Their Individual Learning Styles: A Practical Approach. Virginia, Reston Publishing.
- [DUN 96] Dunn, R., Dunn, K., et Price, G. E. (1996). Learning Style Inventory. Price Systems, Lawrence, KS.
- [DUN 03] Dunn, R., Griggs, S. (2003). Synthesis of the Dunn and Dunn Learning Styles Model Research: Who, What, When, Where and So What The Dunn and Dunn Learning Styles Model and Its Theoretical Cornerstone. New York: St John's University.
- [FEL 93] Felder, R. (1993). Reaching the Second Tier: Learning and Teaching Styles in College Science Education. *College Science Teaching*, 23(5): 286-290.
- [FEL 95] Felder, R., et Henriques (1995). Learning and Teaching Styles In Foreign and Second Language Education. *Foreign Language Annals*, 28(1): 21-31.
- [FEL 88] Felder, R., et Silverman, L.K. (1988). Learning and Teaching Styles In Engineering Education. *Engr. Education* 78(7): 674-681
- [FEL 96] Felder, R., et Soloman, B. A. (1996). "ILS: Index of Learning Styles." <http://www.ncsu.edu/felder-public/ILSpage.html>, visité le 19/12/2009
- [FER 01] J. Ferreira, A. Silva, MySDI: A Generic Architecture to Develop SDI Personalised services, *ICEIS* (1) 2001: 262-270.
- [FIN 00] A Review and Analysis of Commercial User Modeling Servers for Personalization on the World Wide Web, J.Fink, A. Kobsa, *User Modeling and User-Adapted Interaction* 10: pp. 209-249, 2000.
- [FOR 97] Fortin, G., Chevrier, J., et Amyot, É. (1997). Adaptation française du «Learning Styles Questionnaire» de Honey et Mumford. Dans *Mesure et Évaluation en Éducation*, 19(3) : 95-118.
- [FOR 00] Ford, N., et Chen, SY. (2000). Individual differences, hypermedia navigation and learning: an empirical study. *Educ Multimedia Hypermedia*, 9(4) : 281-311.
- [GAU 07] S. Gauch, M. Speretta, A. Chandramouli, and A. Micarelli. User profiles for personalized information access. *The Adaptive Web*, 4321 :54-89, 2007.
- [GAU 03] Gaussier E., Jacquemin C., and Zweigenbaum P. (2003). Traitement automatique des langues et recherche d'information. In E. Gaussier and M.-H. Stéfani, editors, *Assistance Intelligente à la Recherche d'Information*, pages 71-96. Hermes, Paris.

- [GAU 99] Ontology Based Personalized Search, A. Pretschner, S. Gauch, roc 1th IEEE Intl. On Tools with Artificial Intelligence, pp. 391-398, Chicago, November 1999.
- [GOD 05] D. Godoy. Learning User Interests for User Profiling in Personal Information Agents. PhD thesis, Departamento de Computación y Sistemas, Facultad de Ciencias Exactas Univ. Nacional del Centro de la Provincia de Buenos Aires, Tandil, Bs. As., Argentina, 2005.
- [GRA 07] Graf, S. (2007). Adaptativity in Learning Management Systems Focussing on Learning Styles. Thèse de Doctorat, Vienna University of Technology, Austria.
- [GRA 84] Grasha, A. F. (1984). Learning Styles: the Journey from Greenwich Observatory (1796) to the College Classroom (1984). *Improving College and University Teaching*, 32(1) : 46-53.
- [GRE 02] Leonard Greenberg ; LMS et LCMS quelle est la différence, 2002, <http://www.learningcircuit.org>
- [GRE 79] Gregorc, A.F. (1979). Learning / Teaching styles: their nature and effects. In James W. Keefe (Ed.), *Student learning styles: diagnosing and prescribing programs*, Reston, VA: National Association of Secondary School Principals (NASSP), pp. 19-26.
- [GRE 61] Green W., Chomsky C. et Laugherty K. (1961). BASEBALL: An automatic question answerer. In *Proceedings of the Western Joint Computer Conference*: 219-224.
- [HEN 77] HENDRIX G., « Human engineering for applied natural language processing ». In *Proceedings of the 5th International Joint Conference on Artificial Intelligence (IJCAI'1977)*, 1977.
- [HER 07] N. Hernandez, J. Mothe, C. Chrisment, and D. Egret. Modeling context through do-main ontologies. *Journal of Information Retrieval, Contextual Information Retrieval Systems*, 10(2) :143–172, avril 2007
- [HON 86] Honey, P., et Mumford, A. (1986). *The Manual of Learning Styles*. Berkshire, England: Peter Honey
- [HON 92] Honey, P., et Mumford, A. (1992). *The Manual of Learning Styles*. Maidenhead, Peter Honey Publications.
- [HON 06] Honey, P., et Mumford, A. (2006). *The Learning Styles Helper's Guide*.

Peter Honey Publications Ltd., Maidenhead.

- [HUN 99] M. N. Huhn and L. Stephens. Personal ontologies. *Internet Computing*, 3(5), October 1999.
- [JUN] S. Y. Jung, J-H. Hong, T-S. Kim A Formal Model for User Preference, , IEEE International Conference on Data Mining, 2002.
- [KIN 95] Kinsella, K. (1995). Understanding and empowering diverse learners in the ESL classroom. In Joy M. Reid (Ed.), *Learning styles in the ESL/EFL classroom*, New-York: Heinle & Heinle Publishers, pp. 170-194.
- [KOL 74] Kolb, D.A. (1974). On Management and The Learning Process. Dans D.A. Kolb, I.M. Rubin, & J.M. McIntyre (Ed), *Organizational psychology - A book of readings*, Englewood Cliffs, N.J.: Prentice-Hall, pp. 27-42.
- [KOL 84] Kolb, D.A. (1984). *Experiential learning: experience as the source of learning and development*. New Jersey, Prentice Hall.
- [KOL 00] Kolb, D. A. (2000). *Facilitator's guide to learning*. Boston: Hay/McBer.
- [KOL 05] Kolb, A.Y, et Kolb, D.A, (2005). *The Kolb Learning Style Inventory - Version 3.1, 2005, Technical Specifications: 72p*.
- [KOS 03] D. Kostadinov. La personnalisation de l'information, définition de modèle de profil utilisateur. rapport de dea. Master's thesis, Université de Versailles, France, 2003.
- [LAN 97] Landauer, T. K., & Dumais, S. T. 1997. A Solution to Plato's Problem : The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge. *Psychological review*, 104, 211-240.
- [LAN 98] Landauer, T. K., Foltz, P. W., & Laham, D. 1998. Introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259-284.
- [LAM 04] P. Lamarche, V. (2004). *Styles d'apprentissage et rendement académique dans les formations en ligne*. Faculté des études supérieures, Université de Montréal.
- [LEE 05] Lee, C., Cheng, Y. W., Rai, S., et Depickere, A. (2005). What affect student cognitive style in the development of hypermedia learning system? *Comput. Educ.* 45(1) : 1-19.
- [LEM 03] Lemaire, B., & Dessus, P. 2003. Modèles cognitifs issus de l'analyse de la sémantique latente. *Cahiers Romains de Sciences Cognitives*, 1(1), 55-74.
- [LEH 77] LENHERT W., « A conceptual theory of question answering ». In *Proceedings of the 5th International Joint Conference on Artificial Intelligence (IJCAI'1977)*,

1977

- [LIU 04] F. Liu, C. Yu, and W. Meng. Personalized web search by mapping user queries to categories. In Proceedings of the 11th International Conference on Information and Knowledge Management, pages 558–565, Mclean, Virginia, November 4-9 2002. ACM.
- [LEH 81] LENHERT W. – « A computational theory of question answering ». In JOSHI A., WEBBER W. et SAG I. (éds.) Elements of discourse understanding. Cambridge University Press, Cambridge. 1981.
- [LIE 95] H. Lieberman. Letizia : An agent that assists web browsing. In Proceedings of the 14th International Joint Conference On Artificial Intelligence, Montreal, Canada, August 1995.
- [LI 02] X. Li and D. Roth. 2002. Learning question classifiers. In In COLING'02.
- [LIN 03] Lin J., Quan D., Sinha V., Bakshi K., Huynh D., Katz B. et Karger D. (2003). The Role of Context in Question Answering Systems. In Proceedings of the 2003 Conference on Human Factors in Computing Systems (CHI 2003), April 2003, Fort Lauderdale, Florida.
- [LIN 05] C. Lin, G. Xue, H. Zeng, and Y. YU. Using probabilistic latent semantic analysis for personalized web search. In Proceedings of the APWeb Conference, pages 707–717, Berlin Heidelberg, 2005. Springer-Verlag.
- [MAR 97] H. Marais and K. Bharat. Supporting cooperative and personal surfing with a desktop assistant. In Proceedings of the 10th annual ACM symposium on User interface software and technology, pages 129–138, Banff, Alberta, Canada, October 14-17 1997.
- [MER 02] Merrill, M. D. (2002). Instructional Strategies and Learning Styles: Which Takes Precedence? I. Trends and Issues in Instructional Technology. R. Reiser, and J. Dempsey, Columbus, OH, Prentice Hall: 99-106.
- [MIN 96] M. Minio and C. Tasso. User modeling for information filtering on internet services : Exploiting an extended version of the umt shell. In Workshop on User Modeling for Information Filtering on the WWW, Kailua-Kona, Hawaii, January 2-5 1996.
- [MLA 98] D. Mladenic. Personal webwatcher : design and implementation. In Technical Report IIS-DP-7472. J. Stefan Institute, Department for Intelligent Systems, 1998.
- [MCC 81] McCarty, B. (1981). The 4MAT System: teaching to learning styles with right/left mode techniques, Barrington, Ill.: Excel.

- [MCT 93] McTear M. F. 1993. User modelling for adaptative computer systems : a survey of recent developments. *Artificial Intelligence Review*, 7 : P.157-184,1993.
- [MYE 62] Myers, I., et Briggs, K. (1962). *The Myers-Briggs Type Indicator*, Princeton: Educational testing Services.
- [NAN 03] N. Nanas, U. Uren, and A. Deroeck. Building and applying a concept hierarchy re-presentation of a user profile. In *Proceedings of the 26th Annual ACM Conference on Research and Development in Information Retrieval SIGIR*, pages 154–204, 2003
- [ORA 00] ORAVEP, « Etude comparative technique et pédagogique des plateformes pour la formation ouverte et à distance », 2000
- [OUA 01] D. W. Oard and J. Kim. Modeling information content using observable behavior. In *Proceedings of the 64th Annual Meeting of the American Society for Information Science and Technology*, pages 38–45, USA, 2001.
- [PAP 98] Papadimitriou, C. H., Tamaki, H., Raghavan, P., & Vempala, S. 1998. Latent Semanti Indexing : A Probabilisti Analysis. Pages 159–168 of : *Proceedings of the 7th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*. Pentland, A., Piard, R. W., & S laroll , S. 1996. Photob o ok : content-based
- [PAS 76] Pask, G. (1976). *Styles and Strategies of Learning*. *British Journal of Educational Psychology*, 46: 128-148.
- [PAZ 96] M. J. Pazzani, J. Muramatsu, and D. Billsus. Syskill & webert : Identifying interesting web sites. In *Proceedings of the 30th National Conference on Artificial Intelligence*, pages 54–61, Portland, 1996
- [PCB 03] PCBI, Ministère de la jeunesse, de l'éducation nationale et de la recherche (France). *Etude des outils de gestion de ressources numériques pour l'enseignement.* Juin, 2003.
- [PET 99] Petrelli D., De Angeli A. et Convertino G. 1999. A user-centered approach to user modelling. Dans Kay J., éditeur, *UM99, Proceedings of the Seventh International Conference*, pages 255-264. *UM99_User Modelling*, Springer, Wien, NewYork, 1999.
- [PIT 02] Pithers, R.T. (2002). Cognitive learning style: a review of the field dependent-field independent approach. *Journal of Vocational Education and Training*, 54(1) : 117-132.
- [POH 96] Pohl W. 1996. learning about the user_ user modelling and machine learning. Dans Moustakis V. et Herrmann J., éditeur, *Workshop : Machine learning*

meets Human-Computer Interaction,
pages 29-40. ICML 96,1996.

- [POH 97] W. Pohl. Logic-Based Representation and Inference for User Modeling Shell Systems. PhD thesis, 1997.
- [POP 08] Popescu, E. (2008). Dynamic adaptive hypermedia systems for e-learning. Thèse de Doctorat, Crioiva University, Romania.
- [PRE 99] A. Pretschner. Ontology based personalized search. Master's thesis, University of Kansas, June 1999.
- [RID 98] Riding, R., et Rayner, S. (1998). Cognitive Styles and Learning Strategies, London: David Fulton Publishers.
- [RID 91] Riding, R. J., 1991, Cognitive Styles Analysis User Manual. Birmingham: Learning and Training Technology.
- [RUN 00] Rundle, S. M., et Dunn, R. (2000). The guide to individual excellence: A self directed guide to learning and performance solutions. New York: Performance Concepts International.
- [SAL 71] G. Salton. The SMART Retrieval System—Experiments in Automatic Document Processing. Prentice-Hall Inc, NJ, 1971.
- [SCH 77] SCHANK, R.C. et ABELSON, R. – Scripts, Plans, Goals, and Understanding. Hillsdale, NJ: Earlbaum Assoc, 1977.
- [SHE 01] S. Shearin, H. Lieberman, Intelligent Profiling by Example, , MIT Media Lab, Cambridge, 2001.
- [SCH 08] Schiaffino, S., Garcia, P., et Amandi, A. (2008). eTeacher: Providing personalized assistance to e-learning students. Computers & Education, 51: 1744-1754.
- [SOL 98] S J Soltysiak, I B Crabtree Automatic learning of user profiles-towards the personalisation of gent services, , BT Technol J. Vol 16 No 3, July 1998, pp 110-117.
- [SPE 04] S. Speretta and S. Gauch. Personalizing search based user search histories. In Proceedings of the 13th International Conference on Information Knowledge and Management, pages 238–239, 2004.
- [STA 06] Stash, N., Cristea, A., et de Bra, P. (2006). « Adaptation to Learning Styles in ELearning: Approach Evaluation». World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, VA, AACE.

- [STE 01] Stern, M. K. (2001). Using adaptive hypermedia and machine learning to create intelligent web-based courses. Thèse de Doctorat, University of Massachusetts Amherst, 2001, 170 p.
- [TAM 06] Tamine, L., Bahsoun, W., « Définition d'un profil multidimensionnel de l'utilisateur », Actes de CORIA 2006, Lyon, France, 15-17 mars, 2006, p. 225-236
- [TIN 97] Tinajero, D.C., et Paramo, M.F. (1997). Field dependence-independence and academic achievement: a re-examination of their relationship. *British Journal of Educational Psychology*, 67(2): 199-212.
- [TRA 04] J. Trajkova and S. Gauch. Improving ontology-based user profiles. In *Proceedings of the 8th Conference of Recherche d'Information Assistée par Ordinateur*, pages 380-389, University of Avignon (Vaucluse), France, April 26-28 2004.
- [TYN 05] Tyndiuk, F. (2005). *Référentiels Spatiaux des Tâches d'Interaction et Caractéristiques de l'Utilisateur influençant la Performance en Réalité Virtuelle*, Thèse de Doctorat, Université Bordeaux 2 : 234 p.
- [UNI 06] le vocabulaire de l'e-learning ; 2006
(<http://www.universitysoft.net>)
- [WEN 04] J. Wen, N. Lao, and W. Y. Ma. 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, Sheffield, United Kingdom, August 2004
- [WIL 82] Wilensky, R. 1982. Talking to UNI X in English: An Overview of UC. In the *Proceedings of the National Conference on Artificial Intelligence*. Pittsburgh, PA. August, 1982.
- [WIN 72] WINOGRAD T., *Understanding natural language*. Academic Press, New York, 1972.
- [W3C 05] W3C. 2005.
- [WIT 71] Witkin, H. A., Oltman, P. K., Raskin, E., et Karp, S. A. (1971). *A Manual For The Group Embedded Figures Test*. Consulting Psychologists Press, Palo Alto, CA.
- [WOL 07] Wolf, C. (2007). *Construction of an Adaptive E-learning Environment to Address Learning Styles and an Investigation of the Effect of Media Choice*. Thèse de Doctorat en philosophie, RMIT University, Melbourne, Australia.
- [WOO 72] Woods, W., Kaplan, R., and Webber, B. 1972. "The Lunar Sciences Natural Language Information System: Final Report". Bolt Beranek and Newman Inc.

(BBN), Report No. 2378.

- [ZEM 05] W. N. Zemirli, L. Tamine, and M. Boughanem. Accès personnalisé à l'information : vers la définition d'un profil utilisateur multidimensionnel. In International Symposium On Programming Systems, pages 20–28. USTHB, 2005.