



HAL
open science

**Modélisation de l'utilisateur et recherche coopérative
d'information dans les systèmes de recherche
d'informations multimedia en vue de la personnalisation
des réponses**

Amos David

► **To cite this version:**

Amos David. Modélisation de l'utilisateur et recherche coopérative d'information dans les systèmes de recherche d'informations multimedia en vue de la personnalisation des réponses. domain_stic.inge. Université Nancy II, 1999. tel-00144830

HAL Id: tel-00144830

<https://theses.hal.science/tel-00144830>

Submitted on 5 May 2007

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

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

**Modélisation de l'utilisateur et recherche coopérative d'information
dans les systèmes de recherche d'informations multimedia en vue de la
personnalisation des réponses**

MÉMOIRE

Présenté publiquement le 29 Juin 1999

pour l'obtention de l'

**Habilitation à Diriger des Recherches en
Sciences de l'information et de communication**

par

Amos Abayomi DAVID

Composition du jury

Directeur : Richard BOUCHÉ, Professeur en Sciences de l'Information et de la Communication
à l'ENSSIB (Lyon)

Rapporteurs : Jean-Baptiste CARPENTIER, Professeur en Sciences de l'Information et de la Communication à
l'université de la Sorbone, Directeur de CELSA

Jean-Paul METZGER, Professeur en Sciences de l'Information et de la Communication
à ERSICO, Université de Lyon 3

Odile THIERY, Professeur d'Informatique à l'université de Nancy 2.

Examineurs : Louis-Philippe LAPREVOTE, Professeur Sciences de l'Information et de la Communication à
Nancy 2, directeur du GRICP

Mohammed HASSOUN, Maître de conférences, HDR de Sciences de l'Information et de la Com-
munication à l'ENSSIB (Lyon)

Remerciements

Je remercie Monsieur Richard BOUCHE, Professeur à l'ENSSIB, qui a accepté malgré toutes ses responsabilités d'être directeur de cette habilitation. Je lui suis reconnaissant pour ses conseils et pour le temps qu'il a consacré pour la relecture de ce mémoire.

Je tiens à exprimer toute ma reconnaissance à Monsieur Louis-Philippe LAPREVOTE, Professeur à l'Université Nancy 2, pour l'aide et les conseils qu'il m'a apportés depuis ma nomination comme Maître de conférences à l'Université Nancy 2.

Un grand merci à Monsieur Jean-Baptiste CARPENTIER, Professeur à l'université de la Sorbone d'avoir accepté de rapporter sur ce travail.

Je remercie également Monsieur Jean-Paul METZGER, Professeur à l'université de Lyon 3 d'avoir accepté de rapporter sur ce travail et pour ses remarques.

Je remercie Mme Odile THIERY, professeur à l'Université Nancy 2, d'avoir accepté de rapporter ce travail et pour ses relectures de ce mémoire. Je lui suis reconnaissant pour ses conseils.

Je remercie Mme Marion CREHANGE, Professeur Emérite à l'Université Nancy 2 pour ses conseils.

Un grand merci à Monsieur Mohamed HASSOUN, Habilité à Diriger des Recherches à l'ENSSIB d'avoir accepté d'être membre du jury. Je lui suis très reconnaissant de ses encouragements.

Je tiens enfin à remercier les membres du groupe de discussion sur la recherche en IST, Roland DACHLET, Fidelia IBEKWE-SANJ-JUAN, Lyette LACOTE-GABRYSIK, Louis VIERA et David BUENO de l'Université de Malaga Espagne. Merci à Pierre TIAKO et à Abalo CHANGO pour leurs encouragements.



*A Elizabeth,
A Funsho,
A Seun,
A Bunmi,
A mon DIEU.*

Table des matières

Chapitre I	
Introduction générale	9
I.1 Le fonctionnement d'un SRI	10
I.1.1. Etape 1 : Présentation des besoins par l'utilisateur	10
I.1.2. Etape 2 : Traitement de la requête de l'utilisateur	11
I.1.3. Etape 3 : Reformulation de la requête de l'utilisateur	12
I.2 Conclusion et perspectives	13
I.2.1. La modélisation de l'utilisateur	13
I.2.2. La recherche coopérative d'information	14
I.2.3. Perspectives de recherche et de direction de chercheurs	15
I.2.4. Organisation de la suite du mémoire	16
Chapitre II	
La recherche d'information.....	18
II.1 Indexation des documents	18
II.2 Evolution des SRI	19
II.2.1. Evolution des informations gérées.....	19
II.2.2. Evolution des utilisateurs.....	20
II.2.3. Evolution des progrès technologiques	20
II.3 Recherche booléenne	20
II.4 Recherche vectorielle	25
II.5 Conclusion et propositions pour la recherche d'information	27
II.5.1. Recherche classique : les contraintes.....	28
II.5.2. Analyse de données	28
a)Analyse de fréquence	29
b)Analyse-croisée de données intra-champ	30
c)L'analyse-croisée inter-champ	32
Chapitre III	
La modélisation dynamique de l'apprenant : le projet BIRDS	35
III.1 Pourquoi la modélisation de l'utilisateur.....	35
III.2 Le modèle de l'apprenant : un modèle utilisateur particulier	37
III.2.1. Enseignement, apprentissage : quelles différences?.....	38
III.2.2. Approche cognitive en apprentissage	39
Chapitre IV	
La recherche coopérative d'informations	47
IV.1 L'architecture d'un système de recherche coopérative d'information.....	48
IV.2 Les modes de coopération	51
Chapitre V	
Le système METIORE.....	53
V.1 Le schéma global de METIORE	53
V.2 Analyse et accès à l'information	54
V.3 Exemples d'analyse et d'accès à l'information dans METIORE	55
V.3.1. L'interface graphique de METIORE	55
V.3.2. Recherche booléenne.....	57
V.3.3. Analyse de fréquence.....	58
V.3.4. Analyse intra-champ.....	58
V.3.5. Analyse inter-champ.....	59
V.3.6. Calcul de l'évolution des analyses.....	60
V.4 METIORE et la recherche coopérative.....	61

V.4.1.	L'interface de connexion	61
V.4.2.	L'interface d'échange de message textuel	61
V.5	Le modèle de l'utilisateur dans METIORE	62
V.5.1.	Implémentation actuelle	62
Chapitre VI		
	Le modèle de l'utilisateur dans un environnement de recherche coopérative d'information	64
VI.1	Du modèle de l'élève au modèle de l'utilisateur	64
VI.2	Comparaison des activités de l'élève et celles de l'utilisateur en SRI	67
VI.2.1.	La phase d'observation	67
VI.2.2.	La phase d'abstraction élémentaire	67
VI.2.3.	La phase de symbolisation et de raisonnement	68
VI.2.4.	La phase de créativité	68
VI.3	La théorie de catégories fondamentales.....	68
VI.4	Notre modèle de l'utilisateur	70
VI.5	Acquisition de connaissance sur l'utilisateur	73
VI.6	Analyse des connaissances sur un utilisateur	74
VI.6.1.	Fréquences occurrences valeurs	75
VI.6.2.	Fréquences co-occurrences intra-champ	77
VI.6.3.	Fréquences co-occurrences inter-champs.....	78
VI.7	Conclusion	79
VI.7.1.	Implémentation à l'aide de stéréotypes	80
VI.7.2.	Implémentation dynamique	80
VI.7.3.	L'historique - une base d'information à part entière	80
VI.8	Le modèle de l'utilisateur dans un contexte de recherche coopérative	81
VI.9	Conclusion	82
Chapitre VII		
	Représentation formelle de nos modèles et architectures	84
VII.1	Le modèle de session.....	84
VII.2	Représentation des activités de l'utilisateur	86
VII.3	L'architecture d'un système de recherche coopérative d'information.....	87
Chapitre VIII		
	Perspectives et projet de recherche	89
VIII.1	Mes parcours en recherches scientifiques.....	89
VIII.2	Orientations dans mes recherches.....	90
VIII.2.1.	Formalisation de l'interprétation des connaissances sur l'utilisateur.....	90
VIII.2.2.	Automatisation de l'interprétation des connaissances sur l'utilisateur	90
VIII.2.3.	Extraction des connaissances d'un expert	91
VIII.2.4.	Exploitation des connaissances des experts.....	91
VIII.3	Nos outils dans un système d'ingénierie documentaire.....	91
VIII.4	Conclusion	92
	Bibliographie	93
	Index	97
	Annexes	100
Annexe I		
	Le prototype STREEMS	101
I.1	Objectif de STREEMS	101
I.2	Le modèle de données de STREEMS	101
I.3	Architecture de STREEMS	105
I.4	Les interfaces de STREEMS	107

I.5	Outil de dialogue	109
I.5.1.	Classification sans contraintes	109
I.5.2.	Classification avec contraintes	110
Annexe II.	
Curriculum Vitae	113
II.1	Curriculum Vitae	113
II.2	Responsabilités administratives	114
II.3	Animations scientifiques	114
II.4	Activités d'encadrement	115
II.5	Coopérations internationales	115
II.6	Publications	116
II.6.1.	Thèse	116
II.6.2.	Conférences invitées	116
II.6.3.	Revue	117
II.6.4.	Conférences internationales avec comité de sélection	117
II.6.5.	Conférences nationales avec comité de sélection	118
II.6.6.	Rapports pour les projets européens	118
II.6.7.	Rapports internes	118

Table des figures

Figure 1. : Modèle de SRI pour intégrer les données sur les utilisateurs	11
Figure 2. : Architecture du prototype BIRDS	41
Figure 3. : Schéma global de l'architecture d'un SRCI	50
Figure 4. : Diagramme d'états pour la communication entre deux applications	52
Figure 5. : Le schéma global de METIORE	54
Figure 6. : L'interface graphique de METIORE	56
Figure 7. : L'interface de la recherche par analyse-croisée avec contraintes	57
Figure 8. : Une recherche booléenne	57
Figure 9. : La fréquences des publications des auteurs	58
Figure 10. : Analyse de co-auteurs	58
Figure 11. : Présentation graphique des fréquences des co-auteurs	59
Figure 12. : Analyse-croisée inter-champ	59
Figure 13. : Graphique analyse-croisée inter-champ	60
Figure 14. : Evolution des co-auteurs	60
Figure 15. : L'interface d'échange de message textuel entre deux utilisateurs	61
Figure 16. : Algorithme d'exploitation du modèle de l'utilisateur dans METIORE	62
Figure 17. : Modèle objet du modèle de l'utilisateur et de session	85
Figure 18. : Architecture d'un SRCI en formalisme OMT	87
Figure 19. : Le modèle de données de STREEMS en représentation OMT	103
Figure 20. : L'architecture de STREEMS en représentation OMT	106
Figure 21. : Le modèle de l'interface de STREEMS en représentation OMT	107
Figure 22. : Interface de STREEMS	108
Figure 23. : Outil de dialogue textuel	109
Figure 24. : Classification sans contrainte	110
Figure 25. : Classification avec contraintes	111

Résumé

Le fonctionnement d'un SRI est constitué de plusieurs étapes : l'étape de présentation des besoins par l'utilisateur, celle du traitement de la requête de l'utilisateur par le système, celle de l'évaluation des solutions du système par l'utilisateur et celle de la modification ou mise à jour de la requête par le système ou par l'utilisateur. Dans chacune de ces étapes, la prise en compte des particularités de chaque utilisateur est nécessaire pour rendre le système plus efficace, en particulier par rapport à la pertinence des réponses du système. On exige de l'utilisateur de bien définir ses besoins et d'être capable de les traduire dans le langage du système. Le jugement de pertinence d'une réponse par l'utilisateur dépend des paramètres qui lui sont propres.

Dans le cadre de développement de SRI, nos travaux de recherche portent sur deux axes : la modélisation de l'utilisateur pour la représentation explicite des utilisateurs et la définition d'un système de recherche coopérative d'information comme un nouveau type de fonctionnement d'un SRI. En ce qui concerne l'utilisateur, nous travaillons sur trois types de connaissance pour le représenter : ses *habitudes évocatrices*, son *niveau de connaissance* basé sur la notion de *catégorie fondamentale*, et ses *jugements de pertinence* sur les solutions du système en réponse à ses requêtes. Par l'observation du fonctionnement d'un centre de documentation et par l'efficacité des résultats obtenus par le dialogue entre l'utilisateur et un médiateur du centre, nous avons orienté nos travaux vers le développement d'un système de recherche coopérative d'information (SRCI). Dans un SRCI, l'utilisateur pourra effectuer sa recherche en collaboration avec un médiateur sur des postes de travail distants.

Pour valider ces propositions, nous avons développé trois systèmes qui nous ont permis de mieux définir ces objectifs de recherche. Le premier système (BIRDS), un système d'apprentissage a été expérimenté au Lycée Henri Poincaré, Nancy en 1989/1990. Le deuxième système (STREEMS) a été développé dans un cadre d'un projet européen. Le système permet d'accéder et d'analyser des informations sur les arbres autorisés par l'Union Européenne pour le reboisement des forêts. Le troisième système (METIORE) permet d'accéder aux références bibliographiques du laboratoire LORIA et de les analyser. Nos travaux de recherche ont donné lieu à plusieurs publications. Une thèse de doctorat de l'université de Malaga, Espagne est actuellement en cours à la suite de ces travaux.

Les perspectives à court terme sont de représenter chaque utilisateur par l'historique de ses activités. Ces activités sont représentées sous forme de documents qui pourront être analysés par notre méthode d'accès et d'analyse de l'information. Nous comptons employer les techniques de raisonnement à base de cas pour exploiter ces analyses pour personnaliser les réponses du système et exploité par un médiateur dans un contexte de recherche coopérative d'information.

A plus long terme, nous pensons qu'il faudra étudier le processus de communication (ou de dialogue) entre l'utilisateur et le médiateur dans un contexte de SRCI. Les systèmes que nous avons développé pourront servir d'outils d'observation. Il faudra également étudier l'intégration de ces outils dans un environnement de recherche documentaire permettant non seulement d'obtenir et d'analyser les informations d'une base mais permettant d'accéder aux documents primaires.

Chapitre I

Introduction générale

Le besoin en informations économiques, scientifiques, techniques, juridiques, sociales, etc est exprimé de plus en plus fort aussi bien par des personnes physiques, (par exemple, des chercheurs, du grand public) que par des personnes morales (par exemple les entreprises, les gouvernements, les organismes sociaux, etc). Pratiquement tous les acteurs de la vie économique et sociale sont concernés par ce phénomène de besoin en information : les chercheurs dans tous les domaines scientifiques, les agriculteurs, les commerçants, les dirigeants d'entreprise, les économistes, les conseillers d'orientation professionnelle, etc. Parallèlement, les entreprises qui proposent des services d'informations spécialisées se multiplient. On trouve par exemple des entreprises qui proposent des services juridiques pour fournir des informations adaptées à des individus du grand public et des courtiers en informations qui proposent leurs services aux dirigeants d'entreprise pour les aider dans la prise de décision. Les chercheurs ont recours aux responsables des centres de documentation de leurs centres de recherche ou de leurs entreprises pour connaître les travaux les plus récents sur des sujets spécifiques. L'une des particularités des besoins en information exprimés par les différents demandeurs porte sur le caractère de «pertinence» des informations recherchées. Une autre particularité porte sur le caractère «veille» pour trouver les informations demandées. Il s'agit ici non seulement de trouver des informations pertinentes mais aussi les informations les plus récentes et parfois de réaliser une analyse sur l'évolution des informations trouvées sur une certaine durée.

Quelle que soit la façon de voir les utilisateurs (ou demandeurs), nous constatons leur diversité. Nous pouvons remarquer également que les acteurs sont de catégories différentes : les demandeurs d'information, ceux qui effectuent la recherche d'information à proprement parler et les concepteurs de système de recherche d'informations.

Les fournisseurs d'information peuvent être classés en deux catégories. La première catégorie consiste en ceux qui mettent en place une base d'information et qui l'alimentent. Dans cette catégorie se trouvent les centres de documentation, les instituts d'informations scientifiques et techniques, les instituts de propriétés industrielles et les organismes gouvernementaux comme par exemple la commission de l'union européenne. La deuxième catégorie de fournisseurs sont les courtiers d'informations. Ils utilisent essentiellement les sources d'informations proposées par les fournisseurs de la première catégorie. Les techniques employées pour trouver des informations par les organismes fournisseurs de la deuxième catégorie reposent sur des techniques souvent semi-automatiques, à savoir, la recherche d'informations dans des bases de données et l'analyse de ces informations par des techniques souvent manuelles.

Deux domaines d'étude s'intéressent aux problématiques liées à la recherche d'informations pertinentes et adéquates : les études en systèmes de recherche d'informations (SRI) et les systèmes de veilles (SVX) (veille technologique, veille économique, veille concurrentielle, etc.).

Dans le domaine de recherche en SRI, l'intérêt porte sur la mise en correspondance des informations du système et le besoin exprimé par l'utilisateur. Les techniques de mise en correspondance vont de la mise en correspondance lexicale à l'interprétation sémantique du besoin de l'utilisateur. Un certain nombre de techniques employées dans les SRI pour fournir des réponses aux utilisateurs sont développées dans le *Chapitre II (La recherche d'information)*.

Les SVX sont des systèmes spécialisés pour gérer des informations d'un domaine particulier ou pour traiter des demandes d'informations particulières. Par exemple une entreprise de bâtiment peut désirer s'informer et suivre le développement des nouveaux matériaux de construction pour décider de sa stratégie d'investissement ou pour choisir les sujets de recherche à soutenir dans son service de recherche et de développement. Le ministère de l'enseignement et de la recherche peut décider d'analyser les sujets de recherche qui évoluent vers de nouveaux sujets non encore traités afin de décider les axes de recherche à soutenir.

La pertinence d'une réponse peut être évaluée par rapport à la requête de l'utilisateur. Dans un contexte de veille, elle porte sur les objets de la veille. La pertinence concerne la veille d'abord pour retrouver les informations nécessaires pour des analyses et ensuite adapter l'analyse à l'objectif spécifique du demandeur.

Ce n'est pas notre propos ici de développer les recherches en SVX. En revanche, sur le traitement des besoins en information et le développement des SRI, notre intérêt porte sur la notion de **pertinence** d'information ou la **personnalisation** des réponses du système. Pour mieux situer nos travaux de recherche, nous présentons un modèle de SRI dans la *Figure 1. (Modèle de SRI pour intégrer les données sur les utilisateurs)*.

I.1 Le fonctionnement d'un SRI

Le fonctionnement d'un SRI est constitué de plusieurs étapes : l'étape de présentation des besoins par l'utilisateur, celle du traitement de la requête de l'utilisateur et celle de la modification ou mise à jour de la requête de l'utilisateur. Chacune de ces étapes est présentée dans les sections suivantes.

I.1.1. Etape 1 : Présentation des besoins par l'utilisateur

Pour utiliser un SRI, l'utilisateur doit présenter ses besoins au système. Les besoins sont exprimés par des requêtes. Les requêtes sont, en fait, les besoins de l'utilisateur exprimés dans le langage du système. Cette étape constitue une première difficulté pour les utilisateurs. On exige de l'utilisateur de bien définir ses besoins et d'être capable de les traduire dans le langage du système.

En effet, la formulation d'un besoin par un utilisateur dépend de beaucoup de paramètres qui lui sont propres. Par exemple cette expression dépend de son niveau de connaissance sur l'in-

formation recherchée, de ses expériences dans le traitement du même type de besoin et de ses expériences dans l'utilisation du SRI. Cette étape du fonctionnement d'un SRI permet de mettre en évidence la nécessité de prendre en compte les particularités des utilisateurs, ceci pour traiter les besoins de chacun d'eux différemment. Cette étape est illustrée dans la figure 1 par la flèche entre *utilisateur* et *besoins*.

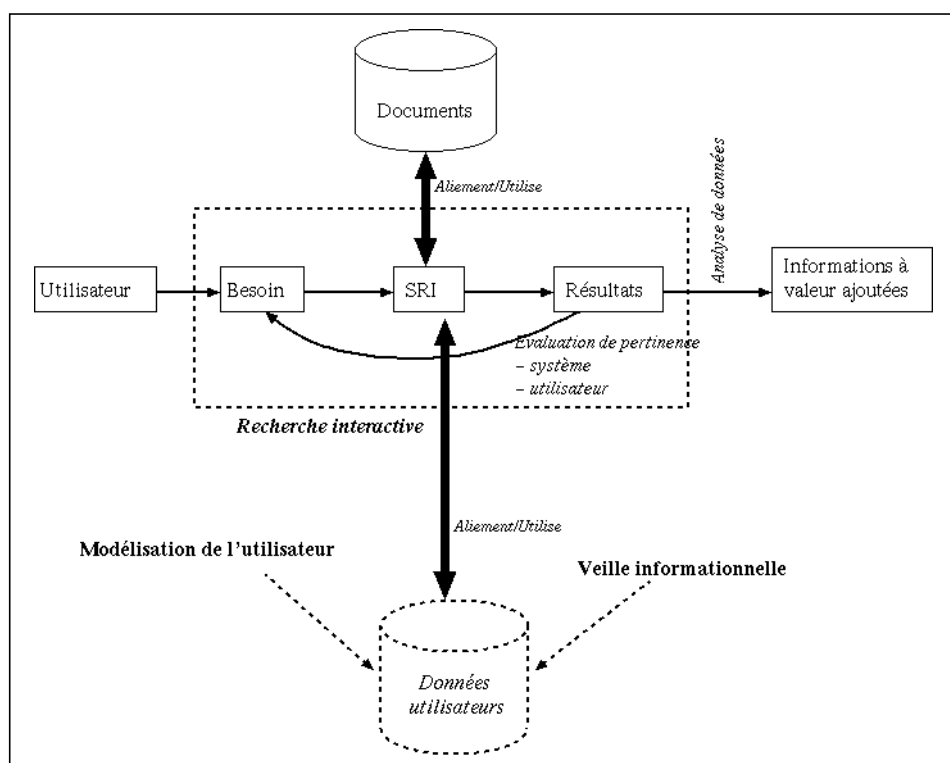


Figure 1. Modèle de SRI pour intégrer les données sur les utilisateurs

Présentation de la figure 1

- *Le SRI utilise ou alimente une «base de documents».*
- *Le SRI utilise ou alimente une base de «données sur l'utilisateur».*
- *La base de données sur l'utilisateur est basée sur un «modèle de l'utilisateur». Elle comprend des données sur les comportements de l'utilisateur ainsi que les synthèses de ces comportements. Les synthèses sont basées sur les techniques de «veille informationnelle».*
- *L'utilisateur exprime un «besoin» ; le besoin est présenté au SRI ; après traitement du besoin, le SRI présente le «résultat» du traitement*
- *La pertinence du résultat est évalué par le système et par l'utilisateur. Le résultat de cette évaluation constitue une partie du besoin reformulé.*
- *Le résultat peut être analysé pour trouver des «informations à valeur ajoutée»*

I.1.2. Etape 2 : Traitement de la requête de l'utilisateur

Cette étape est illustrée dans la figure 1 par les flèches entre *Besoin*, *SRI* et *Résultat*. Le traitement de la requête de l'utilisateur passe par la mise en correspondance de celle-ci avec les

données gérées par le système, illustrée par la flèche entre *SRI et Documents* (la base documentaire). Plusieurs techniques sont utilisées pour la mise en correspondance. Les techniques les plus utilisées sont la technique de recherche booléenne et la recherche vectorielle. L'ensemble des données trouvées à la suite de la mise en correspondance est présenté à l'utilisateur. Certains points du processus de la mise en correspondance doivent être soulignés. Ces points portent sur *l'interprétation des termes utilisés dans la requête de l'utilisateur*.

Afin de rendre le résultat aussi pertinent que possible au besoin de l'utilisateur, certains systèmes transforment des termes utilisés dans la requête. Des critères de transformation sont basés soit sur des expériences du concepteur du système et du spécialiste du domaine des données gérées par le système, soit sur les résultats d'évaluations de pertinence par les utilisateurs.

Nous remarquons encore dans cette étape, l'importance et l'influence des spécificités de chaque utilisateur sur le jugement de pertinence des solutions proposées par le système. Les jugements de pertinence par l'utilisateur constitue une information essentielle dans la compréhension de ses besoins.

I.1.3. Etape 3 : Reformulation de la requête de l'utilisateur

Ainsi que nous l'avons dit dans la section précédente, le résultat du système peut être évalué pour vérifier sa pertinence par rapport au besoin de l'utilisateur. Pour un besoin spécifique, l'arrêt du système dépend du résultat de l'évaluation de la pertinence des propositions du système. En effet, beaucoup de SRI permettent de modifier soit automatiquement, soit manuellement, la requête initiale de l'utilisateur en prenant en compte le résultat de l'évaluation de pertinence des solutions proposées par le système. L'hypothèse de base de cette fonctionnalité est que l'utilisateur ne sait pas forcément ce qu'il cherche ou alors qu'il est incapable d'exprimer son besoin par une seule requête. Les solutions proposées par le système en réponse au besoin exprimé par l'utilisateur sont donc susceptibles d'être polluées par des bruits, c'est-à-dire des informations trouvées qui ne sont pas demandées ou encore par le silence, c'est-à-dire des informations demandées mais qui ne sont pas trouvées. Cette étape est illustrée dans la figure 1 par la flèche entre *Résultat et Besoin* avant une nouvelle mise en correspondance. Nous indiquons dans la figure que l'analyse de pertinence peut être faite par le système et par l'utilisateur.

Nous voyons ici un pas vers la prise en compte des données sur les utilisateurs. On ne se contente plus seulement du processus de mise en correspondance mais de proposer un système capable de s'adapter aux difficultés des utilisateurs. Dans certains systèmes, le jugement de pertinence par l'utilisateur est intégré dans la modification de la requête initiale. L'itération des étapes pour la formulation de requête, la mise en correspondance après transformation éventuelle, de l'évaluation du résultat et la modification de la requête initiale constituent ce qui est appelé «la recherche interactive», étapes que nous avons délimitées par le rectangle pointillé dans la figure 2.

I.2 Conclusion et perspectives

Nous avons montré les différentes phases du fonctionnement d'un SRI où la prise en compte des particularités de chaque utilisateur est nécessaire pour rendre le système plus efficace en particulier par rapport à la pertinence de ses réponses. Ce constat est à la base de nos études sur la représentation explicite des utilisateurs ainsi qu'à la définition d'un nouveau type de fonctionnement d'un SRI. Nos travaux de recherche portent sur deux axes : la modélisation de l'utilisateur pour la représentation explicite des utilisateurs et la définition d'un système de recherche coopérative d'information comme un nouveau type de fonctionnement d'un SRI. Nous avons montré dans les différentes étapes des raisons pour une meilleure prise en compte des particularités des utilisateurs, qui est l'objet de notre premier axe de recherche. Cette prise en compte de particularités des utilisateurs est illustrée par l'intégration des *données utilisateurs* de la figure 1. Le modèle de l'utilisateur permettra d'alimenter cette base et nous comptons utiliser des techniques de veille informationnelle pour synthétiser ces données.

Le nouveau type de fonctionnement d'un SRI, non illustré par la figure 1, la recherche coopérative d'information que nous proposons, est présenté dans la section

I.2.1. La modélisation de l'utilisateur

Le premier axe de notre recherche concerne la modélisation de l'utilisateur dans un système de recherche d'informations. Nous avons montré l'influence des spécificités des utilisateurs dans toutes les phases du processus de recherche d'information : pour interpréter le besoin de l'utilisateur, le système a besoin de connaître son niveau de connaissance et ses préférences ; pour modifier sa requête initiale, le système a besoin de connaître son évaluation des solutions proposées. Compte tenu de ces influences, nous avons orienté nos travaux de recherche vers la représentation explicite des connaissances sur les utilisateurs. Cette orientation donne aux connaissances sur l'utilisateur un statut particulier dans l'architecture du système. Les données sur les utilisateurs sont représentées par le modèle de l'utilisateur. Les techniques de veille informationnelle sont utilisées pour alimenter le modèle de l'utilisateur et analyser les données le concernant. Le résultat de l'analyse sur les données de l'utilisateur permet une meilleure connaissance à son sujet et facilite ainsi le jugement sur son niveau de connaissance, ses préférences et donne des indications sur la manière d'interpréter ses besoins. L'intégration des connaissances sur l'utilisateur, leur conférant un statut aussi important que les autres données du système, est illustrée par la *Figure 1. (Modèle de SRI pour intégrer les données sur les utilisateurs)*

Pour cet axe, les problèmes auxquels nous voulons apporter des réponses sont : de quelle façon représenter un utilisateur d'une manière explicite dans un SRI ? quelles sont les connaissances sur l'utilisateur à représenter dans le système ? comment acquérir ces connaissances ? et comment les exploiter ? Nous travaillons sur trois types de connaissance sur l'utilisateur : ses *habitudes évocatives*, son *niveau de connaissance* basé sur la notion de *catégorie fondamentale*, et ses *jugements de pertinence* sur les solutions du système en réponse à ses requêtes. Pour le concept d'habitudes évocatives et de catégories fondamentales, nous nous sommes basés sur les résultats des études en sciences cognitives [fod86, gar88, den89].

La notion d'habitudes évocatives

L'une des connaissances sur l'utilisateur que nous essayons d'acquérir concerne ce que les cognitiens appellent les habitudes évocatives. Nous avons retenu quatre habitudes évocatives qui sont l'*observation*, l'*abstraction élémentaire* ou l'acquisition de vocabulaire, la *symbolisation et le raisonnement*, et la *créativité*. Nous observons les activités (et l'évolution de ces activités) de l'utilisateur dans les quatre habitudes évocatives pour connaître son niveau de connaissance dans le domaine. Nous observons également les termes utilisés par l'utilisateur et leurs contextes d'utilisation (par rapport aux quatre habitudes évocatives) pour renforcer notre connaissance de son niveau de connaissance. En fonction de ces observations qui se font sur toutes les sessions de recherche d'information par l'utilisateur et par l'utilisation des outils de veille informationnelle, le système propose des réponses adaptées à son niveau de connaissance.

La notion de catégories fondamentales

Nous avons adopté le concept de catégories fondamentales pour connaître le niveau d'abstraction des termes utilisés par l'utilisateur. Ce concept permet de hiérarchiser les niveaux d'abstraction des termes d'un domaine basé sur les termes considérés d'une catégorie dite fondamentale. Les niveaux d'abstraction des termes employés par l'utilisateur, mesurés par rapport aux niveaux des catégories fondamentales, permettent de mieux cerner le niveau de connaissance de l'utilisateur et ainsi d'adapter les termes utilisés dans les réponses du système à ce niveau d'abstraction. Ce concept sera développé dans le *Chapitre III (La modélisation dynamique de l'apprenant : le projet BIRDS)*.

Jugement de pertinence

Le jugement de pertinence diffère d'un individu à un autre. La prise en compte des particularités des utilisateurs nécessite l'acquisition de leurs jugements sur les pertinences des solutions proposées par le système. Cet aspect de notre recherche est développé dans la section *VI.6 (Analyse des connaissances sur un utilisateur)*. L'objectif que nous recherchons est de savoir le plus précisément possible pourquoi un jugement a été émis sur une réponse du système.

I.2.2. La recherche coopérative d'information

Malgré tous les outils et les techniques intégrés dans les SRI pour mieux comprendre les besoins des utilisateurs et pour apporter des réponses adéquates à leurs besoins, il arrive encore que les utilisateurs ne parviennent pas à obtenir des résultats «satisfaisants». Comme nous l'avons présenté dans les sections précédentes, les raisons de cet échec sont multiples. Par exemple l'échec peut être dû au fait que :

- l'utilisateur ne sait pas formuler sa demande,
- l'utilisateur ne sait pas comment utiliser le système. Ici l'utilisateur sait ce qu'il cherche et il obtient des solutions, mais il ne sait pas comment utiliser les fonctions du système pour améliorer les solutions du système,

- les propositions du système ne correspondent pas au niveau de connaissance de l'utilisateur,
- l'utilisateur détient une solution partielle et il cherche des compléments de solution mais le système lui donne toutes les solutions possibles.

Nous remarquons que des échecs dus aux raisons ci-dessus sont beaucoup moins nombreux quand l'utilisateur peut s'adresser à un médiateur humain comme dans le contexte d'un centre de documentation. En effet, un dialogue est souvent engagé entre l'utilisateur et le médiateur, avant d'effectuer la recherche d'information au moyen d'un SRI. Le dialogue porte souvent sur le besoin de l'utilisateur, exprimé en langage naturel, sur ce que l'utilisateur a déjà fait et sur ce qu'il a déjà obtenu comme résultat. Ainsi, le médiateur dispose d'informations sur l'utilisateur lui permettant d'orienter les solutions à proposer à l'utilisateur. Les aides que le médiateur peut apporter à l'utilisateur peuvent porter sur les termes les mieux adaptés à utiliser pour interroger le SRI, la redéfinition de l'objectif global ou, encore, les fonctions du système les mieux adaptées pour rechercher les informations souhaitées.

En raison de l'efficacité des résultats obtenus à l'issue de ce type de dialogue entre l'utilisateur et le médiateur, nous avons orienté nos travaux vers le développement d'un système de recherche coopérative d'information (SRCI). Dans un SRCI, l'utilisateur pourra dialoguer avec un médiateur pour effectuer sa recherche sur des machines distantes.

Comme nous l'avons présenté dans la section précédente, nous travaillons également sur la représentation explicite des utilisateurs afin d'acquérir des connaissances sur chacun d'eux. Nous disposons actuellement d'outils d'analyse des données permettant d'obtenir des synthèses d'information sur les comportements d'un utilisateur. Ces synthèses d'information sont en effet très importantes dans un contexte de recherche coopérative d'information. En effet, outre le dialogue possible entre le médiateur et l'utilisateur, le médiateur peut, à partir de ces synthèses d'information, connaître le passé de l'utilisateur.

La présentation plus détaillée d'un SRCI, son architecture, ses applications, sont présentées dans le *Chapitre IV (La recherche coopérative d'informations)*.

I.2.3. Perspectives de recherche et de direction de chercheurs

Comme présenté dans les sections *I.2.1. (La modélisation de l'utilisateur)* et *I.2.2. (La recherche coopérative d'information)*, nos travaux de recherche portent sur deux axes : la modélisation de l'utilisateur en vue d'une intégration des données personnelles de l'utilisateur et le développement de SRCI. Les résultats du premier axe de recherche seront effectivement exploités dans le cadre d'une recherche coopérative. Par exemple, les synthèses de données sur l'utilisateur seront disponibles pour le médiateur, ce qui permettra au médiateur d'orienter la façon d'aider l'utilisateur.

Nous avons déjà encadré des recherches sur les deux axes (deux thèses de CNAM, un DEA et une co-direction d'une thèse de doctorat). Ces expériences d'encadrement nous ont permis de mieux définir ces projets de recherche et de mieux cerner la façon de les conduire.

Nous sommes en discussion actuellement avec des collègues de la 71e section des Universités de Nancy 2, de Metz et de Strasbourg (cinq maîtres de conférences) pour mettre en place un groupe de recherche sur «les données personnelles et les technologies de l'information». Les premières discussions nous ont permis de définir comme problématique «l'acquisition, l'organisation, la gestion et l'exploitation des données personnelles dans les technologies de l'information». Nous entendons par «données personnelles» les informations nécessaires sur un utilisateur pour personnaliser les réponses du système. Dans ces études, un utilisateur peut être un individu ou un organisme, par exemple dans un contexte de veille technologique.

J'assure également actuellement la co-direction d'une thèse de doctorat de l'Université de Malaga en Espagne. Elle est financée en partie par des fonds d'un projet européen dont je suis le responsable scientifique pour le laboratoire LORIA (Laboratoire lorrain de Recherche en Informatique et ses Applications). Le sujet de la thèse est «La modélisation de l'utilisateur et son exploitation dans un contexte de recherche coopérative d'information».

Nous constatons que, actuellement, il n'y a pas de formation pour préparer une thèse en IST dans la région Nord-Est de la France. Le groupe de recherche que nous sommes en train de mettre en place pourra servir à définir un programme d'étude de niveau DEA et ainsi faciliter la préparation d'une thèse. Il ne sera pas nécessaire de créer une formation distincte dans un premier temps car la formation peut s'insérer dans un DEA existant comme par exemple le DEA information et communication de Nancy-Metz. J'ai présenté une demande pour intégrer une option IST dans ce DEA.

I.2.4. Organisation de la suite du mémoire

Le reste du mémoire est organisé en neuf chapitres. Le *Chapitre II (La recherche d'information)* détaille le fonctionnement générale d'un SRI ainsi que deux des principales techniques de mise en correspondance : la recherche booléenne et la recherche vectorielle. Le *Chapitre III (La modélisation dynamique de l'apprenant : le projet BIRDS)* présente nos premières expériences dans le domaine de la modélisation de l'utilisateur où l'utilisateur est un apprenant dans un système d'apprentissage basée sur l'image. Nous détaillons également les raisons de la modélisation, quoi modéliser, comment acquérir les connaissances sur les utilisateurs et les techniques que nous avons proposé pour l'exploitation de ces connaissances.

Le nouveau type de système de recherche d'information que nous proposons, système de recherche coopérative d'information, est présenté au *Chapitre IV (La recherche coopérative d'informations)*. Son architecture, ses modes de fonctionnement et l'exploitation du modèle de l'utilisateur dans ce contexte y sont présentés. Notre première application concrète de nos propositions dans le cadre d'un projet européen (le prototype STREEMS)¹ est présentée en *Annexe I (Le prototype STREEMS)*. Notre deuxième application concrète (le prototype METIORE) est

présentée dans le *Chapitre V (Le système METIORE)*.

Bien que nos propositions et nos perspectives de recherche soient résumées dans chaque chapitre, une synthèse de ces perspectives est présentée au *Chapitre VIII (Perspectives et projet de recherche)*.

Mes animations scientifiques, responsabilités scientifiques, administratives et pédagogiques, mes coopérations scientifiques internationales et nationales et mes publications sont présentées en *Annexe II (Curriculum Vitae)*.

1. La plupart des exemples que nous donnons dans ce rapport sont de ce prototype

Chapitre II

La recherche d'information

Les objets des applications des SRI sont des documents. Les documents sont des produits ou des entités d'informations. Dans la nouvelle forme numérique de stockage de documents, on distingue deux catégories de modes de représentation : la représentation image et la représentation structurée. En représentation image, les pages des documents sont comme des prises de photo. Les éléments d'une page d'un document ne peuvent pas être dissociés. Par exemple les caractères, les mots, les paragraphes et les images ne sont pas dissociés. En représentation structurée, tous les éléments d'une page d'un documents sont dissociés. On pourra, par exemple, extraire les caractères, les mots, les paragraphes et les images d'une page d'un document.

Dans la plupart de SRI, quel que soit le mode de représentation des documents gérés, les documents primaires ne sont pas gérés directement. Les documents sont dans un premier temps analysés et ensuite représentés sous une forme de base documentaire. Cette analyse porte aussi bien sur la forme que sur les contenus des documents. Compte tenu de l'importance de cette phase d'analyse dans la chaîne de la mise en place d'un SRI, nous présentons brièvement les problèmes liés à l'indexation des documents.

II.1 Indexation des documents

Les informations qui décrivent les documents sont relativement faciles à représenter. L'approche utilisée consiste à identifier les propriétés des documents. Par exemple, pour représenter des livres, on utilisera ses propriétés comme *titre*, *auteurs*, *date édition*, etc. Les informations contenues dans les documents sont beaucoup plus difficiles à représenter. L'approche généralement utilisée consiste à utiliser un ou deux champs pour stocker des valeurs décrivant les contenus des documents. Par exemple, dans certains systèmes, on utilise les champs *mots-clés* et *résumé*. Les études sur les techniques d'indexation démontrent la difficulté de cette représentation.

Pour tout type de documents, il y a un groupe de propriétés qui représentent les contenus des documents où chaque propriété du groupe représente les documents à des niveaux de détail différents. Par exemple, le contenu d'un livre peut être représenté en utilisant son titre, son résumé ou le livre entier. Pour chaque application, il faut déterminer le niveau de détail de la représentation des documents. Dans certaines applications on peut se contenter de représenter les documents par leurs titres et leurs résumés. Par exemple les contenus des brevets sont généralement représentés uniquement par leurs titres et leurs résumés. Dans d'autres applications, on demande une représentation plus détaillée où il faut représenter la totalité du corps de chaque document.

La technique d'indexation peut être manuelle ou automatique. Dans le cas d'une indexa-

tion manuelle, un indexeur est chargé de faire le choix des descripteurs pour les valeurs des propriétés choisies pour la description. En indexation automatique, le système prend en charge le choix des indexations en se basant essentiellement sur les techniques linguistiques pour les documents textuels. Il y a aussi des travaux qui portent sur l'indexation automatique des documents multimedia en se basant sur des études en reconnaissance de forme. Enfin il y a des techniques d'indexation mixtes qui pallient les manques en indexation automatique par l'intervention d'un indexeur.

La stratégie d'indexation ne portent pas uniquement sur le niveau de détail de la description des documents. La stratégie d'indexation concerne également ce que l'on peut appeler la «couverture sémantique» des descripteurs. Une stratégie consiste à ne choisir que les descripteurs utilisés dans les documents. Une autre stratégie consiste à choisir aussi bien les descripteurs utilisés dans les documents et aussi ceux que l'on peut déduire par des liens sémantiques à partir des termes utilisés dans les documents.

Alors que les concepts d'un document textuel sont facilement identifiables et comportent peu d'erreur d'interprétation par les indexeurs des documents, les concepts dans les documents multimedia, en particulier les images, dépendent beaucoup de l'interprétation de l'indexeur. Ce problème constitue en effet un axe de recherche comme le montre les travaux de recherche de Mouaddib [mou89]. Comme nous le présentons ci-dessous, certains facteurs, qui contribuent à l'évolution des SRI, ont des effets sur le processus d'indexation.

II.2 Evolution des SRI

Trois facteurs contribuent d'une manière significative à l'évolution des SRI. Ces facteurs sont :

- 1) les informations gérées ;
- 2) les utilisateurs ;
- 3) les progrès technologiques ;

II.2.1. Evolution des informations gérées

La nature des informations gérées par les SRI a beaucoup évolué. Les premiers SRI avaient pour objectif de gérer des informations textuelles (les documents eux-mêmes et leurs représentations). Désormais, les SRI doivent gérer des informations multimedia et hypermedia. Alors qu'un document hypermedia peut faire référence explicitement à un autre document, un document multimedia ne contient pas de lien explicite vers d'autres documents. Un document hypermedia peut en effet contenir un document multimedia. La «lecture» d'un document hypermedia est basée sur le concept d'hypertexte où la lecture du document se fait non pas linéairement mais par association d'idées. Nos travaux portent sur la gestion des documents multimedia. Dans le reste du rapport, nous employons le terme *information* pour désigner à la fois les informations sur les documents et les informations contenues dans les documents.

II.2.2. Evolution des utilisateurs

Les utilisateurs des SRI ont beaucoup évolué. Les premiers utilisateurs des SRI étaient des personnes compétentes pour formuler leurs besoins en information et qui maîtrisaient l'utilisation de ces systèmes. La prolifération des ordinateurs personnels, le développement d'internet et la disponibilité d'un plus grand nombre de sources d'informations (conséquence du progrès technologique) permettent à un public plus large d'avoir accès aux informations. On trouve donc de plus en plus d'utilisateurs qui sont à la fois moins compétents pour formuler leurs demandes et qui possèdent peu de compétence dans l'utilisation des SRI proposés pour la recherche des informations.

II.2.3. Evolution des progrès technologiques

Les progrès en électronique ont permis de développer des outils de stockage de haute capacité et de produire des ordinateurs de haute vitesse de calcul. Ces progrès permettent de réduire les coûts d'achat des ordinateurs performants et de rendre les ordinateurs accessibles au grand public. Le progrès dans la vitesse de calcul des ordinateurs et des outils de stockage de grande capacité permet de développer des systèmes à interface graphique, intuitive pour les utilisateurs, et de gérer des informations multimedia. Alors que les premiers systèmes étaient à base de commandes systèmes, les nouveaux systèmes sont à base de menus utilisant des graphiques, des images, voire même de la parole.

Quels que soient l'évolution des informations gérées par les SRI, l'évolution des utilisateurs et les progrès technologiques, les techniques de base en SRI reposent sur les méthodes de mise en correspondance de la requête de l'utilisateur et les représentations des documents dans le système. Deux de ces méthodes, qui sont parmi les plus utilisées actuellement dans les SRI, la méthode de recherche booléenne et la méthode de recherche vectorielle, sont présentées dans les sections suivantes.

II.3 Recherche booléenne

Considérons la base d'un SRI constituée de documents d_1, d_2, \dots, d_n . Chacun de ces documents est décrit par des termes t_1, t_2, \dots, t_m . On aura donc :

$$d_k(t_1, t_2, \dots, t_m) \quad \text{où terme } t_i \text{ décrit document } d_k$$

Considérons la liste inverse qui permet d'associer à chaque terme, l'ensemble des documents qu'il décrit. On aura

$$t_i (d_1, d_2, \dots, d_l)$$

Par exemple, dans le cas où les documents sont décrits par des mots clés et le fonds documentaire est constitué des documents suivants :

- d_1 (informatique documentaire, indexation, gestion de documents, thesaurus)
- d_2 (informatique documentaire, système d'information, informations multimedia)
- d_3 (gestion d'information, recherche d'information, indexation, informations multimedia, thesaurus)
- d_4 (système d'information, indexation, thesaurus, informations multimedia)

Les listes inverses sont :

t_1 : informatique documentaire (d_1, d_2)

t_2 : indexation (d_1, d_3, d_4)

t_3 : gestion de documents (d_1)

t_4 : système d'information (d_2, d_4)

t_5 : informations multimedia (d_2, d_3, d_4)

t_6 : gestion d'information (d_3)

t_7 : recherche d'information (d_3)

t_8 : thesaurus (d_1, d_3, d_4)

La mise en correspondance en recherche booléenne est basée sur l'algèbre de boole. On utilise les opérateurs booléens sur des ensembles de valeurs. Ici, les documents décrits par chaque mot clé constituent nos ensembles.

Prenons l'exemple de la requête utilisateur sous forme de la demande suivante :

«Je cherche des documents qui portent sur informations multimedia et thesaurus»

requête documentaire de la demande :

(mot-clé = informations multimedia) ET (mot-clé = thesaurus)

Cette requête est constituée de deux critères de recherche :

c_1 : (mot-clé = informations multimedia)

c_2 : (mot-clé = thesaurus)

liés par l'opérateur ET.

Chaque critère est formé

- du champ de description (*ici mot-clé*)
- d'un opérateur de comparaison (*ici =*)
- d'une valeur («*informations multimedia*» dans c_1 , «*thesaurus*» dans c_2)

Utilisant la liste inverse, nous devons trouver l'intersection de t_5 et t_8 à cause de l'opérateur booléen ET, ce qui correspond à d_3 .

La forme générale d'une requête booléenne peut être exprimée comme :

$$c_1 \text{ opb}_1 c_2 \text{ opb}_2 \dots c_n$$

où

$c_i = \text{champ opc valeur}$

$\text{opb}_i \in \{\text{ET, OU, SAUF, ...}\}$

$\text{opc} \in \{=, <=, >=, <>, \dots\}$

La notion de recherche par le contenu

Comme nous l'avons présenté dans la section précédente, les SRI ne gèrent pas directement les documents primaires. Ce sont les représentations des documents (textuels ou multimedia) qui sont gérées. Néanmoins, le résultat d'une recherche fournit des informations permettant d'accéder aux documents primaires. Si les documents primaires sont stockés dans l'ordinateur, leur accès peut être fait automatiquement. C'est ce qui se fait pour accéder aux images ou aux livres en ligne.

Cette technique de recherche booléenne permet d'effectuer une recherche en exprimant la requête par ce que l'utilisateur possède comme connaissance sur les documents recherchés. C'est ce que l'on appelle «*la recherche par le contenu*». Par exemple, si l'on connaît le titre d'un livre, on peut utiliser cette connaissance pour rechercher l'auteur du livre. Dans cet exemple, l'utilisateur veut compléter sa connaissance sur les propriétés des documents. Dans d'autres contextes, le désir de l'utilisateur peut être d'avoir accès aux documents primaires. Par exemple en connaissant l'auteur d'un livre ou d'un tableau de peinture, l'utilisateur peut désirer emprunter le livre ou accéder au tableau de peinture pour illustrer un concept dans un livre.

Pour faciliter la recherche par le contenu, certains SRI fournissent des listes d'index pour chaque propriété de description. Par exemple, un SRI peut fournir la liste des auteurs, la liste des mots clés, etc. Cette fonctionnalité est un grand pas vers l'assistance de l'utilisateur pour exprimer son besoin. Malheureusement, certains problèmes persistent, comme la transformation de

l'expression du besoin en langage naturel en une requête (ou langage formel). Par exemple, le besoin exprimé en langage naturel

«je cherche des photos prises par Monsieur David»

doit être transformé en

«*auteur = David*» dans certains systèmes, et

«*photographe = David*» dans d'autres.

Ni le mot «*auteur*» ni le mot «*photographe*» n'apparaît dans l'annoncé du besoin en langage naturel. Pourtant on demande à l'utilisateur d'effectuer la transformation. Par mes expériences en enseignement des étudiants en licence module documentation et en DESS information scientifique et technique, où d'entre eux n'ont jamais pratiqué la documentation, la transformation des besoins, même pour un exemple aussi simple que celui ci-dessus est une difficulté majeure.

Un deuxième problème qui persiste malgré les assistances proposées par les SRI actuels concerne le manque d'opérateur de comparaison et booléen pour exprimer la requête de certaines demandes en langage naturel. Voici deux exemples qui sont impossible à formuler dans certains SRI uniquement avec les opérateurs booléens et de comparaison classiques :

- «*je cherche les derniers livres de Mitterrand*»

- «*je cherche un bon livre sur la modélisation de l'utilisateur*»

Comment exprimer «*les derniers livres*» et «*un bon livre*» par les opérateurs habituels ? Même si l'on utilise un champ «*appréciation*» pour représenter les livres, l'affectation de la valeur «*bon*» dépend de l'indexeur et non de l'auteur ou de l'utilisateur. Malheureusement, la plupart des SRI n'intègrent même pas cet attribut de description.

Description d'un document par plusieurs champs (idée de facettes)

La représentation des documents par plusieurs champs permet de les présenter sous des angles différents, par le regroupement de ces champs. Un document peut être ainsi présenté par un sous-ensemble de champs qui le décrivent. Ces différentes façons de représenter les documents sont appelées «*facettes*» [hal89, aig91]. L'idée de facette permet de délimiter les parties d'information sur les documents qui sont d'intérêt pour l'utilisateur. Par exemple, pour des documents multimedia, et plus précisément des photos, on peut classer les informations en trois catégories :

- la catégorie *production* (auteur, date, éditeur)
- la catégorie *morphologie* (prise de vue, couleur)
- la catégorie *dénotation* (les objets qui figurent sur les photos)
- la catégorie *connotation* (significations attribuées aux objets qui figurent sur des photos)

Pour un document textuel, on peut avoir trois catégories :

- la catégorie *identité* (titre, auteur, éditeur, date édition)
- la catégorie *thème*
- la catégorie *mots-clés*

Certes, les facettes peuvent permettre aux utilisateurs de mieux cibler leurs demandes mais dans certains cas, elles peuvent aggraver leurs problèmes à formuler leurs demandes. Prenons le cas d'un utilisateur qui cherche des documents sur «voiture». Dans le cas des documents multimedia, on peut chercher sur les objets qui figurent sur les documents. Mais dans le cas des documents textuels avec les facettes de notre exemple, l'utilisateur aura du mal à décider si il faut chercher dans les mots-clés ou dans les thèmes. Le problème ici est que par les deux catégories «thème» et «mots-clés» (pour les documents textuels) ou «dénotation» et «connotation» (pour les documents multimedia) on essaie de représenter une hiérarchie de concepts. Par exemple les mots-clés seront utilisés pour représenter les concepts moins abstraits et les thèmes les concepts abstraits des mots-clés.

Remarquons que l'implémentation des facettes impose à l'utilisateur de «structurer» sa connaissance dans le domaine de l'information. Malheureusement la **capacité** de l'utilisateur à structurer sa connaissance n'est pas prise en compte dans les méthodes de mise en correspondance. Même si la capacité à structurer les connaissances relève des études en sciences cognitive, je pense qu'il serait nécessaire de la prendre en compte dans le processus de mise en correspondance.

Dans le cadre de nos études sur la modélisation de l'utilisateur, nous proposons d'intégrer les connaissances sur l'utilisateur concernant sa capacité à structurer ses propres connaissances dans le domaine de l'information. Nous proposons par exemple d'observer l'évolution des propriétés utilisées par l'utilisateur pour formuler ses requêtes.

Une égalité empêche une correspondance nuancée

Un reproche que l'on fait à la méthode de recherche booléenne est l'emploi de l'égalité stricte des termes des requêtes et les termes des documents. Il n'est pas possible par exemple de nuancer le degré d'appartenance d'un terme de la requête à un document. Pour résoudre ce problème, plusieurs solutions sont proposées dont l'intégration de thesaurus, l'affectation de poids aux descripteurs des documents et des requêtes ainsi que l'intégration de fonctions de calcul de nuance des termes de la requête et des documents.

Le thesaurus permet d'hiérarchiser les concepts par leurs niveaux d'abstraction. Des termes peuvent être substitués à d'autres qui sont plus abstraits ou qui sont moins abstraits. Dans certains systèmes, l'utilisateur peut décider du niveau d'abstraction pour les substitutions. Par exemple dans le logiciel documentaire Polybase, l'utilisateur peut demander explicitement la substitution par les termes génériques, les termes spécifiques ou par les termes sémantiquement proches en utilisant le lien sémantique «voir aussi».

La technique qui consiste à affecter des poids aux termes qui sont utilisés dans les requêtes et dans les représentations des documents constitue la technique de base en recherche vectorielle. Cette technique est développée dans la section suivante.

Les fonctions de calcul de nuance sont définies par les concepteurs des systèmes. Ces fonctions portent sur les termes des documents et des termes des requêtes. Cette technique est principalement utilisée dans les bases de données nuancées et adaptées aux SRI. Dans certains systèmes, les fonctions de calcul de nuances disposent de paramètres permettant à l'utilisateur de modifier le comportement des fonctions.

Dans les trois approches de solution qui permettent de nuancer le degré d'appartenance d'un terme à un document, on peut remarquer que l'utilisateur est sollicité par sa capacité à préciser sa demande. L'affectation d'un poids à un terme de la requête, permet de préciser le degré de nuance et préciser le niveau d'abstraction d'un terme sont des éléments de cette sollicitation. Nous voyons encore par ces approches la nécessité d'avoir une représentation explicite et individuelle des utilisateurs pour mieux adapter la réponse du système à son niveau de connaissance.

II.4 Recherche vectorielle

L'un des reproches formulé à l'égard de la recherche booléenne est qu'il n'y a pas de différenciation entre tous les documents proposés en réponse à une requête. Par exemple, prenons un ensemble de documents :

d_1 (*thesaurus, recherche d'information*)

d_2 (*informations multimedia, recherche d'information, modèle booléen, thesaurus*)

d_3 (*thesaurus, informatique*)

Et prenons la requête suivante :

(*thesaurus OU recherche d'information OU informatique*)

d_1 , d_2 et d_3 satisfont la requête mais la recherche booléenne n'indique pas que d_1 et d_3 représentent mieux à la requête que d_2 .

Le principe du modèle vectoriel est que les documents et les requêtes sont des vecteurs dans un espace à n dimensions, où à un terme descripteur correspond une dimension de l'espace et l'ensemble T des termes d'indexation constitue la base de l'espace. Le modèle permet de calculer une mesure de similarité entre deux vecteurs, par exemple entre deux vecteurs de requête, entre deux vecteurs de document, et entre un vecteur de documents et un vecteur de requête. Les scores qui résultent des calculs des mesures de similarité entre une requête et les documents permettent de classer les documents trouvés.

Dans ce modèle, le document j peut être représenté comme

$$D_j = w_{j1}t_1, w_{j2}t_2, \dots, w_{jn}t_n$$

où w_{ji} représente le poids de pertinence affecté au terme t_i du document D_j ;

et une requête peut être représentée comme

$$Q = a_1t_1, a_2t_2, \dots, a_nt_n$$

où a_i représente le degré d'importance du terme t_i dans la requête.

La mesure de similarité est calculée avec ces coefficients de degré d'appartenance. L'une des mesures de similarité les plus utilisées est la mesure de similarité cosinus :

$$\cos(D_i, Q) = \frac{\sum_{j=1}^n w_{ij} \cdot a_j}{\sqrt{\left(\sum_{j=1}^n a_j^2\right) \left(\sum_{j=1}^n w_{ij}^2\right)}}$$

où

$\cos(D_i, Q)$ représente la mesure de similarité entre le $i^{\text{ème}}$ document D_i et la requête Q ;

a_i représente le degré d'importance du terme i dans la requête

w_{ij} représente le poids de pertinence du terme j dans le document i

Si nous reprenons l'exemple de documents ci-dessus en affectant des poids de pertinence au documents, 1 représentant la présence du terme et 0 l'absence d'un terme. Nous ne représentons pas les termes qui sont absents dans un document.

d_1 (1:thesaurus, 1:recherche d'information)

d_2 (1:informations multimedia, 1:recherche d'information, 1:modele booléen, 1:thesaurus)

d_3 (1:thesaurus, 1:informatique)

De la même manière, nous pouvons représenter la requête comme ci-dessous avec les degrés d'importance correspondant à chaque terme dans la requête :

(1:thesaurus, 1:recherche d'information, 1:informatique)

Ici, tous les terme ont la même importance 1.

Les scores résultant du calcul de la mesure de similarité pour chaque document est présenté dans le *Tableau 1 : (Score issu de la mesure de similarité)* avec un tri par les scores.

Document	Score
D1	0,82
D3	0,82
D2	0,58

Tableau 1 : Score issu de la mesure de similarité

L'un des avantages de cette mesure est que le résultat fourni par le système peut être trié par ordre de similarité des requêtes et les documents. Malheureusement, cette méthode de recherche demande l'attribution des valeurs de degré d'appartenance aux termes des documents pendant l'indexation et dans les requêtes. Les valeurs sont attribuées par les indexeurs pour décrire les documents et par les utilisateurs pour décrire les requêtes. Dans un cas le plus simple, on attribue la valeur 1 ou 0 aux termes, la valeur 1 signifiant l'appartenance du terme au document ou à la requête et la valeur 0 la non appartenance du terme au document ou à la requête.

Nous pouvons remarquer que l'attribution d'une valeur de degré d'importance d'un terme à une requête par l'utilisateur peut ne pas être facile pour lui. Déjà l'indication des termes de la requête n'est pas évidente comme nous l'avons montré dans l'une des sections précédentes. On sollicite ici la capacité de l'utilisateur à exprimer et à hiérarchiser sa connaissance dans le domaine de l'information.

Nous retenons néanmoins cette méthode de recherche comme un moyen pour comparer les différentes requêtes de l'utilisateur afin de suivre l'évolution de ses requêtes.

II.5 Conclusion et propositions pour la recherche d'information

Nous pouvons remarquer que toutes les méthodes de recherche proposées dans les SRI dépourvus de modélisation de l'utilisateur reposent énormément sur la capacité de l'utilisateur à exprimer ses besoins et en particulier à hiérarchiser ses connaissances. Néanmoins, les techniques de mise en correspondance constituent pour nous des outils pour analyser les connaissances que nous comptons acquérir sur les utilisateurs.

Dans les SRI actuels, des aides sont proposées aux utilisateurs pour leur permettre de mieux formuler leurs requêtes. Par exemple, la liste des index lui permet de prendre connaissance des termes utilisés pour indexer les documents (indexation des titres, du résumé, etc) et l'utilisation d'un thesaurus lui permet de hiérarchiser ses connaissances et d'étendre ou restreindre le champ sémantique de sa recherche.

Nous proposons qu'un SRI dispose, en plus des méthodes classiques de recherche d'information, d'une méthode d'analyse globale des informations du système. Cette analyse globale portera à la fois sur les informations du domaine et sur les connaissances acquises sur les utilisateurs individuellement. Nous appelons cette technique «*analyse-croisée avec contraintes*». Nous présentons d'abord les contraintes (la recherche classique) et ensuite l'analyse des données dans les sections suivantes.

II.5.1. Recherche classique : les contraintes

Dans l'approche mixte que nous adoptons, les contraintes correspondent aux critères dans les requêtes classiques. Il s'agit en effet de la recherche booléenne mais limitée à l'opérateur booléen OU. Limiter ainsi l'opérateur booléen à OU nous permet de trier le résultat d'une recherche selon une mesure de similarité. Par exemple la mesure de cosinus peut être utilisée où les valeurs de degré d'appartenance des termes aux documents et aux requêtes sont limitées à 0 et 1.

Les contraintes peuvent être exprimées comme :

$$c_1 \text{ } opb_1 \text{ } c_2 \text{ } opb_2 \text{ } \dots \text{ } c_n$$

où

$$c_i = \text{champ } opc \text{ valeur}$$

$$opb \in \{ET, OU, SAUF, \dots\}$$

$$opc \in \{=, <=, >=, <>, \dots\}$$

Par exemple dans le prototype STREEMS où nous gérons des informations sur les arbres,

$$\langle \text{maximum_altitude} \geq 1000 \rangle \text{ OU } \langle \text{minimum hauteur} \leq 15 \rangle$$

fournira les arbres qui peuvent être plantés à une altitude maximum d'au moins 1000 mètres et qui ont une hauteur d'au plus 15 mètres. Malheureusement ce type de requête ne donne pas la distribution des arbres (gérés par le système) sur les altitudes. C'est l'une des raisons pour laquelle nous avons introduit la propriété d'analyse de données dans la formulation de requête. La formule de mesure de similarité est celle de la mesure de cosinus (cf section II.4 (*Recherche vectorielle*)).

II.5.2. Analyse de données

Rappelons que dans les SRI, les documents (entités d'informations) sont décrits par des champs. Par exemple un arbre peut être décrit par son *nom scientifique*, son *nom commun*, sa *hauteur*, l'*altitude maximum* où l'on peut le planter, les *produits de bois* ; une publication peut être décrite par le *titre*, l'*auteur*, l'*éditeur*, la *date de publication* et les *mots-clés*.

Nous avons adopté deux types d'analyses pour analyser les informations gérées par le système et les connaissances sur les utilisateurs. Dans le premier type d'analyse, les valeurs d'un seul champ sont analysées. Dans ce cas, il y a deux possibilités :

- la distribution de chaque valeur possible pour le champ
- la co-occurrence de deux (ou plus) valeurs dans le même champ.

Le premier cas correspond à ce que nous appelons «*analyse de fréquence*» et le second «*analyse-croisée intra-champ*».

Dans le second type d'analyse, deux (ou plus) champs sont utilisés pour l'analyse où la co-occurrence des paires de valeurs des deux champs est calculée. C'est cela que nous appelons «*analyse-croisée inter-champ*».

La technique d'analyse-croisée est largement utilisée dans les systèmes d'Information Scientifique et Technique (IST), en infométrie et en fouille de données (data mining) [jak95, cou90, dou95]. L'objectif dans ces domaines est d'obtenir de l'information à valeur-ajoutée pour la prise de décision, fournissant une vue globale des caractéristiques d'une base d'informations. Nous pensons que cette technique d'analyse-croisée permettra à l'utilisateur de mieux exprimer son besoin et surtout de lui donner une meilleure connaissance des informations disponibles dans le système.

Dans nos implantations actuelles de cette méthode, trois champs peuvent être utilisés pour les analyses intra-champ ou inter-champ.

a) Analyse de fréquence

L'analyse de fréquence donne une vue globale de la distribution des valeurs d'un champ, le champ étant un attribut des documents de la base. Nous représentons cette analyse comme

$$f_j = \sum_{i=1}^n |D_i \cap v_j|$$

où

$$|D_i \cap v_j| \in \{0, 1\}$$

D_i = ensemble de termes utilisés pour décrire le document i

V = ensemble de valeurs du champ d'analyse

$$v_j \in V$$

n = le nombre de documents dans le système

Nous représentons la même analyse sous forme matricielle comme

<i>valeurs du champ</i>	v_1	$v_2 \dots v_m$
	f_1	$f_2 \dots f_m$

m = le nombre de valeurs utilisées dans le champ analysé

v_i = une valeur du champ d'analyse

f_i = la fréquence de cette valeur dans la base

Par exemple, on peut obtenir la fréquence des «produits de bois» dans notre prototype STREEMS utilisé pour gérer les informations sur les arbres autorisés par l'union européenne pour le reboisement des forêts

<i>Produit de bois</i>	<i>planche</i>	<i>charbon</i>	<i>bouchon</i>
	100	40	2

Cette analyse sera très utile pour la prise de décision sur les types d'applications possibles pour des arbres autorisés pour le reboisement par l'union européenne. Tous les champs de la base ne sont pas forcément utiles pour l'analyse de fréquence. Par exemple, l'analyse de fréquence sur le champ «*nom scientifique*» ne donnera rien comme information car les arbres ont des noms scientifiques uniques. L'interprétation de cette information à valeur-ajoutée dépendra de l'utilisateur. A noter aussi qu'il faut la contribution des experts dans le domaine de l'information pour déterminer les champs qui ont de l'intérêt pour ce type d'analyse.

Cette technique d'analyse peut être appliquée en IST pour une analyse globale des informations disponibles dans une source particulière d'informations. Par exemple, si une base documentaire spécialisée en pharmacologie dispose d'un champ «techniques de fabrication» des produits, ce type d'analyse donnera l'étendue d'utilisation de chaque technique de fabrication. Une entreprise en quête d'innovation pourra s'intéresser aux techniques les moins utilisées pour étudier leurs limites et les raisons de leur non utilisation. Par contre une entreprise qui cherche à augmenter son chiffre d'affaire s'intéresserait aux techniques les plus utilisées pour garantir la confiance de ses produits. Il faut remarquer que l'interprétation à donner au résultat de l'analyse dépend des experts ainsi que l'objectif fixé par ces experts.

b) Analyse-croisée de données intra-champ

L'analyse de données intra-champ est largement utilisée dans les études de bibliométrie et de scientométrie. L'objectif principal est d'étudier la distribution de co-occurrence de termes

dans un contexte spécifique pour un seul champ (ou critère). Nous représentons l'analyse intra-champ par

$$f_{kj} = \sum_{i=1}^n |D_i \cap \{v_k, v_j\}|$$

$$|D_i \cap \{v_k, v_j\}| \in \{0, 1\}$$

D_i = ensemble de termes utilisés pour décrire le document i

V = ensemble de valeurs du champ d'analyse

$$v_k, v_j \in V$$

n = le nombre de documents dans le système

Nous représentons l'analyse sous forme matricielle par

Valeurs du champ	v_1	v_2	\dots	v_m
Valeurs du champ				
v_1	f_{11}	f_{12}	\dots	f_{1m}
v_2	f_{21}	f_{22}	\dots	f_{2m}
\dots	\dots	\dots	\dots	\dots
v_m	f_{m1}	f_{m2}	\dots	f_{mm}

m = le nombre de valeurs utilisées dans le champ analysé

v_i = une valeur du champ d'analyse

f_{ij} = la fréquence de co-occurrence de deux valeurs v_i et v_j dans la base

Par exemple, nous pouvons considérer le champ «maladie» dans le prototype STREEMS

<i>maladie</i>	<i>phytophthora_cambivora</i>	<i>nectria_ditissima</i>
<i>maladie</i>		
<i>nectria_ditissima</i>		3
<i>ceratocystis_ulmi</i>	1	4

La co-occurrence de termes dans la base d'informations du champ «maladies» peut montrer les liens entre deux types de maladies. Une forte corrélation de deux maladies peut aider un utilisateur à orienter sa recherche, en particulier pour ceux qui ne connaissent pas exactement ce qu'il cherche.

Reprenant l'exemple d'une entreprise pharmaceutique de la *section a* ci-dessus, ce type d'analyse révélera l'utilisation simultanée de deux techniques de fabrication. Une vision globale des «co-utilisations» des techniques de fabrication peut révéler l'importance de certaines associations.

c) L'analyse-croisée inter-champ

L'analyse-croisée inter-champ est aussi largement utilisée dans les études de la bibliométrie et de la scientométrie. Si les champs qui décrivent les documents sont considérés comme des facettes du document, l'analyse de fréquence et l'analyse intra-champ peuvent être considérées comme une analyse mono-critère. Dans une analyse inter-champ, nous sommes intéressés par de multiples critères sur les documents. Nous représentons l'analyse-croisée inter-champ comme

$$f_{kj} = \sum_{i=1}^n |D_i \cap \{v_k, t_j\}|$$

$$|D_i \cap \{v_k, v_j\}| \in \{0, 1\}$$

D_i = ensemble de termes utilisés pour décrire le document i

V = ensemble de valeurs du premier champ d'analyse

V' = ensemble de valeurs du deuxième champ d'analyse

$$v_k \in V$$

$$t_j \in V'$$

n = le nombre de documents du système

Nous représentons l'analyse sous forme matricielle comme

<i>Champ1</i>	v_1	v_2	...	v_m
<i>Champ2</i>				
t_1	f_{11}	f_{12}	...	f_{1m}
t_2	f_{21}	f_{22}	...	f_{2m}
...
t_k	f_{k1}	f_{k2}	...	f_{km}

m = le nombre de valeurs utilisées dans le premier champ d'analyse

k = le nombre de valeurs utilisées dans le deuxième champ d'analyse

f_{ij} = la fréquence de co-occurrence de les valeurs v_i et t_j pour le champs d'analyse champ1 et champ 2 respectivement

v_i = une valeur du champ 1 d'analyse

t_i = une valeur du champ 2 d'analyse

Par exemple, nous pouvons prendre les champs «maladie» et «altitude» dans le prototype STREEMS

<i>maladie</i> <i>altitude</i>	<i>phytophthora</i>	<i>nectria_ditissima</i>	<i>ceratocystis_ulmi</i>
100		5	
500	2		
1000	1	1	8

La co-occurrence des termes dans la base peut nous montrer la distribution des maladies sur les altitudes. Cette information à valeur-ajoutée peut aider un agriculteur à planifier la façon de développer une parcelle particulière dans une altitude particulière. Comme pour le cas de maladies en analyse-croisée intra-champ, l'interprétation de cette information à valeur-ajoutée revient à l'utilisateur.

Reprenant l'exemple d'un entreprise pharmaceutique, s'il y a les champs «technique de fabrication» et «type de produit», ce type d'analyse révélera par exemple, la distribution de l'utilisation des différentes techniques de fabrication sur les types de produits. On connaîtra donc les techniques les plus utilisées pour un certain type de produit, ce qui peut faciliter par exemple le choix d'un type de technique de fabrication pour un produit particulier.

En résumé, nous avons montré la nécessité de prendre en compte les spécificités de chaque utilisateur dans le traitement de sa requête. Nous avons introduit la notion de classification avec contraintes dans le processus de recherche d'information, la proposition qui consiste en une recherche classique (les contraintes), l'analyse intra-champ et l'analyse inter-champ. L'application de cette technique à trois champs permet non seulement d'observer la fréquence des occurrences des valeurs mais elle permet également d'observer l'évolution de ces analyses. Cela est rendu possible en indiquant un champ «date» comme le deuxième ou le troisième champ d'analyse.

Remarquons que cette approche est indépendante du domaine d'application. En effet, dans nos réalisations, l'interface de cette méthode de recherche est dissociée des différentes applications. Cette modularité permet d'employer la même interface dans n'importe quel domaine d'application. Par exemple, l'interface est déjà utilisée dans les deux systèmes que nous avons réalisés : METIORE et STREEMS. Ce qui change d'une application à l'autre est la liste de champs utilisable pour les analyses. Le système METIORE est présenté dans le *Chapitre V (Le système METIORE)* et le système STREEMS dans *Annexe I (Le prototype STREEMS)*.

Cette méthode de recherche et d'analyse de l'information sera utilisée pour analyser le comportement de l'utilisateur ainsi que l'évolution de ses connaissances dans le domaine d'application. Comme nous le montrerons dans le *Chapitre VI (Le modèle de l'utilisateur dans un environnement de recherche coopérative d'information)*, les utilisateurs sont représentés explicitement et leurs activités seront représentées sous forme de documents.

Notre première expérience dans la modélisation de l'utilisateur est présentée dans le chapitre suivant. Il s'agit de la modélisation de l'apprenant. En effet, l'apprenant est un utilisateur particulier et le l'objectif d'un système d'apprentissage n'est pas le forcément même que l'objectif d'un SRI. La dérivation de notre modèle de l'utilisateur à partir du modèle de l'apprenant est détaillée également dans le *Chapitre VI (Le modèle de l'utilisateur dans un environnement de recherche coopérative d'information)*.

Chapitre III

La modélisation dynamique de l'apprenant : le projet BIRDS

Nous présentons dans ce chapitre notre première expérience dans le domaine de la modélisation de l'utilisateur. Il s'agit de la modélisation de l'apprenant. Nous présentons dans la première section des raisons qui poussent à modéliser l'utilisateur et le modèle de l'apprenant (ou modèle de l'élève) est présenté dans la deuxième section.

III.1 Pourquoi la modélisation de l'utilisateur

Pour illustrer l'objectif que nous voulons atteindre dans les SRI, nous allons prendre des activités d'une agence de tourisme comme exemple. De ces activités, l'agence dispose des destinations de vacances et les propose aux clients. Voici des situations possibles dans cette agence :

- *Un client voyage pour la première fois et donc l'hôtesse de l'agence ne le connaît pas.* D'une manière générale, on donne toutes les informations disponibles au client en indiquant les préférences de la majorité des clients pour chaque type d'information.
- *Un client a déjà voyagé plusieurs fois mais c'est la première fois qu'il utilise l'agence.* Dans cette situation, le client présente à l'hôtesse un certain nombre de ses expériences au cours des voyages précédents. L'hôtesse prend en compte ces expériences pour orienter ses propositions au client.
- *Un client a déjà voyagé plusieurs fois en utilisant la même agence mais c'est la première fois que l'hôtesse s'occupe de lui.* L'hôtesse peut adopter deux attitudes :
 - La première attitude consiste à demander au client de présenter certaines de ses expériences avec l'agence et de ses voyages. Dans ce cas, l'hôtesse intégrera les expériences du clients dans ses propositions.
 - La deuxième attitude consiste à ignorer les expériences du clients. Dans ce cas, l'hôtesse proposera toutes les solutions possibles comme pour le premier type de client ci-dessus.
- *Un client a déjà voyagé plusieurs fois en utilisant la même agence et c'est la même hôtesse qui s'occupe de lui.* On suppose que l'hôtesse connaît bien le client. L'hôtesse sait donc quelles sont les destinations où le client a déjà passé des vacances, les problèmes rencontrés ainsi que les satisfactions obtenues par le client lors de ses séjours dans les destinations. En bref, l'hôtesse possède des connaissances spécifiques sur le client en question.

Si nous devons mettre en place un SRI pour fournir les informations aux clients de l'agence, il est indispensable d'avoir une représentation explicite et individuelle de ces clients. Des informations exactes ne suffisent pas. Une information exacte ne correspond pas forcément à la préférence du client. Et même la préférence d'un client peut être ce qu'un autre client déteste.

On peut prendre comme autre exemple, un centre de documentation d'un centre de recherche où les utilisateurs sont essentiellement des enseignants chercheurs permanents. Pour avancer dans sa recherche, le chercheur a besoin de connaître ce qui se fait ailleurs dans son domaine de recherche. Prenons le cas de monsieur Dupont qui fait sa recherche sur la modélisation de l'utilisateur. S'il demande au documentaliste des références sur le terme «modélisation de l'utilisateur», le documentaliste lui donnera *«toutes»* les références sur ce terme. Supposons que le chercheur revienne la semaine suivante et qu'il pose la même question. Donner le même type de réponses que la semaine précédente serait ridicule en particulier s'il s'adresse à un même documentaliste. Il ne serait pas utile de fournir toutes les réponses - les réponses disponibles la semaine dernière et les nouvelles depuis la semaine dernière - sauf si l'utilisateur demande **toutes** les réponses explicitement. Si le chercheur s'adresse au même documentaliste, le documentaliste devrait se rappeler que le chercheur en question a déjà formulé ce type de demande la semaine précédente. Il devrait logiquement lui donner uniquement les nouvelles références depuis la semaine précédente. Cela signifie que le documentaliste garde dans son souvenir les informations spécifiques au chercheur en question, c'est-à-dire :

- qu'il est déjà venu,
- à quelle date,
- qu'elle a été sa demande,
- ce qu'il a obtenu comme réponse.

Malheureusement, ce type de personnalisation n'existe pas dans les centres de documentation, ceci pour plusieurs raisons :

- le nombre d'utilisateurs est très élevé par rapport au nombre de documentalistes. Il n'est donc pas possible que le petit nombre de documentalistes se souvienne de tous les utilisateurs et de leurs demandes ;
- les préférences des utilisateurs sont diverses et variées. Par exemple, pour la première fois, le chercheur peut désirer uniquement les références publiées dans des livres, pour la deuxième fois cela peut être un intérêt pour les revues et encore pour une autre fois des références dans les congrès. Le chercheur peut également spécifier ses préférences par rapport aux attributs des documents. Par exemple avoir une préférence pour les pays où sont publiés les documents ou bien dans les organismes de recherche particuliers. On peut donc imaginer les difficultés qu'auraient les documentalistes à gérer les spécificités de chaque utilisateur.

Dans la plupart des SRI où un modèle de l'utilisateur est représenté, les propositions pour personnaliser les réponses peuvent être résumées par la phrase suivante : «*donner les réponses acceptées comme pertinentes par la plupart des utilisateurs pour les mêmes types de requête*».

Si $A_i =$ *Le nombre de fois que la solution i est acceptée pour une requête*
 $S_i =$ *Le nombre de fois que la solution i est proposée pour la même requête*
 $R_i =$ *Le nombre de fois que la solution i est rejetée pour la même requête*

Alors

$$(A_i + R_i) \leq S_i$$

Le degré d'acceptabilité Q_i de la solution i peut être représenté comme

$$Q_i = \frac{A_i}{S_i}, \quad 0 \leq Q_i \leq 1$$

On suppose que la pertinence de la solution i est directement proportionnelle à Q_i .

Malheureusement cette technique n'est pas applicable pour tous les utilisateurs. En effet la pertinence d'une réponse dépend d'une manière complexe de plusieurs facteurs. L'un des facteurs concerne la fréquence d'utilisation d'un SRI par les utilisateurs. On peut en effet identifier deux catégories d'utilisateurs par rapport à la fréquence d'utilisation d'un SRI : *les utilisateurs occasionnels* et *les utilisateurs réguliers*.

Pour les utilisateurs occasionnels, le système est utilisé pour un nombre de fois réduit, pour des besoins différents et bien espacés dans le temps. La technique de mesure de pertinence d'une solution ci-dessus est acceptable pour ce type d'utilisateurs parce que l'on ne possède pas assez de connaissances sur chacun d'eux pour déduire les préférences individuelles.

Pour les utilisateurs réguliers, le système est utilisé plusieurs fois, assez rapprochées dans le temps et très souvent pour des besoins similaires. Dans ce cas, la technique ci-dessus n'est plus satisfaisante car les particularités de chaque individu doivent être prises en compte.

III.2 Le modèle de l'apprenant : un modèle utilisateur particulier

L'objectif que nous souhaitons atteindre pour un SRI porte principalement sur la personnalisation des réponses du système à chaque utilisateur. J'ai déjà travaillé sur un même type de problématique dans le domaine de l'enseignement assisté par ordinateur, l'objet de ma thèse de doctorat de l'INPL[dav90a]. Il s'agissait de proposer des scénarios différenciés aux élèves en fonction de leurs particularités. Pour cela, nous avons défini un modèle de l'élève [dav91]. Nous

présentons dans cette section le modèle de l'élève et son expérimentation dans un prototype appelé BIRDS.

Notre hypothèse de travail était qu'un environnement d'apprentissage, fondé sur les informations multimedia, en particulier l'image, et le processus cognitif de l'élève, dans lequel est implanté un modèle de l'élève peut être motivant et capable d'entraîner l'élève dans un véritable processus d'apprentissage. Nous faisons une différence entre «enseignement» qui est la transmission de connaissances par un maître et «apprentissage» qui est un *acte volontaire et conscient* par un apprenant pour entreprendre une recherche de connaissance ou une culture de capacités. Notre propos repose sur «l'apprentissage» au sens précédent et, parmi les différentes théories de l'apprentissage qui ont été proposées, nous en avons retenu deux : comportementaliste et cogniviste. Notre approche étant de favoriser l'approche cognitive, nous avons fait un certain nombre de choix : en particulier celui de baser les activités (ou processus) d'apprentissage sur l'utilisation d'images et d'intégrer un modèle de l'élève fondé sur six paramètres détaillés dans la section III.2 (*Le modèle de l'élève : un modèle utilisateur particulier*). Afin de mettre en oeuvre nos propositions, en particulier le modèle de l'élève, nous avons construit le prototype BIRDS. Ce prototype nous a permis de réaliser une expérimentation au lycée H. Poincaré à Nancy, en classe de 5e et 6e. Nous avons validé le bien fondé de notre hypothèse par la comparaison de la représentation des élèves par le système, les adaptations proposées par le système aux élèves en fonction de leurs représentations avec un questionnaire que ceux-ci ont rempli. Nous avons remarqué, par exemple, une cohérence entre la déduction du système sur les comportements des élèves et ce que les élèves ont répondu au questionnaire. Par exemple, quand le système propose aux élèves d'appliquer les connaissances qu'ils ont acquises après avoir remarqué que ces élèves passent la plupart de leurs temps à observer les oiseaux ou à se familiariser avec le vocabulaire du domaine, certains élèves refusent ces propositions. Dans le questionnaire, nous avons demandé ce que les élèves ont «détesté» le plus dans le système. Nous avons été surpris que les élèves qui ont refusé les propositions du système (ce que nous avons pu observer par la représentation de chaque élève par le système) ont indiqué dans le questionnaire que c'est «d'être guidé par le système» qu'ils ont «détesté» le plus.

Nous présentons dans la section suivante deux concepts de base d'un processus d'acquisition de connaissance : le concept d'enseignement et le concept d'apprentissage.

III.2.1. Enseignement, apprentissage : quelles différences?

Examinons des différences fondamentales entre les termes «enseigner» et «apprendre». En enseignement, presque toutes les initiatives viennent de l'enseignant. Par contre, les initiatives viennent, d'une manière générale, de l'élève pendant le processus d'apprentissage. Pour le terme «enseigner», on suppose qu'il faut davantage de connaissance chez l'élève. Ce jugement vient souvent de l'extérieur, c'est-à-dire que l'élève ne se rend pas compte de son manque de connaissance ou bien il ne voit pas la nécessité de les acquérir et, même s'il en reconnaît la nécessité, il ne voit pas ce qu'il faut acquérir comme connaissances. Ce contexte met l'élève dans une situation d'obligation, car qu'il voit ou non la nécessité d'acquérir ces connaissances, il lui faudra se soumettre à la démarche préconisée.

Contrairement au terme «enseigner», «apprendre» suppose qu'il y ait intérêt de la part de l'élève, car c'est lui qui cherche à acquérir des connaissances. Cela met l'élève non pas dans une situation d'obligation mais dans une situation de désir et de responsabilité.

Parmi les théories d'apprentissage, celles qui nous intéressent sont la théorie du comportementaliste et celle du cognitivisme car ces théories abordent le problème d'apprentissage par rapport à la représentation mentale de connaissances par l'apprenant et elles abordent également le rôle de la motivation de l'apprenant dans le processus d'acquisition de connaissance. Nous développons uniquement l'approche cognitive, base de nos réflexions sur la modélisation de l'utilisateur.

III.2.2. Approche cognitive en apprentissage

Nous résumons, dans ce paragraphe, les théories de Decroly, Piaget et de La Garanderie qui ont travaillé sur l'approche cognitive dans le processus d'apprentissage.

Pour Piaget, l'apprentissage est en fait un développement cognitif qui s'inscrit dans le temps[her85]. Le terme «développement cognitif» signifie, dans ce contexte, les processus par lesquels un être vivant acquiert des informations sur son environnement. Selon lui, l'enfant expérimente le monde et acquiert, dans le temps, des compétences intellectuelles complexes. Nous pouvons dire que, selon Piaget, les phases du développement cognitif sont :

- l'observation
- l'acquisition du langage
- la maîtrise de la logique formelle et la capacité de vérifier des hypothèses
- la capacité de développer des idées abstraites.

Selon de La Garanderie, la pédagogie est fondamentalement la science des processus quasi spontanément employés par les individus humains pour retenir, pour comprendre, pour composer, pour chercher ou pour trouver [gar88]. Il défend la thèse selon laquelle la connaissance est représentée sous forme d'images mentales (image visuelle ou image auditive), qui deviennent une habitude évocative. A partir de là, il propose une classification des différentes formes d'images dans les habitudes évocatives. Nous donnons ici la classification des formes d'images visuelles qui a été à la base de notre travail :

- première classe : les images des objets perçus réellement dans le monde visible : choses, êtres, scènes de la vie ;
- deuxième classe : les images des mots eux-mêmes ;
- troisième classe : les images des figures, des croquis, des symboles etc ;
- quatrième classe : l'élaboration d'images visuelles mentales pour se représenter ce qui

n'a pas été perçu ou pour symboliser des objets perçus.

Pour Decroly, l'intérêt est le fondement unique de l'éducation : la condition essentielle pour que l'instruction soit féconde, c'est qu'elle excite l'intérêt - ce qui excite l'appétit de l'esprit [fod86]. Il distingue deux sources fondamentales d'intérêt : d'abord, le sentiment d'attention curieuse qui provoque l'expérience, l'étude de la nature, la recherche des connaissances ; ensuite, l'intérêt qui résulte de la vie sociale, de la fréquentation et du commerce des êtres humains. Les formes de l'intérêt qui se rattachent à la connaissance comprennent trois phases :

- l'intérêt empirique : il naît dans la perception des choses sensibles, de la variété des objets concrets que la nature ou l'enseignement présente,
- l'intérêt spéculatif : il est dérivé de la méditation prolongée ou des objets de l'expérience, du besoin d'explication, de la recherche de liaisons causales (l'intérêt abstrait selon lui), tout en étant basé sur des symboles et en tenant compte des expériences faites,
- l'intérêt esthétique : il est dérivé de la contemplation de la beauté dans la nature, dans les œuvres d'art ou dans les actions morales.

Il y a des points communs dans les théories des trois psychologues présentées ci-dessus. Les trois psychologues distinguent différentes phases dans le processus d'apprentissage et ces différentes phases se succèdent dans le temps de la même manière. Les quatre phases que nous pouvons mettre en évidence sont les suivantes :

- l'observation ou la perception des choses réelles (P1),
- un niveau élémentaire d'abstraction pour désigner les objets observés (ce que nous pouvons appeler l'apprentissage du vocabulaire - un langage élémentaire) (P2),
- une symbolisation et abstraction (qui donne lieu à l'emploi d'un langage plus spécialisé et fournit des objets de raisonnement) (P3),
- la créativité (qui est une extension d'idée souvent basée sur le résultat des trois phases précédentes) (P4).

Le modèle de l'élève que nous proposons prend en compte l'évolution cognitive de l'élève. La plupart des exemples que nous donnons ici pour illustrer la signification des composants du modèle et en faciliter la compréhension provient du prototype BIRDS, dont l'objectif pédagogique est de «donner à l'élève la capacité de reconnaître les caractéristiques morphologiques extérieures des oiseaux et de les lier à leurs modes de nourriture» et qui a été construit à partir du vidéodisque «British Garden Birds» édité par la BBC. Cependant le modèle que nous proposons n'est pas lié au domaine d'application. Il est suffisamment général pour être utilisé dans d'autres domaines.

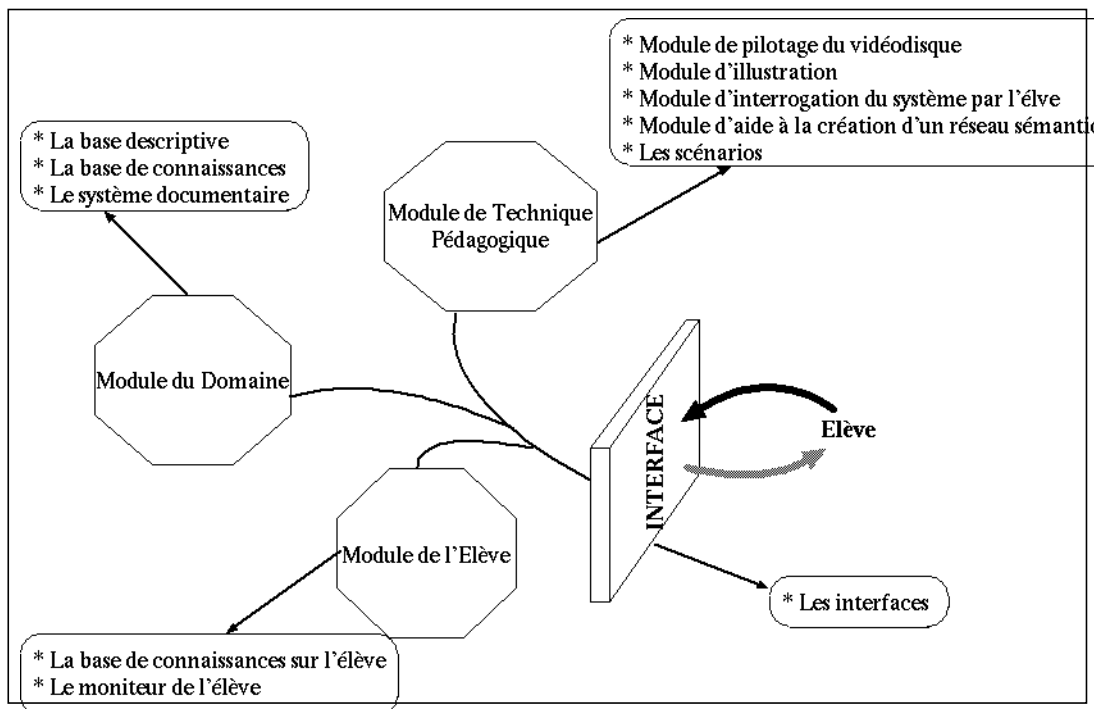


Figure 2. Architecture du prototype BIRDS

Le modèle de l'élève est composé d'un ensemble de six éléments

$$M = (R, A, F, C, E, N)$$

- $R = (N, P, D)$, référence personnelle

Ce composant sert à identifier l'élève :

N son nom,

P son prénom,

D sa date de naissance.

- $A = \{A_i\}$, l'ensemble d'activité de l'élève dans les différentes sessions

$i = 1, n$ (n le nombre de sessions)

A_i = l'activité de l'élève à la date d_i

$$A_i = (P1, P2, P3, P4)^i$$

$P1, P2, P3, P4$ représentent respectivement, la mesure d'activité de l'élève dans les quatre modules de technique pédagogique correspondant aux quatre niveaux d'habitudes évocatives. Il permet d'observer l'évolution du niveau d'habitude évocative de l'élève dans le temps. Par

exemple le P_j correspondant au $\max(P1, P2, P3, P4)$ représente le niveau d'habitude évocative dominant à la date d_i . A chaque session correspondent un ou plusieurs niveaux d'habitude évocative dominants.

Rappelons que la méthode que nous avons adoptée pour connaître les niveaux d'habitude évocative de l'élève est de *construire des modules pédagogiques de manière à refléter chaque niveau d'habitude évocative*. Ainsi nous pouvons connaître le niveau d'habitude évocative en observant les activités de l'élève dans les modules du système. Nous avons créé des fonctions qui permettent de mesurer l'activité de l'élève dans chaque module, prenant en compte l'interaction de l'élève avec le système.

Connaissances factuelles (F)

$$\bullet F = \sum_{i=1}^k f_i$$

k = nombre de sessions

$$f_i = \left\{ \left[cf_{jd_i} \right] \right\}$$

$$cf_i \in CF$$

cf_i = une connaissance factuelle et utilisée par l'utilisateur

CF = les connaissances factuelles du domaine

d_i la date de la session

$j = 1, m$ (le nombre de connaissances utilisées)

Par connaissances factuelles, nous entendons les connaissances concrètes, élémentaires ou les faits (ainsi interprétés du point de vue de l'enseignant) dans le domaine. En fait ce sont, pour nous, les informations visibles sur les images. Par exemple, *insecte, poisson, pattes-palmées, bec-crochu, arbre, eau* sont des connaissances factuelles en ornithologie pour exprimer respectivement la nourriture, la morphologie et l'habitat des oiseaux. La valeur de ce composant pour un élève reflétera son habitude évocative du type P1 et P2, mais particulièrement P2.

Connaissances conceptuelles (C)

$$\bullet C = \sum_{i=1}^k c_i$$

k = nombre de sessions

$$c_i = \left\{ \left\{ cc_{jd_i} \right\} \right\}$$

$cc_i \in CC$ et utilisé par l'élève

$CC = \text{connaissances conceptuelles du domaine}$

$cc_i = \text{une connaissance conceptuelle}$

d_i la date de la session

$j = 1, m$ (le nombre de connaissances utilisées)

Les connaissances conceptuelles signifient, pour nous, les connaissances qui relèvent du concept ou de l'abstraction. Elle ne sont pas toujours visibles (dans le domaine de la science de l'observation) et sont plutôt des abstractions de connaissances. Par exemple *bec*, *piscivore* son, pour nous des connaissances conceptuelles, *bec* est une abstraction de *toutes les formes de becs possibles*. *Piscivore* est une catégorie de *mode de nourriture en ornithologie*. Les **connaissances conceptuelles** dans le domaine de la résolution de problème consistent en les **méthodes de résolution, les axiomes, les méthodes de classification**, etc. L'interprétation d'une connaissance comme factuelle ou conceptuelle est l'affaire de l'enseignant. Une connaissance interprétée comme factuelle dans un domaine peut être interprétée comme conceptuelle dans un autre. Par exemple, nous avons donné comme types de connaissances factuelles les termes «insecte, poisson» alors que ces termes peuvent être considéré comme des connaissances conceptuelles si le domaine d'étude est les insecte ou les poissons. En fait, le domaine d'application et l'objectif pédagogique devant être parfaitement définis, la dichotomie connaissance factuelle - connaissance conceptuelle s'impose d'elle-même.

Résumé des erreurs (E)

- $E = \{(EL, EA)^i\}$

$i = 1, n$ (le nombre de sessions)

$$EL = \sum_{j=1}^n el_j$$

$n = \text{nombre de sessions}$

$el_j = \text{mesure d'erreurs lexicographiques}$

$$EA = \sum_{j=1}^n ea_j$$

n = nombre de sessions

ea_j = mesure d'erreurs d'association

$$el_k = |\{e_j\}|$$

$e_j \in$ erreur lexicographique à la date d_k

$j = 1, m$ (nombre de connaissances utilisées)

$$ea_k = |\{e_j\}|$$

$e_j \in$ erreur d'association à la date d_k

$j = 1, m$ (nombre de connaissances utilisées)

Les types d'erreurs commises peuvent aider à mieux connaître l'élève. Il y a deux formes d'erreurs - les erreurs lexicographiques et les erreurs d'association. Ces deux types d'erreurs se produisent dans tous les domaines.

Les erreurs lexicographiques ont les même formes dans tous les domaines : omission, insertion, remplacement d'un ou plusieurs caractères.

Les erreurs d'associations n'ont pas les mêmes formes dans tous les domaines. Dans certains domaines, elles se produisent par l'utilisation d'un seul terme (par exemple l'utilisation d'un terme qui n'a pas de signification dans le domaine). Le terme est correctement écrit mais il n'appartient pas au dictionnaire. Par exemple, dans le prototype qui a pour domaine l'ornithologie, *bec-palmé* est correctement écrit mais ce n'est pas un terme du domaine, c'est-à-dire qu'il n'est pas dans le dictionnaire. Ces erreurs se produisent quand il y a une contradiction contextuelle. Dans d'autres, ce type d'erreur concerne l'application de mauvaises règles en résolution de problème (en mathématique par exemple) ou une association erronée de deux termes par un lien sémantique (par exemple en sciences naturelles, dire que *faucon crécerelle est une sorte de piscivore alors que c'est un carnivore*).

Le dernier paramètre représente l'ensemble d'activités de l'élève qui permet de déterminer son niveau d'habitude évocative.

Niveau d'habitude évocative (N)

- $N = \{SA_i\}$

$i = 1, n$ (nombre de sessions)

SA_i = l'ensemble d'activités jusqu'à la date d_i

$$SA_i = (SP1, SP2, SP3, SP4)^i$$

$i = 1, n$ (nombre de sessions)

$$SPK = \sum_{j=1}^n PK_j$$

PK = Les activités du niveau d'habitude K pour une session

$j = 1, n$ (nombre de sessions)

Ce paramètre reflète la progression de l'élève dans ses niveaux d'habitude évocative depuis le début d'utilisation du système. Par exemple, dans le prototype, si l'élève ne fait que piloter le vidéodisque (dans le module d'observation) ou bien s'il le privilégie chaque fois qu'il utilise le système, cela peut nous indiquer que l'élève est toujours au niveau P1 de la classification par niveau d'habitude évocative (c'est-à-dire l'observation). Si le module indique que l'élève privilégie le module d'illustrations, cela peut signifier que l'élève est au niveau P2 où l'élève fait un minimum d'abstraction. Si l'élève utilise plutôt le module d'interrogation et que la valeur de C augmente dans le temps, cela peut signifier que l'élève est au niveau P3 où il conceptualise davantage ses connaissances. P4 peut être déduit si l'élève veut construire sa base de connaissances et sa base descriptive.

Ce modèle, tel que nous l'exploitons dans le prototype BIRDS, permet de suivre l'évolution de l'élève dans tous les niveaux auxquels nous nous intéressons. C'est une sorte d'évaluation d'un élève en terme de «meta-connaissance». Nous nous intéressons par exemple aux types d'erreurs, aux types de connaissances et à l'évolution des habitudes évocatives.

Une autre forme d'utilisation du modèle est de s'intéresser aux «connaissances» acquises par l'élève. On peut, par exemple, s'intéresser aux connaissances employées et non pas aux types de connaissances, s'intéresser aux erreurs, et non pas aux types d'erreurs, s'intéresser aux interactions de l'élève avec le système et non pas à ses activités par rapport à ses niveaux d'habitude évocative ou l'évolution de ses habitudes évocatives.

Le prototype BIRDS exploite le modèle de la première façon. DEDALE [mai91], un prototype développé dans notre équipe de recherche, selon l'architecture proposée par un étudiant en thèse au CNAM, exploite le modèle de la deuxième façon.

Rappelons que l'origine de cette première expérience dans le domaine de la modélisation de l'utilisateur a été d'appliquer un processus de recherche d'informations pour l'enseignement.

Suite à cette première expérience, nous avons cadré nos travaux de recherche sur le domaine de SRI.

Tout en continuant nos travaux sur la modélisation de l'utilisateur, nous avons démarré des études sur les moyens à mettre en oeuvre pour dépasser les limites des SRI qui sont basés sur l'automatisation de la gestion des connaissances du domaine d'application et les connaissances sur l'utilisateur. Le résultat principal de ces travaux est la proposition d'un nouvel environnement de recherche d'information que nous appelons un environnement de recherche coopérative d'information. Nos propositions dans le domaine de cet nouvel environnement sont présentées dans le chapitre suivant.

Chapitre IV

La recherche coopérative d'informations

Malgré tous les outils et les techniques intégrés dans les SRI pour mieux comprendre les besoins des utilisateurs et pour apporter des réponses adéquates à leurs besoins, il arrive encore que les utilisateurs ne parviennent pas à obtenir des résultats «satisfaisants». Les raisons de cet échec sont multiples (cf I.2.2. (*La recherche coopérative d'information*)). Nous remarquons que des échecs dus à ces raisons sont beaucoup moins nombreux quand l'utilisateur peut s'adresser à un médiateur humain, comme dans un centre de documentation. En effet, il y a un dialogue qui est souvent engagé entre l'utilisateur et le médiateur avant d'effectuer la recherche d'information à l'aide d'un SRI. Le dialogue porte souvent sur le besoin de l'utilisateur, exprimé en langage naturel, ce que l'utilisateur a déjà fait et ce qu'il a déjà obtenu comme résultat. En raison de l'efficacité des résultats obtenus à l'issue de ce type de dialogue entre l'utilisateur et le médiateur, nous avons orienté nos travaux vers le développement d'un système de recherche coopérative d'information (SRCI). Dans une SRCI, l'utilisateur pourra dialoguer avec un médiateur pour effectuer sa recherche sur des machines distantes.

Comme nous l'avons présenté dans le *Chapitre III (La modélisation dynamique de l'apprenant : le projet BIRDS)*, nous travaillons également sur la représentation explicite des utilisateurs afin d'acquérir des connaissances sur eux de façon individuelle. Nous disposons actuellement d'outils d'analyse des données permettant d'obtenir des synthèses d'information sur les comportements d'un utilisateur. Ces synthèses d'information sont très importantes dans un contexte de recherche coopérative d'information. En effet, outre le dialogue possible entre le médiateur et l'utilisateur, le médiateur peut, à partir de ces synthèses d'information, connaître le passé de l'utilisateur.

Pour situer cet axe de recherche (développement d'un système de recherche coopérative d'information) dans un contexte plus global des SRI, il faut présenter les composants (ou acteurs) et les types de communication entre ces composants. L'utilisation d'un SRI implique deux principaux composants : l'utilisateur et le système. Pour utiliser le système, l'utilisateur doit disposer des outils nécessaires pour communiquer avec le système. C'est donc tout naturellement que la recherche en communication homme-machine est l'un des principaux objectifs dans le développement des SRI. Les outils de communication homme-machine, les interfaces homme-machine, facilitent l'expression des requêtes par les utilisateurs et la présentation des solutions par le système. Outre le développement des interfaces, l'une des préoccupations principales reste l'amélioration des processus de traitement des requêtes des utilisateurs. L'amélioration des méthodes de mise en correspondance, extraction des connaissances sur les sessions d'utilisation des systèmes (mémoire de sessions), intégration des connaissances sur les utilisateurs pour l'interprétation de sa requête sont quelques exemples de domaines sur lesquels portent ces travaux. Ces améliorations sont généralement confiées au système, ce qui explique l'évolution des SRI vers les systèmes dits «intelligents». Ce que nous proposons dans le cadre de nos recherches est de concevoir un système basé sur la notion de **communication homme-homme** où les améliorations ne seront plus confiées uniquement au système mais confiées également à un être humain, par exemple à un médiateur. La notion de médiateur dans ce contexte englobe tout autre personne capable

d'aider dans un contexte de recherche d'information. Le médiateur peut être un utilisateur qui possède déjà de l'expérience dans la recherche d'information similaire, un expert en recherche documentaire, un expert du domaine de l'information recherchée, etc.

Nous avons démarré la réflexion sur l'architecture d'un SRCI par une étude de faisabilité dans le contexte des architectures différentes des ordinateurs. Notre objectif principal était de permettre à deux applications de communiquer quelle que soit l'architecture de l'ordinateur sur lequel l'application s'exécute. Cette étude a été confiée à un étudiant en thèse de CNAM, qui a travaillé sur le sujet pendant un an (1995/96). Nous avons défini comme cahier des charges fonctionnel les fonctions suivantes que l'architecture doit permettre de réaliser :

- l'utilisation des outils courants pour le développement d'autres outils nécessaires à la réalisation de l'architecture,
- la communication à distance entre deux applications (donc une communication entre deux utilisateurs),
- la communication entre deux ordinateurs à architectures différentes,
- les trois modes de fonctionnement (autonomie, observation, coopération),
- l'entrée en communication avec un autre utilisateur qui doit être entièrement sous le contrôle des utilisateurs impliqués (donc pas de communication non autorisée).

Au cours de l'étude de faisabilité, nous avons étudié une architecture existante qui permettrait de remplir les fonctions que nous nous sommes définies. Malheureusement, à l'époque de cette étude, les outils pour développer les spécifications de l'architecture n'étaient pas disponibles et la solution nous a semblée trop lourde à mettre en oeuvre. Il s'agit de l'architecture CORBA (Common Object Request Broker Architecture) [orf97]. Nous avons donc défini une nouvelle architecture permettant de réaliser les mêmes fonctions définies dans le cahier des charges. Le schéma global de l'architecture est présenté dans la *Figure 3. (Schéma global de l'architecture d'un SRCI)* et le formalisme OMT de l'architecture dans la *Figure 18. (Architecture d'un SRCI en formalisme OMT)*.

IV.1 L'architecture d'un système de recherche coopérative d'information

Notre architecture d'un SRCI est composée de quatre modules : *un serveur de localisation (SL), un serveur d'application (SA), une interface de connexion (IC) et l'interface de l'application*. Nous avons adopté la méthode de communication par l'envoi de message entre les applications. Une application envoie un message à une autre application qui le reçoit, l'interprète et le traite.

Un SRI, dans le contexte d'un SRCI, est un système de recherche d'information capable de communiquer avec un autre système de recherche d'information. Nous considérons donc le SRI comme une application avec son interface. Chaque application dispose donc des outils lui permettant de communiquer avec une autre application. Une application est capable d'envoyer

et de recevoir des messages. L'un de ces outils est le serveur d'application (SA). À chaque application est associée SA.

Le **serveur d'application** (SA) reçoit tous les messages des autres applications, effectue les transformations nécessaires et transmet le message à l'application à laquelle il est attaché. Le développeur de l'application a la possibilité d'insérer des contrôles sur la manière d'interpréter et de traiter les messages reçus par son application. Dans les applications que nous avons développées, nous avons utilisé la *socket* comme outil de transmission de message. En effet la socket est disponible sur pratiquement tous les types de système d'exploitation des ordinateurs.

Le SA ressemble au BOA (Basic Object Adapter) de CORBA. Comme dans un BAO, le SA reçoit des messages et les traduit sous la forme compréhensible par l'application. Comme dans CORBA, nous pouvons transmettre des messages en structure complexe de données. Dans le langage de programmation que nous utilisons, TclPro, une structure complexe de données peut être définie par une liste imbriquée, c'est-à-dire, une liste composée de listes.

Un autre composant de l'architecture est le **serveur de localisation** (SL). A noter que le SL est une application distincte au sens génie logiciel. Il dispose de la capacité à envoyer des messages et dispose d'un SA lui permettant de recevoir des messages. Le rôle du SL est d'enregistrer toutes les applications et de diffuser les informations qu'il reçoit à toutes les applications enregistrées. Les informations à diffuser sont définies par le développeur de l'application. Parmi les informations stockées dans le SL on trouve l'identité de chaque utilisateur, leurs états et leurs partenaires. A chaque utilisateur (ou une application) correspond un identificateur unique. Par exemple dans les applications que nous avons développées, les identificateurs des applications sont obtenus par la concaténation du *nom login de l'utilisateur, le nom de sa machine et du numéro de la socket de son SA*.

Le SL ressemble à l'ORB de CORBA. En effet, nous avons défini et réalisé le SL pour qu'il soit indépendant. Comme pour une application CORBA, le client doit connaître l'adresse du SL, connaître les objets (ou services) disponibles dans le serveur ainsi que la modalité de paramétrage de messages et les sémantiques des objets des services.

Contrairement à une application CORBA, pour nous, chaque application est potentiellement à la fois un serveur et un client car elle peut envoyer un message à une autre application comme elle peut recevoir des messages.

La procédure de contrôle d'entrée en communication avec une autre application est gérée par l'**interface de connexion** (IC). Cette interface est identique pour toutes les applications.

L'IC peut être comparé à l'IR (Interface Repository) de CORBA. L'IC dispose à tout moment de la liste des applications (serveurs au sens CORBA) actives ainsi que les informations

nécessaires pour leur envoyer des messages.

Sont également disponibles dans le SL le nom de la machine sur laquelle s'exécute l'application et le *port serveur* (un numéro attaché à une socket) sur la machine.

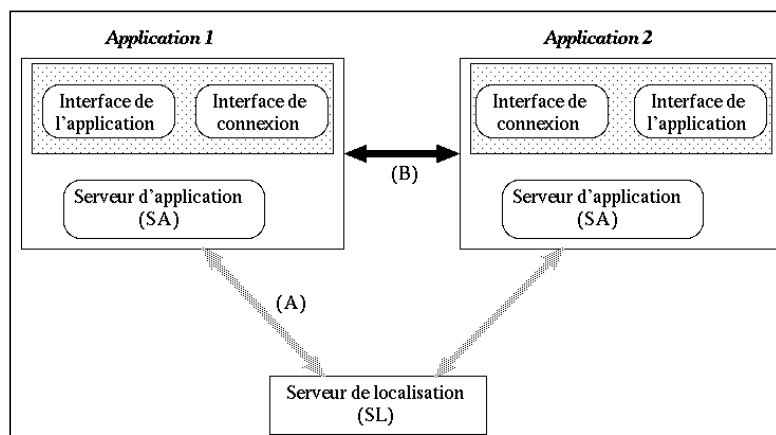


Figure 3. Schéma global de l'architecture d'un SRCI

Dans le contexte de la recherche coopérative, chaque application doit suivre la procédure suivante :

- 1) l'application s'enregistre dans le SL qui est toujours actif,
- 2) le SL informe la nouvelle application de toutes les autres applications déjà enregistrées,
- 3) le SL diffuse la présence de la nouvelle application à toutes les applications déjà enregistrées.

Par cette procédure, toutes les applications enregistrées possèdent toutes les informations nécessaires pour établir une connexion avec une autre application. Il suffit donc pour un utilisateur d'identifier son interlocuteur et demander d'entrer en connexion avec lui.

Comme l'entrée en communication est sous le contrôle des utilisateurs, l'interface dispose de trois types d'opération :

- *sélectionner un partenaire dans la liste des applications* : la liste des applications enregistrées dans le SL est transmise et mise à jour dans chaque application par le SL. Pour chaque enregistrement, départ et changement du statut d'une application, les informations concernant les applications sont mises à jour dans le SL et dans les applications. La mise à jour se fait par l'envoi de messages aux applications par le SL. L'interface contient pour chaque application son partenaire s'il y en a, son état (autonome, attente, demandé et connecté), et son identité. L'état *autonomie* indique que l'utilisateur ne souhaite pas entrer en communication avec une autre application. L'état *attente* indique que l'utilisateur peut accepter une communication. L'état *demandé* signifie que l'établisse-

ment de communication est en cours avec un autre utilisateur et, dans ce cas, le partenaire indique l'utilisateur avec qui il va entrer en communication. L'état *connecté* indique que l'utilisateur et son partenaire travaillent en mode coopératif.

- *demander une connexion ou une déconnexion* : En sélectionnant le partenaire souhaité, il faut solliciter l'autorisation du partenaire après avoir vérifié qu'il n'est pas déjà en communication avec un autre utilisateur. Cette demande est transmise à l'utilisateur demandé qui va répondre par l'opération de la catégorie *accepter* définie ci-dessous. Comme pour une demande de connexion, l'utilisateur qui souhaite arrêter le travail en coopération doit prévenir son partenaire pour mettre fin à la communication.
- *accepter une connexion ou une déconnexion* : Les opérations d'acceptation suivent les opérations de demande ci-dessous. Le partenaire potentiel et actuel d'un utilisateur doit confirmer sa volonté sur les demandes de l'utilisateur.

IV.2 Les modes de coopération

Un utilisateur peut être dans cinq états : *attente*, *demandé*, *connexion*, *observation* et *coopération*, comme le montre la *Figure 4*. (*Diagramme d'états pour la communication entre deux applications*).

En état «attente», l'utilisateur est en mode de fonctionnement autonome. Il n'est pas en connexion avec un autre utilisateur.

En état «demandé», un autre utilisateur demande l'autorisation d'entrer en communication avec l'utilisateur. Puisque le contrôle d'entrée en communication est sous le contrôle des utilisateurs, l'utilisateur demandé peut refuser d'entrer en communication avec le demandeur. Si l'utilisateur demandé accepte la connexion, il passe de l'état *demandé* en état *connexion*.

En état «connexion», les deux utilisateurs en communication peuvent se mettre dans deux types d'état : l'état «observation» ou l'état «coopération».

En état «observation», l'un des deux utilisateurs ne fait que regarder ce que l'autre utilisateur est en train de faire sans pouvoir agir sur ce qu'il fait.

En état «coopération» tout ce que fait un utilisateur est répété sur le poste du deuxième utilisateur et inversement. En effet les deux utilisateurs travaillent sur leurs applications comme s'ils disposaient d'une seule application.

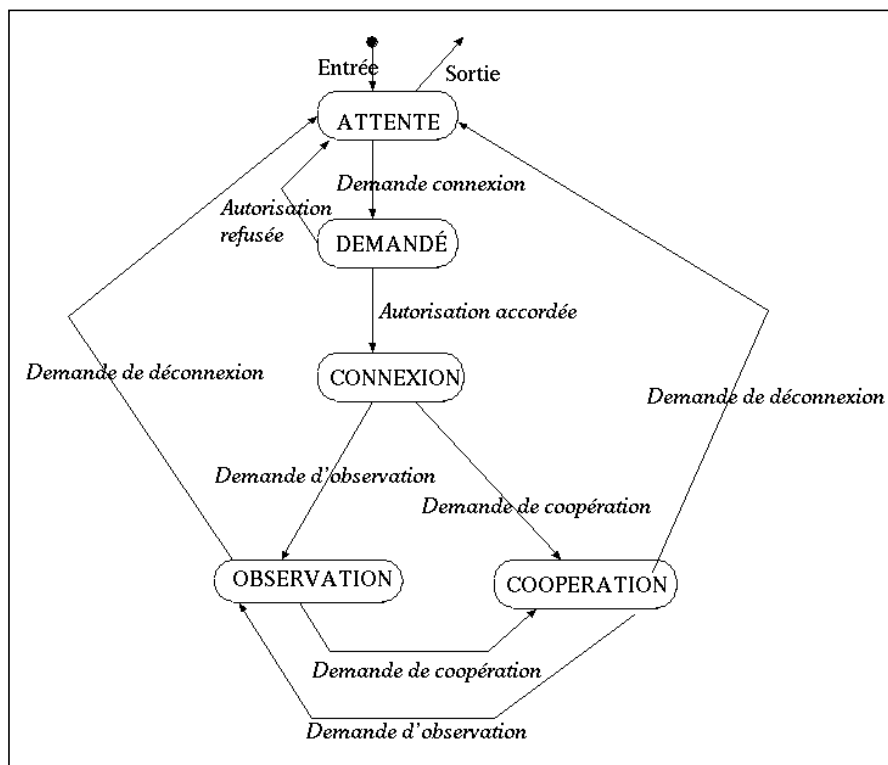


Figure 4. Diagramme d'états pour la communication entre deux applications

Cet environnement de recherche d'information est implémentée dans deux systèmes : STREEMS et METIORE. STREEMS est appliqué sur les informations botaniques portant sur les arbres autorisés par l'union européenne pour le reboisement des forêts. METIORE est appliqué sur les références bibliographiques du laboratoire LORIA.

Nous présentons METIORE dans le *Chapitre V (Le système METIORE)* et STREEMS dans l'*Annexe I (Le prototype STREEMS)*.

Chapitre V

Le système METIORE

Comme nous l'avons montré dans le *Chapitre III (La modélisation dynamique de l'apprenant : le projet BIRDS)*, nous pensons qu'un système de recherche d'information doit permettre à l'utilisateur (comme pour l'apprenant) de pouvoir évoluer dans les différents niveaux d'habitudes évocatives (observation, abstraction élémentaire, symbolisation et raisonnement, et créativité). Nous avons implémenté ce concept dans le prototype BIRDS par les différents modules du système et l'avons appliqué au processus d'apprentissage. METIORE est un système de recherche d'informations appliqué aux références bibliographiques. Le système est actuellement expérimenté sur une collection de 3160 références bibliographiques du laboratoire LORIA. Le système repose sur le même concept que BIRDS, à savoir, permettre à l'utilisateur de pouvoir évoluer dans ses niveaux d'habitudes évocatives. Cette fonctionnalité est implémentée dans le système par la méthode de recherche d'information que nous avons développée (cf section *II.5 (Conclusion et propositions pour la recherche d'information)*), et que nous appelons *analyse-croisée avec contraintes*.

Le nouveau type de recherche d'information et la recherche coopérative d'information sont également implémentés dans le système METIORE. Celui-ci permet donc d'effectuer un processus de recherche d'information en trois modes : autonomie, observation et coopération (cf *Chapitre IV (La recherche coopérative d'informations)*).

Le système METIORE est en fait le deuxième système que nous avons développé en implémentant les concepts d'évolution dans les habitudes évocatives et la recherche coopérative. Une description complète du système est présentée des les sections suivantes. Une présentation détaillée du premier système, STREEMS, se trouve dans l'*Annexe I (Le prototype STREEMS)*.

V.1 Le schéma global de METIORE

Comme toute application qui peut être utilisée dans un environnement de recherche coopérative d'information (cf *Chapitre IV (La recherche coopérative d'informations)*), METIORE est composé d'une interface graphique, une interface de connexion et d'un serveur d'application. Rappelons que l'interface de l'application permet à l'utilisateur d'effectuer des interactions avec le système. L'interface de connexion permet à l'utilisateur de contrôler la connexion avec d'autres utilisateurs. Le serveur de d'application permet de recevoir des messages des autres utilisateurs. Le système dispose d'une interface de messagerie par le texte en mode de coopération.

La *Figure 5. (Le schéma global de METIORE)* montre comment une application METIORE se présente dans un environnement de recherche coopératif. Quand une application METIORE est démarrée, elle est automatiquement en mode autonomie. En passant de ce mode en mode

d'observation ou en mode de coopération, l'application est automatiquement enregistrée sur le serveur de localisation. L'application METIORE(1) peut entrer en communication avec l'application METIORE(2). METIORE prend comme base de données, des fichiers au format SGML ou des fichiers avec les délimiteurs des champs.

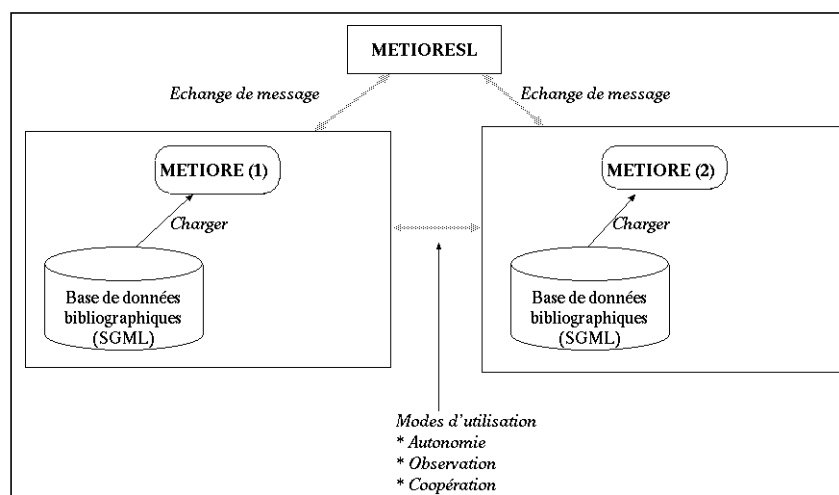


Figure 5. Le schéma global de METIORE

Les différents types d'accès et d'analyse d'information sont présentés dans les sections suivantes.

V.2 Analyse et accès à l'information

La méthode de recherche d'information par analyse croisée avec contraintes permet de formuler une recherche classique et d'indiquer des attributs pour une analyse croisée.

Une recherche classique est composée de critères combinés par des opérateurs booléens (cf II.3 (Recherche booléenne)). Par exemple dans METIORE, (author * david) OU (year >= 1990) indique que l'utilisateur cherche des références où l'auteur contient le nom «david» ou des publications à partir de 1990

L'interface d'analyse croisée avec contraintes permet d'indiquer trois attributs pour une analyse. Les combinaisons possibles des attributs sont présentées dans le *Tableau 2 : (Types de spécification des attributs pour les analyses)*.

Pour le type 1, aucun attribut n'est spécifié par l'utilisateur. Dans ce cas, nous utilisons le champ «référence» comme l'attribut1.

Le type 2 permet à l'utilisateur de spécifier l'un des attributs d'une référence bibliographique. Le système fournit les fréquences des valeurs de l'attribut.

Les types 3 et 4 permettent à l'utilisateur de spécifier deux attributs pour l'analyse. Dans le cas du type 3, les deux attributs sont identiques. Cela permet de calculer les co-occurrences intra-champ. Par exemple auteur, auteur pour connaître les co-auteurs. Dans le cas du type 4, les deux attributs sont différents. Cela permet de calculer les co-occurrences inter-champ. Par exemple auteur, année pour connaître les publications par année pour chaque auteur.

Les types 5 et 6 permettent à l'utilisateur de spécifier trois attributs pour l'analyse. Pour ces deux cas, l'attribut 3 doit être différent de l'attribut1 et attribut 2. Par contre dans le cas du type 5, l'attribut1 et l'attribut 2 sont identiques et dans le cas du type 6 l'attribut1 est différent de l'attribut 2. En spécifiant l'année comme le troisième attribut, l'utilisateur peut observer l'évolution des co-occurrences des valeurs des deux premiers attributs.

Il faut noter que seuls les attributs sont spécifiés par l'utilisateur. En donnant les fréquences des valeurs des attributs spécifiés par l'utilisateur, le système donne une vue globale de la base à l'utilisateur. Le système fournit également la liste des références correspondant aux résultats des analyses, ce qui donne un moyen à l'utilisateur pour découvrir les valeurs possibles pour les attributs. Cette fonctionnalité facilite donc la *recherche par le contenu*.

Type	Attribut1	Attribut2	Attribut3	Variance
1				Utiliser «référence» comme Attribut1
2	x			
3	x	x		Attribut1 == Attribut2
4	x	x		Attribut1 != Attribut2
5	x	x	x	Attribut1 == Attribut2
6	x	x	x	Attribut1 != Attribut2

Tableau 2 : Types de spécification des attributs pour les analyses

Tous ces types d'analyse permettent de suivre l'évolution et l'émergence de concepts. Nous présentons les interfaces avec les différents types de recherche d'informations dans les sections suivantes.

V.3 Exemples d'analyse et d'accès à l'information dans METIORE

Les exemples que nous présentons dans les sections suivantes portent sur les références bibliographiques du laboratoire LORIA entre 1983 et 1996.

V.3.1. L'interface graphique de METIORE

La Figure 6. (L'interface graphique de METIORE) présente l'interface de l'application.

Les valeurs des attributs y sont présentées. Le menu «Fichier» permet de charger une autre base. Le menu «Langue» permet de changer la langue de présentation des informations de l'interface. Seuls tes textes descriptifs sont multilingue et non les valeurs des attributs. Cette fonctionnalité permet à deux utilisateurs de voir les descriptifs de l'interface dans la langue qui leur convient.

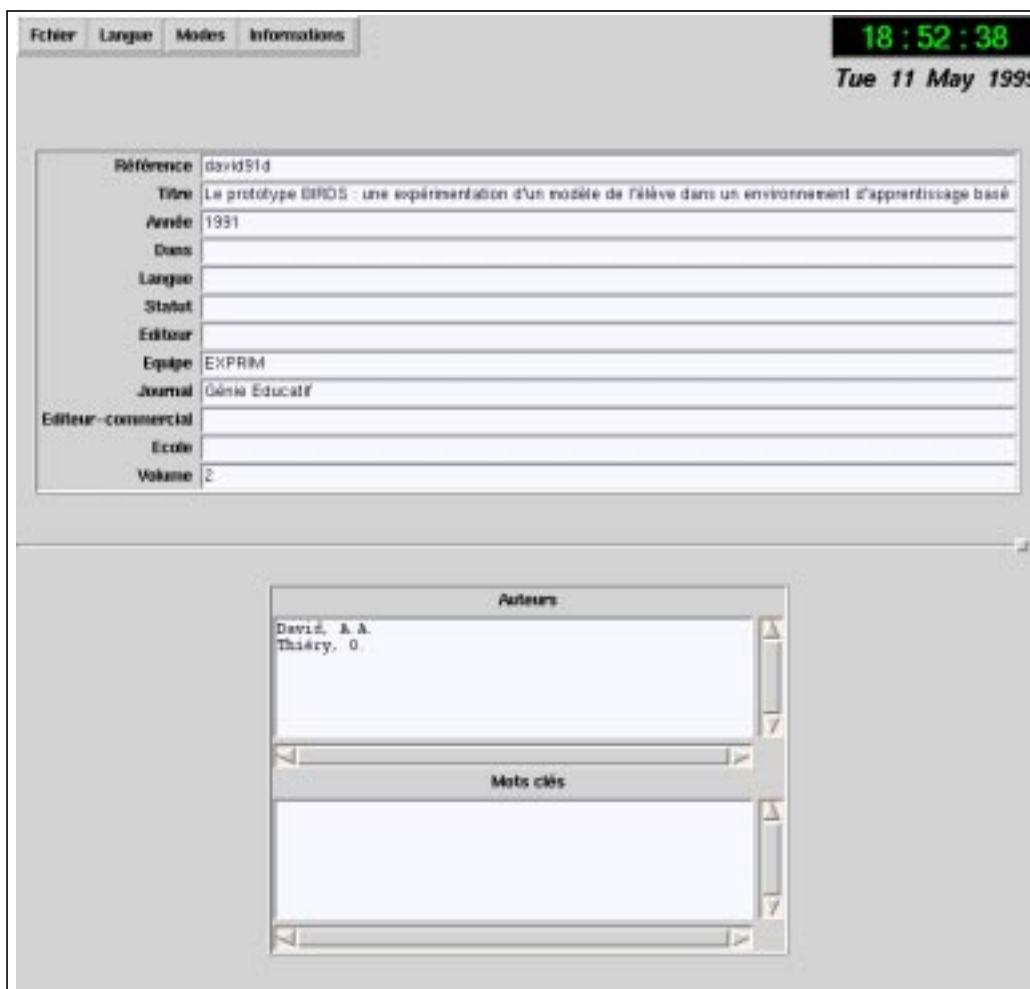


Figure 6. L'interface graphique de METIORE

La Figure 7. (L'interface de la recherche par analyse-croisée avec contraintes) présente l'interface pour l'analyse croisée avec contraintes. Les fréquences et les valeurs de co-occurrences sont présentées dans la zone «Résultats». En cliquant sur l'un des éléments de la liste de cette zone, l'utilisateur verra s'afficher la liste des objets (les références) qui correspondent à cette co-occurrence. En cliquant sur l'un des objets de la zone «objet», la référence sera affichée dans l'interface de la Figure 6. (L'interface graphique de METIORE).

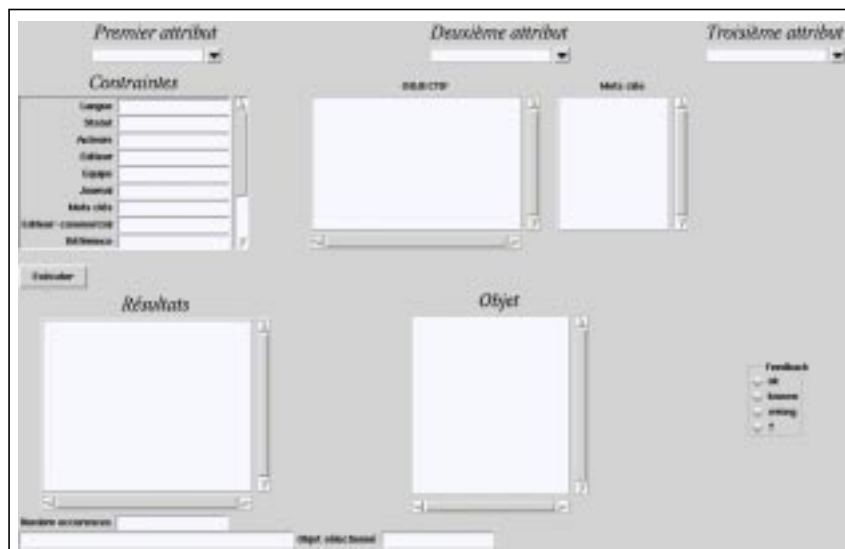


Figure 7. L'interface de la recherche par analyse-croisée avec contraintes

V.3.2. Recherche booléenne

Nous présentons ici un exemple d'une recherche avec uniquement des critères : (Auteur * david) OU (Année >= 1990). Les opérateurs de comparaison possibles sont : * (contient), =, <, <=, >, >=, <>. L'opérateur * est en fait une expression régulière.



Figure 8. Une recherche booléenne

La zone «Feed-back» de la figure montre qu'il est prévu que l'utilisateur puisse donner son évaluation de la solution proposée par le système.

V.3.3. Analyse de fréquence

La figure Figure 9. (*La fréquences des publications des auteurs*) présente une analyse des publications des auteurs.

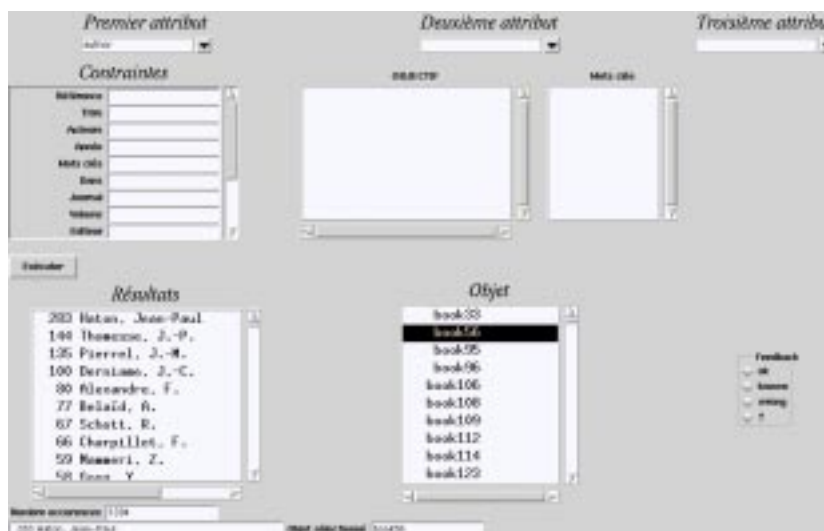


Figure 9. La fréquences des publications des auteurs

V.3.4. Analyse intra-champ

La figure Figure 10. (*Analyse de co-auteurs*) présente l'analyse des co-auteurs. Pour chaque type d'analyse, un fichier du résultat au format excel est généré. Ce fichier permet d'obtenir le résultat sous forme graphique. La Figure 11. (*Présentation graphique des fréquences des co-auteurs*) présente le graphique correspondant au résultat de l'analyse de la Figure 10. (*Analyse de co-auteurs*).

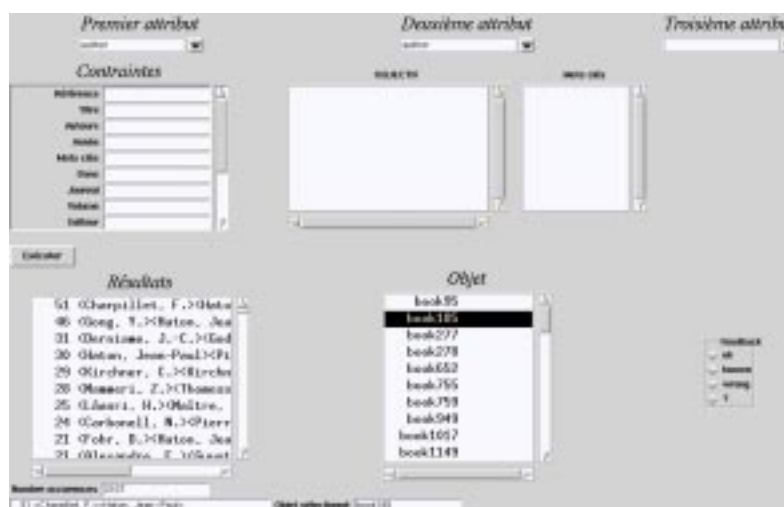


Figure 10. Analyse de co-auteurs

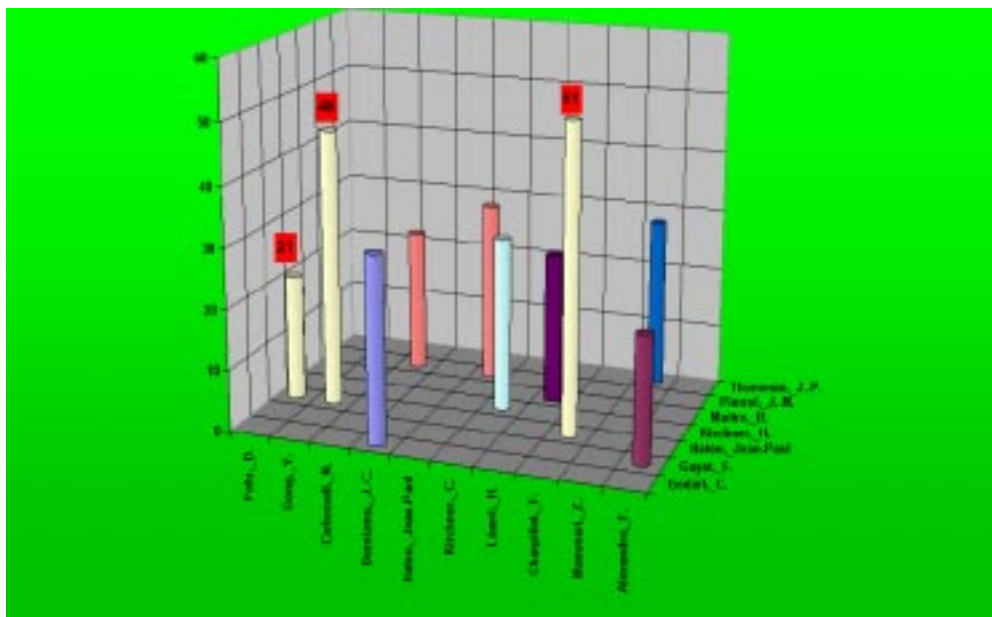


Figure 11. Présentation graphique des fréquences des co-auteurs

V.3.5. Analyse inter-champ

L'analyse inter-champ de la Figure 12. (*Analyse-croisée inter-champ*) montre les mots-clés utilisés par les auteurs. Comme pour tout autre type d'analyse, le graphique peut être obtenu comme le montre la Figure 13. (*Graphique analyse-croisée inter-champ*) pour cette analyse

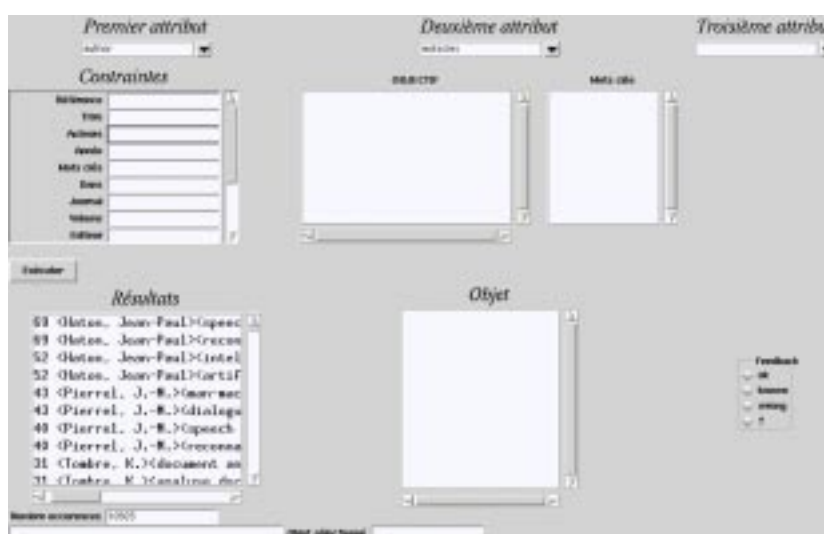


Figure 12. Analyse-croisée inter-champ

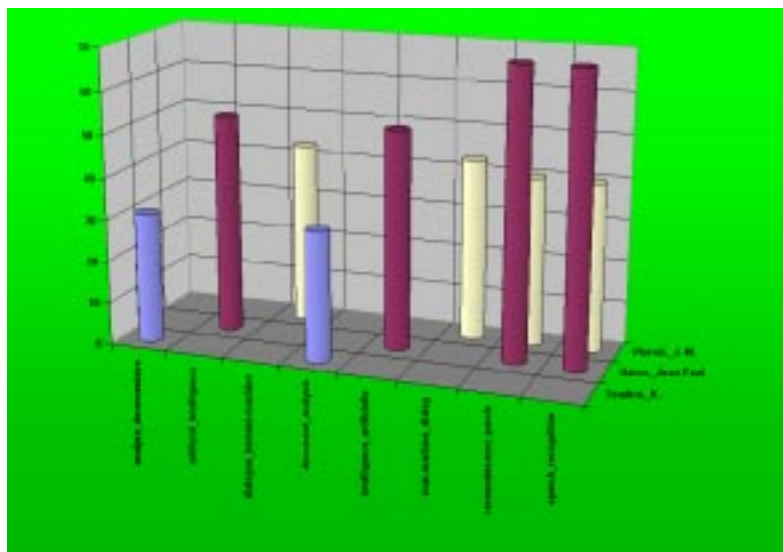


Figure 13. Graphique analyse-croisée inter-champ

V.3.6. Calcul de l'évolution des analyses

Comme nous l'avons présenté dans la V.2 (*Analyse et accès à l'information*), l'évolution des concepts ou l'émergence des concepts peut être obtenue en indiquant année comme le troisième attribut. La Figure 14. (*Evolution des co-auteurs*) présente ce type d'observation. Il s'agit de l'évolution des co-auteurs.

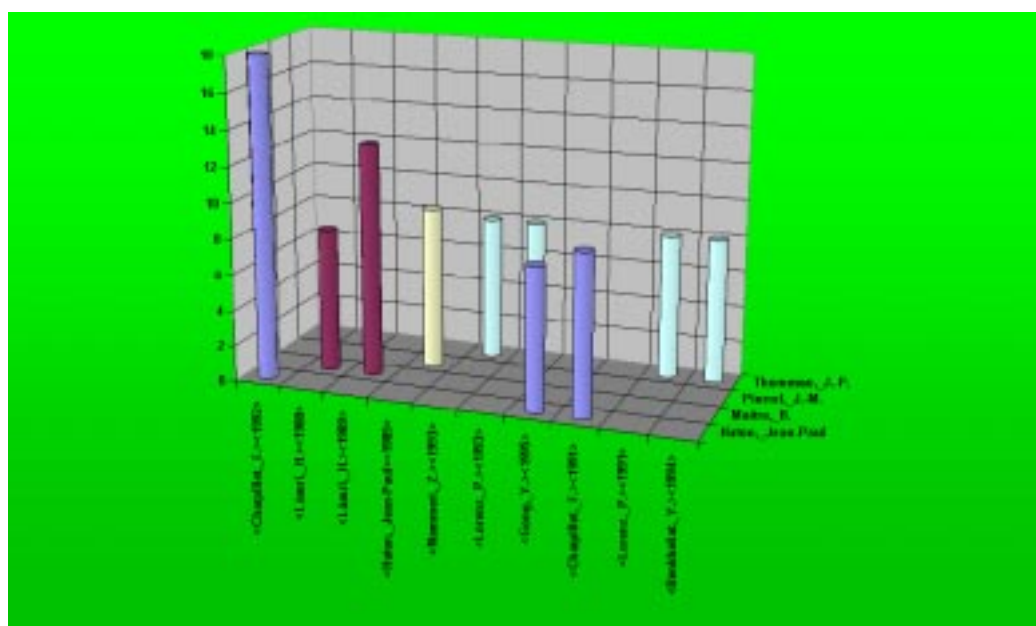


Figure 14. Evolution des co-auteurs

V.4 METIORE et la recherche coopérative

METIORE fonctionne dans un environnement de recherche coopératif. Nous présentons dans les deux sections suivantes les interfaces spécifiques à la recherche coopérative d'information.

V.4.1. L'interface de connexion

Cette interface montre les utilisateurs qui sont enregistrés dans le serveur de localisation. Elle montre aussi que les deux utilisateurs sont déjà entrés en connexion. Les boutons «Request» et «Accept» permettent de contrôler la connexion avec un autre utilisateur.



V.4.2. L'interface d'échange de message textuel

Cette interface est automatiquement créée sur le poste de chaque utilisateur qui sont en communication. Elle permet aux deux utilisateurs en communication d'échanger des messages sous forme de texte.

Comme on peut le remarquer, la date et l'heure des échanges sont affichées. Cela montre que nous obtenons des informations sur les activités de l'utilisateur, y compris les échanges de message entre lui et son interlocuteur.

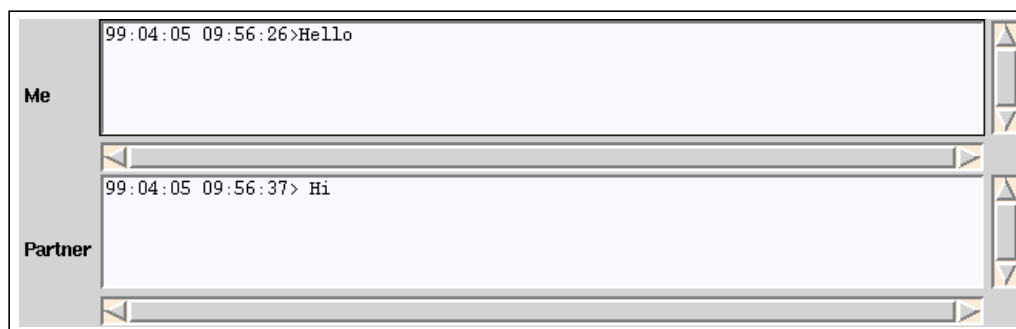


Figure 15. L'interface d'échange de message textuel entre deux utilisateurs

V.5 Le modèle de l'utilisateur dans METIORE

Le modèle de l'utilisateur est exploité selon l'algorithme de la section suivante.

V.5.1. Implémentation actuelle

Quand l'utilisateur démarre l'application, nous lui demandons son objectif principal. Comme nous le montrerons dans la section VI.4 (*Notre modèle de l'utilisateur*), cela est une spécificité de notre modèle et de notre approche du processus de recherche d'information. Nous regardons si l'utilisateur a déjà formulé un objectif similaire. Si oui, le système lui propose les solutions classées selon les types d'évaluation qu'il a données. Il peut changer les évaluations. S'il est satisfait des solutions, alors il peut arrêter sinon il peut modifier son objectif ou effectuer d'autres recherches. Nous introduisons ici la notion de sous-objectif. S'il n'y a pas d'objectif enregistré pour l'utilisateur qui ressemble à celui qu'il vient de présenter, le système regarde s'il y aurait des objectifs similaires qui ont été présentés par d'autres utilisateurs. Si oui alors cette solution servira de point de départ pour le système. L'utilisateur pourra toujours évaluer les solutions du système. S'il n'y a pas d'objectif similaire par les autres utilisateurs, alors l'utilisateur doit effectuer des recherches pour trouver ce qu'il cherche.

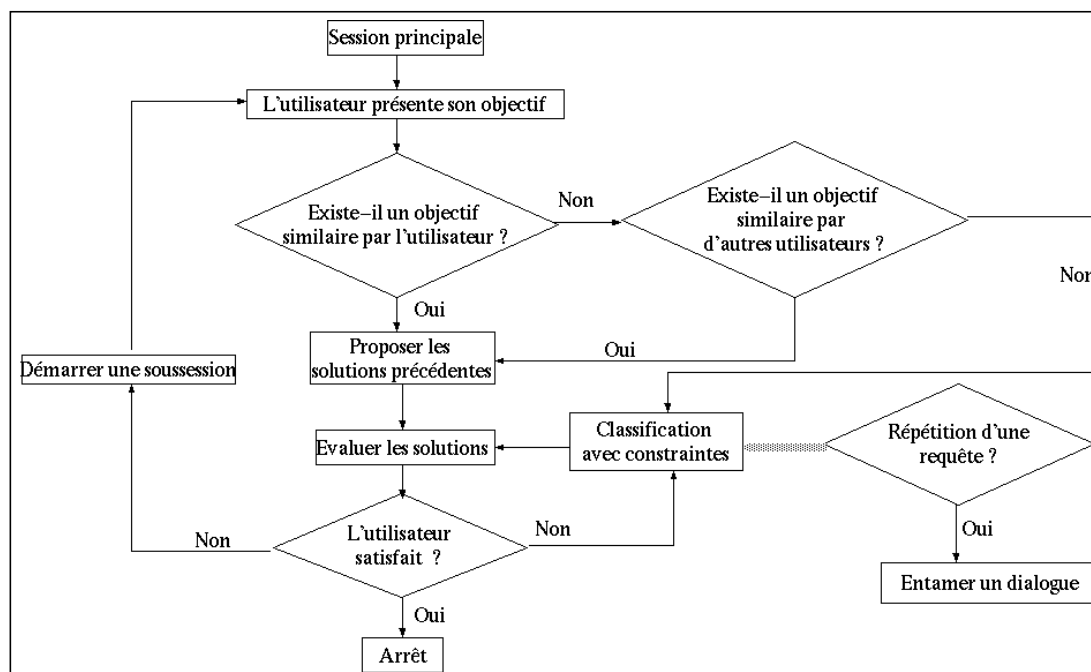


Figure 16. Algorithme d'exploitation du modèle de l'utilisateur dans METIORE

Comme nous l'avons présenté dans la section précédente, le système observe le comportement de l'utilisateur. Si l'utilisateur répète une même requête pour le même objectif alors le système peut déclencher un dialogue pour savoir pourquoi.

L'intégration du paramètre «objectif» dans le modèle de l'utilisateur est déjà prévue dans

l'interface d'accès et d'analyse de l'information par les zones «OBJECTIF» et «Mots clés». La zone «OBJECTIF» donnera le texte complet de l'objectif de l'utilisateur. La zone «mots clés» donnera la liste de mots clés associés à cet objectif.

L'intégration du modèle de l'utilisateur dans l'environnement de recherche coopérative constitue notre perspective de recherche à court terme. Nous présentons dans le chapitre suivant notre modèle de l'utilisateur après avoir détaillé le processus de transformation du modèle de l'apprenant en ce modèle.

Chapitre VI

Le modèle de l'utilisateur dans un environnement de recherche coopérative d'information

Fort des expériences que nous avons acquises dans la représentation de l'utilisateur, le développement de systèmes documentaires et dans la méthode de recherche coopérative d'information, nous avons défini comme objectif de recherche à court terme l'intégration d'un modèle de l'utilisateur et la recherche coopérative d'information dans un système de recherche d'information en vue de la personnalisation des réponses du système. Cet objectif constitue une base de la définition d'un sujet de thèse qui est actuellement menée par David BUENO de l'université de Malaga, sous ma co-direction.

Le cadre de ce travail est présenté dans les sections suivantes. La section *VI.1 (Du modèle de l'élève au modèle de l'utilisateur)* présente comment le modèle de l'utilisateur est dérivé du modèle de l'élève en montrant les similarités et les différences entre les deux types de modèle. Nous présentons la méthode de représentation de connaissances par le concept de catégories fondamentales dans la section *VI.3 (La théorie de catégories fondamentales)* car nous pensons que cette méthode nous permettra de mieux connaître le niveau de connaissance de l'utilisateur. La représentation concrète et les différents types d'exploitation possible du modèle de l'utilisateur sont présentés dans les sections *VI.4 (Notre modèle de l'utilisateur)* et *VI.6 (Analyse des connaissances sur un utilisateur)*. Notre approche pour acquérir les connaissances dont nous avons besoin sur l'utilisateur est présentée dans la section *VI.5 (Acquisition de connaissance sur l'utilisateur)*. Et enfin l'exploitation du modèle dans un contexte de recherche coopérative est présentée dans la section *VI.8 (Le modèle de l'utilisateur dans un contexte de recherche coopérative)*.

VI.1 Du modèle de l'élève au modèle de l'utilisateur

Les systèmes d'apprentissage et les SRI se ressemblent sur beaucoup de points. En effet, la fonctionnalité d'un système d'apprentissage ainsi que celle d'un SRI peuvent être exprimées par la fonction

$$f(O, M, E, C) \rightarrow S$$

Nous présentons dans le *Tableau 3 : (Fonctionnalités des systèmes d'apprentissage et SRI)* les significations des paramètres dans les deux types de système. Rappelons que, pour nous, le modèle de l'élève et le modèle de l'utilisateur permettent de représenter respectivement les apprenants. Une étoile devant un point de comparaison signifie une différence entre les deux types de système.

Dans un SRI, on ne cherche pas à faire évoluer les capacités des utilisateurs à formuler leurs demandes. Pourtant en mettant l'utilisateur au centre des préoccupations, nous pensons qu'il y a des idées employées dans les systèmes d'apprentissage qu'il faut intégrer dans un SRI. Par exemple, considérer qu'un SRI doit aider à «résoudre un problème de recherche d'information» transforme le système en une forme de système d'apprentissage. Dans ce cas, il faut répondre aux questions suivantes :

- est-ce que le problème à résoudre est bien clair pour l'utilisateur?
- est-ce qu'il dispose d'outils pour acquérir les connaissances nécessaires pour bien définir son problème?
- est-ce qu'il dispose d'outils nécessaires pour appliquer ses connaissances pour résoudre le problème?
- est-ce qu'on s'intéresse à ce que l'utilisateur puisse créer sa propre stratégie pour résoudre le même problème et d'autres qui sont similaires?
- est-ce que l'intérêt du système est de cultiver chez l'utilisateur la capacité à résoudre un problème ou bien, résoudre le problème à la place de l'utilisateur?

Paramètre	système d'apprentissage avec modèle de l'élève	système de recherche d'information avec modèle de l'utilisateur
M (Motivation)	La motivation joue un rôle important dans le processus d'apprentissage. Si l'apprenant n'est pas motivé, il peut renoncer à la moindre difficulté. L'objectif pédagogique ne peut pas être atteint par un apprenant démotivé même si l'objectif pédagogique est bien défini et avec un système facile à utiliser.	La motivation joue un rôle important dans le processus de recherche d'information. Si l'utilisateur n'est pas motivé, il peut renoncer à la moindre difficulté. Un utilisateur qui recherche des informations qui ne comportent pas de véritable enjeu pour lui renoncera à la moindre difficulté même si le système est facile d'utilisation.
E (Etat)	Représente l'état de l'apprenant, concrétisé par le modèle de l'élève	Représente l'état de l'utilisateur, concrétisé par le modèle de l'utilisateur
C (Connaissance)	Représente la base de connaissance du système portant sur le domaine d'application	Représente la base de connaissance du système portant sur le domaine d'application

Tableau 3 : Fonctionnalités des systèmes d'apprentissage et SRI

Paramètre	système d'apprentissage avec modèle de l'élève	système de recherche d'information avec modèle de l'utilisateur
<p>O (Objectif)</p>	<p>Représente un objectif pédagogique ou un sous-objectif de l'objectif global. L'objectif global du système est fixé par l'enseignant.</p> <p>L'objectif principal peut être connu ou non pour l'apprenant.</p> <p>Des problèmes sont associés à un objectif pédagogique. Résoudre ces problèmes constitue des étapes pour atteindre l'objectif pédagogique.</p> <p>D'une manière générale, résoudre des problèmes ne constitue pas un objectif en soi. C'est la façon de résoudre le problème qui compte le plus souvent. C'est l'apprenant que l'on cherche à évaluer.</p> <p>La bonne résolution des problèmes associés à un objectif fournit un moyen pour évaluer le degré de la maîtrise de la méthode pour atteindre l'objectif par l'apprenant.</p>	<p>Représente un besoin en information qui correspond à l'objectif principal. Des sous-objectifs peuvent être associés à cet objectif principal. L'utilisateur définit son objectif principal.</p> <p>Le besoin en information peut être bien défini ou non pour l'utilisateur.</p> <p>Exprimer un besoin nécessite un certain nombre de pré-requis comme la maîtrise des termes à utiliser pour formuler le besoin et la façon de combiner les termes dans l'expression du besoin.</p> <p>* D'une manière générale, la façon de formuler des besoins ne constitue pas un objectif en soi c'est le résultat final qui compte le plus pour l'utilisateur. L'objectif d'un SRI n'est pas d'évaluer l'utilisateur, c'est l'utilisateur qui évalue le système par rapport à l'adéquation des solutions proposées.</p> <p>Le niveau d'acquisition des pré-requis par l'utilisateur fournit un moyen au système pour l'évaluer dans l'objectif (pour le système) de s'adapter à ce niveau.</p>
<p>f</p>	<p>Représente la fonction, qui selon l'objectif pédagogique, l'état de l'apprenant, la motivation de l'apprenant et l'ensemble de connaissances du système portant sur le domaine, propose une stratégie pédagogique. Une stratégie pédagogique est un ensemble de scénarios (un ensemble d'actions possibles).</p> <p>Même si l'apprenant connaît l'objectif pédagogique à atteindre, il ne sait pas forcément comment le réaliser.</p>	<p>Représente la fonction, qui selon le besoin en information, l'état de l'utilisateur, la motivation de l'utilisateur et l'ensemble de connaissances du système portant sur le domaine, propose des solutions ou suggère une autre démarche pour rechercher l'information. Pour nous un SRI dispose de plusieurs stratégies pour rechercher les informations.</p> <p>Un besoin en information bien défini ne signifie pas forcément la connaissance des démarches pour trouver les informations.</p>

Tableau 3 : Fonctionnalités des systèmes d'apprentissage et SRI

Paramètre	système d'apprentissage avec modèle de l'élève	système de recherche d'information avec modèle de l'utilisateur
S	<p>Représente une stratégie pédagogique : le système comprend en effet un ensemble de stratégies</p> <p>En fonction des valeurs des paramètres du modèle, une nouvelle stratégie est proposée à l'élève.</p>	<p>Représente un type de démarche pour rechercher des informations. Pour nous, un SRI comprend un ensemble de stratégies pour rechercher des informations.</p> <p>* En fonction des valeurs des paramètres du modèle, le système adapte les solutions à proposer à l'utilisateur mais il peut lui suggérer une stratégie.</p>

Tableau 3 : Fonctionnalités des systèmes d'apprentissage et SRI

VI.2 Comparaison des activités de l'élève et celles de l'utilisateur en SRI

Un SRI ne vise pas uniquement des spécialistes du domaine de l'information recherchée ou uniquement les spécialistes de la recherche documentaire. Par rapport aux utilisateurs «non spécialistes», Cluzeau-Ciry dit dans [clu88] que «personnalisées, leurs demandes correspondent au besoin culturel d'un *autodidacte*». Nous remarquons qu'il y a en effet une similarité entre ce que Cluzeau-Ciry présente comme les quatre types de requêtes et les quatre niveaux d'habitude évocative, que nous utilisons dans notre modèle de l'élève.

Type de requête	Niveau d'habitude évocative
1. Consultation	Observation
2. Demande précise	Abstraction élémentaire
3. Demande thématique	
4. Demande connotative	Symbolisation et raisonnement

Tableau 4 : Comparaison types de requête et niveaux d'habitude évocative

Nous proposons qu'un SRI soit un véritable environnement de recherche d'information où plusieurs approches de recherche d'informations sont intégrées. Nous développons au minimum, quatre types d'activité qui correspondent aux quatre niveaux d'habitude évocatives.

VI.2.1. La phase d'observation

Dans cette phase, l'apprenant prend connaissance du domaine étudié par le processus d'observation. Pour un utilisateur en recherche d'information, cette phase correspond à une simple consultation. Dans notre exemple d'une agence de voyage, on peut présenter des séquences vidéos à l'utilisateur.

VI.2.2. La phase d'abstraction élémentaire

Dans cette phase, l'apprenant acquiert le vocabulaire du domaine, désigne des objets et commence ainsi un processus élémentaire d'abstraction. Pour un utilisateur en recherche d'infor-

mation, cette phase correspond à des demandes en information précises ou des demandes thématiques. Par rapport à notre agence de voyage, le client peut demander ce que signifie le terme «*formule d'un voyage*».

VI.2.3. La phase de symbolisation et de raisonnement

Dans cette phase, l'apprenant emploie un vocabulaire spécialisé du domaine, il est capable de développer des idées abstraites et de raisonner avec des connaissances déjà acquises, soit pour exprimer ses connaissances, soit pour résoudre des problèmes. Cette phase correspond pour un utilisateur en recherche d'information, à des demandes connotatives. Par rapport à notre agence de voyage, le client peut demander des informations précises sur une formule de voyage avec ses préférences.

VI.2.4. La phase de créativité

Dans cette phase, l'apprenant exploite ses connaissances pour en développer d'autres qui ne lui ont pas été transmises. Pour un utilisateur en recherche d'information, cette phase correspond à ses jugements, par exemple pour exprimer ses préférences à la suite d'une analyse des réponses proposées par le système. Par rapport à notre agence de voyage, le client peut donner à l'hôtesse de l'agence ses préférences en fonction de ses analyses des voyages précédents.

Dans la réalisation du modèle de l'élève et de l'utilisateur, nous observons les activités de l'utilisateur dans les différentes phases. Nous observons également les termes utilisés dans chacune des phases. Pour une plus grande efficacité, ces observations doivent avoir lieu dans le temps et ne pas être limitées à une seule session d'utilisation du système.

Notre observation des termes utilisés est basée sur la théorie de catégories fondamentales étudiée en science cognitive. Nous résumons cette théorie dans la section suivante.

VI.3 La théorie de catégories fondamentales

La description de la théorie des catégories fondamentales est extraite de [fod86] (pages 123 - 127). L'auteur précise que l'une des idées les plus intéressantes qu'ont produit les recherches récentes en sciences cognitives est qu'il existe un niveau «fondamental» d'objets perceptuels (ou de catégories perceptuelles fondamentales).

Considérons une hiérarchie de catégorie comme *caniche, chien, mammifère, animal, objet physique, chose*. Sur ce type de hiérarchie, on peut remarquer que les objets sont classés par degré croissant d'abstraction, de sorte qu'une entité donnée peut appartenir à n'importe laquelle ou à toutes les catégories, et que l'extension de la catégorie est de plus en plus grande à mesure que l'on s'élève dans la hiérarchie. Par exemple, il y a plus de chiens que de caniches possibles.

On remarque également que cette hiérarchie est *implicative*, ce qui veut dire qu'il est né-

cessaire que tout ce qui appartient à une catégorie de niveau n appartienne à toutes les autres catégories de niveau plus grand que n . Par exemple, tous les caniches sont des chiens. L'implication dans cette hiérarchie peut être analytique ou linguistique.

L'idée des catégories fondamentales découle de l'observation que, dans ce type de hiérarchies implicatives, certains niveaux d'abstraction sont particulièrement proéminents d'un point de vue psychologique. Intuitivement, les niveaux les plus saillants sont les niveaux «moyens» d'abstraction. En utilisant l'exemple ci-dessus, le niveau *chien* serait le niveau de la catégorie fondamentale et non les niveaux de *caniche* ou *chose*.

Il faut remarquer qu'il n'y a pas de définition indépendante de «moyen» (cf paragraphe précédent) et que la notion de proéminence est plutôt relative. Néanmoins les propriétés psychologiques suivantes convergent sur le même niveau de chaque hiérarchie implicative. Un membre d'une hiérarchie qui possède une de ces propriétés risque de posséder toutes les autres. Une catégorie qui les possède toutes est une catégorie «fondamentale».

- ⇒ (a) La catégorie fondamentale d'une hiérarchie correspond à l'item lexical le plus fréquent ; «chien» est un item lexical plus fréquent que «animal» ou «caniche»
- ⇒ (b) En général, le mot qui correspond à la catégorie fondamentale d'une hiérarchie est appris avant les mots qui exprime d'autres niveaux de la hiérarchie.
- ⇒ (c) Souvent, la catégorie fondamentale est le membre le moins abstrait de la hiérarchie auquel correspond un item lexical monomorphémique. Par exemple, dans la hiérarchie «fauteuil à bascule», «fauteuil Empire», «fauteuil», «meuble», «artefact», «objet physique», «fauteuil» est l'item le moins abstrait des trois autres items composés («fauteuil à bascule», «fauteuil Empire», «fauteuil»). Pour certains domaines, il semble que la lexicalisation monomorphémique des catégories fondamentales soit universelle. Par exemple, peu de langues auront à la fois un mot morphologiquement simple pour «rouge rosâtre délavé» et un mot morphologiquement complexe pour «rouge». Il est considéré que (a) et (b), (c) sont reflet linguistique de la proéminence psychologique des catégories fondamentales.
- ⇒ (d) Les catégories fondamentales sont faciles à apprendre par ostension. Un enfant qui ne connaît pas le mot «caniche» peut apprendre le mot «chien» par ostension, mais il sera probablement incapable d'apprendre le mot «animal» par ostension s'il n'a pas appris au moins quelques noms d'animaux du niveau de «chien».
- ⇒ (e) Les catégories fondamentales sont des «pic d'information», au sens suivant. Demandez à quelqu'un d'énumérer toutes les propriétés qui lui viennent à l'esprit en pensant à des *animaux* ; demandez-lui ensuite d'énumérer toutes les propriétés qui lui viennent à l'esprit en pensant à des *chiens* ; enfin demandez-lui d'énumérer toutes les propriétés qui lui viennent à l'esprit en pensant à des «caniches». Vous obtiendrez beaucoup plus de propriétés pour *caniche* que pour *animal*, et la liste proposée pour *caniche* contiendra assez peu de propriétés qui n'auront pas été données pour *chien*. Il semble que les catégories de base soient celles qui contiennent le plus d'information «par unité de jugement». Avec la «maxime de quantité» (Soyez

informatif) et la «maxime de manière» (Soyez bref) de Paul Grice, cela accrédite les faits pragmatiques suivants :

- ⇒ (f) Les catégories fondamentales sont celles qu'il est naturel d'utiliser pour décrire les choses «en l'absence d'exigence particulière». Vous me demandez : «Que voyez-vous par la fenêtre?» ; Je réponds : «Une femme qui promène un chien» (et non : «Une femme qui promène un animal» ou «Une femme qui promène un petit caniche gris argenté»). Ici, l'important est que, toutes choses étant égales, le niveau de description préféré est le premier niveau de description, même lorsqu'on dispose d'informations suffisantes pour une description plus détaillée. On peut aussi remarquer que ces faits linguistiques semblent être la manifestation superficielle d'une réalité psychologique plus profonde, qui est :
- ⇒ (g) Les catégories fondamentales sont *données* phénoménologiquement ; elles représentent, en quelque sorte, le niveau naturel pour *se décrire* les choses à soi-même. Donc, un coup d'oeil par la fenêtre nous révèle une *femme* qui promène un *chien*, et non une femme qui promène un caniche gris etc. Evidemment, une inspection plus détaillée modifie tout cela. Mais la prééminence phénoménologique, c'est l'accessibilité sans inspection. On est tenté de prédire à partir de ces intuitions que l'identification perceptuelle des catégories fondamentales sera plus rapide que l'identification de membres plus ou moins abstraits de la hiérarchie implicite.
- ⇒ (h) Typiquement, les catégories fondamentales sont les catégories les plus abstraites de la hiérarchie qui recouvrent des individus d'apparence à peu près semblable. Dans, en gros, vous pouvez dessiner quelque chose qui ne sera qu'un chien, mais vous ne pouvez pas dessiner quelque chose qui ne sera qu'un animal ; vous pouvez dessiner quelque chose qui ne sera qu'une chaise, mais vous ne pouvez pas dessiner quelque chose qui ne sera qu'un meuble.

Cela laisse penser que les catégories fondamentales (à la différence des catégories plus abstraites) peuvent être déterminées, de manière fiable, à partir des propriétés visuelles des objets.

VI.4 Notre modèle de l'utilisateur

L'activité observation : La demande de classification porte uniquement sur un seul attribut. Le système donne seulement la liste de valeurs de cet attribut. Cela signifie, dans un contexte d'information multimedia, que l'utilisateur a un moyen de visualiser les occurrences de ces valeurs.

L'activité de renseignement : L'utilisateur possède quelques idées sur ce qu'il veut mais cette idée est très limitée, portant sur des contraintes par exemple. Ici l'utilisateur ne demande pas explicitement la liste des informations par les attributs. Il formule sa demande par une seule contrainte, c'est-à-dire un seul critère de recherche.

L'activité de recherche : L'utilisateur emploie ses connaissances du domaine pour recher-

cher d'autres informations. Ce type d'activité porte sur la définition de plusieurs contraintes.

L'activité de synthèse : Elle concerne la demande de classification portant sur au moins deux attributs avec éventuellement des contraintes. Les techniques d'analyse utilisées sont l'analyse-croisée intra-champ et l'analyse-croisée inter-champ (cf II.5.2. (Analyse de données)).

L'activité d'annotation : Elle permet à l'utilisateur de préciser ses préférences en particulier par l'évaluation des solutions du système.

Nous avons adopté un modèle explicite de l'utilisateur pour représenter explicitement chaque utilisateur dans notre SRI. Chaque activité de l'utilisateur est considérée comme un document. Chaque activité est représentée par des champs. La représentation du modèle de l'utilisateur pour un utilisateur particulier est représentée par :

$$A = R \cup S ,$$

où

$R = \{r_i\}$, r_i la i ème requête, R représentant l'ensemble des requêtes

$S = \{s_i\}$, s_i la i ème solution, S représentant l'ensemble des solutions

$A = L$ ensemble des activités de l'utilisateur

Chaque utilisateur possède une identité. L'utilisation du système pour un besoin particulier est considérée comme une session. L'utilisateur peut employer plusieurs types d'activité dans une seule session : observation, renseignement, recherche, et annotation. La classification des activités en ces catégories est inspirée par les processus employés dans un contexte d'apprentissage humain. L'utilisateur peut créer des sous-sessions pour définir un nouvel objectif. Nous supposons que l'objectif d'une sous-session est logiquement associé à l'objectif principal de la session du point de vue de l'utilisateur même si cela peut apparaître incohérent par un observateur. Ce phénomène est flagrant pendant la navigation sur internet où l'utilisateur peut décider de changer temporairement son objectif avant de revenir à l'objectif principal de sa recherche.

Chaque activité de l'utilisateur est considérée comme un enregistrement d'un document. Une activité est représentée par :

objectif :

type d'activité :

date :

heure :

requête :

Par exemple si l'utilisateur a pour objectif la *recherche des livres sur la modélisation de*

l'utilisateur et qu'il demande au système une *analyse de fréquences des titres*, cette activité sera représentée par :

objectif : *modélisation de l'utilisateur*

type d'activité : *observation*

date : *98/09/02*

heure : *10:30*

requête : *année > 1995*

Dans cet exemple, il y a une requête, qui est en fait une contrainte sur l'analyse qu'il demande au système. On peut remarquer également que cette contrainte n'est pas exprimée dans son objectif.

Il y aura autant d'enregistrements qu'il y aura d'activités. Le choix que nous avons adopté est de représenter les objectifs et les requêtes par des listes de termes. Nous comptons poursuivre nos travaux de recherche afin que l'utilisateur puisse avoir la possibilité de présenter ses objectifs en langage naturel.

Les réponses fournies par le système en réponse aux requêtes sont considérées comme des solutions. Il est nécessaire que l'utilisateur évalue chaque solution proposée par le système. Cette évaluation permet au système d'avoir l'avis de l'utilisateur (le feed-back). En effet, le degré de pertinence d'une réponse dépend beaucoup de la particularité de l'utilisateur. Par exemple, l'utilisateur peut donner les jugements suivants sur l'évaluation de la solution proposé par le système :

- 1) la réponse est pertinente ;
- 2) la réponse n'est pas pertinente parce que l'utilisateur possède déjà la solution (*ce jugement peut être interprété comme une connaissance du domaine par l'utilisateur*) ;
- 3) la réponse n'est pas pertinente parce qu'elle ne correspond pas à la requête et ce jugement est correct (*ce jugement peut être interprété comme du bruit dans la solution et comme une connaissance du domaine par l'utilisateur*) ;
- 4) la solution n'est pas pertinente parce que la réponse ne correspond pas à la requête, mais ce jugement est erroné (*ce jugement peut être interprété comme un manque de connaissance du domaine par l'utilisateur*) ;
- 5) la réponse n'est pas pertinente mais l'utilisateur ne sait pas pourquoi.

L'information fournie par ce feed-back nous permet de faire certaines inférences sur le niveau de connaissance de l'utilisateur dans le domaine. Cette connaissance peut être exploitée sous forme d'heuristiques pour reformuler les requêtes de l'utilisateur et pour adapter les termes à utiliser pour présenter les solutions.

Chaque solution est présentée comme

solution :
objectif :
évaluation utilisateur :
date :

Les champs «solution» et «objectif» sont utilisés pour stocker les termes utilisés dans les solutions et dans les objectifs correspondants. Le champ *évaluation utilisateur* est l'une des valeurs de jugement ci-dessus.

Par exemple, si le système donne le résultat suivant en réponse à l'activité de l'utilisateur :

Fréquence	Titre
4	La modélisation de l'utilisateur
2	Modéliser l'utilisateur dans un SRI
2	Aspects cognitifs dans la modélisation de l'utilisateur
1	Une approche de modélisation de l'utilisateur pour un SRI

Tableau 5 : Résultat d'une demande d'analyse de fréquences

Evaluation	Titre
1	La modélisation de l'utilisateur
1	Modéliser l'utilisateur dans un SRI
3	Aspects cognitifs dans la modélisation de l'utilisateur
3	Une approche de modélisation de l'utilisateur pour un SRI

L'évaluation de l'utilisateur sur le premier titre sera représentée par :

solution : La modélisation de l'utilisateur
objectif : modélisation de l'utilisateur
évaluation utilisateur : 1
date : 98/09/02

VI.5 Acquisition de connaissance sur l'utilisateur

Notre approche pour l'acquisition des connaissances sur chaque utilisateur est basée sur l'observation de ses comportements. Nous proposons d'observer les types d'activités ainsi que les termes utilisés pour exprimer ces besoins. Ces termes porteront aussi bien sur les objectifs de l'utilisateur que sur les termes utilisés dans les différentes activités.

Nous proposons d'intégrer un «moniteur» de l'utilisateur qui aura pour rôle la mise à jour du modèle de l'utilisateur.

VI.6 Analyse des connaissances sur un utilisateur

Comme nous l'avons présenté en section VI.4 (*Notre modèle de l'utilisateur*), les activités de l'utilisateur et ses évaluations des solutions proposées par le système sont considérées comme des documents. Cette approche nous permet d'appliquer les méthodes utilisées pour l'analyse de l'information sur les connaissances acquises sur l'utilisateur. Nous présentons dans les trois sections suivantes les types d'analyse que nous comptons faire et des réflexions sur les manières d'exploiter les informations à valeur-ajoutée résultant de ces analyses. Les tableaux *Tableau 6 : (Interprétation des évaluations de l'utilisateur)*, *Tableau 7 : (Interprétation des activités de l'utilisateur)* et *Tableau 8 : (Interprétation des activités de l'utilisateur (suite))* présentent les deux catégories d'analyse que nous comptons mener : l'analyse des évaluations de l'utilisateur et l'analyse des ses activités.

Dans le *Tableau 6 : (Interprétation des évaluations de l'utilisateur)*, les colonnes représentent les jugements d'évaluation des solutions proposées par le système. La ligne représente le terme qui a permis de trouver la solution évaluée. Les cellules correspondent à la façon dont nous comptons interpréter les jugements d'évaluation de l'utilisateur. Par exemple,

- *cas 1* : si une solution est acceptée, alors le terme qui a permis de trouver cette solution doit être retenu ;
- *cas 2* : si une solution est rejetée parce que la solution est déjà connue, alors il faut éviter le terme dans le processus de mise en correspondance ;
- *cas 3* : si une solution est rejetée parce que la solution a été introduite suite à une reformulation de la requête, qui a introduit un terme que l'utilisateur n'a pas donné, alors cette solution est considérée comme du bruit et le terme qui a permis de trouver cette solution doit être évité dans le processus de mise en correspondance ;
- *cas 4* : si une solution est rejetée parce que l'utilisateur pense que c'est du bruit, alors il faut lui expliquer le terme qui a permis de trouver la solution ;
- *cas 5* : si une solution est rejetée mais l'utilisateur ne sait pourquoi elle ne lui convient pas, alors il faut lui expliquer le terme qui a permis de trouver la solution ;

solution	terme de la requête
acceptée <1>	Terme à retenir
rejetée (ok) <2>	Termes à éviter dans la recherche
rejetée (ok) <3>	Termes à éviter dans la recherche
rejetée (faux) <4>	Termes à expliquer à l'utilisateur
ne sais pas <5>	Termes à expliquer à l'utilisateur

Tableau 6 : Interprétation des évaluations de l'utilisateur

Dans le *tableau 6* et le *tableau 7*, les colonnes représentent la date d'une activité, les colonnes observation, renseignement, recherche et annotation les différents types d'activité et la colonne terme(j) correspond à un terme utilisé dans l'activité. Les lignes représentent les mêmes informations. Les cellules présentent les types d'information à valeur-ajoutée que nous pouvons obtenir en faisant une analyse inter-champs (par exemple date et observation) ou une analyse intra-champ (par exemple terme et terme).

	date	observation	renseignement
observation	Evolution des activités d'observation		
renseignement	Evolution des connaissances élémentaires acquises		
recherche	Evolution des connaissances appliquées		
annotation	Evolution des préférences		
terme(i)	Evolution des termes employés	Les termes du processus d'acquisition de connaissance	Les termes du processus d'acquisition de connaissance

Tableau 7 : Interprétation des activités de l'utilisateur

	recherche	annotation	terme(j)
terme(i)	Connaissances acquises	Termes des préférences	Co-occurrence des termes dans les activités

Tableau 8 : Interprétation des activités de l'utilisateur (suite)

Les trois sections suivantes présentent, par type d'analyse, comment nous comptons exploiter les résultats des analyses possibles à partir du modèle de l'utilisateur.

VI.6.1. Fréquences occurrences valeurs

Nous proposons d'utiliser la technique d'analyse de fréquence comme présentée dans la section II.5.2. (*Analyse de données*) pour observer les paramètres suivants des activités de l'utilisateur :

- le champ «type d'activité» : la distribution des types d'activité (les quatre que nous proposons) nous fournira une vue globale sur les activités privilégiées par l'utilisateur.

Fréquence	Activité
10	Observation
1	Renseignement
2	Recherche
0	Analyse

Tableau 9 : Exemple d'activités d'observation dominante

Des interprétations possibles du résultat de cette analyse sont :

- ⇒ Si les activités d'observation et de renseignement sont privilégiées, alors l'utilisateur ne possède pas beaucoup de connaissance dans le domaine.
- ⇒ Si nous considérons le SRI comme un système d'aide à la résolution de problème et si l'utilisateur privilégie les activités d'observation et de renseignement, cela peut être interprété comme une méconnaissance de la stratégie pour la résolution du problème.

Fréquence	Activité
10	Observation
8	Renseignement
2	Recherche
1	Analyse

Tableau 10 : Exemple d'activités d'observation et de renseignement dominants

- ⇒ Privilégier les activités de recherche et d'annotation comme nous les avons définies, peut être interprété comme un utilisateur qui possédant une bonne connaissance du domaine.

Fréquence	Activité
2	Observation
2	Renseignement
8	Recherche
1	Analyse
5	Annotation

Tableau 11 : Exemple d'activités de recherche et d'annotation dominantes

- ⇒ Si nous nous plaçons dans une situation de résolution de problème, privilégier les activités de recherche et d'annotation ne signifie pas forcément une maîtrise de la stratégie nécessaire pour résoudre le problème. Il pourrait s'agir d'une stratégie «d'essai et erreur», pour explorer les différentes possibilités pour résoudre le problème. Il faut en effet contrôler ce résultat avec le résultat d'une autre type d'analyse, par exemple le résultat de l'analyse des évaluations des solutions proposées par le système.

- le champ «requête» : la distribution des termes des requêtes nous fournira les termes les plus utilisés.

Des interprétations possibles du résultat de cette analyse sont :

- ⇒ Les fréquences des termes utilisés dans les requêtes peuvent nous fournir des indications sur les domaines d'intérêt de l'utilisateur. Par exemple, les termes les plus utilisés peuvent appartenir à un seul champ sémantique.

⇒ Également, les termes les plus utilisés peuvent nous donner des indications sur le niveau de connaissance de l'utilisateur dans le domaine. Par exemple, en évaluant les «niveaux de catégories fondamentales» des termes utilisés, on peut déduire le niveau de connaissances de l'utilisateur (cf VI.3 (*La théorie de catégories fondamentales*)).

- le champ «objectif» : comme pour le champ requête, la distribution des termes des objectifs nous fournira les termes les plus utilisés pour formuler les besoins. Les interprétations sont les mêmes que pour le champ requête.

VI.6.2. Fréquences co-occurrences intra-champ

Rappelons que l'analyse de fréquences co-occurrence intra-champ concerne l'analyse de co-occurrence de deux ou plusieurs valeurs dans un même champ. Pour illustrer nos propositions, considérons une base documentaire pour gérer des documents du domaine informatique, considérons également la hiérarchie des connaissances en informatique suivante :

Systèmes, **système d'exploitation**, Unix, Solaris

Systèmes, **système d'exploitation**, Unix, Sunos

Systèmes, **système d'exploitation**, Windows, Windows NT

Systèmes, **système d'exploitation**, Windows, Windows 98

Systèmes, **réseau**, réseau local

Systèmes, **réseau**, gestion réseau

Nous considérons que les termes *système d'exploitation* et *réseau* appartiennent à la catégorie fondamentale.

Nous proposons de faire ce type d'analyse sur les champs suivants :

- le champ «objectif» : l'analyse sur ce champ nous montrera les termes qui sont utilisés le plus souvent ensemble dans la formulation des besoins.

Des interprétations possibles du résultat de cette analyse sont :

⇒ Si seulement l'un des deux termes appartient à la catégorie fondamentale, le terme moins abstrait (ou plus abstrait) peut être interprété comme une précision de la demande. Cela peut signifier également une meilleure connaissance du domaine. Par exemple si les termes *réseau* et *réseau local* apparaissent très souvent dans les objectifs, alors on peut considérer que le terme *réseau local* est utilisé pour préciser la demande, mais également que l'utilisateur possède des connaissances sur le terme *réseau*, en particulier des concepts informatiques liés au terme.

⇒ Si les deux termes appartiennent à la catégorie fondamentale, on peut déduire que l'utilisateur ne possède pas assez de connaissance du domaine. Par exemple si les termes *réseau* et *système d'exploitation* apparaissent très souvent dans l'objectif, on peut considérer que l'utilisateur ne peut pas préciser son objectif. Remarquons en-

core ici que la notion de catégorie fondamentale dépendent du type d'utilisateur auquel on s'intéresse. Les termes réseau et système d'exploitation peut être considérés comme appartenant à la catégorie fondamentale pour les étudiants en informatique mais pas pour le grand public.

- le champ «requête» : le résultat et l'interprétation de l'analyse sur ce champ sont identiques à ceux obtenus sur le champ «objectif».
- le champ «solution» : l'analyse sur ce champ nous montrera les termes qui sont utilisés le plus souvent ensemble dans les solutions acceptées ou rejetées.

Des interprétation possibles du résultat de cette analyse sont :

- ⇒ Si seulement l'un des deux termes appartient à la catégorie fondamentale, le terme moins abstrait (ou plus abstrait) peut être interprété comme une précision de la demande. Cela peut signifier également une meilleure compréhension du domaine contrairement à une interprétation comme une «meilleure connaissance» dans une requête.
- ⇒ Si les deux termes appartiennent à la catégorie fondamentale, on peut déduire que l'utilisateur ne comprend toujours pas les différents niveaux d'abstractions des termes du domaine.

VI.6.3. Fréquences co-occurrences inter-champs

Rappelons que l'analyse de fréquences de co-occurrences inter-champs concerne l'analyse de co-occurrence de deux ou plusieurs valeurs dans deux (ou plus) champs différents. Nous proposons de faire ce type d'analyse sur les paires de champs suivants :

- objectif - requête : l'analyse sur cete paire de champs montrera les termes qui sont utilisés pour exprimer des besoins.
- objectif - solution : l'analyse sur cette paire de champs montrera les termes des solutions associés aux besoins.
- solution - évaluation utilisateur : l'analyse sur cette paire de champs utilisé avec la paire objectif- solution montrera les termes associés aux besoins par l'utilisateur.
- requête - solution : dans un cas de reformulation des requêtes, cette analyse montrera l'adéquation ou l'inadéquation des termes introduits dans les requêtes à la suite des reformulations.
- objectif - date : l'analyse sur cette paire de champs montrera l'insistance de l'utilisateur sur l'utilisation d'un terme pour formuler des besoins. Une grande fréquence peut être interprétée comme une non-satisfaction du besoin de l'utilisateur.
- requête - date : l'analyse et l'interprétation sur cette paire de champs sont identiques à la paire objectif-date.
- solution - date : l'analyse sur cette paire de champs montrera le degré de pertinence d'un terme de la solution. Néanmoins, une grande fréquence peut être interprétée comme un

manque de compréhension d'objectif principal par l'utilisateur.

VI.7 Conclusion

Nous avons justifié la nécessité d'une représentation explicite de l'utilisateur par un modèle de l'utilisateur. Nos travaux antérieurs dans ce domaine ont été présentés également : le modèle de l'élève et la représentation de connaissance basée sur la théorie de catégories fondamentales. Nous avons montré la continuité de ces travaux pour aboutir au modèle de l'utilisateur.

Des choix faits ont été présentés. Par exemple le choix du mode de représentation de l'utilisateur par ses activités et en considérant chaque activité comme un document. Nous avons également présenté le choix des types d'analyse où il s'agit essentiellement de la méthode d'analyse que nous avons définie : la classification avec contraintes. Nous avons présenté aussi, avec des exemples, les types d'analyse possibles à partir de la représentation de l'utilisateur.

Je poursuis ces travaux de recherche pour

- étudier d'autres méthodes d'analyse,
- intégrer l'interprétation automatique des analyses sur les activités de l'utilisateur pour la reformulation des requêtes,
- intégrer l'interprétation automatique des analyses sur les activités de l'utilisateur pour adapter les aides qu'un expert peut donner à l'utilisateur,
- étudier l'apport du langage naturel dans la formulation des objectifs et des requêtes.

L'état de mes réflexions sur ces perspectives sont présentées dans les sections suivantes. Ces travaux sont basés sur l'analyse de l'historique des sessions, à la fois l'historiques des sessions d'un utilisateur particulier et celle de l'ensemble des utilisateurs pour un certain type de traitement. Nous comptons obtenir une analyse de l'historique sous deux formes :

Forme générale

(Objectif, Solution, Evaluation, Fréquence)

- ⇒ Objectif représente l'un des objectifs formulés par les utilisateurs,
- ⇒ Solution représente l'une des solutions proposées par le système pour l'objectif,
- ⇒ Evaluation représente l'un des cinq types de jugement de l'utilisateur sur une solution du système,
- ⇒ Fréquence représente le pourcentage de Fréquence d'un jugement prononcé pour une solution concernant l'objectif, par rapport au nombre total de solutions pour le

même objectif et pour le même jugement.

Pour donner des réponses pertinentes à l'utilisateur, nous allons calculer le degré de similarité (T) entre l'objectif actuel de l'utilisateur et les objectifs stockés dans l'historique de toutes les sessions. Nous allons choisir le meilleur critère de jugement, par exemple le jugement du type 1 (réponse pertinente). Nous utilisons Fréquence pour déterminer le seuil de tolérance (N) qui indique un seuil à partir duquel la solution peut être acceptable.

Forme particulière

(Utilisateur, Objectif, Solution, Evaluation, Fréquence)

Cette forme est similaire à la forme générale sauf que le paramètre Utilisateur est intégré. Ce paramètre restreint l'analyse à l'historique des sessions de l'utilisateur. Dans ce cas, T et N sont relatifs à l'historique de l'utilisateur.

VI.7.1. Implémentation à l'aide de stéréotypes

Nous comptons exploiter les deux formes d'analyse ci-dessus comme des stéréotypes. Si nous ne disposons pas de donnée sur un utilisateur, nous utiliserons la forme générale de l'analyse. Si nous disposons de données sur l'utilisateur, selon les valeurs attribuées à T et à N, nous pouvons utiliser soit uniquement la forme particulière soit encore la forme générale. Nous pouvons utiliser la forme générale par exemple si la valeur de T ne permet pas de trouver un objectif dans l'historique de l'utilisateur. Si les valeurs de T et de N sont fixées par le système et cela d'une façon statique, ce type d'analyse est identique à ce qui est utilisé dans l'implémentation des stéréotypes. Dans certains articles [ric83, jam92], ce type d'implémentation est considéré comme un modèle dynamique. Dynamique dans le sens qu'en absence de donnée sur l'utilisateur, ou en cas de données insuffisantes pour un utilisateur, on se base sur les historiques des autres.

VI.7.2. Implémentation dynamique

Nous comptons donner à l'utilisateur la possibilité d'attribuer à T et à N les valeurs qui lui conviennent. Nous comptons également lui donner la possibilité de les changer en cours d'une session. L'utilisateur pourra déterminer quelle forme d'analyse il souhaite utiliser : la forme générale ou la forme spécifique.

Dans ce type d'implémentation, tout est à l'initiative de l'utilisateur. Nous ne sommes plus dans le cas de stéréotypes car tout dépendra de l'utilisateur et de son initiative à des instants différents.

VI.7.3. L'historique - une base d'information à part entière

Nous avons remarqué que l'historique peut être utilisé comme un accélérateur du processus de recherche. Cette approche ressemble à l'approche qui consiste à utiliser les FAQ (Frequently Asked Question). Très souvent, en particulier pour ceux qui développent des

applications, quand un problème surgit, on recherche dans les FAQ pour savoir si un autre développeur a déjà rencontré ce même type de problème. Si personne n'a jamais rencontré ce type de problème, alors on peut le soumettre à la communauté de développeurs ou à l'équipe d'assistance technique pour solliciter des solutions.

Dans le cas d'un SRI, l'historique peut être utilisé d'abord à l'instar du développeur pour savoir si quelqu'un a déjà utilisé le système pour le même type d'objectif. S'il n'y a jamais eu d'objectif similaire, alors l'utilisateur peut utiliser le SRI avec toutes les fonctionnalités offertes. Si l'utilisateur n'obtient toujours pas de solution satisfaisante, il pourra contacter un expert par le processus de recherche coopérative.

Cette possibilité de recherche coopérative que nous avons conçue est présentée dans le chapitre suivant. Dans ce chapitre présentons le pourquoi d'une recherche coopérative, l'architecture que nous avons développée et la suite de notre travail dans ce domaine.

VI.8 Le modèle de l'utilisateur dans un contexte de recherche coopérative

Comme présenté dans le *Chapitre III (La modélisation dynamique de l'apprenant : le projet BIRDS)*, nous disposons d'outils d'analyse des comportements des utilisateurs. Dans le contexte d'une recherche coopérative, notre objectif est de proposer à un médiateur (un expert par exemple), une synthèse d'information sur les activités de l'utilisateur pour que le médiateur puisse avoir une idée plus précise sur ce que l'utilisateur a déjà fait, de son niveau de connaissance et ainsi faciliter la définition de la meilleure manière de l'assister. Dans ce cas, l'expert et l'utilisateur peuvent utiliser le SRI en mode coopératif.

La méthode de synthèse par l'analyse de fréquences de co-occurrence intra-champ ou inter-champs peut donner au médiateur les observations que nous présentons dans les tableaux *Tableau 12 : (Synthèse des activités de l'utilisateur)* et *Tableau 13 : (Synthèse des activités de l'utilisateur (suite))*

Type activité	date
observation	Evolution des activités d'observation
renseignement	Evolution des connaissances élémentaires acquises
recherche	Evolution des connaissances appliquées
annotation	Evolution des préférences
terme	Evolution des termes employés

Tableau 12 : Synthèse des activités de l'utilisateur

Le *Tableau 12 : (Synthèse des activités de l'utilisateur)* montre qu'en analysant les activités d'observation de l'utilisateur par rapport aux dates nous obtenons l'évolution de l'utilisateur par rapport aux types d'activité et aux termes qu'il emploie. Les lignes de *renseignement*, *recherche* et *annotation* représentent les types d'activité. La ligne de *terme* représente les termes que l'utilisateur emploie.

Type activité	terme(j)
observation	Les termes du processus d'acquisition de connaissance
renseignement	Les termes du processus d'acquisition de connaissance
recherche	Connaissances acquises
annotation	Termes des préférences
terme(j)	Co-occurrence des termes dans les activités

Tableau 13 : Synthèse des activités de l'utilisateur (suite)

Dans le *Tableau 13 : (Synthèse des activités de l'utilisateur (suite))*, les lignes *observation*, *renseignement*, *recherche* et *annotation* nous permettent d'observer les termes utilisés dans les différents types d'activité. La ligne *terme* nous permet d'observer la co-occurrence des termes dans les différentes activités.

L'évolution des termes employés par l'utilisateur pour formuler ses requêtes peut être exploitée pour connaître son niveau de connaissance en utilisant la théorie de catégories fondamentales (cf. section VI.3 (*La théorie de catégories fondamentales*)). Prenons les deux listes suivantes, du concept animal :

animal, mammifère, chien, caniche

animal, mammifère, être humain, femme, Julie

En se basant sur la théorie de catégorie fondamentales, si l'on demande à une personne de regarder par la fenêtre pour nous décrire ce qu'il voit, la majorité de personne dira «je vois une femme avec un chien» et rarement «je vois Julie avec son caniche». Les termes «chien» et «femme» appartiennent à la catégorie fondamentale dans ce domaine.

Il faut remarquer que les termes utilisés par une personne en situation de dialogue dépend d'une part de sa connaissance dans le domaine et d'autre part de sa connaissance sur son interlocuteur. Par exemple la deuxième réponse ci-dessus peut être donnée si les interlocuteurs connaissent bien Julie et le fait qu'elle possède un caniche.

VI.9 Conclusion

Nous avons présenté le schéma global de l'architecture d'un SRCI que nous proposons. Nous avons présenté certains choix que nous avons faits dans le développement des applications pour réaliser les spécifications de l'architecture. L'architecture est en fait assez générale pour permettre la réalisation de communication entre deux applications complètement différentes. Par exemple dans le contexte du projet LEONARDO, nous avons utilisé l'architecture pour développer une interaction entre des prototypes développés : le prototype STREEMS et le prototype GIS. Alors que le prototype STREEMS gère des informations botaniques sur les arbres, le prototype GIS gère des informations géographiques sur des régions. L'architecture nous a permis de réaliser une interface entre les deux système pour, par exemple, obtenir des informations géographiques nécessaires pour le développement de certains arbres.

Le résultat de la synthèse d'information sur l'utilisateur peut être utilisé pour définir des heuristiques en vue d'automatiser l'interprétation des connaissances sur l'utilisateur. Par exemple :

heuristique 1 : si les activités de l'utilisateur sont constituées principalement d'observation, c'est-à-dire les activités d'observation des images ou de simple parcours des objets de la base, l'utilisateur peut être considéré comme quelqu'un de peu de connaissance du domaine.

heuristique 2 : si les bruits introduits dans les solutions suite à la reformulation des requêtes par le système sont correctement identifiés (le rejet des réponses par l'utilisateur est valide), il ne faut plus reformuler la requête de l'utilisateur.

Après avoir présenté nos réalisations et nos modèles dans ce chapitre et les deux derniers, nous présentons dans le chapitre suivant les représentations formelles de ces objets. Il s'agit de leur représentation en formalisme OMT (Object Modelling Technique).

Chapitre VII

Représentation formelle de nos modèles et architectures

Nous présentons dans ce chapitre deux principaux objets de nos réalisations : le modèle de l'utilisateur et de session, et l'architecture d'un SRCI.

VII.1 Le modèle de session

Rappelons que notre objectif est de représenter explicitement chaque utilisateur par ses activités. Les activités de l'utilisateur sont basées sur la notion de session.

Le modèle de l'utilisateur ainsi que le modèle de session que nous proposons est représenté par le modèle objet (formalisme OMT [rum97]) de la *Figure 17. (Modèle objet du modèle de l'utilisateur et de session)*.

Une application (oApplication) est utilisée par un utilisateur (oUser).

Une session d'utilisation sur une application par un utilisateur est composée d'une session principale (oSessionP) qui peut être composée à son tour par des sous-sessions (oSubSession). Puisqu'un utilisateur peut utiliser l'application plusieurs fois, plusieurs sessions principales sont associées à l'utilisateur. Les classes oSessionP et oSubSession sont dérivées de la classe oSession. Les sous-sessions sont employées pour des digressions momentanées de l'objectif principal ou pour des sous-objectifs de l'objectif principal. Ces sous-objectifs sont censés suivre une évolution cognitive de l'utilisateur. Les liens des sous-sessions avec la session principale ne sont pas forcément évidents pour un observateur de l'historique des activités de l'utilisateur.

Une session (la session principale ou les sous-sessions) d'utilisation de l'application est composé d'un objectif et de sa date, ainsi définis dans la classe oSession.

Pour chaque session, c'est-à-dire pour chaque objectif, l'utilisateur emploiera des activités (oActivity) fournies dans l'application pour parvenir à son objectif. Les activités définies actuellement sont classées en quatre catégories en nous basant sur des études en sciences cognitives relatives aux processus cognitifs pour la résolution de problèmes. Ces catégories sont *observation* : *oObservation*, *demande de renseignement* (ou acquisition de vocabulaire) : *oInformation*, *recherche d'information* (ou application de connaissance) : *oResearch*, *synthèse d'information* (*oSynthesis*) et *annotation* (ou créativité) : *oAnnotation*.

La technique de base utilisée pour réaliser les activités (à l'exception de l'activité d'anno-

tions) concerne le concept d'*analyse-croisée avec contrainte* (cf II.5 (Conclusion et propositions pour la recherche d'information)). L'analyse-croisée est en fait une forme de classification.

Une classification (oClassification) porte sur un, deux ou trois attributs des informations recherchées. Il est possible d'intégrer des contraintes à une demande de classification, où une contrainte porte sur un triplet *attribut, opérateur, valeur*. La composition d'une demande de classification détermine le type d'activité de l'utilisateur.

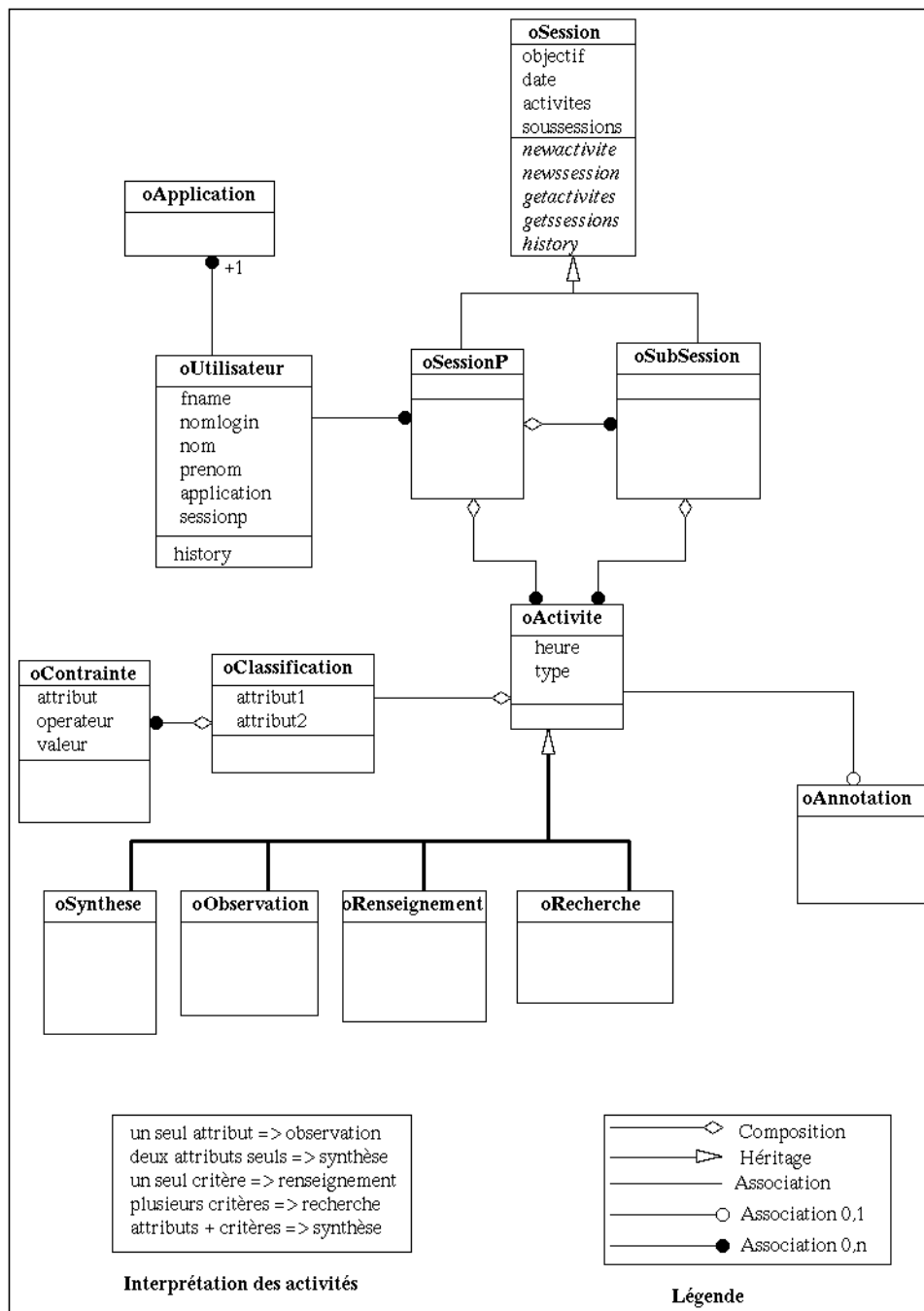


Figure 17. Modèle objet du modèle de l'utilisateur et de session

VII.2 Représentation des activités de l'utilisateur

Nous avons indiqué que les activités de l'utilisateur sont considérées comme des documents, ou plus précisément des enregistrements de documents. Nous disposons de plusieurs modèles de représentation : modèle objet, modèle relationnel ou modèle SGML. Il faut noter que SGML peut être utilisé pour la représentation d'un modèle de données en particulier le modèles hiérarchiques, mais il est avant tout une norme de codification. Nous avons décidé de représenter les activités de l'utilisateur au format SGML parce que la réalisation informatique du modèle SGML peut se faire avec un modèle objet. Les systèmes METIORE et STREEMS sont réalisés avec un outil de développement orienté objet. Donc représenter les données en SGML se prête à l'application.

L'exemple ci-dessous représente une partie d'une session de recherche d'information. Rappelons qu'une session de recherche est déterminée par un objectif, de sous-objectifs et des activités de recherche. Dans cet exemple, on voit les types de classification employés dans les activités.

```
<user>
<nomlogin> adavid </nomlogin>
<sessionp>
  <objectif> Des arbres que l'on peut planter dans la ville de sluch </objectif>
  <activité>
    <heure> 10 :30 </heure>
    <type> observation </type>
    <classification>
      <attribut1> noms scientifique </attribut1>
    </classification>
  </activité>
  <activité>
    <heure> 10 :50 </heure>
    <type> renseignement </type>
    <classification>
      <contrainte>
        <attribut> Altitude maximum >= 1000 </attribut>
      </contrainte>
    </classification>
  </activité>
  <activité>
    <heure> 10 :55 </heure>
    <type> recherche </type>
    <classification>
      <contrainte>
        <attribut> Hauteur maximum </attribut>
        <operator> <= </operator>
        <value> 20</value>
      </contrainte>
    </classification>
  </activité>
  <activité>
    <heure> </heure>
    <type> </type>
    <classification>
      <attribut1> Altitude maximum</attribut1>
      <attribut2> Hauteur maximum</attribut1>
    </classification>
  </activité>
</sessionp>
```

```

    <heure> </heure>
    <type> </type>
    <classification>
        <attribut1> Altitude maximum</attribut1>
        <attribut2> Hauteur maximum</attribut1>
        <contrainte>
            <attribut> Altitude maximum</attribut>
            <operator> >=</operator>
            <value> 1000</value>
        </contrainte>
        <contrainte>
            <attribut>Hauteur maximum </attribut>
            <operator> <=</operator>
            <value> 20</value>
        </contrainte>
    </classification>
</activité>
</sessionp>
</user>

```

VII.3 L'architecture d'un système de recherche coopérative d'information

La réalisation de l'architecture d'un SRCI est basée sur l'utilisation de socket. Les trois principales classes sont : le processus, l'application et le serveur de localisation

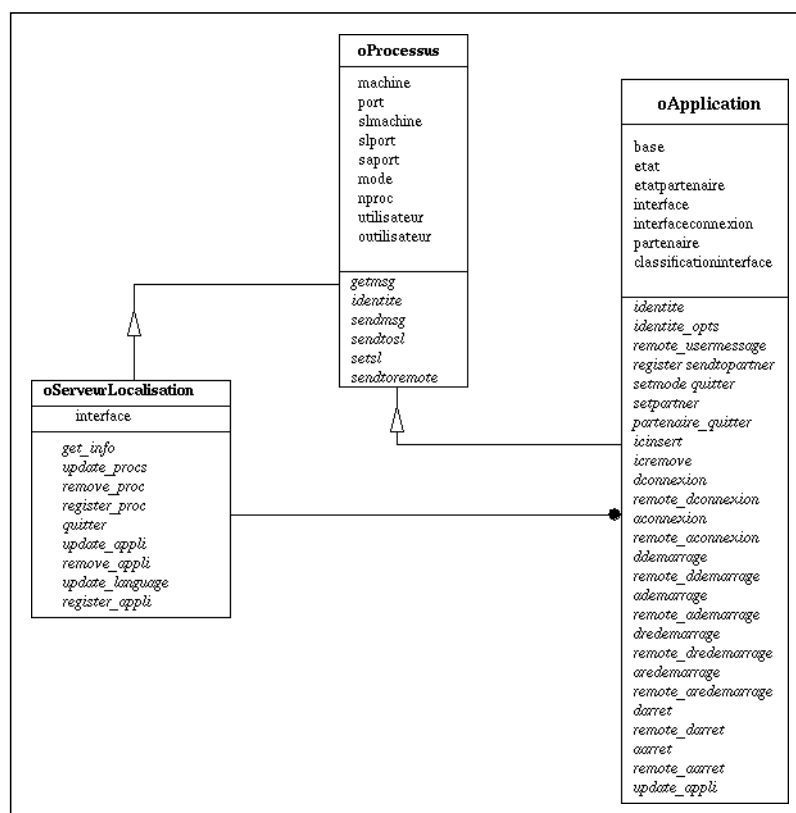


Figure 18. Architecture d'un SRCI en formalisme OMT

La classe processus (oProcessus dans le schéma) dispose des outils de communication in-

ter-application. Un processus peut envoyer et recevoir des message.

Une application hérite de la classe processus. Elle dispose de la base de données et il est associé au serveur de localisation. L'association entre l'application et le serveur de localisation est matérialisée quand l'application est en mode d'observation ou en mode de coopération.

Le serveur de localisation hérite de la classe processus. Il dispose donc des outils pour communiquer avec toutes les applications enregistrées et il peut envoyer des messages aux applications.

Chapitre VIII

Perspectives et projet de recherche

VIII.1 Mes parcours en recherches scientifiques

J'ai démarré mes études de recherche pendant mon DEA informatique en 1986 sur le sujet «Application d'EXPRIM en EAO basé sur l'image». Dans cette étude, nous avons mis en évidence les avantages que peuvent apporter les concepts des systèmes experts, le système de recherche d'information, l'ordinateur, les images et le support de stockage qui était le vidéodisque. Nous avons fait des propositions qui serait possible d'appliquer dans les systèmes d'EAO existants et qui avaient pour support pédagogique les images.

Ce travail a été poursuivi pendant ma thèse de doctorat [dav90a]. Le travail de la thèse a permis :

- de définir une modèle de l'élève qui intègre les aspects cognitifs du comportement de l'élève. Ce modèle a été développé au *Chapitre III (La modélisation dynamique de l'apprenant : le projet BIRDS)*.
- de proposer un mode de représentation permettant d'identifier les niveaux d'abstraction des termes du domaine d'application qui est basé sur la notion de catégories fondamentales
- d'expérimenter le modèle proposé dans une situation réelle pour valider mes propositions. Cette expérimentation a eu lieu au Lycée Henri Poincaré de Nancy dans les classes de 6e et de 5e en 1989.

Après la thèse, je disposais de deux orientations possibles pour la suite de mes travaux de recherche : soit orienter mes recherches vers les systèmes d'EAO, soit recadrer mes recherches vers les systèmes de recherche d'information, plus proche des préoccupations de l'équipe EXPRIM. J'ai décidé d'orienter mes recherches vers les systèmes de recherche d'information avec une redéfinition de certains modèles proposés, en particulier le modèle de l'élève.

En m'inspirant des rôles des documentalistes dans les centres de documentation, les différents types de dialogues entre les utilisateurs et les documentalistes et compte tenu des difficultés à interpréter efficacement les connaissances sur l'utilisateur d'une manière automatique, j'ai élargi mes études vers les systèmes pour la recherche coopérative d'information.

VIII.2 Orientations dans mes recherches

J'ai présenté l'architecture d'un SRCI et son application dans les systèmes METIORE et STREEMS. Je travaille actuellement sur l'intégration du modèle utilisateur dans les systèmes de recherche d'information et sur la présentation tri-dimensionnelle des résultats de la synthèse des connaissances sur l'utilisateur.

Les résultats de cette étude peuvent être exploités dans plusieurs autres domaines (outre l'IST) comme l'apprentissage machine, l'enseignement assisté par ordinateur, domaine de tourisme ou dans des applications de vente ou de marketing. Dans le domaine d'apprentissage machine, les propositions d'un expert peuvent être considérées comme des connaissances expertes pour constituer une base de connaissances experte dans le système. En enseignement assisté par ordinateur, la caractéristique coopérative du système permet à deux étudiants ou à un étudiant avec son professeur de participer à la résolution d'un problème. Dans les applications de tourisme, les préférences d'un client peuvent être analysées pour donner des réponses plus satisfaisantes. Et enfin l'architecture s'adapte parfaitement à un système de recherche d'information. L'aspect le plus difficile sur lequel je travaille actuellement est l'automatisation de l'interprétation des connaissances sur l'utilisateur à l'aide du modèle de l'utilisateur, pour que le système puisse adapter ses réponses aux spécificités de chaque utilisateur.

VIII.2.1. Formalisation de l'interprétation des connaissances sur l'utilisateur

Nous avons adopté la méthode qui consiste à représenter les requêtes de l'utilisateur ainsi que les solutions et les évaluations de ces solutions par l'utilisateur sous forme de documents. Notre hypothèse est qu'en utilisant des techniques d'analyse d'information, nous pourrions obtenir des résultats intéressants sur le comportement, le niveau de connaissance, les besoins et les préférences de l'utilisateur.

Cette approche nous ouvre deux orientations :

- étudier les techniques en IST (bibliométrie, scientométrie) pour analyser les données du système. On pourra analyser aussi bien les données du domaine d'application et les données sur l'utilisateur.
- étudier les techniques de «fouille de données» (data mining, knowledge discovery), technique utilisée en particulier en intelligence artificielle, pour extraire des connaissances sur les particularités de l'utilisateur et les données de l'application.

VIII.2.2. Automatisation de l'interprétation des connaissances sur l'utilisateur

Notre objectif dans les analyses des données, en particulier les données sur l'utilisateur est de découvrir des invariants sur les lesquels nous pourrions nous baser pour interpréter «correctement» les connaissances du système sur l'utilisateur. Nous disposons actuellement d'un seul

moyen d'interprétation, à savoir, l'interprétation que fait un expert des résultats des analyses. Ceci est possible compte tenu du système coopératif que nous proposons. Notre objectif à terme est de pouvoir intégrer un certain nombre d'éléments d'interprétation dans le système d'une manière automatique, pour prendre en compte les spécificités des utilisateurs d'une manière individuelle.

Cet objectif d'étude nécessitera certainement des coopérations avec des psychologues, des cognitivistes et une forte collaboration avec des experts des domaines d'application.

VIII.2.3. Extraction des connaissances d'un expert

A défaut de techniques «fiables» d'interprétation des résultats des analyses sur l'utilisateur, notre système coopératif nous permet d'extraire les interprétations d'un expert dans la situation où l'utilisateur le sollicite pour apporter de l'aide dans son processus de recherche d'information (dans une situation de recherche coopérative d'information). Les apports de l'expert constituent pour nous, des connaissances expertes.

Remarquons qu'il y a deux types d'expertise dans notre contexte de travail : l'expertise dans le domaine d'application du système et l'expertise dans le domaine de la recherche d'information. Les apports des deux types d'expert doivent être étudiés pour faciliter une meilleure connaissance de la façon d'assister l'utilisateur dans son processus de recherche d'information.

VIII.2.4. Exploitation des connaissances des experts

Comme nous l'avons présenté dans la section précédente, les connaissances expertes doivent fournir les façons «idéales» d'interprétation des résultats des analyses des informations sur l'utilisateur et les données de l'application. Notre approche consiste dans un premier temps à considérer les interprétations des experts comme des documents de la même manière que les requêtes et les évaluations des solutions du système par l'utilisateur.

Il faut mener une étude sur les techniques à utiliser pour analyser les connaissances expertes dans le contexte de recherche d'information.

VIII.3 Nos outils dans un système d'ingénierie documentaire

Le besoin de l'utilisateur évolue actuellement vers l'accès aux documents primaires. Il faut donc trouver des moyens pour représenter les documents déjà produits sous forme électronique. Nous pensons que l'utilisateur doit pouvoir trouver des informations tertiaires sur les documents, de les analyser et d'accéder aux documents primaires.

Nous sommes en discussion avec l'équipe READ du LORIA qui travaille sur la retroconversion et la conversion de documents [dav99a]. Par cette collaboration, nous fournirons nos compétences en matière de recherche et d'analyse de l'information et l'équipe nous fournira les compétences en matière de conversion de documents.

VIII.4 Conclusion

Ces perspectives que nous avons énumérées placent en effet l'utilisateur au centre de notre préoccupation, que ce soit par la prise en compte de ses particularités ou par le développement d'un système coopératif de recherche d'information. Nous constatons une évolution dans le rôle attribué à un SRI. En effet un SRI n'a plus seulement pour rôle le traitement d'une série de requêtes. On considère que les utilisateurs cherchent à résoudre des problèmes et que les requêtes sont des expressions ou des démarches pour résoudre ces problèmes.

Dans notre perspective de recherche, nous considérerons les besoins de l'utilisateur comme un problème à résoudre. Des questions auxquelles il faut apporter des réponses avec ce nouveau rôle d'un SRI sont les suivantes :

- Est-ce qu'un SRI doit résoudre d'une manière automatique le problème de l'utilisateur à sa place ?
- Est-ce qu'il faut cultiver chez l'utilisateur sa capacité à définir ses problèmes et à les résoudre ?
- Est-ce qu'il faut observer le comportement de l'utilisateur et lui proposer une stratégie parmi les deux ci-dessus ?

Notre choix est pour la troisième approche. En effet, il n'est pratiquement pas possible de développer un système général de résolution de problème. Les problèmes sont aussi variés qu'il peut y avoir d'utilisateurs. La deuxième approche n'est pas non plus envisageable car un SRI n'est pas un système d'enseignement assisté par ordinateur. L'objectif de l'utilisateur n'est pas forcément de se former. Il faut donc marier l'observation du comportement de l'utilisateur, l'aider à résoudre une partie de son problème d'une manière automatique si possible, et enfin le former à s'approprier certaines compétences pour bien définir et résoudre son problème.

En conclusion, mes acquis dans le domaine de la modélisation de l'utilisateur par une représentation explicite de celui-ci, mes premières approches dans l'intégration des connaissances sur les utilisateurs et les évaluations de mes propositions dans les prototypes me permettent de mieux définir mes axes de recherche. Le deuxième prototype que nous sommes en train de développer nous permettra d'obtenir une meilleure évaluation de mes propositions car plusieurs utilisateurs pourront utiliser le prototype, ce qui nous donnera un échantillon d'utilisateurs représentatif. En outre, mes encadrements, mes initiatives dans la mise en place d'un groupe de recherche et la création d'une option d'IST en DEA nous permettront d'assurer nos objectifs de recherche.

Bibliographie

- [aig91] AIGRAIN P., LONGUEVILLE V. A conception graph for user navigation in large image bank, In Proceedings of RIAO, Volume 1, P 25-44, Barcelona, Spain, Avril 1991..
- [and92] ANDLER D. *Introduction aux sciences cognitives*. Folio. Gallimard, 1992.
- [bin96] BINI J-P. Communication inter-application dans un Système de Recherche d'Information Multimedia, Thèse de CNAM, CAMOS, 1996.
- [fod86] FODOR Jerry A. *La modularité de l'esprit*. Les Edition de Minuit, 1986. Essai sur la psychologie des facultés.
- [cah92] CAHOUR B. How do experts categorize the interlocuteur during consultation dialog ? In *UM92, Third International Workshop on User Modeling*, Aout 1992.
- [chi92] Y. CHIARAMELLA and J.P. CHEVALLET. About Retrieval Models and Logic. *The Computer Journal*, 35(3) :233–242, 1992.
- [chi86] CHIN David N., KNAME : Modeling What the User Knows in UC. In *UM 86 : First International Workshop on User Modeling*, 1986.
- [chin89] CHIN David N., *User Models in dialog systems*. Berlin : Springer, 1989.
- [clu88] CIRY-CLUZEAU M. Typologie des Utilisateurs et des Utilisations d'une banque d'images : Application au projet EXPRIM. *ADBS*, 25, mai-juin 1988.
- [com94] Commissariat Général du Plan. Intelligence économique et stratégique des entreprises. La documentation française, 1994.
- [dav99c] DAVID A. A., BUENO D., Towards cooperative information retrieval system with user modeling, ISAS'99 (5th International conference on information systems analysis and synthesis, July-August, Orlando, USA, 1999.
- [dav99b] DAVID A. A., BUENO D., Personalisation of information based on the concept of relevance using a user model, ISKO-SPAIN'97, pages 419-426, Granada, Spain, March 1999.
- [dav99a] BELAÏD A., DAVID A., The Use of Information Retrieval Tools in Automatic Document Modeling and Recognition, DAUDD'99, September 1999
- [dav97] DAVID A. A. Modélisation de l'utilisateur et Recherche Coopérative dans les Systèmes de Recherche d'Informations, ISKO'97, pages 341-354, Lille, France, October 1997.
- [dav96] DAVID A. A. Vers une recherche coopérative dans les systèmes de recherche d'informations. In ORSTOM, éditeur, *CARI'96*, pages 217– 226, Libreville GABON, Octobre 1996.
- [dav90a] DAVID A. A. *Processus EXPRIM, Image et IA pour un EIIAO individualisé (Enseignement par l'Image Intelligemment Assisté par Ordinateur) : Le prototype BIRDS*. Doctorat INPL, spécialité informatique, INPL (Institut National Polytechnique de Lorraine), CRIN, Nancy, France, Janvier 1990.
- [dav89c] DAVID A. A., Un Environnement d'Apprentissage basé sur l'Image : Notre Prototype BIRDS. In *Open learning en Agriculture*, I.R.F.A.T.A Bologne, Italie, Juin 1989. papier invité.

- [dav89d] DAVID A. A. and THIERY O., Image et Pédagogie. In *A la rencontre des Sciences et des Techniques, quelle culture pour demain?*, Pont-à-Mousson, Octobre 1989. papier invité.
- [dav91b] DAVID A. A., MAIROT N., THIERY O., Student modelling and cognition in education knowledge based tools, PEG'91, International conference on Knowledge Based Environment for Teaching and Learning, Genova, Italie, June, 1991.
- [dav90] DAVID A. A., THIERY O., Taking into consideration the Student's Cognitive Evolution in CAI. In *International Conference on Technology of Education*, Bruxelles, Belgique, Mars 1990.
- [dav89e] DAVID A. A. and THIERY O.. Le Prototype BIRDS : Une expérimentation d'un Modèle de l'Elève dans un Environnement d'Apprentissage basé sur l'Image. In *Journées de travail EIAO-PRC IA*, Paris, Décembre 1989.
- [dav89a] DAVID A. A., THIERY O., and CREHANGE M., Intelligent Hypermedia in Education. *ICCAL*, Mai 1989.
- [dav89b] DAVID A. A., THIERY O., and CREHANGE M., Intelligent Image Based Computer Aided Education (IICAE). In *International conference on «Applications of Artificial Intelligence»*, Orlando-USA, Mars 1989. SPIE et IEEE.
- [den87] DENIS Michel. *Formes imagées de la représentation cognitive*. PhD thesis, Université de Paris VIII, Paris, 1987.
- [den89] DENIS Michel. *Les images mentales*. P.U.F., Paris, 1989.
- [den92] DENIS Michel. *Le cerveau et l'Intelligence*. PhD thesis, Université Henri Poincaré, Nancy I, Décembre 1991. Directeur de Recherche au CNRS, Centre d'études de psychologie cognitive, Université de Paris-Sud.
- [dou95] DOU H., Veille technologique et compétitivité. Dunod, 1995.
- [des92] Sous la direction de H. DESVALS and H. DOU, editors. *La veille technologique*. Dunod, 1992.
- [fal86] Pierre FALZON. Human-Computer Interaction : Lessons from Human-Human communication. In *Third european conference on cognitive ergonomics, ECCE-3*, 1986.
- [fol92] Peter W. FOLTZ and Susan T. DUMAIS. Personalized Information Delivery : An Analysis of Information Filtering methods. *Communications of ACM*, 35(12) :51–60, 1992.
- [gar88] A. DE LA GARANDERIE. *Les Profils Pédagogiques : Discerner les Aptitudes Scolaires*. Le Centurion Collection Paidoguides, 1988.
- [Hal89] HALIN G. Apprentissage pour la recherche interactive et progressive d'images : processus EXPRIM et prototype RIVAGE. - These d'université, Université Henri Poincaré, Nancy 1, France, Octobre 1989.
- [ing92] INGWERSEN P. *Information retrieval interaction*. Taylor Graham, 1992.
- [jak88] F. JAKOBIAK. *Maîtriser l'information critique*. Les Edition d'Organisation, 1988.
- [jak91] F. JAKOBIAK. *Pratique de la veille technologique*. Les éditions d'Organisation, 1991.
- [jak92] F. JAKOBIAK. *Exemples commentés de veille technologique*. Les Edition d'Organisation, 1992.

- [jak95] F. JAKOBIAK. L'information scientifique et technique. Presses Universitaires de France, 1995.
- [jam92] Anthony JAMESON. Generalizing the Double-Stereotype Approach : A Psychological Perspective. In *UM 92 : Third International Workshop on User Modeling*, 1992.
- [kas91] Robert KASS. Building a User Model Implicitly from Cooperative Advisory Dialog. *User modeling and User-Adapted Interaction*, 1 :203–258, 1991.
- [loe92] S. LOEB. Architecturing personalized delivery of multimedia information. *ACM*, 35(12), Dcembre 1992.
- [mai96] Noël MAIROT Conception et réalisation d'un environnement d'EAO, utilisant l'image appliqué à la mécanique de vol - La maquette DEDALE, Thèse de CNAM, 1991.
- [mar95] B. MARTINET. L'intelligence économique. Les Editions d'Organisation, 1995.
- [mar89] B. MARTINET and J.M. RIBAUT. La veille technologique, concurrentielle et commerciale. Les Editions d'Organisation, 1989.
- [mou89] N. MOUADDIB, Gestion des information nuancées : Une proposition de modèle et de méthode pour l'identification nuancée d'un phénomène, Thèse de Doctorat de l'Université de Nancy 1, Mai 1989.
- [nie89] Jianyun NIE. An Information Retrieval Model based on Modal Logic. *Information Processing and Management*, 25(5) :477–491, 1989.
- [par86] Cécile L. PARIS. The use of Explicit User Models in a Generation System for Tailoring Answers to the User's Level of Expertise. In *UM 86 : First International Workshop on User Modeling*, 1986.
- [pia69] J. PIAGET. *Psychologie et Pédagogie*. PUF, 1969.
- [pia74] J. PIAGET. *Adaptation vitale et Psychologie de l'Intelligence*. HERMANN, Paris, 1974.
- [ras91] François RASTIER. *Sémantique et Recherche Cognitives*. Formes sémiotiques. puf, France, 1991.
- [ric79] Elaine RICH. User Modeling via Steretypes. *Cognitive Science*, 3 :329–354, 1979.
- [ric83] Elaine RICH. Users are individuals : individualizing user models. *International Journal Man-Machine Studies*, 18 :199 – 214, 1983.
- [rij86] C. J. Van Rijsbergen. A non-classical logic for information retrieval. *The computer journal*, 29(6) :481–485, 1986.
- [ros93] H. ROSTAING. Veille technologique. Concept - Outils - Applications. PhD thesis, Université Aix-Marseille III, 1993.
- [rul95] RULQUIN V., DAVID A., and THIERY O. Personnalisation des réponses dans un système de recherche d'information. In Actes Journée d'étude sur les systèmes de recherche d'information élaborés, Ile Rousse, Corse, France, 1995.
- [rum97] J. RUMBAUGH and al. La modélisation et conception orientée objet (OMT), Prentice Hall, 1997.
- [sal68] SALTON G. Automatic Information Organisation and Retrieval, McGraw-Hill, 1968
- [sal83] SALTON G, McGill M.J. Introduction to Modern Information Retrieval, McGraw Hill,

1983

- [sal90] SALTON G., BUCKLEY C. Improving Retrieval Performance by Relevance Feedback, *Journal of American Society for Information Science*, 41(4) : 288-297, 1990.
- [thi88] THIERY O., A. A. DAVID, and CREHANGE M., Un Enseignement par l'Image Intelligentement Assisté par Ordinateur (EIIAO) : un exemple concret en ornithologie. In *Premières Journées Internationales Application de l'Intelligence Artificielle à l'Agriculture, à l'Agrochimie et aux Industries Agro-Alimentaires*, Caen, France, 29-30 1988.
- [thib89] THIERY O., A. A. DAVID, and CREHANGE M., A knowledge and Image based system as an Intelligent Educational Tool : Our Prototype BIRDS. In *Third International Conference on Children in the Information Age - Human Development and Emerging Technologies*, Sofia, Bulgaria, 20 - 23 May 1989.
- [thia89] THIERY O., A. A. DAVID, and CREHANGE M., Images and Artificial Intelligence in Computer Aided Education : Our Prototype BIRDS. In *International Conference on Technology of Education*, Orlando-USA, Mars 1989.

Index

A	Abstraction élémentaire	68
	acquisition du langage	40
	activité d'annotation	72
	activité de recherche	71
	activité de renseignement	71
	activité de synthèse	72
	activité observation	71
	activités de l'utilisateur	87
	analyse des co-auteurs	59
	analyse inter-champ	60
	annotation	76, 82
	apprendre	39
	architecture d'un SRCI	49, 88
B	Basic Object Adapter	50
	BBC	41
	BIRDS	39, 41, 46
	British Garden Birds	41
C	catégories fondamentales	69
	catégories perceptuelles fondamentales	69
	COMETT	113
	communication homme-homme	48
	Connaissances acquises	76
	Connaissances conceptuelles	43
	Connaissances factuelles	43
	Consultation	68
	CORBA	50
	créativité	41
	CSA	114
D	DEDALE	46
	Demande connotative	68
	Demande précise	68
	Demande thématique	68
	DESS IST	112
	données sur les utilisateurs	12
E	EAO	90
	émergence des concepts	61
	enseigner	39
	environnement d'apprentissage	39
	erreurs	44
	erreurs d'association	45
	erreurs lexicographiques	44

	évolution des co-auteurs	61
	évolution des concepts	61
	Evolution des informations gérées	20
	Evolution des progrès technologiques	21
	Evolution des SRI	20
	Evolution des utilisateurs	21
	Exploitation des connaissances	92
	EXPRIM	90
	Extraction des connaissances	92
F		
	Feed-back	58
	fonctionnement d'un SRI	11
	FORCE	114
	Fréquences co-occurrences inter-champs	79
	Fréquences co-occurrences intra-champ	78
	Fréquences occurrences	76
	Frequently Asked Question	81
G		
	GIS	83
H		
	habitudes évocatives	14
	heuristiques	73
I		
	Indexation des documents	19
	informations multimedia	39
	ingénierie documentaire	92
	intérêt empirique	41
	intérêt esthétique	41
	intérêt spéculatif	41
	interface de connexion	50
	Interface Repository	50
	IRFATA	113
J		
	jugements d'évaluation	75
	jugements de pertinence	14
L		
	LEONARDO	83, 114
M		
	mesure d'activité de l'élève	42
	METIORE	34, 53, 54, 87
	modèle de l'élève	41, 42
	modèle de l'utilisateur	63, 82
	modèle de session	85
	modélisation de l'utilisateur	14, 36
	Motivation	66
N		
	Niveau d'habitude évocative	45
	niveau élémentaire d'abstraction	41

O	Object Modelling Technique	84
	observation	40, 76, 82
	OMT	84
	ORB	50
P	personnalisation	11, 37
	personnalisation des réponses	38
	Perspectives de recherche	16
	pertinence	11
	préférences des utilisateurs	37
	processus cognitif	39
	processus d'apprentissage	39
R	recherche coopérative	82
	recherche coopérative d'information	15, 48
	recherche d'information	19
	Reformulation	13
	renseignement	76, 82
	requête de l'utilisateur	13
S	serveur d'application	50
	serveur de localisation	50
	session	85
	SGML	55, 87
	SRCI	48
	SRI	10, 12
	stéréotypes	81
	STREEMS	34, 53, 54, 83, 87
	SVX	10
	symbolisation et abstraction	41
	Symbolisation et raisonnement	68
T	technique pédagogique	42
	Termes des préférences	76
	théorie des catégories fondamentales	69
	Traitement de la requête de l'utilisateur	12
	Type activité	82
U	utilisateurs occasionnels	38
	utilisateurs réguliers	38
V	veille technologique	10

Annexes

Annexe I

Le prototype STREEMS

I.1 Objectif de STREEMS

Le prototype STREEMS est développé dans le cadre du projet LEONARDO qui a pour objectif de proposer des outils pour faciliter la réhabilitation des forêts dans les pays méditerranéens. Le prototype STREEMS est l'un des systèmes développés dans ce cadre du projet qui a pour objectif de proposer un système multimedia pour gérer des données botaniques et des données d'exploitation sur les arbres. Les fonctions suivantes doivent être réalisées par le système :

- permettre de retrouver des arbres par des requêtes
- permettre une analyse globale des données du système
- permettre une recherche coopérative d'information (RCI) entre deux utilisateurs
- intégrer un modèle de l'utilisateur pour suivre l'évolution du comportement de l'utilisateur

Dans le même cadre du projet, l'université de Malaga a pour objectif de réaliser un système ATRIS pour gérer des informations géographiques en relations avec des arbres. L'architecture de RCI que nous avons développé a été utilisée pour développer un système coopératif entre les deux prototypes. La coopération entre les deux systèmes permet de rechercher des arbres à partir des informations géographiques et retrouver les informations géographiques nécessaires pour le développement d'un arbre à partir des informations botaniques.

Nous présentons dans les sections suivantes uniquement le prototype STREEMS. Le modèle de données utilisé, l'architecture global du système et quelques images écrans sont présentés dans les sections suivantes.

I.2 Le modèle de données de STREEMS

Voici la description des de gestion du domaine d'application que nous gérons dans le prototype STREEMS. Cette description a été donnée par le CSA Italie, l'un des participants du projet qui se spécialise dans le domaine d'agronomie.

- Un arbre (ou un arbuste) peut être planté dans plusieurs **habitats**.
- Un arbre (ou un arbuste) peut avoir plusieurs (ou aucun) types de **résistance**. Les résis-

tances connues sont *pollution, sol argileux, sol calcaire, pH intense*.

- Un arbre (ou un arbuste) peut avoir une seule **orientation**. Les orientations connues sont «fastigiata», *étendue, ovale, sphérique, pyramidale, arbuste*.
- Un arbre (ou un arbuste) peut pousser dans plusieurs types d'**exposition solaire**. Les types d'exposition solaires connus sont *lumière, mi-ombre, ombre*.
- Un arbre (ou un arbuste) peut avoir plusieurs (ou non) **saisons de floraison**.
- Un arbre (ou un arbuste) peut avoir plusieurs (ou non) **saisons de production de fruit**
- Un arbre (ou un arbuste) peut convenir pour plusieurs **utilités**. Les utilités connues sont *haies, coupes-vent, protection, «pioneer»*.
- Un arbre (ou un arbuste) peut nécessiter un plusieurs types d'**environnement**.
- Un arbre (ou un arbuste) peut convenir pour plusieurs (ou aucun) types de **produit du bois**. Les types de produit de bois connus sont *travaux du bois, industrie du bois, chauffage*.
- Un arbre (ou un arbuste) peut avoir un ou plusieurs types d'**endurance**. Les types d'endurance et sensibilité connus sont *parasite et physiophates*.
- Un arbre (ou un arbuste) peut avoir un ou plusieurs types de **vitesse de croissance**. Les vitesses de croissance connues sont *rapide, moyenne et lente*.
- Un arbre (ou un arbuste) peut avoir une seule **forme de feuille**.
- Un arbre (ou un arbuste) peut avoir une ou plusieurs **caractéristiques spécifiques**. Les caractéristiques spécifiques connues sont *mellifère, attirance fauna, comestible, poisons, épineux, hallier, couvrir*.

Dans la représentation objet, pour les attributs qui sont des structures (à l'exception des listes), nous en créons de nouveaux objets. Par exemples, pour l'attribut «habitat» qui est composé du nom de l'habitat et d'une image de l'habitat, nous créons un objet oHabitat. Par contre l'attribut *résistance* ne constitue pas un nouvel objet car c'est une simple liste.

Tous les attributs qui ont des représentations visuelles ont leurs images associées Par exemple l'attribut *forme de feuille* est défini par *lfname* et *lfimage*.

Dans la représentation objet, les attributs qui commencent par *w* : sont des pointeurs vers les objets correspondants. Ces attributs ne sont pas nécessaires dans une représentation objet «pure» mais nous les indiquons pour le besoin de réalisation.

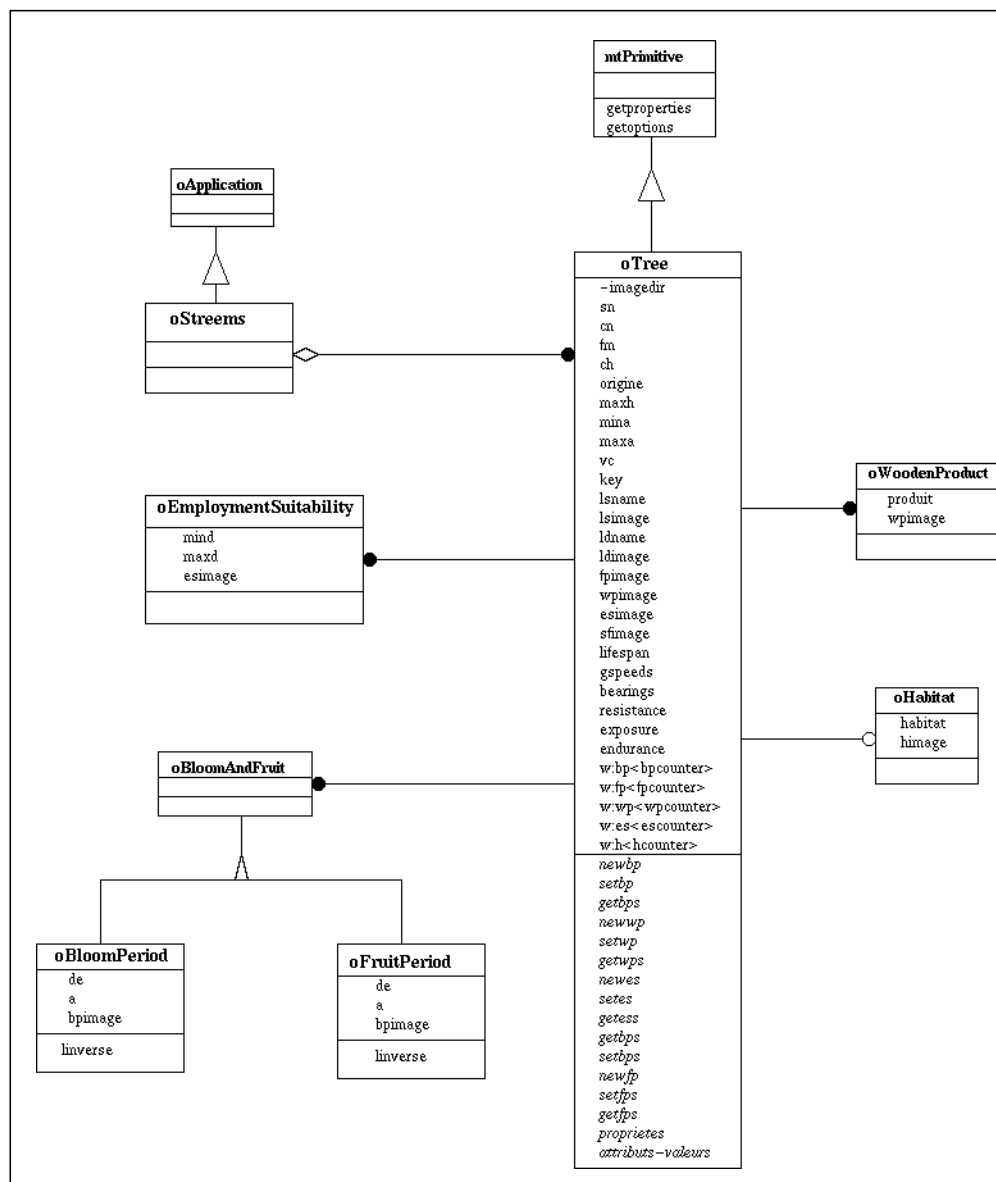


Figure 19. Le modèle de données de STREEMS en représentation OMT

Ce modèle de donnée a été réalisé dans le langage de programmation Tixwish, une extension du langage TclTk. C’est un langage orienté objet. Nous présentons ci-dessous un exemple de la définition d’une classe.


```
# La classe oFruitPeriod
tixWidgetClass oFruitPeriod {
  -classname oFruitPeriod
  -superclass mtPrimitive
  -flag {de a fpimage}
  -configspec {
    {de de De {}}
    {a a A {}}
    {fpimage fpimage Fpimage {} }
  }
  -method {proprietes attributs-valeurs linverse getValueOf}
}

proc oFruitPeriod:linverse {w key} {
  upvar #0 $w data
  set proprietes [$w proprietes]
  # Pour chaque attribut
  foreach e $proprietes {
    global linversefp$e
    # Pour chaque valeur d'un attribut
    foreach b $data($e) {
      # Un array n'autorise pas de blanc comme indice
      regsub -all " " $b "_" b
      # Attacher la clé à la valeur
      # La liste inverse commence par linversebp
      lappend linversefp${e}($b) $key
    } }
}

proc oFruitPeriod:InitWidgetRec {w} {
  upvar #0 $w data
  tixChainMethod $w InitWidgetRec
}

proc oFruitPeriod:ConstructWidget {w} {
  upvar #0 $w data
  tixChainMethod $w ConstructWidget
}

proc oFruitPeriod:SetBindings {w} {
  upvar #0 $w data
  tixChainMethod $w SetBindings
}
```

I.3 Architecture de STREEMS

Le prototype STREEMS est composé de trois principaux modules : le gestionnaire de données représenté par la classe `oTree`, l'interface de connexion pour la recherche coopérative représentée par la classe `mtInterfaceConnexion` et les interfaces de l'application représentées par la classe `mtStreemsInterface`.

L'application elle-même est représentée par la classe `oStreems` qui hérite de la classe `oApplication`. Toute application hérite en effet de la classe `oApplication` qui est elle-même une sous-classe de la classe `oProcessus`. En effet, nous considérons chaque application comme un processus. La classe `oProcessus` dispose de méthodes qui permettent de recevoir et d'envoyer des messages. Chaque processus possède une identité qui permet de l'identifier d'une manière unique. Dans un système réseau internet par exemple, une identité est composé du nom de la machine sur laquelle s'exécute le processus et un numéro de socket. Puisque sur une seule machine un numéro de socket ne peut être affecté qu'une seule fois, nous avons la garantie qu'une identité d'un processus sera unique.

Nous avons défini la superclasse `oProcessus` dont hérite toutes les applications pour leurs donner toutes les capacités nécessaires pour la communication inter-application. De cette sorte, toutes nos applications possèdent des capacités à communiquer avec une autre application. Cette possibilité permet de communiquer non seulement avec une application de même type mais avec toute autre application qui possède la capacité de communiquer avec une autre application. Par exemple, c'est cette capacité qui nous a permis de créer l'interface entre le prototype ATRIS qui gère les données géographiques et STREEMS.

Rappelons que l'architecture que nous proposons pour un SRCI est définie par trois composantes : le serveur de localisation, le serveur d'application et l'application, où l'application est définie par ses interfaces et une interface de connexion. Chaque application s'enregistre au démarrage auprès du serveur de localisation. Après cet enregistrement, le processus (et donc l'application) dispose de l'identité du serveur de l'application ainsi que les identités de toutes les applications déjà enregistrées.

Les interfaces de STREEMS sont également définies par un ensemble de classes, que nous présentons dans la section suivante.

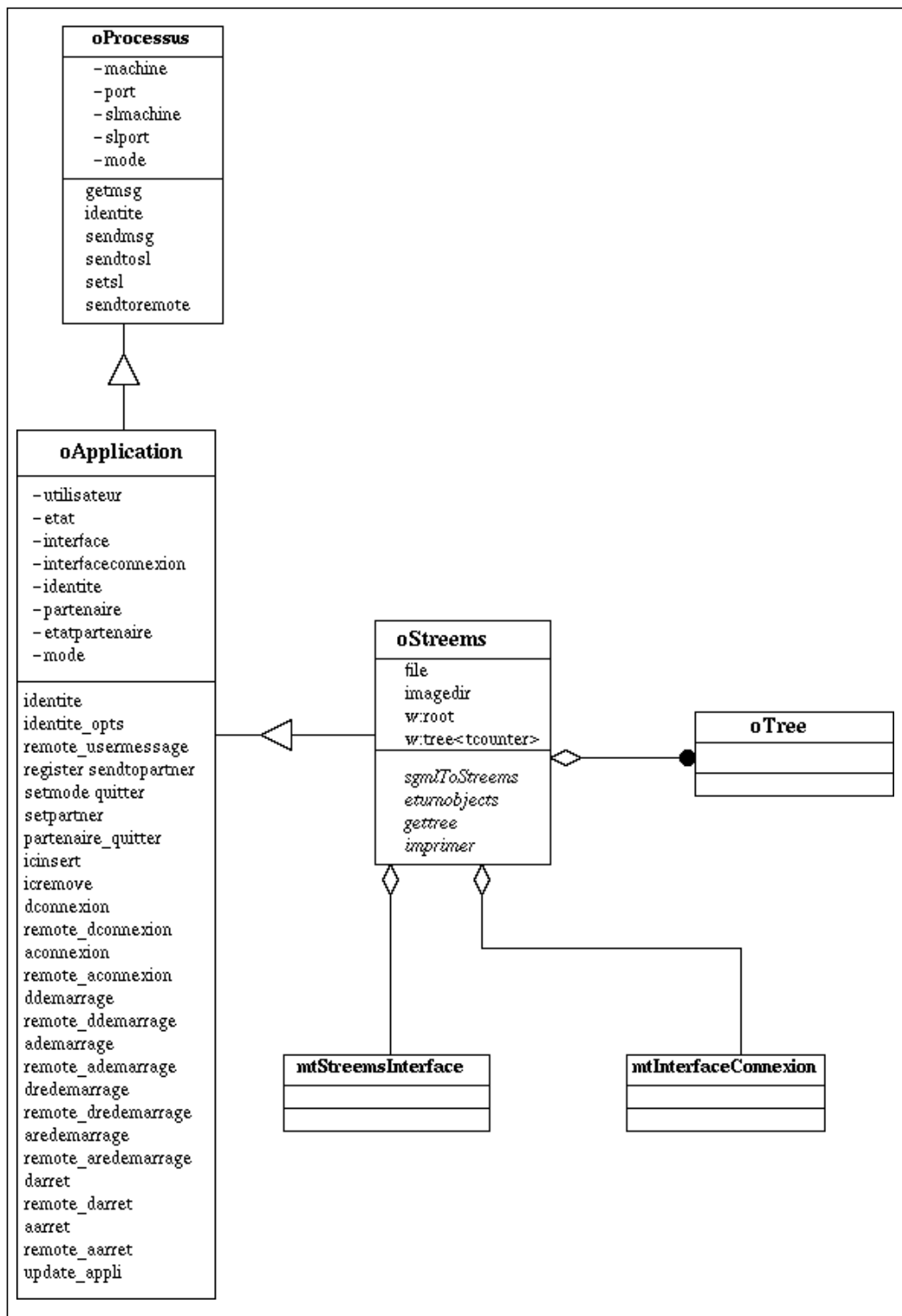


Figure 20. L'architecture de STREEMS en représentation OMT

I.4 Les interfaces de STREEMS

STREEMS est composé d'une interface pour visualiser les arbres et une interface de classification qui permet de mener des activités de recherche et de synthèse d'informations comme présenté dans la section II.5 (*Conclusion et propositions pour la recherche d'information*).

L'interface pour la recherche et de synthèse d'information est représentée par la classe mt-Cluster. Cette classe est définie de sorte à pouvoir être utilisée dans d'autres application où le principe de classification avec contraintes est employé. Par exemple la classe est utilisée pour analyser les activités de l'utilisateur.

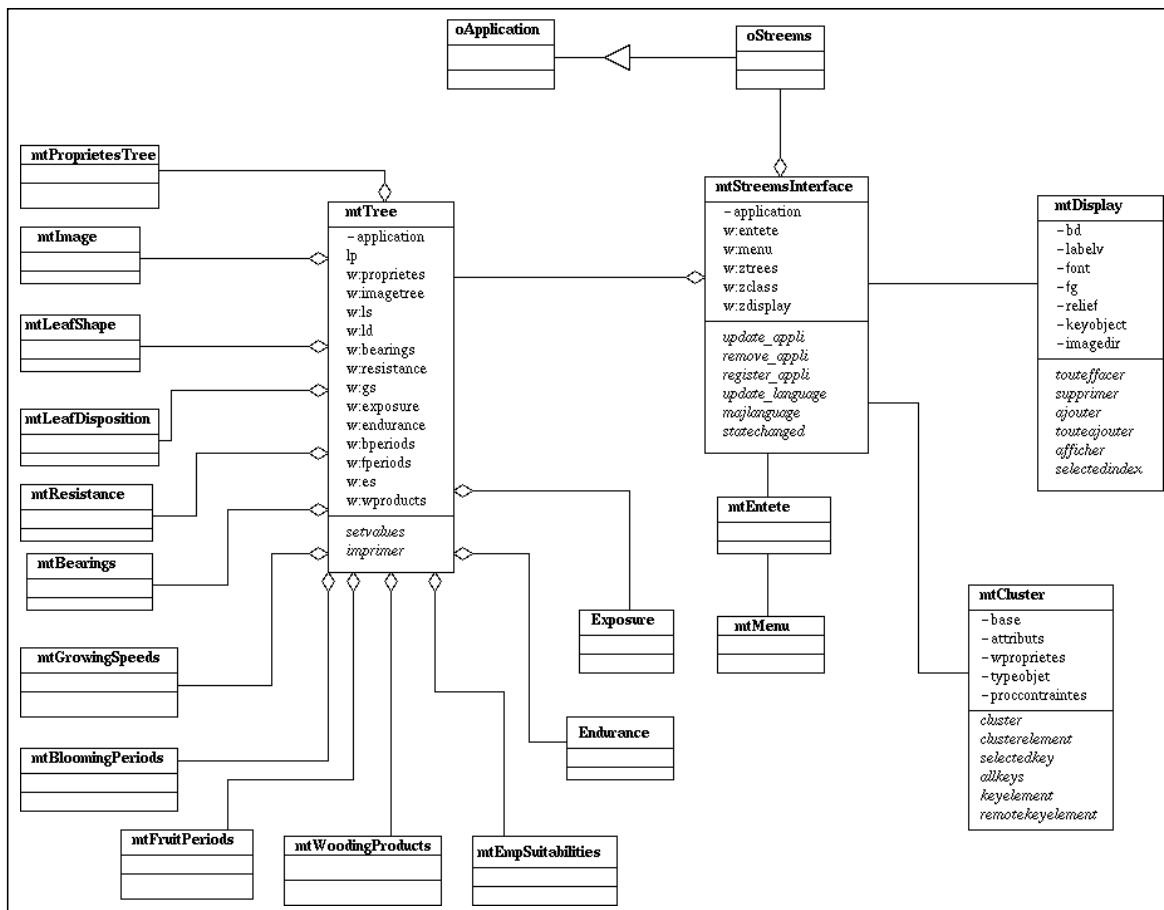


Figure 21. Le modèle de l'interface de STREEMS en représentation OMT

L'interface d'analyse des activités de l'utilisateur est en cours de développement par David Bueno dans le cadre de sa thèse que je co-dirige.

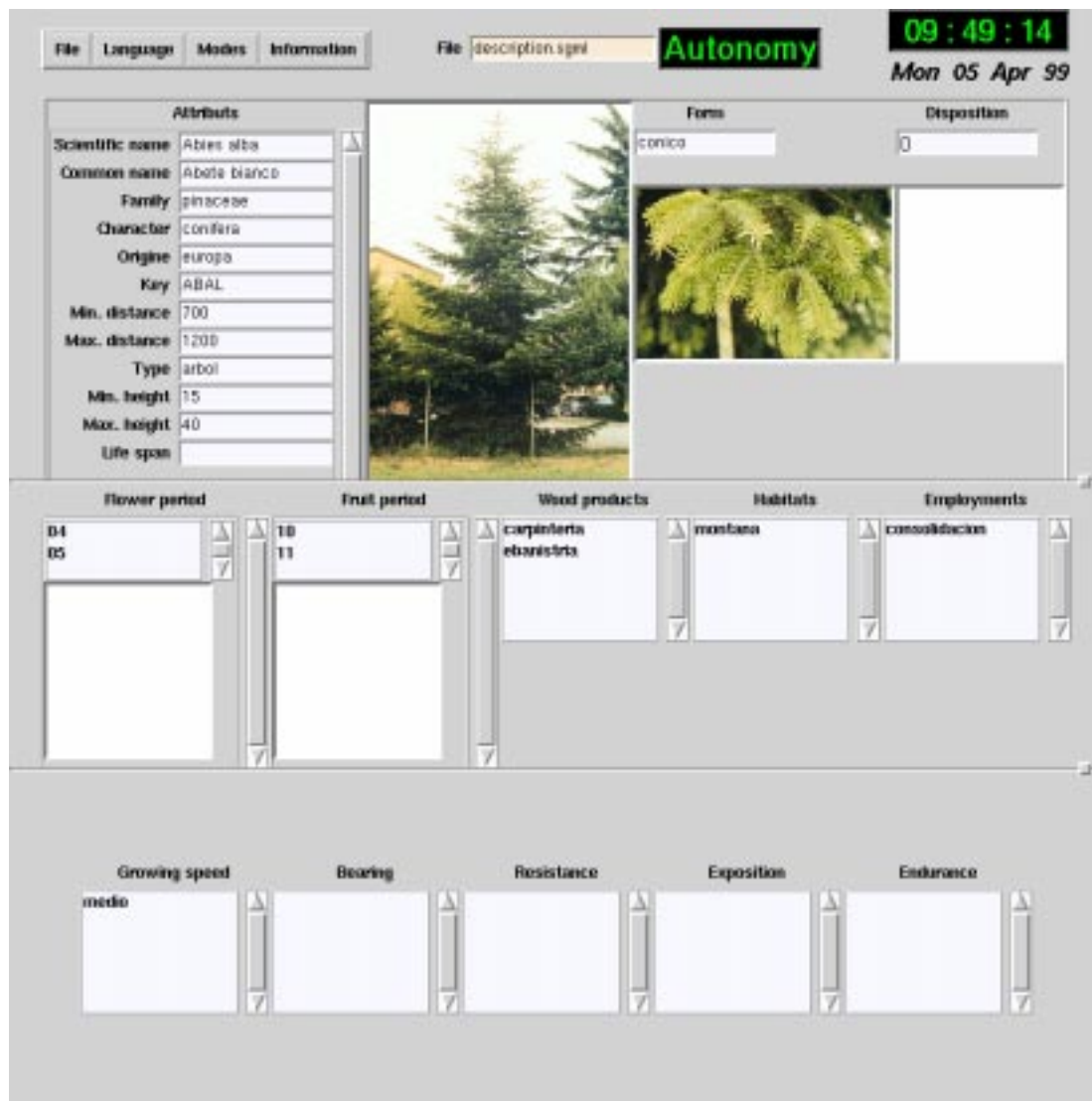


Figure 22. Interface de STREEMS

Le mode de l'interface est également assez ouvert et peut être appliqué dans d'autres domaines d'application. Par exemple, nous sommes en train d'utiliser le modèle de la classe mtStreemsInterface et le modèle de la classe oTree pour gérer des informations bibliographiques. Dans ce système documentaire, le modèle de la classe mtCluster sera utilisé pour analyser les informations bibliographiques ainsi que les informations sur les activités de l'utilisateur.

La figure *Figure 22. (Interface de STREEMS)* présente une image écran de l'interface qui présente les informations sur un arbre.

I.5 Outil de dialogue

Dans un contexte de recherche coopérative d'information, nous avons développé un outil qui permet aux deux utilisateur d'échanger des informations en langage naturel. L'outil est divisé en deux parties : la partie moi et la partie partenaire. Tout ce qu'un utilisateur tape sur le clavier sera automatiquement affiché dans la partie partenaire de l'autre utilisateur, et inversement.

Cet outil est utile pour des dialogues qui portent sur des aspects qui ne sont pas pris en compte dans les autres interfaces du système, par exemple l'explication du fonctionnement du système ou des raisons pour certaines activités.

L'outil est automatiquement affiché quand deux utilisateurs entre en communication et seulement dans ce cas. Rappelons qu'en cas de communication, les deux types de fonctionnement du système à savoir l'observation et la coopération installent cet outil. En cas de situation de coopération, les opérations des deux utilisateurs sont reflétées dans les deux applications.

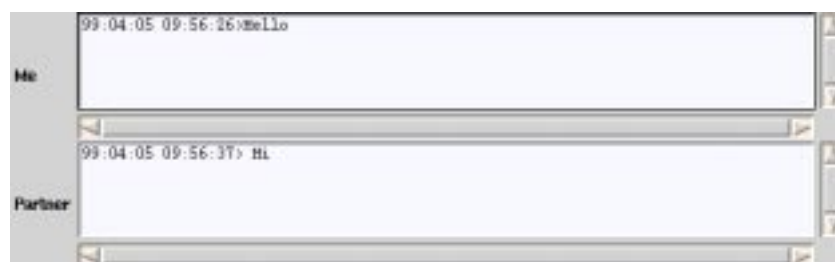


Figure 23. Outil de dialogue textuel

L'interface de recherche et de synthèse d'information est présentée dans les sections suivantes. Rappelons que ces activités de recherche et de synthèse sont basées sur le concept de classification avec contraintes.

I.5.1. Classification sans contraintes

Comme nous l'avons présenté dans la section *II.5 (Conclusion et propositions pour la recherche d'information)*, nous proposons un système de recherche basé sur le concept de «classification avec contraintes». Ce concept comporte deux attributs et une liste de contraintes. Les deux attributs permettent de faire une analyse croisée intra-champ ou inter-champ. Une analyse intra-champ permet d'obtenir la fréquence de co-occurrence de deux termes dans un seul champ pour toute la base. Une analyse inter-champ donne la fréquence de co-occurrences des valeurs des deux attributs.

L'image écran ci-dessous présente une classification sans contraintes de type analyse in-

ter-champ. L'utilisateur souhaite une analyse portant sur les deux attributs maxh (hauteur maximum des arbres) et mina (altitude minimum pour la plantation des arbres). Nous obtenons dans le prototype 2 (40,0) ; 2 (20,0) qui signifie qu'il y a deux arbres qui ont un hauteur maximum de 40 mètres et d'une altitude minimum de 0 mètre ; qu'il y a également 2 arbres qui ont un hauteur maximum de 20 mètres et d'une altitude minimum de 0 mètre. En cliquant sur un triplet, par exemple 2 (40,0), l'utilisateur aura les arbres qui correspondent à ce résultat.

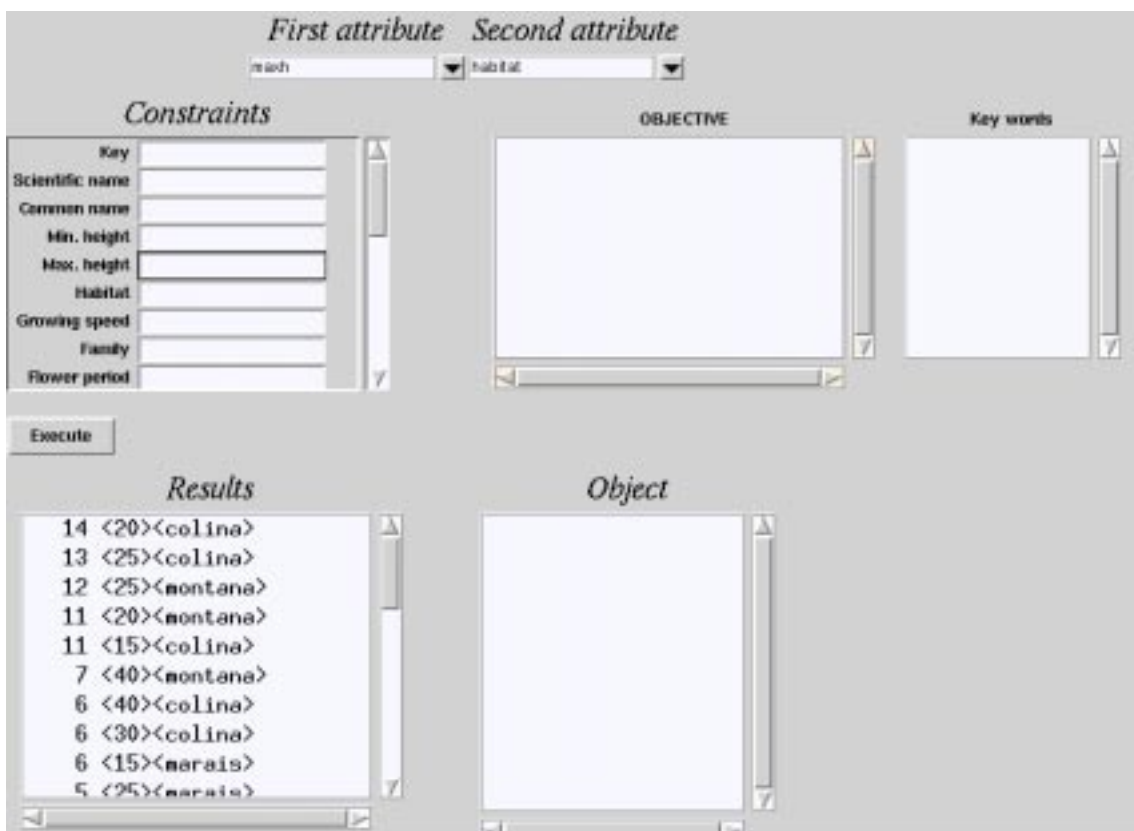


Figure 24. Classification sans contrainte

I.5.2. Classification avec contraintes

L'image écran de la *Figure 25. (Classification avec contraintes)* montre comment formuler une classification avec contraintes. Dans l'exemple, l'utilisateur demande une analyse inter-champ en utilisant les attributs sn et maxh mais avec la contraintes Altitude max. ≥ 1000 qui signifie une demande d'analyse uniquement sur les arbres que l'on peut planter dans les altitudes d'au moins 1000 mètres. Nous obtenons le même type de résultat que dans une classification sans contrainte (présenté dans la zone «cluster»).

Remarquons que l'utilisateur aurait pu demander une classification uniquement avec contrainte. En effet donner uniquement la contrainte (Altitude max. ≥ 1000) donnera le même résultat. Cela est du au fait que classifier avec l'attribut sn (nom scientifique) donnera toujours la fréquence 1 quel que soit le deuxième attribut de l'analyse.

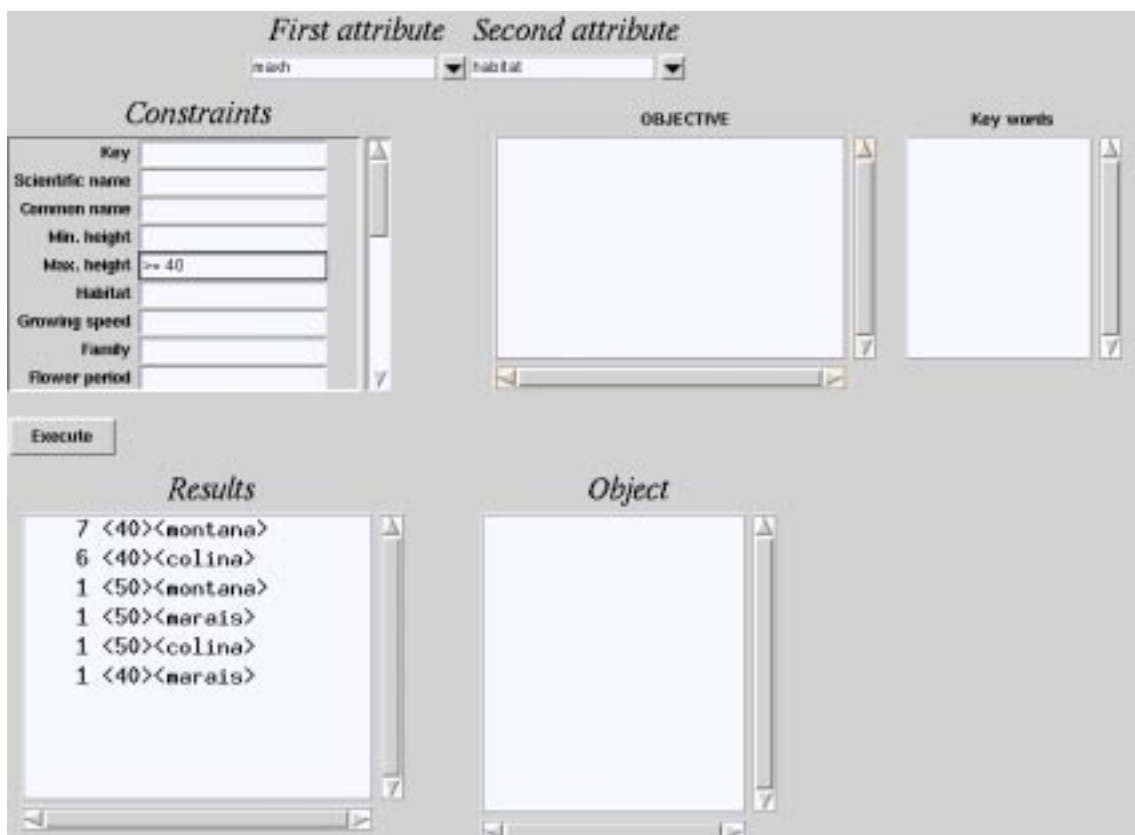


Figure 25. Classification avec contraintes

Les quatre autres type de classification sont la classification uniquement avec contraintes, la classification avec un seul attribut sans analyse intra-champ et la classification avec un seul attribut avec analyse intra-champ.

La classification uniquement avec contraintes correspond à la formulation d'une requête classique avec des opérateurs de comparaison. La différence ici est que le seul opérateur booléen utilisé est l'opérateur OU. Nous appliquons en effet la méthode de la recherche vectorielle et présentons le résultat par ordre décroissant de degré de pertinence.

La classification avec un seul attribut sans analyse permet d'obtenir la liste de tous les arbres.

Rappelons également que pour représenter un utilisateur, nous conservons les activités de l'utilisateur. Ces activités sont classées par catégories fondamentales comme nous l'avons présenté dans la section VI.4 (*Notre modèle de l'utilisateur*). Nous développons actuellement l'interface qui permet d'analyser les activités de l'utilisateur. La même technique de classification avec contraintes est utilisée. En effet la méthode de classification avec contraintes peut être utilisée pour plusieurs types d'analyse. Seulement les attributs d'analyse et des contraintes chan-

gent.

L'architecture générale de communication inter-application et le concept de classification avec contraintes sont suffisamment abstraits pour permettre leurs applications sur d'autres domaines. Ces propositions sont par exemple utilisées (outre l'application dans le prototype STREEMS) dans la gestion des informations géographiques et pour l'analyse des connaissances sur l'utilisateur.

Annexe II

Curriculum Vitae

II.1 Curriculum Vitae

Nom : DAVID

Prénoms : Amos Abayomi

Date de naissance : 19 Décembre 1956

Lieu de naissance : Ejiba-Nigeria

Nationalité : Française

Situation familiale : Marié, 3 enfants

Adresse de bureau : LORIA, UMR 7503, BP 239, CNRS,
54506 Vandoeuvre-les-Nancy Cedex, France.

Tel bureau : (33) 03-83-59-20-60

Adresse domicile : 21 rue Catherine Kuttinger, 54180 Houdemont, France.

Tel domicile : (33) 03-83-51-19-02

Email : Amos.David@loria.fr

Diplômes

- Doctorat de l'INPL, Janvier 1990, INPL, Nancy, France sur «Processus EXPRIM, Image et IA pour un EIAO individualisé (Enseignement par l'Image Intelligemment Assisté par Ordinateur) : Le prototype BIRDS». Les membres du jury sont :
 - Professeur Daniel COULON (président) ;
 - Professeur Monique GRANBASTIEN (rapporteur);
 - Professeur Martial VIVET (rapporteur) ;
 - Professeur Marion CREHANGE (examineur) ,
 - Professeur Jean-Claude DERNIAME (examineur) ,
 - Jean-François NICAUD (examineur) ;
 - Professeur Odile THIERY (directeur de thèse) ;
 - Colette ROYNETTE (invitée).
- DEA Informatique, Juin 1986, INPL, Nancy, France sur «Application d'EXPRIM en EAO»
- B. Sc. Computer Science, Juin 1981, University of Ibadan, Ibadan, Nigeria.

Emploi à l'université

- Je suis maître de conférences 71e section, UFR Lettres, Université de Nancy 2, sur un poste d'Information Scientifique et Technique, depuis 1991, titulaire en 1992.
- J'ai fait partie de l'équipe EXPRIM du CRIN de 1985 à 1997, dirigée par professeur Marion CREHANGE. Actuellement, je fais partie du groupe LORIASI du LORIA, dirigé par Professeur Jean-Claude DERNIAME. Actuellement je suis en discussion avec l'équipe READ pour étudier la possibilité de créer un groupe de recherche au LORIA sur le thème Ingénierie documentaire.

II.2 Responsabilités administratives

Depuis la rentrée 1992, je suis le responsable pédagogique du DESS Information Scientifique et Technique, un DESS cohabilité des trois universités de Nancy 2 : INPL, UHP Nancy 1 et Nancy 2. Les tâches que je remplis par cette responsabilité concernent l'organisation du recrutement des étudiants, le suivi des stages, la gestion des intervenants extérieurs, la gestion de l'emploi du temps et la coordination de l'ensemble des enseignements. Je dois, de plus, suivre l'insertion professionnelle des étudiants après la formation, observer l'évolution des demandes en stages et en embauches par les entreprises pour faire une proposition d'aménagement du contenu de la formation. C'est en effet les trois responsables des trois universités et moi même qui sommes chargés de rédiger le dossier de demande d'habilitation du DESS IST.

Depuis 1991, je suis également le responsable pédagogique du module documentation pour toutes les licences à l'Université Nancy 2. Le module prépare les étudiants en licence pour la maîtrise science de l'information et de la documentation. Les tâches de cette responsabilité sont identiques à celles du DESS IST mais les étudiants ne font pas de stage.

II.3 Animations scientifiques

- Membre du comité de pilotage du colloque «Intelligence économique : Quelles stratégies de formation», Metz 1999.
- Lecteur pour la conférence ISAS'99, USA.
- Membre du comité scientifique de ISKO-France.
- Membre du comité de programme : conférence Hypermedia, Lille, 1992
- Membre de l'ISKO (International Society for Knowledge Organisation)
- En discussion avec les autres maîtres de conférences en IST à Nancy, à Strasbourg et à Metz pour former un groupe de recherche IST

II.4 Activités d'encadrement

En 1991 j'ai encadré une thèse de CNAM sur l'application de l'architecture que j'ai proposé pour développer un système d'EAO basé sur l'image [mai96]. L'étudiant a réalisé le prototype DEDALE qui appliqué à la mécanique du vol. Ces travaux ont donné lieu à une publication dans une conférence internationale.

En 1994, j'ai encadré un DEA sur la personnalisation des réponses dans un système de recherche d'information. Le travail a été bien mené et l'étudiante a eu une bourse de MESR pour préparer une thèse sur le même sujet. Le travail a donné lieu à une publication dans une conférence nationale. J'ai reçu l'autorisation à encadrer cette thèse en octobre 1995. Malheureusement, pour convenance personnelle, l'étudiante a souhaité arrêter sa thèse en fin de la deuxième année.

Dans le cadre de mon deuxième axe de recherche sur la communication inter-application, j'ai encadré une thèse de CNAM sur le sujet en 1996.

J'ai encadré au CRIN un stage post-ingénieur en 1997 pendant deux mois dans le cadre de projet LEONARDO en 1997, de l'université de Malaga, Espagne. L'ingénieur a réalisé une interface entre deux prototypes (STREEMS et GIS), le premier pour la gestion des informations multimedia des arbres autorisés pour le reboisement en France et le deuxième la gestion des informations géographiques dans les pays impliqués dans le projet (Espagne, France, Grèce, Italie, Portugal). J'ai également encadré un stagiaire du DESS IST pour la recherche d'information, la mise en forme de ces informations et leur intégration dans STREEMS en 1997.

Actuellement, je codirige la thèse de David BUENO qui prépare sa thèse à l'université de Malaga, Espagne. C'est moi qui ai défini son sujet de thèse qui est «la modélisation de l'utilisateur et la recherche d'informations dans les systèmes de recherche d'information». La thèse est financée en partie par le projet LEONARDO dont je suis le responsable scientifique pour le laboratoire LORIA.

II.5 Coopérations internationales

J'ai participé à deux projets européens de 1990 à 1995. Le premier est le projet COMETT II de 1990 à 1993, pour le développement d'un système de formation des techniciens, intégrant une base de données multimedia dans le domaine botanique. Ce système devait permettre la consultation, la planification et la simulation du développement d'un espace vert (privé ou public). Les autres partenaires de ce projet sont l'IRFATA Italie, le département informatique de l'université de Malaga Espagne, Agrorolles France, Computer Data Corp. Grèce.

Le deuxième projet a été le projet FORCE de 1992 à 1995 sur une étude de faisabilité et le développement d'un système de communication commerciale pour les fruits et légumes utili-

sant les outils courant du réseau informatique. Les partenaires de ce projet étaient les mêmes que pour le projet COMETT II.

Suite à ces participations, j'ai été sollicité pour participer à un projet d'étude de faisabilité lancé par la région de Bologne Italie.

Je travaille sur un projet européen LEONARDO depuis 1996, sur l'exploitation des techniques de communication inter-application pour développer des systèmes de recherche coopérative d'information entre deux ou plusieurs utilisateurs. Le domaine d'application est l'agriculture.

Remarques

Ma participation aux projets européens m'a permis d'affiner mes réflexions sur mes axes de recherche et permet aux autres partenaires de ces projets d'appliquer les résultats de ma recherche. Par exemple le CSA (Centre des Etudes Agricoles) Italie a pu réaliser, grâce à mes contributions, deux prototypes qui sont expérimentés pour la formation des techniciens en agriculture dans la région de Bologne en Italie.

Ma contribution dans le cadre du projet FORCE a permis au gouvernement régional de Bologne Italie de réaliser un projet d'étude de faisabilité basée sur les idées que j'ai développées pour la communication inter-application dans le cadre de la commercialisation des fruits et légumes dans la région. Nous avons travaillé en collaboration avec le CSA sur ce projet d'étude de faisabilité.

Les projets européens s'ils sont riches en expérience donnent souvent lieu à peu de publication, l'essentiel des résultats étant consignés dans des rapports internes.

II.6 Publications

II.6.1. Thèse

[dav90a] DAVID A. A. *Processus EXPRIM, Image et IA pour un EIHAO individualisé (Enseignement par l'Image Intelligemment Assisté par Ordinateur) : Le prototype BIRDS*. Doctorat INPL, spécialité informatique, INPL (Institut National Polytechnique de Lorraine), CRIN, Nancy, France, Janvier 1990.

II.6.2. Conférences invitées

[dav89c] DAVID A. A., Un Environnement d'Apprentissage basé sur l'Image : Notre Prototype BIRDS. In *Open learning en Agriculture*, I.R.F.A.T.A Bologne, Italie, Juin 1989. papier invité.

[dav89d] DAVID A. A. and THIERY O., Image et Pédagogie. In *A la rencontre des Sciences et des Techniques, quelle culture pour demain?*, Pont-à-Mousson, Octobre 1989. papier invité.

II.6.3. Revues

- [dav99d] DAVID A. and BUENO D., User modeling and cooperative information retrieval in IRS, *International Journal of Knowledge Organization*, A paraître.
- dav91] DAVID A. A. and THIERY O. Le prototype BIRDS : Une expérimentation d'un modèle de l'élève dans un environnement d'apprentissage basé sur l'image. *Génie éducatif*, 1(2) :11 – 19, 1991.
- [dav90b] DAVID A. A., CREHANGE M., THIERY O., An Intelligent Image-Based Computer-Aided Education system : The prototype BIRDS. *International Journal of Pattern Recognition and Artificial Intelligence*, 4(3) :305 – 314, 1990.

II.6.4. Conférences internationales avec comité de sélection

- [dav99c] DAVID A. A., BUENO D., Towards cooperative information retrieval system with user modeling, ISAS'99 (5th International conference on information systems analysis and synthesis, July-August, Orlando, USA, 1999.
- [dav99b] DAVID A. A., BUENO D., Personalisation of information based on the concept of relevance using a user model, ISKO-SPAIN'97, pages 419-426, Granada, Spain, March 1999.
- [dav99a] BELAÏD A., DAVID A., The Use of Information Retrieval Tools in Automatic Document Modeling and Recognition, DAUDD'99, September 1999
- [dav97] DAVID A. A. Modélisation de l'utilisateur et Recherche Coopérative dans les Systèmes de Recherche d'Informations, ISKO'97, pages 341-354, Lille, France, October 1997.
- [dav96] DAVID A. A. Vers une recherche coopérative dans les systèmes de recherche d'informations. In ORSTOM, éditeur, *CARI'96*, pages 217– 226, Libreville GABON, Octobre 1996.
- [dav91b] DAVID A. A., MAIROT N., THIERY O., Student modelling and cognition in education knowledge based tools, PEG'91, International conference on Knowledge Based Environment for Teaching and Learning, Genova, Italie, June, 1991.
- [dav90] DAVID A. A., THIERY O., Taking into consideration the Student's Cognitive Evolution in CAI. In *International Conference on Technology of Education*, Bruxelles, Belgique, Mars 1990.
- [dav89e] DAVID A. A. and THIERY O.. Le Prototype BIRDS : Une expérimentation d'un Modèle de l'Elève dans un Environnement d'Apprentissage basé sur l'Image. In *Journées de travail EIAO-PRC IA*, Paris, Décembre 1989.
- [dav89a] DAVID A. A., THIERY O., and CREHANGE M., Intelligent Hypermedia in Education. *IC-CAL*, Mai 1989.
- [dav89b] DAVID A. A., THIERY O., and CREHANGE M., Intelligent Image Based Computer Aided Education (IICAE). In *International conference on «Applications of Artificial Intelligence»*, Orlando-USA, Mars 1989. SPIE et IEEE.
- [thi88] THIERY O., A. A. DAVID, and CREHANGE M., Un Enseignement par l'Image Intelligemment Assisté par Ordinateur (EIIAO) : un exemple concret en ornithologie. In *Premières Journées Internationales Application de l'Intelligence Artificielle à l'Agriculture, à l'Agrochimie et aux Industries Agro-Alimentaires*, Caen, France, 29-30 1988.
- [thib89] THIERY O., A. A. DAVID, and CREHANGE M., A knowledge and Image based system as an Intelligent Educational Tool : Our Prototype BIRDS. In *Third International Conference on Children in the Information Age - Human Development and Emerging Technologies*, So-

fia, Bulgaria, 20 - 23 May 1989.

- [thia89] THIERY O., A. A. DAVID, and CREHANGE M., Images and Artificial Intelligence in Computer Aided Education : Our Prototype BIRDS. In *International Conference on Technology of Education*, Orlando-USA, Mars 1989.

II.6.5. Conférences nationales avec comité de sélection

- [rul95] RULQUIN V., DAVID A., and THIERY O. Personnalisation des réponses dans un système de recherche d'information. In Actes Journée d'étude sur les systèmes de recherche d'information élaborés, Ile Rousse, Corse, France, 1995.

II.6.6. Rapports pour les projets européens

- [dav97] DAVID A. A. «STREEMS Shrubs and Trees in Mediterranean region Multimedia Database System» rapport pour le projet LEONARDO, Juin 1997.
- [dav94] DAVID A. A. «LOGI-COM2 : LOGIstique et COMunication COMmerciale (Commercial Logistics and Communication)» rapport pour le projet FORCE, Novembre 1994.
- [dav93] DAVID A. A., «Multimedia tools for CIA systems», rapport pour le projet FORCE, Mai, 1993.

II.6.7. Rapports internes

- [dav86] DAVID A. A.. *Application d'EXPRIM en EAO.*, INPL, Nancy, France, Juillet 1986. Rapport de DEA.
- [dav88] DAVID A. A.. Des Scénarios d'EIIAO. Rapport interne, CRIN, 1988.
- [tia98] TIAKO, P. F., DAVID A. - «Environnement centré procédés pour supporter la négociation» Mars 1998