

République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique
Université de Mentouri - Constantine
Faculté des sciences de l'Ingénierie
Département d'Informatique

N° d'ordre : 0 6 4 M A J 2 0 0 5
Série : 008/inf/2005

Mémoire

*Pour obtenir le diplôme de magistère en informatique
Option : Computation & Information*

Personnalisation de l'information et gestion des profils utilisateurs : une approche fondée sur les ontologies

Présenté par : M^{elle} Ilhem Boussaid

Dirigé par : Pr. Zizette Boufaïda

Soutenu le : 18 Avril 2005

Devant le jury composé de :

Mr Nacereddine Zarour	Président	Maître de conférence	Univ.Constantine
Mme Zizette Boufaïda	Rapporteur	Professeur	Univ.Constantine
Mme Souham Meshoul	Examineur	Maître de conférence	Univ.Constantine
Mme Faïza Belala	Examineur	Maître de conférence	Univ.Constantine

A LA MEMOIRE DE MON PERE

Remerciements

Qu'il me soit permis ici de remercier, tout d'abord, Dieu qui m'a aidé à réaliser ce travail, toutes les personnes qui, toutes d'une manière différente, mais toujours dans un but constructif, ont contribué à ce que je puisse aboutir à la réalisation de ce mémoire.

J'aimerais remercier très chaleureusement Madame Zizette Boufaïda, Professeur à l'Université de Constantine, pour la qualité de l'encadrement dont elle m'a fait bénéficier, pour avoir guidé ce travail en conjuguant habilement, disponibilité, conseils et critiques constructives tout en laissant libre cours à ma créativité. Son intégrité est un exemple. Je sais avoir beaucoup appris grâce à elle.

Les personnes qui m'ont fait l'honneur de participer au jury de cette thèse.

Je tiens à remercier, Monsieur Nacereddine Zarour, Maître de Conférence à l'Université de Constantine, pour l'intérêt qu'il a bien voulu porter à ce travail en acceptant de présider ce jury.

Merci à Madame Souham Meshoul, Maître de Conférences à l'Université de Constantine, pour avoir fait preuve d'une très grande disponibilité en acceptant d'être examinateur.

Madame Faïza Belala pour avoir montré un réel intérêt pour mon travail et l'honneur qu'elle me fait en acceptant d'en être examinateur et membre du jury.

J'adresse un merci tout particulier à l'ensemble du personnel du laboratoire LIRE ; enseignants, chercheurs ou étudiants, ils ont contribué à ce que mon travail se déroule dans une atmosphère des plus agréables. Tous ceux et toutes celles qui par leur présence, leur sourire, leur soutien et leur confiance ont fait que je n'étais pas seule pour parcourir ce chemin. J'ai ainsi une pensée pour les membres de l'équipe SI & BC qui ont toujours été présents et prêts à m'apporter leur aide. Je ne peux les citer tous, mais que chacun d'eux sache combien je l'apprécie et combien il m'est agréable de faire partie de cette équipe. Un merci tout particulier à Monsieur Mahmoud Boufaïda.

Du fond du coeur je remercie ma maman. Aucun mot n'est assez fort pour lui exprimer la reconnaissance sincère que je lui porte. Sa force est notre force à tous. Son attention, ses conseils, ses paroles, sa protection sont des trésors inestimables. Mes frères et soeurs pour leur patience et leur tendresse et qui ont toujours été là, pour partager la peur et les doutes ; la joie et l'allégresse aussi. J'embrasse tendrement, et par ordre chronologique, Saber, Wissem, Zouleikha et Zineddine (Mon p'tit frère adoré). Vers eux va tout mon amour...

Denis, je t'adresse un grand, non, un énorme, non non, un gigantesque « merci », pour ta relecture enrichissante et tes encouragements, pour ton soutien et surtout pour avoir à chaque fois trouvé les mots que j'avais besoin d'entendre. Tu as quelque chose de magique, savoir faire renaître le soleil où tout semble si noir.

Ce paragraphe est aussi particulièrement dédié à mes amis proches et à tous ceux que j'aime, présents ou partis trop tôt, et à qui je pense et je penserai toujours.

Merci enfin à tous ceux qui ont fait ce que je suis actuellement : parents, famille, amis, professeurs. Il y a un peu de chacun de vous dans ce mémoire.

Tables Des Matières

Introduction Générale	2
Contexte de ce travail.....	3
Problématique.....	4
Objectifs et approches.....	6
Notre contribution	8
Organisation du mémoire.....	12
Chapitre 1 : Modélisation des utilisateurs	15
1. Introduction.....	16
2. La modélisation des utilisateurs.....	16
2.1 Classification des modèles de l'utilisateur.....	17
2.2 Techniques de représentation des modèles utilisateur	18
2.2.1 Le modèle individuel.....	18
2.2.2 Le modèle de recouvrement (overlay model).....	19
2.2.4 Le stéréotype.....	19
2.2.5 Le modèle partagé.....	19
3. Construction et maintenance des profils utilisateurs.....	20
3.1 La représentation des profils utilisateurs.....	21
3.2 La génération du profil initial.....	22
3.2.1 L'observation directe.....	22
3.2.2 Les interviews.....	23
3.2.3 Les questionnaires.....	23
3.3 Techniques d'apprentissage des profils.....	23
3.3.1 Analyse statistique de termes.....	24
3.3.2 Les techniques de classification.....	26
3.3.3 Les méthodes de clustering.....	27
3.4 Retour de pertinence (Relevance feedback).....	27
4. Sécurité des données du profil	31
5. Discussion.....	31

Chapitre 2 : Personnalisation de l'information –Concepts et précisions..33

1. Introduction.....	34
2. Définition de la personnalisation.....	35
3. Les éléments sujets à la personnalisation.....	36
3.1 Personnalisation du contenu informationnel.....	36
3.2 Personnalisation de la navigation.....	38
3.3 Personnalisation de la présentation.....	39
4. Les services de personnalisation.....	40
4.1 Le filtrage des résultats	40
4.2 Le re-ordonnement des résultats de la requête.....	41
4.3 La recommandation.....	42
4.4 Les extensions des langages pour supporter la personnalisation.....	43
5. Méthode de personnalisation.....	43
5.1 Détermination Explicite.....	44
5.2 Détermination Implicite.....	44
5.3 Customisation.....	45
6. Conclusion.....	46

Chapitre 3 : Représentation des ontologies pour le Web

sémantique.....	48
1. Introduction.....	49
2. Le Web Sémantique.....	50
2.1 L'architecture du Web Sémantique.....	50
3. Les ontologies : origine et définitions	51
3.1 Les ontologies en Ingénierie des Connaissances.....	52
4. Quels sont les composants principaux d'une ontologie?.....	53
5. Le rôle des ontologies	54
6. Types d'ontologie.....	55
7. La construction des ontologies.....	56
7.1 Quelques méthodologie de construction.....	56
7.2 Methontology.....	58
8. Quelques formalismes de représentation des connaissances	59
8.1 Les réseaux sémantiques.....	60

8.2 Les langages de frames	60
8.3 Les Graphes Conceptuels (GC).....	61
8.4 Les logiques de description (LD).....	62
8.4.1 Sémantique.....	63
8.4.2 Les constructeurs des LDs.....	63
8.4.3 Problèmes et procédures de décision pour les LDs.....	65
9. La représentation des connaissances sur le Web.....	65
9.1 RDF (Ressource Description Framework).....	66
9.2 RDF Schema (RDF Vocabulary Description Language).....	67
9.3 OIL (Ontology Inference Layer).....	67
9.4 DAML+OIL (DARPA Agent Markup Language).....	67
9.5 OWL (Ontology Web Language).....	68
10. Environnements et outils de modélisation.....	69
10.1 Protégé2000	71
11. L'exploitation des ontologies.....	72
11.1 Les moteurs d'inférences.....	72
12. Synthèse.....	73
Chapitre 4 : Modèle de l'utilisateur.....	75
1. Introduction.....	76
2. Un modèle utilisateur, pourquoi faire?.....	78
3. Dégagement de la définition du modèle utilisateur.....	78
3.1 Les données personnelles.....	79
3.2 Profession.....	79
3.3 Rôle de l'utilisateur.	81
3.4 Les préférences	83
3.5 Le centre d'intérêt.....	84
3.5.1 La modélisation d'intérêts.....	84
3.6 Le modèle de connaissance.....	85
3.6.1 La génération du modèle de recouvrement.....	86
3.6.2 L'expérience et les compétences.....	88
3.7 L'historique.....	88
3.7.1 Gestion de bookmarks	89

3.8	Modèle de groupe.....	89
3.9	La sécurité.....	90
3.9.1	P3P: Platform for Privacy Preferences Project.....	91
3.10	Gestion des conditions matérielles.....	91
4.	Les propriétés des caractéristiques utilisateurs.....	94
4.1	Granularité, quel est le sujet modélisé ?.....	94
4.2	La donnée est-elle stable ?.....	95
4.3	Quelle est sa durée de vie ?.....	95
4.4	Dans quel cadre est utilisée la caractéristique ?.....	95
4.5	L'acquisition. Qui renseigne la caractéristique ?.....	95
4.6	Quelles sont les principales sources de données ?.....	96
4.6.1	Les formulaires (ou questionnaires).....	96
4.6.2	Les fichiers Log et les clickstreams.....	96
4.6.3	Le retour de préférence ou Relevance feed-backs.....	97
4.7	Identification des utilisateurs.....	98
4.7.1	Adresse IP.....	98
4.7.2	Identifiants de session.....	98
4.7.3	Les Cookies.....	99
4.7.4	Combinaison d'un pseudonyme et d'un mot de passe	99
4.8	Implication de l'utilisateur dans la génération des profils.....	99
5.	Critères d'évaluation.....	100
6.	Conclusion.....	102

Chapitre 5 : Une approche fondée sur les ontologies.....103

1.	Introduction.....	104
2.	La modélisation de la connaissance : pourquoi opter pour les ontologies ?.....	105
3.	Une approche fondée sur les ontologies.....	105
3.1	Une ontologie du domaine.....	106
3.2	Ontologie des utilisateurs.....	107
3.2.1	Prise en compte des caractéristiques des ressources.....	107
3.3	L'ontologie des documents.....	107
3.3.1	Les publications scientifiques.....	108
3.3.2	Schéma Dublin Core.....	111

3.3.3 Un schéma de métadonnées.....	112
3.4 Une ontologie de l'application	113
3.4.1 A qui s'adresse la personnalisation ?.....	114
3.4.2 Choix du niveau de personnalisation.....	114
3.4.3 L'évaluation des documents.....	115
3.5 Liens entre les ontologies.....	117
4. Cycle de vie de l'ontologie.....	118
4.1 La spécification	119
4.1.1 L'évaluation des besoins.....	120
4.2 Acquisition des connaissances	122
4.2.1 Techniques de recueil de données.....	123
4.2.2 Envisager une éventuelle réutilisation des ontologies existantes.....	124
4.3 La conceptualisation.....	125
4.3.1 Création de concepts.....	125
4.3.2 Création d'une taxonomie.....	126
4.3.3 Représentations intermédiaires.....	128
4.4 La formalisation de l'ontologie.....	130
4.4.1 Un formalisme pour représenter les ontologies : les Logiques de Description.....	131
4.4.2 Le langage de description SHIQ.....	131
4.4.3 La classification.....	132
4.5 L'implémentation.....	134
4.5.1 Le langage ontologique OWL.....	136
4.5.2 La structure d'une ontologie OWL.....	137
4.5.3 Visualisation de l'ontologie.....	145
4.5.4 Les requêtes	146
4.5.5 Le raisonnement.....	147
5. Conclusion.....	148
Conclusion générale.....	150
Bibliographie	154

Liste des figures

Figure 1.1 – Génération et maintenance des profils d'utilisateur.....	20
Figure 1.2 – Exemple de deux documents et de leurs vecteurs représentatifs	25
Figure 2.1 – Positionnement de la personnalisation par rapport à l'adaptation [Gavray, 2002].....	36
Figure 2.2 – Représentation d'un système de filtrage.....	41
Figure 2.3 – Architecture générale d'un système de recommandation [Trousse, 2001].....	42
Figure 3.1 – Le « mille-feuilles » du Web sémantique (<i>Layer cake</i>) [Berners-Lee et al., 2001].....	51
Figure. 3.2 – Recensement des objets qui ont pu, à un moment donné, être classifié comme des ontologies, d'après [Smith et Wetly, 2001].....	53
Figure 3.3– Cycle de vie de l'ontologie [Fernández et al., 1997].....	59
Figure 3.4 – La pyramide des langages basés Web.....	66
Figure 4.1 – Exemple graphique des règles de génération.....	87
Figure 4.2 – Exemple d'un profil CC/PP décrit en XML.....	92
Figure 5.1 – Structure des ontologies.....	106
Figure 5.2 – Cycle de vie de l'ontologie.....	119
Figure 5.3 – Le document de spécification.....	119
Figure 5.4 – L'approche ascendante.....	126
Figure 5.5 – l'approche descendante.....	127
Figure 5.6 – L'approche centrifuge.....	127
Figure 5.7 – La classification d'un nouveau concept.....	133
Figure 5.8 – Illustre Protégé 2000 avec une ontologie d'exemple.....	135
Figure 5.9 – Editeur d'expressions OWL.....	135
Figure 5.10 – OntoViz- Visualisation des attributs des classes.....	146
Figure 5.11 – Formulation de la requête : y-t-il une personne de LIRE ?.....	147

Liste des tableaux

Tableau 1.1– Techniques de construction et de maintenance des profils utilisateurs.....	30
Tableau 3.1 – Quelques constructeurs d’une LD.....	64
Tableau 3.2– Liste des constructeurs disponibles dans le langage OWL Lite.....	68
Tableau 3.3– Liste des constructeurs supplémentaires disponibles dans les langages OWL DL et OWL Full.....	69
Tableau 4.1 – Niveaux éducationnels.....	80
Tableau 4.2– Grades des enseignants.....	81
Tableau 4.3– Grades des chercheurs.....	81
Tableau 4.4 – Liste non exhaustive des activités par profession.....	82
Tableau 4.5 – Description du modèle utilisateur.....	94
Tableau 4.6 – l’acquisition des données.....	96
Tableau 4.7 – Format d’un fichier log.....	97
Tableau 4.8 – Critères d’évaluation relatifs à la modélisation des utilisateurs.....	101
Tableau 5.1 – Schéma de métadonnées.....	113
Tableau 5.2 – Ontologie d’application.....	117
Tableau 5.3 –Table des concepts.....	129
Tableau 5.4 – Tables des relations.....	129
Tableau 5.5 – Tables des attributs.....	130
Tableau 5.6 – La relation binaire OrganiseEvenement.....	125
Tableau 5.7 – La relation binaire CommiteDeProgramme.....	125
Tableau 5.8 – La grammaire du langage de description de concepts SHIQ.....	132
Tableau 5.9 – Relation entre la conceptualisation et l’implémentation.....	136
Tableau 5.10 - Quelques différences entre DAML+OIL et OWL.....	137

Introduction générale

« L'esprit scientifique nous interdit d'avoir une opinion sur des questions que nous ne comprenons pas, sur des questions que nous ne savons pas formuler clairement. Avant tout, il faut savoir poser des problèmes. Et quoi qu'on dise, dans la vie scientifique, les problèmes ne se posent pas d'eux-mêmes. C'est précisément ce sens du problème qui donne la marque du véritable esprit scientifique. Pour un esprit scientifique, toute connaissance est une réponse à une question. S'il n'y a pas eu de question, il ne peut y avoir une connaissance scientifique. Rien ne va de soi. Rien n'est donné. Tout est construit »

G. Bachelard, Formation de l'esprit scientifique

De plus en plus de systèmes d'information sont accessibles à travers Internet ou intranet. Ils permettent aux utilisateurs d'accéder à une masse énorme d'information provenant d'une ou de plusieurs sources. Toutes ces informations peuvent ne pas être homogènes. Les utilisateurs ne sont pas tous intéressés par les mêmes informations ; ils ont des profils très différents, des rôles différents, différents centres d'intérêt et différents besoins. De plus, les utilisateurs ne connaissent pas forcément des personnes qui partagent leurs centres d'intérêt et qui ont une expérience utile dans le contexte de leur travail. C'est pourquoi, il est nécessaire d'organiser et de gérer les ressources d'une organisation.

Aujourd'hui, l'objectif du Web de troisième génération ou « Web Sémantique », présentés par [Berners-Lee, 2001], est d'apporter de meilleures précisions et de favoriser le partage et la réutilisation des connaissances et des documents par l'utilisation d'ontologies, de mécanismes d'inférence et l'association de métadonnées sémantiques aux connaissances et documents. Le Web Sémantique est un Web interprétable par les machines de façon à automatiser ou semi-automatiser certains processus et par conséquent aider l'utilisateur dans sa tâche. Il propose une architecture et des standards pour atteindre cet objectif.

Automatiser des processus pour aider les utilisateurs nécessite de caractériser ces derniers. L'ontologie peut jouer un rôle important pour la représentation et la gestion des connaissances relatives aux utilisateurs. Elle fournit un cadre unificateur pour réduire et éliminer les confusions conceptuelles et terminologiques et assurer une compréhension partagée par la communauté visée. Elle peut donc permettre la communication entre des utilisateurs, même s'ils travaillent dans des contextes différents, avec des besoins et des points de vue différents.

1. Contexte de ce travail

Le cadre de notre travail se situe à la croisée de trois domaines : la modélisation des utilisateurs, la personnalisation de l'information et la représentation des connaissances. Ce champ est relativement vaste dans la mesure où ces termes recouvrent de nombreux domaines et notamment l'Intelligence Artificielle, les systèmes d'information, les interfaces Homme/Machine, les hypermédia adaptatifs, les sciences de l'éducation, etc.

Nous nous situons dans le cadre du Web Sémantique, ou plutôt d'un Web Sémantique d'une taille limitée (à celle d'une organisation) pour la modélisation des utilisateurs d'un laboratoire de recherche. Nous souhaitons notamment fournir aux chercheurs, ou aux spécialistes de l'information scientifique, un environnement dans lequel ils puissent exploiter les données de leur domaine, pour des besoins de recherches bibliographiques ou d'analyses du domaine.

Il est essentiel que les utilisateurs puissent accéder facilement aux ressources qui les intéressent. Différents types d'acteurs interviennent, il s'agit principalement de chercheurs, d'étudiants, de personnels administratifs et de techniciens. Ces acteurs utilisent des ressources différentes pour les tâches qu'ils ont à accomplir. Pour partager des informations, particulièrement lorsque les acteurs sont géographiquement distants, il est nécessaire d'utiliser un vocabulaire commun dont les termes ont une signification commune. C'est une des raisons pour laquelle nous proposons dans ce cadre une approche ontologique pour l'exploitation des connaissances et permettre de représenter les ressources, les informations, les connaissances d'une organisation. La personnalisation pour un utilisateur donné est favorisée par l'utilisation d'un modèle qui représente ces utilisateurs.

2. Problématique

Le besoin en information est primordial dans de nombreux domaines, comme celui de la recherche ou celui de la veille scientifique et technique. Mais retrouver des informations répondant à nos besoins devient de plus en plus difficile. En effet, les informations électroniques tendent à croître de façon exponentielle et le problème n'est pas trop l'existence des informations mais bel et bien le moyen de les localiser. Ce problème est d'autant plus important sur le World Wide Web (Web) que la masse d'informations disponible est énorme. Toutefois cette quantité croissante de données disponibles nécessite de mettre en oeuvre des moyens particuliers pour les exploiter.

Parmi les problèmes qui se posent, on peut citer : l'accès à l'information pertinente, la navigation dans un grand espace d'information et la diversité des besoins des utilisateurs.

Tout d'abord, **l'information est difficilement accessible**. Le temps nécessité par les tâches d'accès à l'information représente désormais un coût considérable. Les problématiques de l'accès à l'information et de l'expression du besoin informationnel remplacent alors les préoccupations techniques de déploiement des infrastructures matérielles. En plus, l'information est inscrite sur de nombreux supports à la fois différents dans la forme (rapports, ouvrages, thèse, etc.) et dispersés dans l'espace.

Ensuite, **l'information est trop abondante**. Le problème inverse du précédent se vérifie encore plus fréquemment. Il arrive souvent que pour une requête bien précise on obtienne en retour une pléthore d'informations, que l'on appelle le « bruit ». Autrement dit, l'information reçue est souvent trop abondante et peu pertinente.

La navigation dans un grand espace de ressources : dès que l'espace d'information accessible à un utilisateur par navigation est important, la perte de repères et la désorientation ne permettent pas à l'utilisateur de retrouver les informations recherchées ni de savoir d'où il vient, où il peut aller et comment y aller. Il est important ici de pouvoir guider l'utilisateur dans cet espace d'information, en d'autres termes d'adapter l'accès à l'information en fonction de ses

besoins. Les informations et les liens non pertinents pour un utilisateur mais présentés à l'écran n'entraînent qu'une surcharge cognitive de celui-ci.

Enfin, **les utilisateurs et leurs besoins changent**. Ceux-ci ne veulent plus seulement avoir accès à une information parcellaire, ils souhaitent avoir le maximum d'information pertinente. La plupart du temps, les utilisateurs noyés sous d'énormes volumes d'information de nature diverses et plus ou moins adaptées ont des attentes qui se diversifient. Ils souhaitent toujours bien sûr avoir des documents les plus pertinents possibles, et de ce point de vue ils sont parfois déçus.

Actuellement, la plupart des **méthodes de représentation des utilisateurs** pose le problème du choix des critères pertinents à analyser. En fait, on peut disposer d'énormes quantités de données, mais il est difficile de savoir comment les analyser et lesquelles sont pertinentes pour un objectif donné des utilisateurs. Il y a donc un réel problème pour l'acquisition et la maintenance des modèles utilisateur.

Différentes techniques sont utilisées pour représenter les utilisateurs. Ce sont souvent des modèles simplifiés qui manquent de structuration. Par ailleurs, elles se contentent de catégoriser les informations de profil sans expliciter les corrélations qui existent entre elles.

En plus, les méthodes « traditionnelles » de génération de profils d'utilisateurs, à l'aide de mots clefs pondérés, d'interviews d'utilisateurs ou de stéréotypes (modèles de groupes d'utilisateurs), présentent l'inconvénient de ne pas maîtriser la sémantique des termes utilisés pour caractériser un utilisateur.

Tout ceci exposé, nous nous sommes en outre plus particulièrement intéressés à trois aspects principaux, bien complémentaires. Ceux-ci sont, successivement, la personnalisation de l'information, la construction proprement dite d'un modèle d'utilisateur et la représentation de caractéristiques relatives aux utilisateurs ainsi que la modélisation de mécanismes de raisonnement à partir des diverses données recueillies sur chaque utilisateur.

3. Objectifs et approches

L'objectif de notre travail est d'apporter au mieux l'information à l'utilisateur en fonction de son contexte et de ses besoins. On essaye donc de regrouper et sauvegarder des éléments le concernant dans un *profil*. Ce profil utilisateur a en général un caractère invariant, il doit rester relativement stable pour éviter de fausser totalement la représentation de l'utilisateur. Chaque modification du profil par le système doit être foncièrement légitimée par les actions de l'utilisateur (directes ou non).

Il nous semble important de souligner deux aspects complémentaires: la représentation des connaissances des utilisateurs et la personnalisation de l'information. Nous pensons que ces deux aspects imposent certaines contraintes que les ontologies peuvent aider à respecter.

La personnalisation de l'information

Personnaliser l'information, cela signifie s'adapter aux buts, préférences et capacités de l'utilisateur. La personnalisation a pour objectif, d'une part, de faciliter l'expression du besoin utilisateur et de rechercher des informations sur un sujet en écartant l'information non-pertinente et donc de réduire considérablement l'espace de recherche et, d'autre part, de rendre cette information sélectionnée intelligible à l'utilisateur et exploitable. La pertinence de l'information n'est cependant pas une mesure objective, généralisable à tous les utilisateurs. Elle se définit par un ensemble de critères et de préférences personnalisables spécifiques à chaque utilisateur ou communauté d'utilisateurs. La personnalisation de l'information conduit ainsi nécessairement à modéliser l'utilisateur.

La modélisation utilisateur

La modélisation de l'utilisateur, c'est-à-dire une réelle spécification de l'utilisateur ou la considération de l'utilisateur comme une entité instable que le système essaie de maîtriser, paraît vraiment essentielle pour parvenir à la prise en compte de la diversité des utilisateurs, de leurs besoins, attentes, compétences et centres d'intérêts particuliers.

Il nous apparaît indispensable, en premier lieu, d'aborder et d'étudier plus avant différentes méthodes d'élaboration et d'implémentation des modèles utilisateurs, de même que plusieurs procédés possibles de représentation existent dans la littérature.

Notre objectif est donc de définir un modèle formalisé pour la description des profils des utilisateurs d'un laboratoire de recherche (chercheurs, enseignant/chercheur, étudiants, personnels administratifs et techniciens, etc.) afin de pouvoir l'évaluer, mieux gérer son évolution, le « comparer » aux autres profils et l'interpréter.

Pour réussir à apporter une solution adéquate à cet objectif, nous allons nous focaliser sur le sujet de la représentation des connaissances à travers les ontologies.

La représentation sémantique des connaissances

Notre deuxième objectif concerne le choix d'un mode de représentation des connaissances doté d'une bonne expressivité afin d'améliorer la communication entre différents acteurs, humains ou non, et de favoriser un traitement basé sur la sémantique.

Les ontologies sont une solution pour répondre à un besoin de modélisation et de représentation. Elles constituent un élément clé pour une structuration sémantique de données hétérogènes.

4. Notre contribution

Si nos recherches se résument à la tentative constante de simplification automatique de la vie de l'utilisateur et de la réduction aussi de l'effort cognitif fourni par celui-ci devant une activité de recherche ou d'exploration d'un domaine, nous devons impérativement nous poser la question à suivre. De quelles manières et selon quels procédés pouvons-nous exhaustivement définir chacun de nos utilisateurs, chacune de leurs caractéristiques, actions et réactions ?

Nous devons apporter des innovations supplémentaires au niveau des modélisations d'utilisateur existantes. Ceci contribue, selon nous, à tendre vers de meilleures capacités de personnalisation et de prise en compte des caractéristiques et des besoins des différents utilisateurs.

C'est dans cette optique que nous souhaitons maintenant définir un concept de modélisation de l'utilisateur. Après avoir alors exploré différentes techniques de conception et d'implémentation des modèles d'utilisateurs déjà existants, nous prolongerons nos travaux de recherche par une définition d'un modèle de l'utilisateur, suivie de l'étude d'une méthode appropriée de représentation de ce modèle. Nous voulons maintenant proposer une organisation des connaissances qui répondent le mieux possible à l'ensemble des contraintes que nous avons précisé dans la problématique.

A la différence de nombreux systèmes, nous proposons d'utiliser un profil hiérarchisé, plutôt que des combinaisons de vecteurs comme le font la plupart des systèmes antérieurs, pour avoir accès à différentes catégories du modèle utilisateur qui ne seront sans doute pas sans intersection commune. Nous avons soutenu qu'une présentation de la structure et du contenu des profils utilisateurs est nécessaire. Par structure, nous entendons l'organisation des éléments décrivant l'utilisateur, tandis que le contenu est une structure conceptuelle, c'est-à-dire une caractérisation de ces éléments. En plus, les ressources d'information d'un profil utilisateur contiennent normalement beaucoup de termes différents ayant la même sémantique. Il faut utiliser, par conséquent, un modèle qui peut normaliser tous les termes utilisés afin de lever toute ambiguïté.

L'approche adoptée par [Moukas, 1997], dans laquelle il combine jugement de l'utilisateur et détection automatique de ses centres d'intérêt nous semble très intéressante, sauf que dans notre approche nous en préconisons deux techniques de représentation des centres d'intérêts. La première permet la création de modèles qui sont dépendants du domaine est déterminée en restreignant à une ontologie du domaine la possibilité offerte à un utilisateur de donner ses centres d'intérêts. La seconde technique, qui est indépendante du domaine est accomplie à travers l'analyse des documents personnels de l'utilisateur. Dans ce cas, les centres d'intérêt sont les documents eux-mêmes. Ainsi, la création du modèle d'intérêt de l'utilisateur consiste à trouver les bonnes caractéristiques à partir des documents. Pour cela nous utilisons la mesure TFIDF (Term Frequency - Inverse Document Frequency) [Salton et Buckley, 1988] qui propose de pondérer chaque terme par sa fréquence dans le document

Quant à la modélisation des connaissances de l'utilisateur, nous nous sommes inspirés essentiellement du modèle overlay [Brusilovsky et Eklund, 1998] - qui associe pour un utilisateur particulier, une valeur à chaque concept du modèle de domaine.

Notre contribution ne portant pas sur la définition d'une nouvelle ontologie de domaine, nous avons choisi d'utiliser dans nos expériences une ontologie existante.

L'idée de concevoir de modèles indépendants du domaine d'application et de modèles dépendants du domaine est inspirée de la classification des modèles des utilisateurs donnée par [Koch, 2000].

Pour respecter certaines structures dans les laboratoires de recherche, nous intégrons, dans le modèle utilisateur, la profession ou la fonction de ce dernier. Un autre aspect à prendre en compte est l'activité professionnelle de l'utilisateur et les objectifs et rôles de ce dernier.

En plus de ces catégories, nous considérons d'autres caractéristiques, telles que les données personnelles, les préférences des utilisateurs, les conditions matérielles, l'historique des activités et les données de sécurité.

Le problème que nous nous posons est alors de trouver le formalisme adéquat, à la fois expressif et optimal en terme de manipulation, pour représenter ces descriptions. Une façon possible de résoudre ce problème est d'avoir un cadre unificateur pour réduire et éliminer les confusions conceptuelles et terminologiques et assurer une compréhension partagée par la communauté visée. Les résultats de recherche dans le domaine des ontologies peuvent ici

contribuer à la résolution de ce problème. En plus, la référence à une ontologie commune à tous les acteurs d'une communauté représente un gain de productivité considérable.

Le choix d'utilisation des ontologies et du Web sémantique pour la gestion du savoir n'est pas arbitraire. Le Web sémantique, non seulement est un moyen idéal pour stocker les informations et leurs descriptions (les métainformations) mais il nous permet aussi de raisonner sur ces données.

Au terme de ce mémoire, nous pouvons dire que notre approche préconise une ontologie de niveau abstrait qui fédère sous une seule et même structure quatre ontologies : l'ontologie des utilisateurs, l'ontologie des documents, l'ontologie du domaine et l'ontologie d'application, qui ont des liens entre elles.

L'ontologie utilisateur, est basée sur le modèle utilisateur proposé et utilise une ontologie du domaine. Ceci d'une part permet à un utilisateur d'exprimer ses centres d'intérêts en retenant des mots et des expressions qui lui sont présentées, plutôt que d'avoir à les saisir, elle permet aussi de décrire les connaissances des utilisateurs.

L'ontologie des documents a pour objectif de compléter l'état de connaissance des utilisateurs par l'acquisition d'information contenue par des documents pertinents, en plus, l'activité principale de la communauté scientifique d'un laboratoire de recherche est la production de documents scientifiques, là aussi, une normalisation minimale des données à exploiter est indispensable. Nous avons donc jugé nécessaire que les éléments de métadonnées rentrant dans la description de ressources comportent des éléments pertinents de profil d'utilisateur, car la structuration des documents influe sur la quantité d'information transmise aux utilisateurs demandeur d'information.

Enfin, l'ontologie de l'application se compose notamment des informations relatives à la gestion de la personnalisation, comme la liste des stratégies de personnalisation avec les paramètres nécessaires à leur fonctionnement.

La séparation de ces différentes catégories de connaissance facilite leurs évolutions et leur maintenance. Elle permet aussi d'améliorer leur compréhension et de définir des rôles différents pour chaque ontologie, c'est-à-dire de les spécialiser.

La définition du profil d'un utilisateur particulier revient donc à sélectionner les éléments utiles de l'ontologie et à les instancier, en fonction de ses besoins, du type d'application, de l'environnement d'exécution de cette application, etc.

Nous avons largement étudié les différentes méthodologies de construction d'ontologies, les langages dans lesquels elles sont exprimées ainsi que les outils permettant de les manipuler. Particulièrement nous nous sommes inscrit dans le cadre de la méthodologie de M. Fernández et ses condisciples : METHONTOLOGY [Fernández et al., 1997].

Nous nous sommes placés, dans un premier temps, dans le cadre d'extraction d'une taxonomie de termes à partir de données hétérogènes et issues de sources différentes comprenant à la fois des informations structurelles et textuelles. L'idée consiste à dégager les caractéristiques des utilisateurs, les caractériser par des termes pertinents et les organiser en hiérarchie.

Dans un deuxième temps, nous avons choisis un formalisme permettant d'enrichir l'ontologie ainsi construite et de tenir compte de certaines contraintes. Les logiques de descriptions sont bien adaptées à la représentation d'ontologies car elles disposent d'une sémantique claire, un haut pouvoir de description de concepts et des mécanismes de raisonnement permettant de les valider.

L'ontologie utilisateur (comme les autres) est rédigée en OWL (Ontology Web Language) - un langage de représentation de la connaissance qui est développé pour le Web sémantique et implémentée dans l'outil logiciel Protégé2000.

En se limitant à un sous-ensemble du langage OWL (OWL-DL), un type de raisonnement particulier, la classification automatique, devient faisable. Cela peut être réalisé avec l'outil d'inférence RACER.

5. Organisation du mémoire

La présentation de notre travail est divisée en deux parties. La première présente un état de l'art, la seconde présente à la fois les aspects théoriques et les implémentations qui en découlent. Ces deux parties peuvent être lues de façon indépendante, mais l'état de l'art présente cependant des définitions qui peuvent faciliter la compréhension de certains concepts et de la démarche intellectuelle adoptée dans la deuxième partie.

Nous revenons ici sur le contenu de chacune des parties de manière détaillée. Cette thèse est composée de 5 chapitres regroupés en 2 parties :

Première partie État de l'Art: chapitres 1 à 3

Pour une meilleure compréhension du contexte de notre étude, nous avons tenu à faire un état de l'art dans différents domaines voisins de nos recherches. La diversité et l'étendue des travaux qui ont été menés, par exemple dans le domaine des modélisations des utilisateurs ou dans celui des systèmes de personnalisation nous ont contraint à fixer certaines limites. Pour cette raison la présentation que nous en faisons n'est pas exhaustive, mais nous espérons que le lecteur en retirera une vision assez précise des tendances actuelles et des évolutions technologiques qui ont conduit jusqu'à elles.

De plus ce survol de plusieurs domaines permettra d'introduire le vocabulaire nouveau qui sera utilisé dans la suite de ce mémoire. On pourra comparer la lecture de cette première partie à la construction d'une « ontologie mentale » sur laquelle sera basée la suite de notre présentation.

L'état de l'art du mémoire est organisé autour de trois axes principaux : La personnalisation de l'information, la modélisation des utilisateurs qui constituent l'un des piliers de la personnalisation, et enfin la représentation des connaissances à travers les ontologies.

Il s'agit donc de présenter le cadre scientifique de l'étude menée en revenant en détail sur la problématique et les solutions apportées pour chaque domaine.

Chapitre 2 : Modélisation des utilisateurs

La modélisation de l'utilisateur est le centre d'intérêt principal de cette thèse et sera discutée en détail dans le premier chapitre dans lequel nous proposons un état de l'art sur les modèles des utilisateurs, le contenu des profils et les techniques de construction de ces derniers.

Chapitre 1 : Personnalisation de l'information –Concepts et précisions

Liée aux états de l'art développés dans le contexte des modèles utilisateurs, ce chapitre va permettre d'aboutir à une compréhension des concepts propres à la personnalisation tout en les classifiant en fonction de quatre grandes interrogations : *Qu'est ce que la personnalisation ? ; Pourquoi personnaliser ? ; Que personnaliser ? ; Comment personnaliser ?*.

Chapitre 3 : Les langages de représentation des connaissances et les ontologies

Le chapitre 3 a pour objectif de répondre aux questions suivantes : qu'est ce que le Web Sémantique ? Quelle est la définition d'une ontologie ? Quels sont les composants d'une ontologie? Comment est-il possible de les modéliser ? Y a-t-il des méthodes efficaces de construction d'ontologies, des langages établis pour les utiliser ? Quels apports offrent-elles en comparaison avec une représentation plus classique de la connaissance ? Comment les outils logiciels soutiennent-ils les processus de construction et d'utilisation des ontologies ? Etc.

Deuxième partie : chapitres 4 et 5

Chapitre 4 : Le modèle de l'utilisateur

L'objectif de ce chapitre est de permettre de décider comment les utilisateurs vont être représentés ? Quelles sont les caractéristiques des différents utilisateurs ? C'est sur ces quelques caractéristiques que la personnalisation reposera. Ainsi, tous les types d'utilisateurs potentiels doivent être analysés de manière à découvrir leurs caractéristiques types et à élaborer ensuite un « modèle d'utilisateur ».

Chapitre 5 : Une approche fondée sur des ontologies

Dans ce chapitre nous indiquerons à quels types de besoins nous souhaitons pouvoir répondre et détaillerons notre solution basée sur la définition d'une ontologie pour la modélisation des

utilisateurs en relation avec d'autres ontologies (domaine, documents et application). Le cadre applicatif choisi est celui de la communauté scientifique dans un laboratoire de recherche.

Conclusion générale

Le mémoire s'achève avec une conclusion récapitulant la contribution de cette thèse. Nous présentons également les limites du développement actuel ainsi que les perspectives de cette recherche.

Après avoir dressé les conclusions et perspectives de nos travaux, nous exposons les références bibliographiques que nous avons utilisées.

CHAPITRE 1	15
1. Introduction	16
2. La modélisation des utilisateurs	16
2.1 Classification des modèles de l'utilisateur	17
2.2 Techniques de représentation des modèles utilisateur	18
2.2.1 Le modèle individuel	18
2.2.2 Le modèle de recouvrement (overlay model)	19
2.2.3 Le stéréotype	19
2.2.4 Le modèle partagé	19
3. Construction et maintenance des profils utilisateurs	20
3.1 La représentation des profils utilisateurs	21
3.2 La génération du profil initial	22
3.2.1 L'observation directe	22
3.2.2 Les interviews	23
3.2.3 Les questionnaires	23
3.3 Techniques d'apprentissage des profils	23
3.3.1 Analyse statistique de termes	24
3.3.2 Les techniques de classification	26
3.3.3 Les méthodes de clustering	27
3.4 Retour de pertinence (Relevance feedback)	27
4. Sécurité des données du profil	31
5. Discussion	31

L'objectif de ce chapitre consiste à étudier les différentes méthodes d'élaboration et d'implémentation des modèles utilisateurs, de même que plusieurs procédés possibles de représentation existent dans la littérature. L'acquisition puis la gestion des profils est alors une préoccupation majeure.

1. Introduction

Adapter, personnaliser un document ou une application pour un utilisateur particulier nécessite une description plus élaborée de l'utilisateur et de sa représentation en tant qu'objet à part entière du système¹. Cette représentation de l'utilisateur vise à procurer au système les moyens de procéder aux adaptations souhaitées, d'évaluer la pertinence des objets disponibles (documents, pages Web, etc.) ou d'aider le système à faire des choix.

En amont de la représentation effective d'un utilisateur dans le système, et de l'exploitation de cette représentation, il est primordial de définir *comment* représenter l'utilisateur. En d'autres termes, il faut choisir les informations à retenir pour en construire un modèle. Plusieurs approches sont proposées en section 2 pour guider l'élaboration de tels modèles et permettre de décrire une classification de ceux-ci.

La section 3 décrit les techniques d'obtention et de mise à jour des informations du profil. La section 4 couvre la nature confidentielle des informations du profil et la sécurité des données. Nous concluons enfin ce chapitre par une discussion qui motivera le choix des ontologies pour la modélisation des utilisateurs.

2. La modélisation des utilisateurs

La modélisation des utilisateurs a fait l'objet de très nombreux travaux. Il s'avère que c'est là une problématique fort complexe qu'attestent les emprunts à différentes branches des sciences de l'information et des sciences de l'homme comme la psychologie, l'intelligence artificielle, la pédagogie, les interfaces Homme-Machine, etc. Les résultats restent mitigés avec indéniablement des échecs et des réussites.

Le modèle utilisateur est défini par Mc Tear comme « *une source de connaissances, une base de données sur un utilisateur* » [McTear, 1993]. Plus précisément, il s'agit d'un ensemble de données persistantes qui caractérisent un utilisateur ou un groupe d'utilisateurs particuliers. Un tel modèle peut contenir des caractéristiques sur les connaissances, les préférences, les objectifs, les centres d'intérêts, etc. d'un utilisateur.

¹ Le terme « système » désigne ici un système au sens large. La modélisation de l'utilisateur et la gestion de celle-ci peuvent en effet être déléguées par une application à un serveur uniquement dédié à ces activités (voir [Kobsa, 2001] pour une revue des travaux autour des systèmes « génériques » de modélisation de l'utilisateur (*i.e.* indépendants de toute application) et [Fink et Kobsa, 2000] pour une comparaison des approches centralisées (*i.e.* la modélisation de l'utilisateur est embarquée dans l'application) et décentralisée (elle est déléguée à un serveur).

Mais, pourquoi avons nous besoin de modéliser les utilisateurs ? André Kok pose la question différemment [Kok, 1991]: « *How is adapting to individual users affecting the system's processing* ». En effet, la modélisation des utilisateurs est souvent utilisée pour adapter un système à un utilisateur, mais en quoi l'adaptation affecte, ou plutôt améliore le fonctionnement d'un système ? La première réponse est aussi la plus évidente. L'adaptation permet d'améliorer la convivialité du système. Par exemple, le système peut mettre en avant les informations les plus intéressantes pour l'utilisateur ou plus simplement l'aider dans sa tâche.

Nous présentons dans ce qui suit un rapide survol sur les principales approches proposées pour guider l'élaboration de tels modèles et qui permettent de décrire une classification de ceux-ci (section 2.1), ensuite, nous décrivons les techniques de représentations associées (section 2.2).

2.1 Classification des modèles de l'utilisateur

Daniels [Daniels, 1986] oppose deux classes de modèles de l'utilisateur (ou modèle utilisateur). 1) Les modèles *quantitatifs et empiriques* dont l'objectif est l'étude du comportement externe de l'utilisateur. 2) Les *modèles analytiques et cognitifs* s'intéressent à la modélisation du comportement interne de celui-ci. Ils favorisent l'identification des connaissances et processus cognitifs de l'utilisateur. La représentation de l'utilisateur doit, selon Daniels, tenir compte de ces deux catégories de modèle en raison de la complémentarité qu'elles offrent. Une telle approche permet en effet d'acquérir des informations relatives aux *savoir-faire* et *savoir* de l'utilisateur. Ces deux types d'information peuvent être exploités pour adapter le système.

Quant à Rich [Rich, 1999], il classe les modèles d'utilisateurs suivant trois dimensions principales :

- le *modèle canonique* opposé à une collection de modèles d'utilisateurs individuels
- Le *modèle d'utilisateurs explicite* construit par l'utilisateur ou le concepteur du système opposé au modèle implicite construit par l'ordinateur sur la base du comportement de l'utilisateur.
- Le *modèle basé sur des caractéristiques persistantes* (à long terme) opposé au modèle basé sur les caractéristiques ponctuelles.

Une autre classification possible que nous citons distingue les modèles des utilisateurs en fonction de la connaissance qu'ils véhiculent [Koch, 2000]. Trois modèles sont ainsi généralement mis en évidence :

- les *modèles des connaissances du domaine* qui maintiennent la connaissance que l'utilisateur a du domaine ;
- les *modèles des connaissances indépendantes du domaine* qui concernent les connaissances générales ou non spécifiques au domaine, les centres d'intérêts, acquis (*background*) de l'utilisateur.
- les *modèles psychologiques ou cognitifs* qui font référence aux préférences, capacités ou handicaps, et traits de personnalité de l'utilisateur. De tels modèles sont notamment utilisés dans le domaine éducatif pour représenter les styles d'apprentissage, types de raisonnement, capacités de concentration, etc., des utilisateurs).

D'autres méthodes, comme la reconnaissance de plan (séquence ordonnée d'actions) ou l'apprentissage symbolique, sont présentées dans [PY, 1998]. L'utilisation des réseaux bayésiens est décrite dans [Gonzales et Wuillemin, 1998].

Après avoir décrit différentes classifications de modèles utilisateur issues de la littérature, nous présentons dans la section suivante les techniques les plus couramment utilisées pour représenter de tels modèles.

2.2 Techniques de représentation des modèles utilisateur

Différentes techniques sont utilisées pour représenter les utilisateurs et structurer la connaissance à leur propos. Nous décrivons brièvement les approches les plus utilisées [Villanova, 2002]. Les modèles que nous allons présenter ici peuvent être combinés afin de réaliser le modèle utilisateur qui conviendra à l'application développée :

2.2.1 Le modèle individuel

Ce modèle regroupe les caractéristiques propres à un individu. Il s'agit d'informations pouvant être soit renseignées par l'utilisateur, soit déduites par le système, soit acquises lors de l'utilisation de l'application.

2.2.2 Le modèle de recouvrement (overlay model)

Il ne peut exister que si le système possède un modèle de domaine car le modèle de recouvrement associe pour un utilisateur particulier, une valeur à chaque concept du modèle de domaine. Chaque valeur correspond à une estimation du niveau de connaissance de l'utilisateur pour le concept. On peut donner des précisions sur l'état des connaissances : binaires (su, pas su), qualitatives (bon, moyen, pauvre) ou quantitatives (la probabilité qu'il connaisse ce concept est de...). Le modèle de recouvrement est facile à mettre à jour mais souvent difficile à initialiser surtout au niveau de la mesure du niveau de connaissance. Les actions de l'utilisateur sont analysées pour augmenter ou réduire le niveau de connaissance des concepts du domaine [Brusilovsky et Eklund, 1998].

2.2.3 Le stéréotype

Le modèle de stéréotype correspond à un condensé des caractéristiques les plus représentatives d'un groupe - ou classe - d'individus, que nous pouvons qualifier de valeurs par défaut [Rich, 1989]. Des combinaisons de couples (items, valeurs) définissent différents stéréotypes tels que *novice*, *avancé*, *expert*. L'utilisateur est donc associé à une des catégories élaborées et hérite de ses propriétés.

L'utilisateur dispose, par la suite, des personnalisations réalisées pour le stéréotype. Ce manque d'individualisation de la personnalisation peut être contré par le recours à des approches mixtes qui consistent à initialiser un modèle utilisateur à l'aide d'un stéréotype, puis à l'affiner en utilisant un modèle de recouvrement par exemple. Un exemple est donné par les travaux de [Cannataro et al., 2001] dans lesquels l'association d'un utilisateur à un stéréotype est réévaluée, en fonction de ses actions, par le biais d'une méthode probabiliste.

2.2.4 Le modèle partagé

Il s'agit d'un modèle réutilisable dans plusieurs applications. En effet, de nombreuses caractéristiques d'un utilisateur sont utilisées systématiquement par toutes les applications adaptatives. L'idée du modèle partagé est de disposer d'une partie commune et de parties spécifiques à chacune des applications ou tâches à réaliser. Cette approche permet d'une part de partager des informations entre applications et d'autre part d'obtenir plus facilement une participation des utilisateurs qui n'ont pas à redéfinir un modèle pour chacune des applications. Kobsa et Wahlster [Kobsa et Wahlster, 1989] proposent de disposer d'un modèle central et de plusieurs sous-modèles disjoints.

La modélisation des utilisateurs a donc pour but de déterminer les caractéristiques à retenir pour construire des profils utilisateurs. Dans la section suivante, nous allons présenter les différentes techniques de construction et de mise à jour des profils d'utilisateurs en établissant un cadre de référence pour caractériser les différentes connaissances qui peuvent constituer un profil.

3. Construction et maintenance des profils utilisateurs

La constitution d'un profil d'utilisateur ou d'une communauté d'utilisateurs peut se faire de différentes façons selon la nature des informations constituant le profil mais aussi selon la nature des applications. Elle requiert plusieurs techniques de conception: les techniques de représentation et d'adaptation des profils, les techniques de construction d'un profil initial, la source de feedback qui représente les intérêts de l'utilisateur et les techniques d'apprentissage des profils. La figure 1.1 montre les relations entre ces techniques dans la génération et la maintenance des profils d'utilisateur.

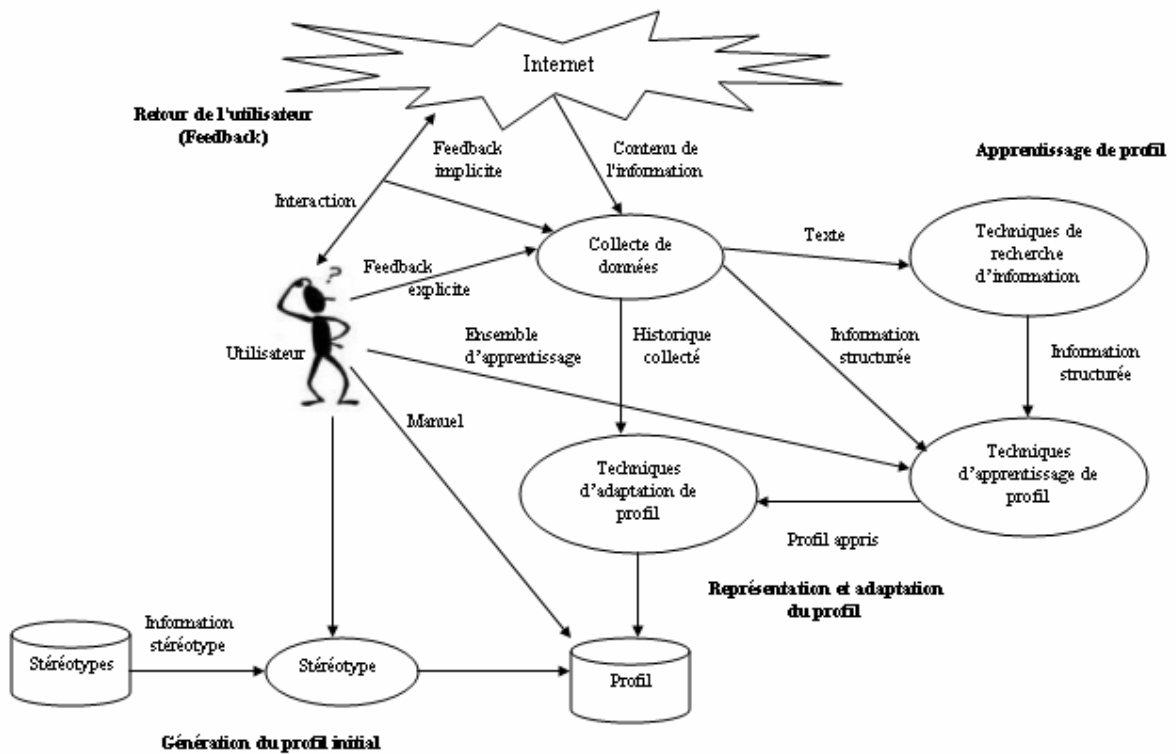


Figure 1.1 – Génération et maintenance des profils d'utilisateur

Nous allons essayer de décrire le fonctionnement de base des approches de construction et de mise à jour des données du profil et de donner des exemples utilisés pour la personnalisation.

3.1 La représentation des profils utilisateurs

Un profil utilisateur regroupe toutes les données liées à un utilisateur, dans les limites de l'intérêt porté à ces données. C'est la finalité du système qui va régir la sélection des attributs collectés. De fait, le contenu et la représentation d'un profil dépendent énormément de l'application qui l'utilise, et c'est aussi ce qui fait que les travaux sur la personnalisation d'information prennent des formes si variées. On s'en rend assez facilement compte en parcourant la littérature à ce sujet (cf. chapitre 2).

Dans cette section, nous allons voir les informations que les systèmes stockent le plus souvent sous forme de profils afin de personnaliser leurs services. Le profil peut contenir des informations brutes, caractéristiques de l'utilisateur (nom, prénom, pseudonyme, adresse, ...). Généralement, ce sont des informations qui viennent directement de l'utilisateur, par un procédé de saisie. Outre les informations brutes, plusieurs techniques peuvent être adoptées pour la construction des profils des utilisateurs.

Le profil utilisateur dans [Salton, 1971], par exemple, s'appuie sur le Vector Space Model. Il est constitué de un ou plusieurs vecteurs définis dans un espace de termes ; les coordonnées correspondent aux poids associés aux termes retenus dans le profil. L'utilisation de plusieurs vecteurs correspond à deux préoccupations : pouvoir prendre en compte des centres d'intérêt multiples et gérer leur évolution dans le temps. Dans ce dernier cas, le système utilise généralement deux groupes de vecteurs, dont l'évolution est gouvernée par des règles différentes (c'est le cas de News Dude [Billsus et Pazzani, 1999]).

Le profil utilisé dans Letizia [Lieberman, 1995] est constitué d'une liste de mots-clés qui représente les éléments pertinents pour un utilisateur.

Dans Syskill & Webert [Pazzani et al., 1996], un ensemble de triplets constitués par un mot-clé, la probabilité que ce mot appartienne à un document pertinent, et la probabilité qu'il appartienne à un document non-pertinent sont déterminées pour chaque utilisateur. Cet ensemble constitue le profil et caractérise la pertinence des documents pour un utilisateur.

Le système Amalthea [Moukas, 1997] combine jugement de l'utilisateur et détection automatique de ses centres d'intérêt. Il construit et fait évoluer un profil d'intérêt constitué de mots-clés pondérés aussi bien à partir des actions passées et en cours de l'utilisateur, des informations contenues dans les listes de sites préférés et/ou d'historique, que d'un *feedback* explicite de la part de l'utilisateur.

Pretschner et Gauch [Pretschner et Gauch, 1999b] proposent une représentation plus élaborée de l'intérêt de l'utilisateur qui s'appuie sur une hiérarchie de concepts construite à partir des pages visitées et du temps passé sur chacune d'elles. Certaines approches ajoutent

une évaluation fournie explicitement par l'utilisateur auquel il est demandé de classer des pages visitées par ordre de préférence [Pazzani et al., 1996].

Jugeant cette façon de faire trop contraignante, Chan propose une technique d'estimation de l'intérêt d'une page pour un utilisateur basée sur l'observation de différents critères (*i.e.* rien n'est demandé à l'utilisateur) [Chan, 1999]. Les critères retenus sont la fréquence à laquelle l'utilisateur accède à la page, le fait que cette page appartienne ou non aux «favoris», la durée de la visite de cette page, la date de la dernière visite, le pourcentage de liens de cette page qu'a déjà visités l'utilisateur.

La représentation concrète du profil peut aussi prendre d'autres formes : l'historique des achats [Amazon, 2004], la navigation Web ou les E-mails [Goldberg et al., 1992], les réseaux sémantiques [Potter et Trueblood 1988], les réseaux associatifs [Riordan et Sorensen 1995], les classificateurs comprenant les réseaux de neurones [Boone, 1998], les arbres de décision [Krulwich et Burkey 1996], les règles d'induction ou les réseaux bayésiens [Jensen, 1996] (cf. section 3.3.2), le modèle des N-Grammes, utilisé pour étudier la prédictibilité d'apparition de suites de chaîne [Sorensen et McElligot, 1995], des matrices d'estimations [Hayes et Cunningham, 2000], un ensemble de caractéristiques démographiques [Kobsa et al. 2001], etc.

Le tableau 1.1 montre les techniques de représentation de profil d'utilisateur employées par les différents systèmes analysés. Chaque système propose un modèle de profil adapté au service qu'il offre.

3.2 La génération du profil initial

Intuitivement la manière la plus simple d'initialiser un profil est la saisie manuelle des paramètres par l'utilisateur. Il est considéré que l'utilisateur connaît mieux ses exigences et de ce fait, il peut saisir les paramètres dont a besoin le système. On peut exposer quelques méthodes d'obtention des informations sur les utilisateurs :

3.2.1 L'observation directe

Il s'agit de la méthode la plus précise. Elle permet d'identifier des classes d'utilisateurs ainsi que les tâches de ces derniers. Malheureusement, il s'agit d'une méthode très coûteuse qui nécessite des personnes qualifiées derrière chacun des individus observés.

3.2.2 Les interviews

Cette technique permet d'obtenir un autre type d'information, l'expérience, les opinions, les motivations comportementales mais surtout les avis sur les outils existants. Ils sont plus courts et moins coûteux que la technique d'observation, néanmoins, ils nécessitent aussi du personnel qualifié.

3.2.3 Les questionnaires

Les questionnaires permettent d'obtenir, à moindre coût, un maximum de données. Les résultats obtenus permettent des études statistiques et des généralisations plus fortes que les interviews. Les questionnaires peuvent être collectés par des personnes non expérimentées. Ils permettent d'avoir à la fois un aperçu de la situation et des points d'information plus précis.

Cependant, la tendance est à minimiser les actions de l'utilisateur parce que le processus de saisie manuelle peut être long et bien souvent l'utilisateur a des idées floues sur ses demandes et par conséquent a du mal à exprimer clairement ses intentions, et il est très possible que cet utilisateur fausse lui-même les données le concernant. C'est en cela que la collection informatisée d'informations sur l'utilisateur est également délicate.

Afin de résoudre ces problèmes, ils existent plusieurs méthodes de capture des paramètres du profil de façon semi-automatique (à travers des stéréotypes et des données d'apprentissage) ou automatique (méthodes liées à l'interprétation des activités de l'utilisateur).

3.3 Techniques d'apprentissage des profils

Il est possible d'obtenir de l'information par l'intermédiaire d'un outil d'apprentissage. En effet, il est intéressant, et très utile d'y ajouter un algorithme d'apprentissage pour obtenir des informations essentiellement comportementales sur l'utilisateur.

Mais certains systèmes maintiennent l'information acquise directement par le système pour construire un profil utilisateur, donc, ils n'ont pas besoin d'une technique d'apprentissage de profil. Il y a trois principaux types de ces systèmes:

- Systèmes qui acquièrent l'information du profil utilisateur d'une base de données. Par exemple, les systèmes de commerce électronique ([Amazon, 2001]; [CDNow, 2001]; [Cunningham et al., 2001]), qui extraient l'information à partir d'une base de données des produits et gardent une liste d'achat comme profil.

- Systèmes de filtrage collaboratif ([Goldberg et al., 1992]; [Resnick et al., 1994]; [Shardanand et Maes, 1995]) qui gardent une matrice avec les estimations utilisateur/article sous forme de profil.
- Les systèmes qui créent un profil initial à partir d'un stéréotype (cf. section 2.2.3) et ne le modifient pas [Krulwich, 1997].

Dans ce qui suit, nous passerons en revue, quelques techniques de construction de profils telles que l'analyse statistique des termes (section 3.3.1), les méthodes de classification (section 3.3.2), les méthodes de clustering (section 3.3.3), etc. Ces approches se différencient suivant les problèmes qui sont les plus appropriés à résoudre. Certaines applications peuvent étudier les éléments qui ont un impact sur une variable donnée (par exemple, les caractéristiques démographiques des individus qui répondent à un mailing). D'autres applications permettent de segmenter la population en différentes classes, de regrouper les individus similaires ou de détecter les comportements atypiques.

3.3.1 Analyse statistique de termes

Ce processus peut être vu comme un pré-traitement des données et leur mise en forme normale pour pouvoir les manipuler ensuite. L'idée principale consiste à analyser le contenu d'un document et d'extraire des mots clés significatifs qui décrivent son contenu. Ces mots clés sont stockés pour être utilisés ensuite afin de comparer des éléments entre eux ou avec les préférences de l'utilisateur lorsqu'elles sont exprimées sous la forme de mots clés.

Il existe différentes structures de stockage et de représentation des mots clés en fonction du contexte dans lequel le profil est utilisé [Soltysiak et Crabtree, 1998]. En plus dans certains cas, on peut rajouter un poids qui exprime l'importance de chaque terme et qui est souvent associé à la fréquence d'apparition du terme.

Un terme est ici une suite de caractères alphanumériques délimités par des espaces ou des ponctuations. Les majuscules sont souvent converties en minuscules. Après avoir extrait ces termes du corpus, chaque document est alors représenté par un vecteur où chaque terme est pondéré selon une fonction de poids. La Figure 1.2 montre deux documents et leurs vecteurs correspondants où le poids de chaque terme est la fréquence d'apparition de ce terme dans un document donné.

Document 1 : Le grand oiseau mange le petit insecte
Document 2 : Le grand insecte mange de l'herbe

	Document 1	Document 2
Le	2	1
grand	1	1
oiseau	1	0
mange	1	1
petit	1	0
insecte	1	1
de	0	1
l	0	1
Herbe	0	1

Figure 1.2 - Exemple de deux documents et de leurs vecteurs représentatifs correspondants où les termes sont pondérés par leur fréquence dans chaque document.

On remarque ensuite que beaucoup de mots ont des formes légèrement différentes, mais leur sens restent le même ou très similaire. C'est notamment le cas des mots conjugués. Par exemple, les mots suivants ont des sens très similaires: *transformer*, *transforme*, *transforment*, *transformation*, *transformateur*, ... Ces mots ont la même racine (lemme). Ainsi, si on arrive à éliminer les terminaisons de mots, et garder seulement la racine, on a donc une forme identique pour eux. C'est l'idée qui conduit à utiliser la lemmatisation. Plusieurs techniques permettent de retrouver la racine d'un mot. La plus utilisée a été développée par M.F Porter [Porter, 1980].

3.3.1.1 Pondération des termes dans un vecteur

Plusieurs mesures ont été proposées pour caractériser un terme donné dans un document. Soit TF , la fréquence d'apparition d'un terme t dans un document d . Le plus simple est de représenter chaque terme par sa fréquence TF . Robertson et Spark Jones proposent d'utiliser $\log(TF+1)$ [Robertson et Spark-Jones, 1976]. Le système Scatter/Gather utilise la racine carrée de TF [Cutting et al., 1992].

Parmi toutes les fonctions proposées, la plus connue est la mesure $TDIDF$ proposée par G. Salton [Salton et Buckley, 1988]. Cette mesure utilise non seulement la fréquence TF d'un mot dans un document, mais également l'inverse de la fréquence du terme dans le corpus IDF . Plus formellement, IDF est donnée par :

$$IDF(t) = \log\left(\frac{N}{n_t}\right).$$

Où N est le nombre de documents du corpus et n_t est le nombre de documents dans lesquels le terme t apparaît au moins une fois. Le poids $TFIDF$ pour un terme t dans un document d est le produit de TF par IDF :

$$TFIDF(t, d) = TF(t, d) * IDF(t)$$

Un mot est donc d'autant plus représentatif qu'il apparaît fréquemment dans le document et rarement dans les autres.

3.3.1.2 Contrôler la dimension des vecteurs de documents

La dimension des vecteurs est énorme si l'on considère chaque terme trouvé dans les documents à classer. Ce problème est d'autant plus important lorsqu'on traite des documents très hétérogènes comme ceux présents sur le Web.

Une solution consiste à éliminer les mots usuels comme les déterminants, articles et conjonctions. En français, les mots « de », « un », « les », etc. sont les plus fréquents. En anglais, ce sont « of », « the », etc. Afin d'éliminer ces mots de force, on utilise une liste, appelée stop-list (ou parfois anti-dictionnaire) qui contient tous les mots qu'on ne veut pas garder.

L'analyse statistique des mots clés est la méthode la plus utilisée parce qu'elle est basée sur des techniques bien comprises d'extraction de mots clés. Les inconvénients de cette approche se trouvent dans le fait qu'elle ne peut être appliquée que sur des éléments textuels et que les mots sont analysés en isolation avec le reste du document ce qui entraîne une perte d'information contextuelle pouvant dégrader l'exactitude des données du profil. Une fois les mots clés extraits, on peut comparer deux éléments sur la base de la distance entre les vecteurs les représentant.

3.3.2 Les techniques de classification

Le but de la classification est d'attribuer un élément donné à un groupe existant. Les groupes sont connus à l'avance et sont donnés en paramètre à l'algorithme qui ensuite attribue les objets à un groupe selon certains critères.

Les algorithmes de classification les plus utilisés dans la littérature sont le raisonnement par cas [Bradley et al., 2000], les classificateurs bayésiens [Fink et Kobsa, 2000], les réseaux de neurones et les règles d'associations[Mobasher et al., 2000]...

3.3.3 Les méthodes de clustering

A la différence de la classification, dans les techniques de clustering les groupes d'objets ne sont pas connus à l'avance et c'est l'algorithme qui se charge de la répartition des éléments en essayant de minimiser la similarité entre les éléments de deux clusters différents et de maximiser la similarité entre les éléments du même cluster.

Un exemple de système qui utilise le clustering est donné par [Mobasher et al., 2000]. L'article présente deux techniques de clustering. La première PACT (Profile Aggregation based on Clustering Transactions) consiste à regrouper les transactions similaires d'un utilisateur. Chaque transaction est un vecteur multidimensionnel de pageviews (vues sur les pages Web), et le regroupement est fait à la base de la distance ou de la similarité entre les vecteurs. Tandis que ARHP (Association Rule Hypergraph Partitioning) prend en compte les pageviews apparaissant souvent ensemble. ARHP peut être utilisé lorsqu'on veut produire un petit ensemble de recommandations spécifiques.

Les deux techniques (PACT et ARHP) peuvent être utilisées pour une personnalisation basée sur les données de navigation des utilisateurs.

3.4 Retour de pertinence (Relevance feedback)

Par un processus de retour de pertinence (« Relevance feedback » en anglais), l'utilisateur n'indique pas seulement l'information pertinente mais aussi l'information non-pertinente (figure 1.1). Le système utilise ces informations pour ajuster la description du profil utilisateur qui reflètera les nouvelles préférences. Cette procédure provoque une évolution constante du profil utilisateur. Ce profil se stabilise après plusieurs recherches, une fois que le profil est défini au plus proche des préférences utilisateur.

Le tableau 1.1 montre les techniques de construction et de maintenance des profils d'utilisateurs, employées par différents systèmes.

Bien qu'il existe déjà plusieurs techniques permettant de capturer les préférences d'un utilisateur, leur utilisation n'est pas encore bien comprise pour permettre la construction et la mise à jour de son profil de façon automatique et complètement transparente pour lui. Actuellement nous pouvons espérer que ceci se fasse avec une implication minimale de l'utilisateur. Le principe est que l'utilisateur doit avoir le contrôle de son profil à tout moment afin de pouvoir invalider les mises à jour incorrectes du profil.

Nom du système	Références	Technique de représentation	Génération du Profil initial	Technique d'apprentissage	Retour de pertinence
ACR News	[Mobasher et al., 2000]	Ensemble de couples (page, poids) où page est un objet de l'espace de recherche et poids est l'intérêt de l'utilisateur pour cette page.	Ensemble d'apprentissage	Induction de règles, clustering	Implicite (historique de navigation)
Amazon	[Amazon, 2004]	Historique des achats avec estimations	Manuel	Aucun apprentissage	Explicite (évaluations), implicite (historique des achats)
Amalthea	[Moukas, 1997]	Vecteur de mots-clés pondérés	Vide	Indexation d'information (TF-IDF), lemmatisation	Explicite (évaluations)
Beehive	[Huberman et Kaminsky, 1996]	Clusters (Vecteur de mots-clés pondérés)	Vide	Clustering	Implicite
Bellcore Video Recom	[Hill et al., 1995]	Matrice d'estimations d'Utilisateur/article	Ensemble d'apprentissage	Aucun apprentissage	Explicite (évaluations)
CDNow	[CDNow, 2001]	Historique d'achat avec estimations	Vide	Aucun apprentissage	Explicite (évaluations), Implicite (historique des achats)
Fab	[Balabanovic et Shokam, 1997]	Vecteur de mots-clés pondérés	Vide	Indexation d'information (TF-IDF)	Explicite (évaluations)
GroupLens	[Resnick et al., 1994]	Couples (objet, note). Matrice d'estimations utilisateur/article.	Vide	Aucun apprentissage	Explicite (évaluations, commentaires des textes), implicite (temps passé)
IfWeb	[Minio et Tasso, 1996] [Asnicar et Tasso, 1997]	Attributs pondérés à valeurs multiples, réseau sémantique pondéré	Ensemble d'apprentissage, stéréotype	Sélection de caractéristiques (stop-list, lemmatisation)	Explicite (évaluations)
InfoFinder	[Krulwich et Burkey, 1995]	Arbre de décision	Ensemble d'apprentissage	Sélection de caractéristiques (heuristiques), arbre de décision	Explicite (évaluations)
INForm	[Riordan et Sorensen, 1995]; [Sorensen et al., 1997]	Réseau associatif pondéré	Ensemble d'apprentissage	Sélection de caractéristiques (stop-list, lemmatisation)	Explicite (évaluations)

Nom du système	Références	Technique de représentation	Génération du Profil initial	Technique d'apprentissage	Retour de pertinence
LaboUr	[Schwab et al., 2001]	Vecteur probabiliste de caractéristiques, vecteur booléen de caractéristiques	Ensemble d'apprentissage	Indexation d'information (Booléen)	Implicite (liens, temps passé)
Let's Browse	[Lieberman et al., 1999]	Vecteur de mots-clés pondérés	Ensemble d'apprentissage	Indexation d'information (TF-IDF)	Implicite (liens, temps passé)
Letizia	[Lieberman, 1995]	Vecteur de mots-clés pondérés	Vide	Indexation d'information (TF-IDF)	Implicite (liens, temps passé)
LifeStyle Finder	[Krulwich, 1997]	Caractéristiques démographiques	Stéréotype	Aucun apprentissage	Explicite (évaluations), implicite (historique des achats)
MovieLens	[Good et al., 1999]	Vecteur de mots-clés pondérés, règles d'induction	Ensemble d'apprentissage	Indexation d'information (TF-IDF), induction de règles	Explicite (évaluations)
News Dude	[Billsus et Pazzani, 1999]	Vecteur booléen de caractéristiques	Ensemble d'apprentissage	Indexation d'information (TF-IDF)	Explicite (aimer/détester, déjà connu, connaître encore plus)
NewsWeeder	[Lang, 1995]	Vecteur de mots-clés pondérés	Ensemble d'apprentissage	Indexation d'information (TF-IDF)	Explicite (évaluations)
NewT	[Sheth et Maes, 1993]	Vecteur de mots-clés pondérés Exemples de situation-action	Ensemble d'apprentissage	Sélection des caractéristiques (stop-list, lemmatisation), Indexation d'information (TF-IDF)	Explicite (aimer/détester)
Personal WebWatcher	[Mladenic, 1996]	Vecteur probabiliste de caractéristiques	Manuel	Indexation d'information (TF-IDF)	Implicite (liens)
PSUN	[Sorensen et McElligot, 1995]	Entités simples (S-entities) et des supervisors	Ensemble d'apprentissage	Sélection des caractéristiques (lemmatisation), modèle des N-Grammes	Explicite (évaluations)
Re:Agent	[Boone, 1998]	Vecteur de mots-clés pondérés, réseau de neurones	Manuel, Ensemble d'apprentissage	Sélection de caractéristiques, Indexation d'information (TF-IDF), clustering, réseaux de neurones	Pas de feedback
Recommender	[Basu et al., 1998]	Règles d'induction	Ensemble d'apprentissage	Induction de règles	Explicite (évaluations)
SIFT Netnews	[Yan et Garcia-Molina, 1995]	Vecteur booléen de caractéristiques, Vecteur de mots-clés pondérés, arbre de décision	Ensemble d'apprentissage	Indexation d'information (Booléen), Indexation d'information (TF-IDF)	Explicite (aimer/détester)

Nom du système	Références	Technique de représentation	Génération du Profil initial	Technique d'apprentissage	Retour de pertinence
SiteIF	[Stefani et Strappavara, 1998]	Réseaux sémantiques pondérés	Vide	Sélection de caractéristiques (stop-list, lemmatisation)	Implicite (liens)
Smart Radio	[Hayes et Cunningham, 2000]	Matrice d'estimations utilisateur/article	Ensemble d'apprentissage	Aucun apprentissage	Explicite (évaluations), implicite (sauvegarde)
Syskill & Webert	[Pazzani et al., 1996]	Vecteurs de mots clés et les probabilités d'apparition de chaque mot dans les pages intéressantes et non intéressantes	Manuel, stéréotype	Sélection de caractéristiques (stop-list), Indexation d'information (Boolean), Indexation d'information (TF-IDF), arbre de décision	Explicite (évaluations)
Tapestry	[Goldberg et al., 1992]	Messages et annotations classés	Vide	Aucun apprentissage	Explicite (aimer/détester, commentaires), implicite
Webmate	[Chen et Sycara, 1998]	Vecteur de mots-clés pondérés	Vide	Indexation d'information (TF-IDF)	Explicite (like/dislike)
WebSell	[Cunningham et al., 2001]	Produits intéressants/ non interesting	Vide	Aucun apprentissage	Explicite
Websift	[Cooley et al., 1999]	Induction de règles, modèles, statistiques	Ensemble d'apprentissage	Induction de règles	Implicite (historique de navigation)
WebWatcher	[Armstrong et al., 1995]	Vecteur booléen de caractéristiques	Manuel	Indexation d'information (TF-IDF), Winnow (Chapitre2, section 4.4)	Explicite (but de recherche), implicite (liens)

Tableau 1.1- Techniques de construction et de maintenance des profils utilisateurs

4. Sécurité des données du profil

La nature personnelle des informations du profil nécessite leur sécurisation. La sécurisation des données se fait sur trois niveaux : le contrôle de l'accès aux informations du profil, la sécurisation des échanges des données et la confidentialité de l'identité de l'utilisateur. Un des paramètres qui jouent un rôle très important pour la sécurité des données est le support utilisé pour le stockage et le traitement des informations.

5. Discussion

Il existe, comme nous l'avons vu, de nombreux moyens de définir un profil. Les données contenues dans celui-ci sont dépendantes des applications qui l'utilisent. Ce sont souvent des modèles simplifiés qui manquent de structuration et ne sont pas généralisables et réutilisables par d'autres approches. Par ailleurs, elles se contentent de catégoriser les informations de profil sans expliciter les corrélations qui existent entre elles.

Une approche simpliste pour la construction d'un profil utilisateur est de construire ce profil en utilisant, pour le décrire, des mots-clés qui seront fournis par l'utilisateur lui-même. Cette approche est simple car elle fait appel à l'utilisateur à tous les niveaux. En fait, l'utilisateur pourrait trouver assez difficile de fournir les mots-clés qui décrivent convenablement ses préférences. Ainsi, bien que ceci soit une possibilité, nous ne pouvons demander à l'utilisateur de décrire précisément son profil, car ceci serait laborieux.

En plus, l'utilisation très généralisée de profils utilisateurs à l'aide de mots clefs pondérés présente l'inconvénient de ne pas maîtriser les effets de synonymie, d'homonymie, et d'effets de sens contextuels. Autrement dit, les mots ignorent la sémantique de leur contexte d'utilisation. Par exemple un profil d'utilisateur qui mentionne « réseaux de neurones » peut être compris comme un utilisateur qui s'intéresse aux réseaux et aux neurones. Les moteurs de recherche permettent de palier en partie à cet inconvénient proposant la recherche de chaînes de caractères ou bien en autorisant d'effectuer des requêtes à l'aide d'opérateurs booléens (par exemple : réseaux + neurones).

Enfin, le problème de capture des paramètres du profil reste non résolu. Les données contenues dans le profil doivent être correctes et justes car elles ont un impact direct sur les résultats qui vont être obtenus. Le fait que l'utilisateur a souvent des idées floues sur ses préférences complique l'extraction des caractéristiques pertinentes et nécessite un processus de découverte de ses préférences. Ce processus doit être fait avec l'implication minimale du côté de l'utilisateur et en même temps de fournir des données correctes par rapport à

l'utilisateur.

En plus, les profils des utilisateurs sont fondés sur des informations provenant de sources hétérogènes, la réutilisation et le partage sont des enjeux qui peuvent être assurés par les ontologies, qui sont très utilisées dans les travaux sur le Web Sémantique. L'ontologie peut jouer un rôle important pour la représentation et la gestion des connaissances relatives aux utilisateurs. Elle fournit un cadre unificateur pour réduire et éliminer les confusions conceptuelles et terminologiques et assurer une compréhension partagée par la communauté visée. Elle peut donc permettre la communication entre des utilisateurs, même s'ils travaillent dans des contextes différents, avec des besoins et des points de vue différents.

On peut donc imaginer des améliorations à ces systèmes. En effet, ces systèmes n'exploitent pas la notion d'ontologie. Les ontologies, qui vont faire l'objet du troisième chapitre, sont des graphes de concepts qui permettent de structurer l'information, et de la placer dans des catégories de concepts ou sous-catégories en fonction de son contenu.

Avec ce chapitre, nous avons traité le premier aspect constituant l'état de l'art de ce mémoire. Un second aspect de l'état de l'art est abordé à travers l'étude des différentes façons d'envisager la personnalisation en fonction des modèles des utilisateurs.

 Personnalisation de l'information – Concepts et précisions

CHAPITRE 2 : Personnalisation de l'information	33
<i>1. Introduction</i>	34
<i>2. Définition de la personnalisation</i>	35
<i>3. Les éléments sujets à la personnalisation</i>	36
3.1 Personnalisation du contenu informationnel	36
3.2 Personnalisation de la navigation	38
3.3 Personnalisation de la présentation	39
<i>4. Les services de personnalisation</i>	40
4.1 Le filtrage des résultats	40
4.2 Le re-ordonnement des résultats de la requête	41
4.3 La recommandation	42
4.4 Les extensions des langages pour supporter la personnalisation	43
<i>5. Méthode de personnalisation</i>	43
5.1 Détermination Explicite	44
5.2 Détermination Implicite	44
5.3 Customisation	45
<i>6. Conclusion</i>	46

*L'*objectif de ce chapitre consiste à représenter ce que nous entendons par le terme « Personnalisation » et comment le processus de personnalisation est abordé aujourd'hui de façon générale. L'importance de la prise en compte de l'utilisateur est également mise en évidence.

1. Introduction

L'accès à une information pertinente, adaptée aux besoins et au contexte de l'utilisateur est un challenge dans un environnement Internet, caractérisé par une prolifération des ressources hétérogènes (données structurées, documents textuels, composants logiciels, images), conduisant à des volumes considérables. Au fur et à mesure que ce volume s'accroît et que les données se diversifient, les systèmes de recherche d'informations (moteurs Web, Système de Gestion de Bases de Données, etc.) délivrent des résultats massifs en réponse aux requêtes des utilisateurs, générant ainsi une surcharge informationnelle dans laquelle il est souvent difficile de distinguer l'information pertinente d'une information secondaire ou même du bruit.

La personnalisation de l'information constitue ainsi un enjeu majeur pour l'industrie informatique. Que ce soit dans le contexte des systèmes d'information d'entreprise, du commerce électronique, de l'accès au savoir et aux connaissances ou même des loisirs, la pertinence de l'information délivrée, son intelligibilité et son adaptation aux usages et préférences des utilisateurs constituent des facteurs clés du succès ou du rejet de ces systèmes.

Dans ce chapitre, nous présentons le concept de personnalisation et les grandes questions qui le sous-tendent. Lors de notre recherche bibliographique, nous avons pu lire de nombreux ouvrages et articles traitant tous de la personnalisation, mais chacun abordant le sujet sous un angle différent. Notre objectif est de présenter chaque concept rencontré dans la littérature comme une réponse possible à l'une des quatre grandes questions suivantes : *Qu'est-ce que la personnalisation ? ; Que personnaliser ? ; et Comment personnaliser ?* Nous avons souhaité suivre, à ce stade du mémoire, une approche très générale, indépendante des aspects propres à la mise en œuvre du processus de personnalisation.

La deuxième section est consacrée à la description de la notion de personnalisation, à partir des différents termes employés dans la littérature.

Après avoir défini la personnalisation, nous proposons de parcourir, dans la section 3, les différents éléments pouvant être sujets à la personnalisation. Nous identifions dans la section 4 plusieurs services de personnalisation, ils diffèrent principalement dans la façon dont ils mettent en œuvre les personnalisations : le processus est plus ou moins dynamique, et laisse une plus ou moins grande place au contrôle à l'utilisateur. La section 5 aborde les principales méthodes de personnalisation. Nous concluons ce chapitre en montrant la place prépondérante d'une modélisation de l'utilisateur en vue de la personnalisation.

2. Définition de la personnalisation

Le terme « Personnalisation » connaît un vif succès dans la littérature. Ce mot figure dans un grand nombre d'articles et de publications récentes traitant pourtant des sujets très variés. La conséquence immédiate de cet engouement s'est traduite par une confusion grandissante sur la signification concrète de ce concept. Dans ce contexte, notre première tâche consiste à clarifier ce vaste sujet en considérant diverses définitions proposées dans la littérature.

Le processus de personnalisation est décrit par [Rosenberg, 2001] comme « *une technologie cognitive* » dans la mesure où il s'appuie sur un apprentissage, par la machine, de ce que veut l'utilisateur. Cet apprentissage est basé sur une observation de son comportement au cours des sessions. Le système a, par ailleurs, la capacité de s'adapter au cours du temps.

Pour Kostadinov [Kostadinov, 2003], la personnalisation de l'information se définit par un ensemble de préférences individuelles, par des ordonnancements de critères ou par des règles sémantiques spécifiques à chaque utilisateur ou communauté d'utilisateurs. Ces modes de spécification servent à décrire le centre d'intérêt de l'utilisateur, le niveau de qualité des données qu'il désire ou des modalités de présentation de ces données.

Le Gartner Group [Janowski et Sarner, 2001] définit la personnalisation comme « *toute interaction avec l'utilisateur dans laquelle le message, l'offre ou le contenu a été taillé sur mesure pour un utilisateur ou groupe d'utilisateur spécifiques* ». Cette définition est concise mais déjà très riche. Elle met l'accent sur le caractère bidirectionnel d'une démarche personnalisée (elle parle d'interaction) et spécifie qu'elle porte sur divers aspects : sur les messages adressés directement à l'utilisateur, sur l'offre de services ou de produits, ou encore sur le contenu informationnel.

Si, pour certains auteurs, personnalisation, customisation et adaptation sont synonymes [Mobasher et al., 2000][Kappel et al., 2000][Rossi et al., 2001], d'autres emploient une terminologie différente.

La customisation est un processus contrôlé par l'utilisateur qui effectue un choix entre plusieurs options (cf. section 5.3). Le choix des options guide complètement la réponse fournie par le système [Rosenberg, 2001]. Par opposition à la customisation, la personnalisation véhicule l'idée d'un processus guidé par le système lui-même. Le système gère de la connaissance relative à l'utilisateur et exploite cette connaissance pour décider ce qui doit être présenté à l'utilisateur.

Le processus de personnalisation est également distingué de celui d'adaptation, mais cette distinction n'est pas évidente. Le principal élément différent de la définition de la personnalisation est la prise en compte d'éléments propres à l'environnement de l'application. L'adaptation se charge également des modifications techniques nécessaires à l'emploi des éléments. Nous déduisons que la personnalisation n'est pas un concept distinct de celui de l'adaptation, mais plutôt une sous catégorie de ce dernier (comme illustré dans la Figure 2.1). Toute personnalisation est donc de l'adaptation mais l'inverse n'est pas exact.

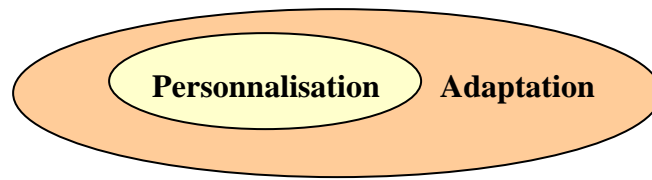


Figure 2.1 - Positionnement de la personnalisation par rapport à l'adaptation [Gavray, 2002]

3. Les éléments sujets à la personnalisation

Dans cette section, nous présentons les différents axes qui peuvent faire l'objet de personnalisations. Cette présentation s'appuie principalement sur les travaux de la communauté hypermédia adaptatifs [Brusilovsky, 2001].

Nous commençons par aborder les possibilités de personnalisation en présentant les éléments personnalisables du contenu informationnel (section 3.1), suivis de ceux propres à la navigation (section 3.2). Nous terminons par les éléments liés à la présentation (section 3.3).

3.1 Personnalisation du contenu informationnel

La personnalisation du contenu consiste principalement à proposer une information qui corresponde aux connaissances de l'utilisateur ou aux activités qu'il doit mener à bien. Le contenu peut être personnalisé de plusieurs façons. Parmi les plus courantes, on distingue les méthodes visant à :

- Fournir un supplément d'information par rapport à un contenu de base défini pour l'ensemble des utilisateurs ;
- Cacher une partie de l'information jugée non pertinente pour un utilisateur donné ; cette méthode peut être vue comme l'inverse de la précédente ;

- Choisir, parmi plusieurs alternatives prédéfinies pour proposer une information, celle qui est la plus adéquate pour un utilisateur donné.

Plusieurs techniques existent pour mettre en oeuvre de telles personnalisations de contenu. Elles portent majoritairement sur des données de type texte, mais peuvent être transposées à des données multimédias [Koch, 2000]. Des exemples de systèmes utilisant les techniques ci-dessous peuvent être trouvés dans [Brusilovsky, 1997], [Brusilovsky et Eklund, 1998]. Les termes en italique sont repris de l'auteur.

Proposer des informations supplémentaires ou, au contraire, en cacher, peut être réalisé par l'association de conditions aux éléments d'information (*conditional text*). Ces conditions expriment généralement le(s) critère(s) requis pour y accéder. Par comparaison avec les valeurs affectées à l'utilisateur pour ces critères, le système décide de montrer ou non l'information. Le même objectif est atteint par le recours à une technique basée sur un principe d'expansion/réduction du texte dans une page Web (*stretchtext*). Une partie de texte (en général un mot ou groupe de mots correspondant à un concept) est associée à une information additionnelle qu'il est possible de faire apparaître. Le système choisit de dévoiler ou non l'information en référence à des spécifications données par un modèle utilisateur.

La méthode basée sur le choix de contenus alternatifs est traduite de deux façons au niveau technique. La première consiste à créer autant de versions de pages Web que nécessaire (*page variants*). Cette technique, simple mais relativement coûteuse, n'est utilisée que dans les systèmes destinés à quelques groupes distincts d'utilisateurs. Une personnalisation individuelle par ce biais est en effet difficilement envisageable. La seconde technique adopte un principe similaire mais à un niveau de granularité plus fin : différentes versions relatives à un concept sont créées (*fragment variants*). Une sélection de la version adéquate est opérée pour construire la page présentée à l'utilisateur.

Enfin, une technique est basée sur une organisation de l'information exploitant la notion de cadres (*frame-based technique*). Le principe est d'insérer dans de tels cadres, toujours en fonction de l'utilisateur, des représentations différentes de l'information à laquelle il vient d'accéder dans la fenêtre principale. Des informations supplémentaires reliées à celle-ci, ou liens vers ces informations, peuvent également être proposé(e)s. Le cadre peut être masqué à la demande de l'utilisateur.

De nombreuses possibilités de personnalisation se situent à ce niveau, parmi elles le filtrage de contenu et les recommandations (cf. sections 4.1 et 4.3).

3.2 Personnalisation de la navigation

La personnalisation de la navigation a pour but d'aider les utilisateurs à trouver leur chemin dans l'espace d'information en adaptant la présentation des liens en fonction des objectifs, de la connaissance ou encore d'autres caractéristiques de l'utilisateur. Ces caractéristiques sont utilisées afin de classer les pages Web - appelées aussi nœuds - qui correspondent aux destinations des liens, pour éviter à l'utilisateur de suivre des chemins le menant à des informations non pertinentes [Brusilovsky, 1997].

Nous décrivons les objectifs les plus courants et les méthodes associées d'après [Brusilovsky et Eklund, 1998]:

- le *guidage global* consiste à aider l'utilisateur à suivre le plus court chemin pour atteindre son « but¹ » d'information. Deux méthodes sont envisagées. L'une, appelée *guidage direct*, est de suggérer étape par étape (i.e. après chaque suivi de lien) le lien suivant à actionner. L'autre suppose de trier les liens potentiels par ordre décroissant de pertinence.
- le *guidage local* vise à assister l'utilisateur dans le choix du lien à suivre à partir du nœud courant. Alors que le guidage global construit un chemin complet de navigation vers un but, le guidage local se résume à une suggestion élaborée à partir des préférences, connaissances, etc. de l'utilisateur. La méthode privilégiée à cet égard est également le tri des liens en fonction de leur pertinence, évaluée ici par rapport aux caractéristiques de l'utilisateur.
- l'*orientation globale* aide l'utilisateur à comprendre la structure complète de l'hypermédia et à situer sa position absolue actuelle dans cette structure. Des méthodes basées sur l'insertion de repères visuels ou d'affichage de la cartographie de l'hypermédia sont utilisées et consistent à agrémenter les liens d'annotations qui introduisent une sémantique facilitant la compréhension de la structure. Les *annotations* de liens dans le domaine hypermédia sont principalement de nature visuelle [Koch, 2000]. Un exemple classique est celui de la couleur utilisée pour un hyperlien textuel qui varie selon que le lien a été déjà visité ou non. Dans une

¹ Brusilovsky parle de but général d'information (*global information goal*). C'est à partir de ce but qu'est réalisé le processus de personnalisation.

démarche de personnalisation à l'utilisateur, la couleur peut être utilisée pour traduire la pertinence du lien pour un utilisateur donné.

- *l'orientation locale* porte sur l'introduction d'informations permettant à l'utilisateur de mieux percevoir sa position relative et surtout d'avoir une meilleure représentation de ce qu'il peut atteindre à partir de celle-ci. Les méthodes d'annotations de liens sont utilisées, ainsi que des méthodes qui cachent les informations non pertinentes (i.e. les liens vers des noeuds non pertinents).

- Les *vues personnalisées* ont comme objectif de réduire l'hyperespace à la partie que les utilisateurs exploitent réellement dans leur tâche quotidienne. Dans [Brusilovsky et Eklund, 1998], la création de telles vues est présentée comme relevant d'une méthode définissant la liste des liens vers les hyperdocuments qu'est amené à manipuler l'utilisateur. Il s'agit globalement de créer des structures de type favoris (ou bookmarks) qui permettent un accès rapide à un espace restreint.

Les techniques utilisées pour implémenter les méthodes précédemment citées ont en commun le recours à des algorithmes d'évaluation et de tri des liens candidats. Elles exploitent également soit une représentation des buts de l'utilisateur, soit de ses caractéristiques, ou encore une combinaison des deux. Enfin, les techniques implémentent certains aspects spécifiques des méthodes tels que la définition des « clés » visuelles utilisées dans les annotations de liens (spécification de la sémantique associée aux couleurs par exemple).

Le cas des personnalisations par annotation de liens peut être considéré comme relevant d'une personnalisation non pas de la navigation, mais de la présentation, puisqu'il repose sur des aspects visuels. Les objectifs de la personnalisation des présentations sont décrits ci-dessous, ainsi que les techniques employées à ces fins.

3.3 Personnalisation de la présentation

La personnalisation de la présentation permet, au même titre que celle du contenu informationnel et indépendamment de celle-ci, de rapprocher le système des préférences de l'utilisateur.

L'interface constitue un aspect essentiel, être en mesure de la personnaliser permet d'accroître sa valeur et de la rendre plus attractive. À ce niveau, ce sont les caractéristiques visuelles qui sont concernées : les personnalisations portent donc sur la *forme* (ou *apparence*).

Nous avons décrit dans cette section les différentes dimensions qui peuvent faire l'objet de personnalisation – contenu, navigation et présentation–. Nous abordons à présent la question de la mise en œuvre de ces personnalisations. Différentes façons d'envisager le processus de personnalisation du système pour l'utilisateur sont décrites dans la littérature. Nous en donnons une description dans la section suivante.

4. Les services de personnalisation

Nous présentons dans cette section les services de personnalisation les plus couramment utilisés : le filtrage du résultat (sections 4.1), le re-ordonnement des éléments du résultat (sections 4.2), la recommandation d'éléments (sections 4.3), et l'extension des requêtes avec les intérêts de l'utilisateur pour permettre la prise en compte de la personnalisation (section 4.4).

4.1 Le filtrage des résultats

Le filtrage d'information est l'expression utilisée pour décrire une variété de processus se rapportant à la fourniture de l'information adéquate aux personnes qui en ont besoin [Belkin et Croft, 1992]. Le filtrage est souvent interprété comme l'élimination de données indésirables sur un flux entrant, plutôt que la recherche de données spécifiques sur ce flux.

Dans la tâche de filtrage [Belkin et Croft 1992], un *profil utilisateur* doit décrire les préférences de l'utilisateur. Un tel profil est alors comparé aux documents entrants afin d'essayer de déterminer ceux qui pourraient être intéressants pour cet utilisateur particulier.

A l'arrivée de nouveaux documents, le système utilise le profil utilisateur pour choisir les documents pouvant potentiellement l'intéresser. Par un processus de retour de pertinence (« Relevance feedback » en anglais, cf. chapitre 1, section 3.4), l'utilisateur n'indique pas seulement les documents pertinents mais aussi les documents non-pertinents (figure 2.2). Le système utilise ces informations pour ajuster la description du profil utilisateur qui reflètera les nouvelles préférences.

L'avantage du filtrage des résultats est sa simplicité parce qu'il ne nécessite aucune modification du fonctionnement des fournisseurs d'information. Tout le traitement est fait après

l'exécution de la requête. Les inconvénients sont le volume de données échangées entre le serveur et le client et le risque d'élimination d'éléments pertinents.

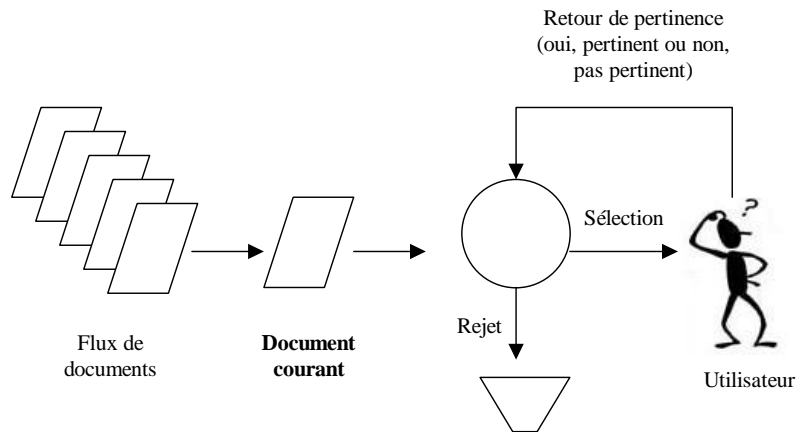


Figure 2.2 - Représentation d'un système de filtrage

Un exemple de système qui filtre les résultats une fois la requête exécutée est CASPER (Case-Based Profiling for Electronic Recruitment) [Bradley et al., 2000].

4.2 Le re-ordonnement des résultats de la requête

Le principe du re-ordonnement est de modifier l'ordre de l'affichage des résultats à l'utilisateur. Il s'agit d'un post traitement qui, étant donné les éléments retournés par une requête, essaie de trouver une manière d'échanger leurs emplacements en fonction des préférences de l'utilisateur sans pour autant négliger l'ordre qui a été attribué aux objets du résultat par le moteur de recherche. L'échange de l'ordre d'apparition des éléments du résultat est fait généralement en appliquant une fonction qui permet de calculer le nouveau rang de l'objet.

Un exemple de fonction permettant d'effectuer le re-ordonnement est donné par [Pretschner et Gauch, 1999].

Le re-ordonnement des résultats de la requête a le grand avantage de ne pas exclure des éléments du résultat. C'est en quelque sorte une garantie que l'utilisateur va trouver ce qui l'intéresse et le seul objectif de cette approche est de faire gagner du temps aux utilisateurs. C'est une technique très peu utilisée dans les systèmes de personnalisation parce qu'elle nécessite le calcul du nouveau rang de chaque élément du résultat ce qui demande beaucoup de temps.

4.3 La recommandation

L'objectif d'un système de recommandations est d'aider les utilisateurs à faire leurs choix dans un domaine où ils disposent de peu d'informations pour trier et évaluer les alternatives possibles. Un système de recommandations peut être décomposé en trois entités de base (figure 2.3) : le groupe d'agents *producteurs* de données brutes, le module de *calcul de recommandations* et le groupe de *consommateurs* des recommandations [Trousse, 2001].

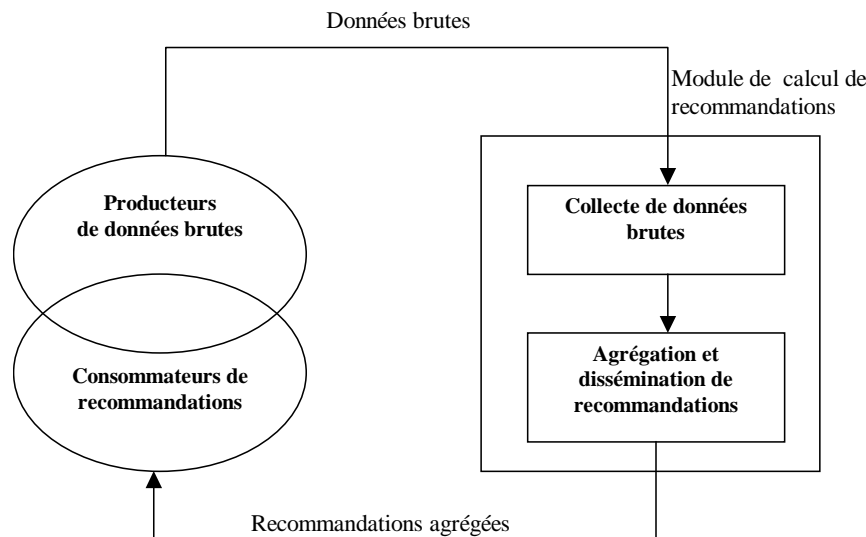


Figure 2.3 - Architecture générale d'un système de recommandation [Trousse, 2001]

GroupLens [Fink et Kobsa, 2000] est un exemple de système de recommandation. Il offre des services de recommandation sur la base de la gestion de groupes d'utilisateurs ayant les mêmes intérêts. Il propose trois types de recommandations : personnelle, anonyme et consultation rapide.

Une autre forme de recommandation consiste à formuler des propositions en fonction du profil de l'utilisateur et du contenu visualisé par ce dernier. C'est notamment cette technique qui est utilisée par **Amazon.com** [Amazon, 2004] pour conseiller ses clients : si le visiteur s'intéresse à un livre, le site lui propose les ouvrages ayant également plu aux acheteurs de ce même livre. Pour Amazon, ce système vise à remplacer le travail que ferait un vendeur dans une librairie traditionnelle.

Concluons ce point en remarquant que les recommandations n'appartiennent pas qu'au domaine de la personnalisation. Certaines pratiques de marketing en délivrent de manière

« aveugle ». Bazsalicz et Naim [Bazsalicz et Naim, 2001] démontrent toutefois qu'un système de recommandations personnalisées fournit de meilleurs résultats qu'un système aléatoire.

4.4 Les extensions des langages pour supporter la personnalisation

L'idée principale de la personnalisation est de faciliter l'écriture des requêtes de l'utilisateur en les enrichissant avec des données invariantes communes à toutes les requêtes. Dans ce contexte, l'utilisateur doit être capable d'exprimer toutes ses préférences et exigences de façon simple. Ces exigences doivent être traduites dans un langage afin d'être ajoutées aux requêtes et exécutées. Les langages classiques comme SQL, ont une syntaxe précise et ne supportent pas la notion de préférence. Pour cette raison, ces dernières années ont vu naître des extensions du langage qui supportent l'évaluation de fonctions externes ou utilisent d'autres clauses permettant d'exprimer des préférences ([Chomicki, 2002], [Borzsonyi et al., 2001]). L'ajout de nouveaux opérateurs au langage entraîne des ajouts aussi au niveau de l'exécuteur des requêtes tandis que la permission d'utilisation de fonctions externes laisse une plus grande liberté de l'implémentation de ces fonctions. Dans ce deuxième cas, la manière d'après laquelle une fonction sera exécutée dépend de l'utilisateur et par conséquent le sens que l'utilisateur donne à ces opérateurs sera contenu dans le profil. Dans les deux cas, les informations du profil vont être utilisées comme des paramètres ou des critères de sélection et vont être intégrées aux requêtes.

Comme nous avons vu dans cette section, les services de personnalisation ne sont pas encore bien compris. Ceux qui sont les plus utilisés ne modifient que très peu le fonctionnement des systèmes de base (qui ne prennent pas en compte la personnalisation) comme par exemple le filtrage des résultats ou le re-ordonnement des éléments du résultat ou ne font pas des restrictions dans l'espace de recherche comme la recommandation d'objets.

5. Méthode de personnalisation

Les différents éléments (contenu informationnel, structuration des liens ou présentation) peuvent être personnalisés pour s'adapter aux utilisateurs. Chaque élément personnalisable peut adopter diverses formes en fonction de l'utilisateur. Mais comment déterminer quelle forme spécifique doit être envoyée à un utilisateur particulier ? Comment déterminer, pour un utilisateur donné, quel contenu lui envoyer, sous quelle présentation et avec quelle structure de liens ? Nous

avons besoin de techniques nous permettant d'effectuer un tri fondé sur des critères tangibles. Il en existe principalement trois : la technique de détermination explicite, la technique de détermination implicite et la customisation. Pour présenter ces trois techniques, on s'est principalement appuyé sur [Bazsalicza et Naim, 2001] et [Allen et al., 2001].

5.1 Détermination Explicite

Cette première technique utilise des données explicitement attribuables à chacun des utilisateurs pour effectuer le travail de détermination. Elles peuvent être fournies par l'utilisateur ou collectées à son insu, il s'agit généralement des données de son clickstream (séquence de clics). Dans les deux cas, elles sont stockées dans un profil. A l'aide de ces données, on peut ensuite établir des règles de manière à déterminer quelles informations doivent être envoyées à un utilisateur particulier. Lorsque l'information utilisée pour la personnalisation est en relation directe avec les données du profil de l'utilisateur, il s'agit de détermination explicite qui porte généralement le nom de « *Rule-based* ». Ceci signifie que les données disponibles sont soumises à des règles logiques afin de fournir un contenu personnalisé pour l'utilisateur.

La pertinence de la détermination explicite repose sur l'exactitude de ses règles. Construire des règles valides n'est pas une tâche aisée car elle nécessite une excellente connaissance et une compréhension à la fois des destinataires de la personnalisation et du contexte dans lequel celle-ci s'inscrit. Une grande partie de l'effort de personnalisation par des règles sera donc consacrée à la révision de ces dernières pour s'assurer de leur pertinence.

5.2 Détermination Implicite

Cette deuxième technique utilise des données déduites à partir d'autres utilisateurs pour effectuer le travail de détermination. Avec cette technique, les caractéristiques de l'utilisateur sont déduites à partir des caractéristiques du groupe auquel il est assimilé. Cette méthode emploie des techniques statistiques utilisant des données sur l'utilisateur dans le but d'estimer ce qu'il désire ou ce qu'il va effectuer.

La technique de personnalisation implicite la plus connue se dénomme le « Filtrage Collaboratif » (En anglais « *Collaborative Filtering* » [Shardanand et Maes, 1995]). Le filtrage collaboratif se base sur l'hypothèse que les gens à la recherche d'information devraient pouvoir se servir de ce que d'autres ont déjà trouvé et évalué. Pour ce faire, pour chaque utilisateur d'un

système de filtrage collaboratif, un ensemble de proches voisins est identifié, et la décision de proposer ou non un document à un utilisateur dépendra des appréciations des membres de son voisinage.

Cette technique de détermination peut être mise en œuvre sans nécessiter pour autant l'intervention de l'utilisateur. Il peut donc profiter des recommandations sans être obligé de fournir des informations le concernant. Elle comporte en revanche deux limites majeures. D'une part, la complexité de mise en œuvre et de l'autre part, pour être véritablement efficace, l'algorithme de filtrage doit disposer d'un échantillon statistique de grande taille au risque de produire des recommandations absurdes.

5.3 Customisation

La différence principale entre la customisation et les deux techniques de détermination vues ci-dessus se situe dans la localisation du pouvoir de décision des éléments à envoyer. Dans la customisation, c'est l'utilisateur qui choisit lui-même ses propres préférences alors que cette tâche revient au serveur dans les deux autres techniques. La sélection des préférences par l'utilisateur, la customisation, possède l'avantage d'être simple et transparente pour l'utilisateur. Cependant seuls les utilisateurs qui désirent effectuer cette démarche vont entreprendre un tel travail de personnalisation.

Grâce à cette technique, on est sûr que les préférences choisies par l'utilisateur sont réellement les siennes. On ne doit produire aucun effort pour estimer ces paramètres. De plus, ceci constitue une source d'informations intéressante pour une éventuelle combinaison avec d'autres techniques de détermination.

Chaque méthode de personnalisation nécessite cependant d'être alimentée par des données propres aux utilisateurs : pour alimenter les règles (détermination explicite), pour alimenter les filtres (détermination implicite) ou pour désigner les préférences à appliquer (customisation). Nous avons donc besoin de réunir et de stocker des données sur les utilisateurs de manière à pouvoir en disposer en temps voulu. Les réceptacles de ces données se nomment « *profils* ».

Le processus de personnalisation est également influencé par le positionnement du pouvoir de décision. Il peut être localisé du côté des développeurs. L'utilisateur est une source d'informations et n'a pas d'influence directe sur le système, mais permet de tester l'application,

donc son rôle se limite à celui d'utilisateur ; il peut également être localisé du côté de l'utilisateur, dans ce cas ce dernier a un rôle actif car on lui demande de comprendre les problèmes et de proposer des solutions.

6. Conclusion

Nous avons débuté ce chapitre en illustrant une terminologie riche associée à la personnalisation. Nous avons mis en relation les termes employés (personnalisation, customisation et adaptation). Ceux-ci correspondent le plus souvent à des façons différentes d'envisager le processus visant à personnaliser le système à l'utilisateur. Parfois, deux termes peuvent être considérés comme des synonymes, la préférence allant à l'un ou à l'autre en fonction du domaine d'application qui l'emploie.

Nous avons ensuite mis en évidence la diversité des dimensions susceptibles d'être personnalisées à l'utilisateur. Trois dimensions principales en vue de la personnalisation ont été identifiées: le contenu, la navigation et la présentation. De nombreux services de personnalisation existent, nous avons choisi d'en présenter quatre parmi les plus fréquemment rencontrés : le filtrage de contenu, le re-ordonnement des résultats de la requête, les recommandations et l'extension des langages pour supporter la personnalisation. Généralement, la recommandation d'éléments est un service interactif et peut répondre aux besoins instantanés d'un utilisateur. Le filtrage et le re-ordonnement sont des processus de post traitement qui ont pour but d'adapter le contenu du résultat à un utilisateur donné, tandis que l'ajout des informations complémentaires aux requêtes a pour but le ciblage précis des informations intéressantes et peut se révéler dangereux dans le cas où les informations ajoutées ne seraient pas pertinentes. Nous avons aussi abordé les principales méthodes de personnalisation (méthodes de détermination implicite et explicite et la customisation).

En plus des questions relatives à la personnalisation, le contexte général de ce mémoire est également celui de la prise en compte des caractéristiques des utilisateurs pour la conception de tels systèmes. Il est important pour améliorer la compréhension de l'utilisateur et l'aider dans sa tâche, d'adapter la représentation, le contenu et la navigation à ses besoins.

Les stratégies de personnalisation doivent être différentes selon les utilisateurs. En effet, pour un utilisateur considéré comme novice, le masquage est une stratégie pertinente, alors que pour un utilisateur expert, c'est l'annotation qui est la plus intéressante.

Il est également nécessaire de pouvoir agir sur les stratégies de personnalisation pour deux raisons, tout d'abord parce que les caractéristiques et les objectifs des utilisateurs peuvent évoluer dans le temps, et ensuite intervenir facilement sur les stratégies permet d'évaluer leurs influences et leurs pertinences sur les utilisateurs. Nous pouvons alors décider quelle technique de personnalisation employer et quels éléments du profil utiliser.

Ainsi, tout processus de personnalisation, quel qu'il soit, repose sur un modèle de l'utilisateur, c'est-à-dire une représentation de ses caractéristiques dont le système tient compte dans la mise en œuvre des personnalisations.

Dans le prochain chapitre nous abordons un autre aspect de l'état de l'art à travers l'étude des ontologies et des différents formalismes de représentation des connaissances.

 Représentation des ontologies pour le Web Sémantique

CHAPITRE 3 : Représentation des ontologies pour le Web Sémantique	48
1. Introduction	49
2. Le Web Sémantique	50
2.1 L'architecture du Web Sémantique	50
3. Les ontologies : origine et définitions	51
3.1 Les ontologies en Ingénierie des Connaissances	52
4. Quels sont les composants principaux d'une ontologie?	53
5. Le rôle des ontologies	54
6. Types d'ontologie	55
7. La construction des ontologies	56
7.1 Quelques méthodologies de construction	56
7.2 Methontology	58
8. Quelques formalismes de représentation des connaissances	59
8.1 Les réseaux sémantiques	60
8.2 Les langages de frames	60
8.3 Les Graphes Conceptuels (GC)	61
8.4 Les logiques de description (LD)	62
8.4.1 Sémantique	63
8.4.2 Les constructeurs des LDs	63
8.4.3 Problèmes et procédures de décision pour les LDs	65
9. La représentation des connaissances sur le Web	65
9.1 RDF (Ressource Description Framework)	66
9.2 RDF Schema (RDF Vocabulary Description Language)	67
9.3 OIL (Ontology Inference Layer)	67
9.4 DAML+OIL (DARPA Agent Markup Langage)	67
9.5 OWL (Ontology Web Language)	68
10. Environnements et outils de modélisation	69
10.1 Protégé2000	71
11. L'exploitation des ontologies	72
11.1 Les moteurs d'inférences	72
12. Synthèse	73

Dans ce chapitre nous décrivons en détail les formalismes de description des ontologies. Nous exposons ensuite quelques méthodologies de construction des ontologies et les différents standards proposés comme solutions (partielles) au Web sémantique : RDF(S), OIL, DAML et OWL, un langage dont l'expressivité correspond à celle des logiques de description. Nous terminons ce chapitre en présentant quelques outils permettant de construire et d'utiliser les ontologies.

1. Introduction

Le premier réflexe lorsque nous parlons de structuration et de représentation de données est de penser aux bases de données et à leurs modèles usuels (entité association, relationnel, orienté objet, ..). Il existe toutefois d'autres approches (notamment au travers des ontologies) qui ne sont pas forcément concurrentes et qui peuvent venir enrichir et renforcer l'effort de structuration et de représentation.

Les ontologies occupent une place centrale dans la réalisation du Web sémantique parce qu'elles donnent un sens à l'information. Nous commençons donc ce chapitre par une définition du concept de Web sémantique (section 2). Nous reviendrons sur ses principes, puis nous détaillerons son architecture. Nous abordons ensuite la notion d'ontologie avec ses origines dans la philosophie occidentale et en Ingénierie des Connaissances (section 3), ses composants de base (section 4), ses apports en comparaison avec une représentation plus classique de la connaissance (section 5) et nous passerons en revue quelques types d'ontologies (section 6). Nous étudierons également le problème délicat de la construction des ontologies en proposant une revue des méthodologies existantes. Nous verrons pourquoi nous nous sommes tournés vers la méthodologie METHONTOLOGY (section 7). Nous présentons ensuite quelques formalismes de représentation des connaissances qui sont à l'origine des langages permettant d'exprimer des ontologies (section 8). Nous verrons alors comment ces formalismes ont évolué tout récemment, sous l'influence du langage XML, pour être adaptés au Web qu'on nous promet sémantique (section 9). Nous compléterons cet état de l'art en présentant quelques outils permettant d'utiliser et de raisonner sur les ontologies (section 10).

2. Le Web Sémantique

Le World Wide Web a été conçu initialement comme un espace d'information partageable entre humains. Il est constitué d'une masse d'informations et de connaissances si vaste que l'homme seul éprouve des difficultés à s'y retrouver. Le succès du Web est essentiellement dû à sa simplicité. Le revers de cette simplicité est le frein qu'elle constitue pour le développement de services intelligents. En effet, les machines sont incapables d'interpréter l'information du Web et par conséquent d'aider l'homme dans sa tâche.

Le Web sémantique, proposé initialement par le W3C (World Wide Web Consortium) [Berners-Lee et al., 2001], est une extension du Web actuel. Il a pour objectif d'être « un Web qui parle aux machines », c'est-à-dire dans lequel les ordinateurs offrent une meilleure aide aux utilisateurs de par le fait qu'ils se servent du contenu du Web.

L'interprétation et la manipulation sont facilitées par l'ajout de descriptions sémantiques, appelées métadonnées. Les métadonnées sont « des données sur des données ». Il s'agit « d'informations structurées qui décrivent une ressource ». Ces descriptions rendent la recherche d'information plus efficace et plus précise, par conséquent elles permettent une utilisation plus pertinente de la ressource. Le plus souvent, les métadonnées sont définies en fonction d'un schéma de métadonnées. Ce schéma spécifie pour chacune des métadonnées son domaine de valeur et son rôle. Quelques exemples de schémas sont disponibles comme Dublin Core [Powell, 1998], XMLNews¹ ou encore LOM (Learning Object Metadata) [IEEE, 2002] pour l'enseignement.

2.1 L'architecture du Web Sémantique

Le Web Sémantique n'est pas seulement une vision de l'avenir du Web. Certaines technologies sont déjà disponibles et figurent dans l'architecture illustrée dans la figure 3.1.

L'architecture repose sur des technologies de structuration de documents fondées sur XML. Sur cette base, des langages de métadonnées sémantiques de haut niveau ont été développés et permettent la description de ressources sur le Web. Afin de fournir un cadre interprétatif à ces métadonnées sémantiques, le Web Sémantique utilise des ontologies. Au niveau supérieur, le raisonnement sur les données est assuré par des mécanismes d'inférence,

¹ <http://www.xmlnews.org>

qui permettent d'une part de construire de la connaissance, mais aussi d'en maintenir la cohérence. Au niveau le plus haut, nous retrouvons la notion de confiance dont la question principale est : « Sur quelles bases peut on considérer comme vraie de l'information disponible sur le Web ? ». Les langages et outils développés dans le cadre du Web Sémantique ne sont que des bases technologiques. Ils ne résolvent pas à eux seuls les problèmes de présentation et d'utilisation de ces données sémantiques.

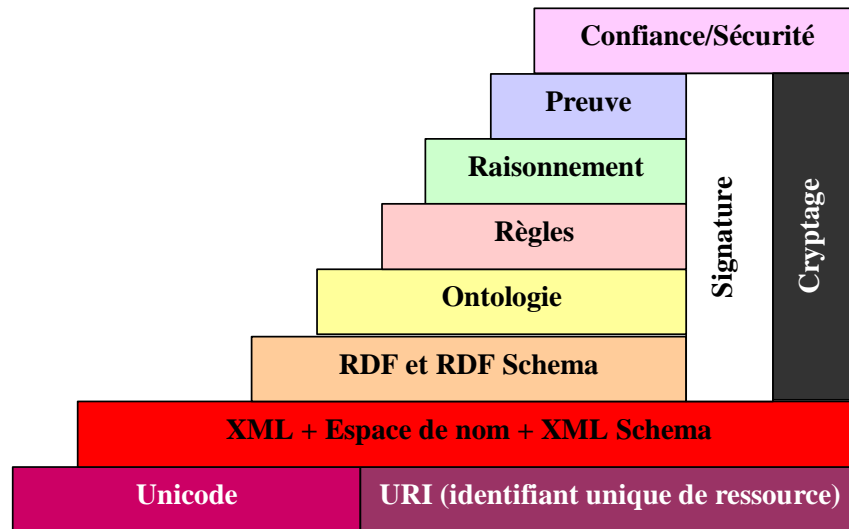


Figure 3.1 - Le « mille-feuilles » du Web sémantique (*Layer cake*) - d'après Tim Berners-Lee [Berners-Lee et al., 2001]

3. Les ontologies : origine et définitions

Le terme ontologie est emprunté du domaine de la philosophie où il signifie l'étude sur l'existence de l'être en tant qu'être. Au premier abord, cette définition philosophique paraît éloignée de celles qui vont suivre, et qui sont utilisées en informatique. Pourtant ce que nous cherchons à modéliser concerne bien les connaissances sur des parties du monde réel, c'est à dire des connaissances qui ne sont modélisables que parce qu'elles représentent des entités qui *existent*. Les deux points de vue, philosophiques et Intelligence Artificielle (IA), ne sont donc pas si éloignés l'un de l'autre.

Nous présentons dans la section 3.1 les principales définitions disponibles dans la littérature informatique. Elles sont souvent très proches ou complémentaires. Certaines de ces définitions peuvent être retrouvées dans [Mizoguchi, 1998].

3.1 Les ontologies en Ingénierie des Connaissances

La définition la plus souvent référencée dans les écrits en IA et aussi la plus synthétique est sans doute celle de Gruber [Gruber, 1993]: « *une ontologie est une spécification formelle explicite d'une conceptualisation partagée* ». C'est cette définition que nous allons conserver dans la suite. Les attributs « formelle » et « explicite » signifient qu'une ontologie permet une interprétation automatisée par la machine de la conceptualisation. Il s'agit d'un ensemble de définitions, de primitives, de représentation de connaissance spécifique au contenu : classes, relations, fonctions et constantes d'objet. L'ontologie est donc le fait de conceptualiser l'univers du discours et les relations pertinentes dans cet univers. Mais cette définition laisse déjà la porte ouverte à de nombreuses interprétations.

Guarino [Guarino, 1997] souligne l'ambiguïté du terme conceptualisation qui doit être prit dans son sens intuitif et propose la définition suivante pour tenir compte du caractère subjectif: « *l'ontologie est une spécification partielle et formelle d'une conceptualisation* ». L'ontologie est partielle car une conceptualisation ne peut pas toujours être entièrement formalisée dans un cadre logique, du fait d'ambiguïtés ou du fait qu'aucune représentation de leur sémantique n'existe dans le langage de représentation d'ontologies choisi.

Pour [Grüninger et Fox, 1995] une ontologie *est une description formelle d'entités et de leurs propriétés, relations, contraintes, comportement*. Elle est simplifiée dans [Ikeda et al., 1999] où une ontologie est définie comme *un ensemble de définitions de concepts et leurs relations* à ne pas confondre avec un modèle qui *est un ensemble d'instances de ces concepts*.

Mais avant de conclure ces quelques définitions, il faut nous interroger sur l'utilisation – pour certain abusive – du terme « ontologie ». Force est de constater que ce terme est de plus en plus utilisé, tout en renvoyant à des objets assez différents dans leur nature et dans leur but. En témoignent les travaux de C. Wetly, M. Gruninger et M. Uschold rapportés dans [Smith et Wetly, 2001] et illustrés par la figure 3.2.

Un système d'information aussi simple qu'un catalogue peut être considéré comme une ontologie des produits vendus par une société. Légèrement plus complexe, le système peut fournir un ensemble de textes en langue naturelle sur lesquels on pourra établir des correspondances de chaînes de caractères. Le glossaire, en classant les termes référencés, impose une structure au système d'information. Les thésaurus permettent d'organiser les termes décrits en hiérarchie via la relation de spécialisation. La représentation des

connaissances et l'ingénierie logiciel orienté objet utilisent ces taxinomies pour faire hériter des propriétés dans les hiérarchies de classes. Les systèmes de *frames* ajoutent des relations entre les objets et des restrictions sur quoi et comment classes et objets peuvent être liés. Finalement, le système d'information « ontologique » le plus complexe utilise des axiomes de la logique du premier ordre, d'ordre supérieur et de la logique modale.

Derrière cet apparent chaos, il y a quand même des points communs qui se dessinent : la nécessité d'avoir une classification partagée des entités manipulées par exemple.

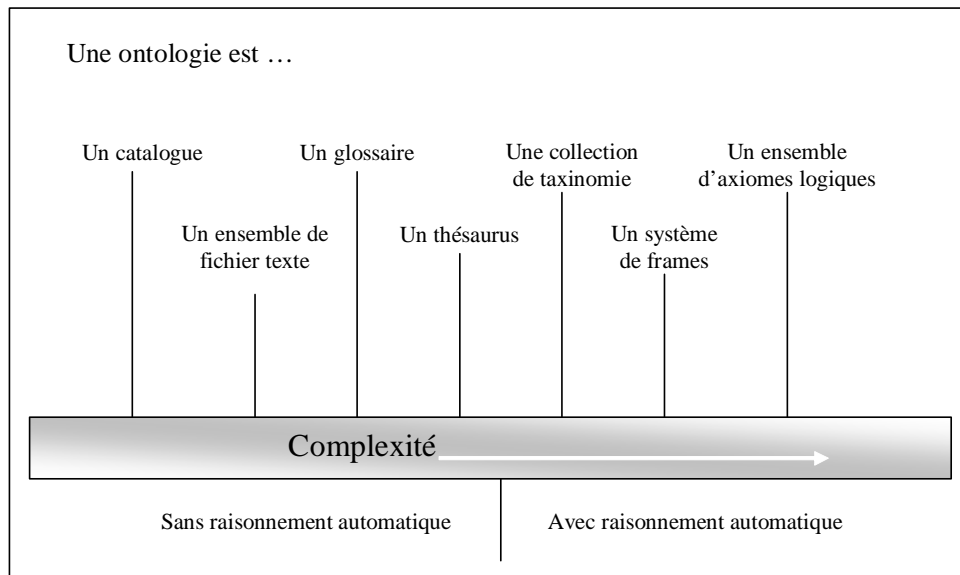


Figure. 3.2- Recensement des objets qui ont pu, à un moment donné, être classifié comme des ontologies, d'après [Smith et Wetly, 2001]

Pour conclure cette section, nous pouvons donc affirmer que les définitions du terme ontologie abondent dans la littérature scientifique. Les définitions, dans leur diversité, offrent des points de vue à la fois différents et complémentaires sur un même concept.

Dans la section suivante, nous présentons les briques de base qui peuvent constituer une ontologie.

4. Quels sont les composants principaux d'une ontologie?

Les connaissances intégrées dans les ontologies sont formalisées en mettant en jeu cinq types de composants: les concepts, les relations, les fonctions, les axiomes et les instances [Gruber, 1993].

- **Les concepts**, aussi appelés termes ou classes de l'ontologie, correspondent à tout ce qui peut être évoqué comme la description d'une tâche, d'une fonction, d'une action, d'une stratégie ou d'un processus de raisonnement, etc. Selon [Gómez-Pérez, 1999] ces concepts peuvent être classifiés selon plusieurs dimensions : 1) *niveau d'abstraction* : concret ou abstrait ; 2) *atomicité* : élémentaire (électron) ou composée (atome) ; 3) *niveau de réalité* : réel ou fictif.
- **Les relations** unissent les concepts ensemble pour construire des représentations conceptuelles complexes. Elles sont formellement définies comme tout sous-ensemble d'un produit de n ensembles, c'est-à-dire $R: C_1 \times C_2 \times \dots \times C_n$. Des exemples de relations binaires incluent: *sous-classe-de* (généralisation/spécialisation) ; *Partie-de* (agrégation ou composition) ; *Associée-à* ; *Instance de*, etc.
- **Les fonctions** sont des cas particuliers de relations dans lesquelles le n ème élément (extrant) de la relation est défini de manière unique à partir des $n-1$ premiers (intrants). Formellement, les fonctions sont définies ainsi : $F: C_1 \times C_2 \times \dots \times C_{n-1} \rightarrow C_n$. Comme exemple de fonctions binaires, nous avons la fonction *mère de*, et comme exemple de fonction ternaire, *Prix-voiture-occasion*, qui calcule le prix d'une voiture d'occasion selon le modèle de la voiture, la date de construction et le nombre de kilomètres.
- **Les axiomes** constituent des assertions, acceptées comme vraies, à propos des abstractions du domaine traduites par l'ontologie. Ils peuvent intervenir dans la définition des concepts et des relations, dans la vérification de l'exactitude d'informations indiquées dans l'ontologie ou dans la déduction d'une nouvelle information.
- **Les instances** (ou individus) constituent la définition extensionnelle de l'ontologie ; ces objets véhiculent les connaissances (statiques, factuelles) à propos du domaine du problème.

5. Le rôle des ontologies

On utilise l'ontologie dans différents domaines : la représentation d'informations et de connaissances, l'intégration des systèmes d'informations, etc. mais aussi dans :

- **La communication.** Il existe trois types de communication dans un projet : communication homme-homme, homme-système ou entre les différents modules du

système. Ces trois types possèdent tous des caractéristiques particulières qui engendrent certains problèmes auxquels les ontologies peuvent apporter des solutions.

- **L'interopérabilité.** L'interopérabilité est une spécialisation de la communication. L'ontologie répertorie les concepts que des applications peuvent s'échanger même si elles sont distantes et développées sur des bases différentes [Charlet, 2002].
- **Modularité et réutilisabilité des connaissances.** Les ontologies possèdent des caractéristiques qui favorisent la réutilisation et le partage de données [Gruber, 1991]. En effet, elles permettent l'étude de conceptualisations, indépendamment du formalisme choisi pour les représenter et doivent être définies indépendamment du langage utilisé pour la programmation des applications, de la plate-forme utilisée et des protocoles de communication (protocoles réseaux). Cependant, Gruber souligne les difficultés techniques occasionnées par la conception d'ontologies communes. Ces idées ont été beaucoup approfondies et développées dans [Gruber, 1993] et [Gruber, 1995].
- **L'indexation et la recherche d'information.** Plus récemment, les travaux autour du Web sémantique ont réactivé la problématique et l'utilisation des ontologies : en sus d'un rôle de médiateur, les ontologies y sont utilisées pour l'indexation, fournissant les index conceptuels décrivant les ressources sur le Web [Charlet, 2002].

Les différents avantages qu'offrent les ontologies ne doivent pas occulter leurs limites. La principale est la notion de *point de vue* qui reste un point obscur et peu discuté dans la littérature de ce domaine. Cette notion traduit le fait qu'un concept n'est pas considéré de la même manière suivant le contexte dans lequel il est utilisé.

6. Types d'ontologie

Cette section ne cherche pas à donner une typologie approfondie des ontologies. Cependant, elle présente les types le plus généralement utilisés d'ontologies, afin d'avoir une idée de la connaissance qui est incluse dans chaque type.

Tout d'abord on peut catégoriser une ontologie selon le type de langage utilisé pour la construire. Une ontologie est **formelle** si elle est décrite par un langage artificiel défini de

façon formelle par opposition à une ontologie **informelle** qui utilise des langages naturels. Dans ce mémoire nous ne traiterons que d'ontologie formelle.

Les ontologies peuvent aussi être classifiées en fonction de deux dimensions : leur niveau de détail et leur niveau de dépendance par rapport à une tâche particulière. Nicola Guarino [Guarino, 1997] distingue plusieurs niveaux dans les ontologies :

- Les ontologies de haut niveau contiennent les concepts généraux, commun à tous les domaines (temps, espace, objet, évènement). Les travaux de J. Sowa et de N. Guarino tentent de formaliser une telle ontologie « universelle » ;
- Les ontologies liées à un domaine particulier sont de deux sortes : i) soit elles contiennent le vocabulaire spécifique à un domaine bien défini et sont des spécialisations d'une ontologie de haut niveau. De nombreuses ontologies de domaine existent déjà, telles que MENELAS dans le domaine médical [Zweigenbaum, 1999], ENGMATH pour les mathématiques [Gruber, 1994], TOVE dans le domaine de la gestion des entreprises [Grüninger et Fox, 1995], etc. ii) soit il s'agit d'ontologies de tâche qui contiennent l'ensemble des tâches réalisées dans un domaine donné ;
- Les ontologies d'application dépendent à la fois d'un domaine et d'une tâche. L'exemple le plus représentatif de ce type d'ontologie est l'*ontologie de Frame* (*frame-ontology* d'ONTOLINGUA- [Gruber, 1993]) qui intègre les primitives de représentation des langages à base de *frames* (classes, instances, facettes, propriétés/*slots*, relations, restrictions, valeurs permises, etc.).

7. La construction des ontologies

Dans cette section nous proposons une revue de quelques méthodologies de construction existantes (section 7.1). Nous montrons alors pourquoi nous nous sommes orientés vers la méthodologie proposée par M. Fernández et ses condisciples [Fernández et al., 1997] que nous détaillons dans la section 7.2.

7.1 Quelques méthodologies de construction

À l'heure actuelle, il n'existe pas encore de consensus à propos des meilleures pratiques à adopter lors du processus de construction ou même des normes techniques régissant le

processus de développement des ontologies, bien que certaines contributions dans cette direction soient déjà disponibles dans les écrits.

Les ontologies produites sont très diverses comme en témoigne une des premières études comparatives consacrées aux méthodologies de construction [Noy et Hafner, 1997]. Nous présentons ci-après quelques éléments méthodologiques.

La méthodologie d'Uschold et Kings [Uschold et King, 1995] se fonde sur l'expérience de la construction de l'Ontologie d'entreprise, qui inclut un ensemble d'ontologies pour la modélisation d'entreprises. Ils proposent les étapes suivantes :

- (1) identification du but et de l'étendue de l'ontologie,
- (2) construction de l'ontologie en consignnant et en codant des connaissances, ainsi qu'en les intégrant à des ontologies existantes,
- (3) évaluation,
- (4) documentation, et
- (5) établissement de directives pour chaque étape.

Quant à la méthodologie de Grüninger et Fox [Grüninger et Fox, 1995], elle se base sur l'expérience de la construction d'une ontologie de modélisation d'entreprise dans le cadre du projet TOVE (TOronto Virtual Enterprise). Il s'agit essentiellement de la construction d'un modèle logique des connaissances à inclure dans l'ontologie. Ce modèle n'est pas construit directement. Les spécifications que doit comprendre l'ontologie sont décrites de façon informelle en identifiant un ensemble de questions de compétence, et cette description est ensuite formalisée dans un langage basé sur la logique des prédicats. Les questions de compétence constituent l'élément clé qui permet de caractériser de façon rigoureuse les connaissances que doit inclure une ontologie et elles spécifient le problème ainsi que ce qui constituerait une bonne solution au problème.

D'autres approches cherchant à plus automatiser la structuration de l'ontologie via des algorithmes d'apprentissage ont été développées comme celle de Maedche et Staab [Maedche et Staab, 2000] ou ASIUM développé par Faure et Nedellec [Faure et Nedellec, 1999]. La première approche cherche à minimiser les interactions et construit automatiquement une taxinomie de concepts à partir d'un dictionnaire du domaine. La seconde approche, ASIUM, utilise des techniques d'apprentissage pour construire interactivement une ontologie.

7.2 Methontology

Le cadre METHOTOLOGY insère l'activité de construction d'ontologies dans un processus de gestion de projet [Fernández et al., 1997]. Les ontologies étant destinées à être utilisées comme des composants logiciels dans des systèmes répondant à des objectifs opérationnels différents, leur développement doit s'appuyer sur les mêmes principes que ceux appliqués en génie logiciel. En particulier, les ontologies doivent être considérées comme des objets techniques évolutifs et possédant un cycle de vie qui nécessite d'être spécifié. Les activités liées aux ontologies sont d'une part des activités de gestion de projet (planification, contrôle, assurance qualité), et d'autre part des activités de développement (spécification, conceptualisation, formalisation, implémentation, maintenance); s'y ajoutent des activités transversales de support telles que l'évaluation, la documentation, la gestion de la configuration. La méthodologie elle-même spécifie les étapes à suivre pour réaliser chaque activité, les techniques utilisées, les produits à être mis au point ainsi que la façon de les évaluer (figure 3.3).

La phase de **conceptualisation** constitue la principale phase du processus d'élaboration de l'ontologie selon l'approche de *Methontology*. Au cours des phases de **spécification** et de **conceptualisation**, un processus d'**intégration** est achevé en utilisant des ontologies réalisées en interne ou à l'extérieur. Ce cadre est soutenu en partie par l'*Ontology Design Environment* (ODE) [Blazquez et al., 1998], un environnement logiciel. Plusieurs ontologies ont été mises au point en ayant recours à cette méthodologie: *Chemicals* [Fernández et al., 1999], une ontologie spécialisée dans les produits chimiques; la *Reference-Ontology* [Arpirez et al., 1998], une ontologie qui constitue une sorte de pages jaunes des ontologies; ainsi que la version restructurée de l'ontologie (KA)2 [Blazquez et al., 1998]. La *Foundation for Intelligent Physical Agents (Fipa)*², qui promeut l'interopérabilité entre les applications programmées à l'aide d'agents, a proposé de recourir à cette méthodologie pour construire des ontologies.

² <http://www.fipa.org>

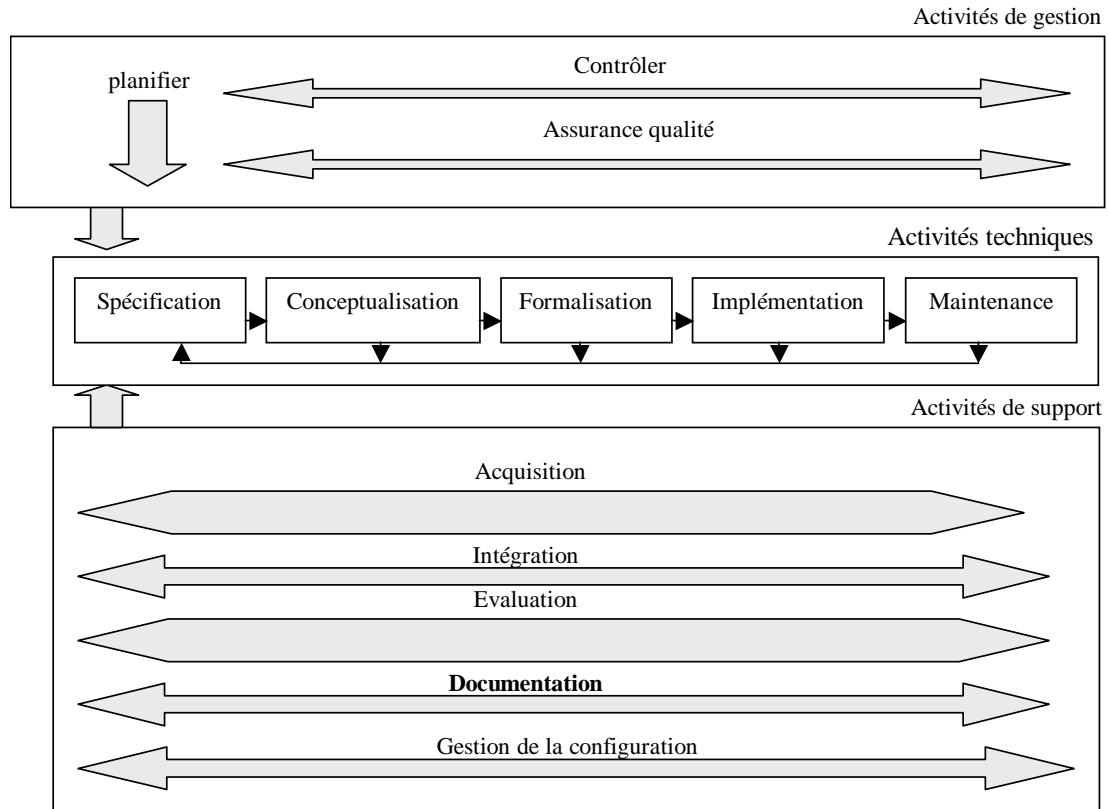


Figure 3.3- Cycle de vie de l'ontologie [Fernandez et al., 1997]

Les différentes expériences menées jusque-là avec les méthodologies décrites ici montrent que l'intervention humaine est indispensable et même primordiale dans les choix de conceptualisation à faire. Il est alors plus important d'outiller cette intervention que de chercher à la minimiser.

8. Quelques formalismes de représentation des connaissances

En toute généralité, représenter des connaissances propres à un domaine particulier consiste à décrire et à coder les entités de ce domaine de manière à ce qu'une machine puisse les manipuler afin de raisonner ou de résoudre des problèmes [Kayser, 1997]; [Euzenat et al., 2000]. Cette définition met en évidence deux composantes complémentaires de la représentation des connaissances (RC), à savoir l'expression et la manipulation des connaissances. D'une part, les connaissances sont exprimées à l'aide d'un langage formel, dit de description des connaissances. Le langage est doté d'une syntaxe, précisant l'ensemble des expressions admissibles du langage et d'une sémantique qui permet de fournir un sens aux

formules justifiant ainsi la validité des opérations effectuées. D'autre part, le but est de mécaniser un certain nombre de manipulations sur les connaissances exprimées. Ainsi, il sera nécessaire de modifier, compléter, inférer de nouvelles connaissances. Ces manipulations possibles sont spécifiées sous forme de mécanismes respectant la sémantique et opérant sur les éléments de la représentation.

Comme alternative à la logique classique, l'IA a proposé divers formalismes de représentation des connaissances, afin de permettre une représentation plus intuitive des connaissances [Baader et al., 2003]. Les travaux en RC ont donné naissance à de nombreux formalismes. Nous nous contenterons de décrire succinctement les principaux formalismes. Une représentation plus détaillée peut être trouvée dans [Euzenat, 1999].

8.1 Les réseaux sémantiques

Les réseaux sémantiques [Quillian, 1968] sont fondés sur un modèle graphique permettant de combiner la représentation des concepts (sous forme de nœuds) et des relations entre concepts (sous forme d'arcs). Un mécanisme de raisonnement fondé sur le parcours de la structure graphique permet d'établir des liens entre des concepts du même réseau.

La structuration de la connaissance en nœuds et arcs rend visibles les diverses relations existantes entre les objets, ainsi que la notion de « distance » entre deux concepts (nombre de liens du chemin connectant les deux concepts). Mais les principales critiques émises par ce formalisme sont son manque de sémantique précise, notamment en ce qui concerne l'héritage par défaut des propriétés. De plus, les réseaux sémantiques offrent une représentation statique du monde, ce qui rend difficile la modélisation de l'évolution de l'information. Diverses évolutions ont alors vu le jour pour l'améliorer, dont un nouveau formalisme graphique – les réseaux d'héritage structurés – implémenté dans le système KL-ONE [Brachman et Schmolze, 1985].

8.2 Les langages de frames

Le langage de frames ou schémas a été introduit par M. Minsky [Minsky, 1975]. Un des buts de ce formalisme est de représenter toutes les connaissances relatives à une situation (ou un objet) dans un schéma plutôt que de distribuer la connaissance à travers de nombreux axiomes.

Une frame contient des attributs (« slots »), pour décrire les propriétés de l'objet représenté, auxquels on peut attribuer des valeurs (« facets »). Le formalisme permet de classer les frames (indiquer si une frame est plus spécifique ou plus générale qu'une autre) ou de rechercher l'ensemble des valeurs des propriétés pour un objet particulier.

Les systèmes de frame sont très utiles pour modéliser les domaines où les objets sont complexes et riches ; le regroupement de toutes les informations d'un objet dans une unité fournit une représentation structurelle concise et facilement exploitable. Mais ils manquent d'un formalisme de base ; ce qui entraîne un manque de rigueur dans l'utilisation des différentes notions.

En 1995, Kifer, Lausen et Wu [Kifer et al.,1995] ont proposé de formaliser la structure des langages orientés objets et des langages de frame, sous la notion de Logique de Frame (abrégée en F-logic).

8.3 Les Graphes Conceptuels (GC)

Les GCs de Sowa [Sowa, 1984] sont un formalisme général de représentation des connaissances fondé sur la logique des prédicats du premier ordre, avec extension au second ordre. Un GC est décomposé en deux niveaux: le niveau terminologique où sont décrits les concepts, les relations et les instances de concepts, ainsi que les liens de subsomption entre concepts et entre relations et le niveau assertionnel où sont représentés les faits, les règles et les contraintes sous forme de graphes où les sommets sont des instances de concepts et les arcs des relations.

La non ambiguïté et la facilité d'utilisation font que ce formalisme est utilisé dans plusieurs applications comme l'acquisition des connaissances, la recherche d'information et le raisonnement sur la connaissance conceptuelle. Mais le choix d'un GC comme l'unité de représentation pose des problèmes, à savoir que la connaissance concernant un objet est disséminée dans les différents graphes qui en parlent; ce problème est spécialement grave quand le domaine possède des objets complexes et quand la base de connaissances atteint une grande taille.

8.4 Les logiques de description (LD)

Les LDs (ou logiques terminologiques, ou encore langages basés sur les termes) constituent l'aboutissement d'une longue période de recherche sur la formalisation dans les langages de représentation de connaissances. La brève introduction représentée ici est principalement inspirée des travaux de Nebel [Nebel, 1990] et plus récemment de Napoli [Napoli, 1997].

Une LD est composée de deux parties : un langage terminologique (appelé T-Box) et un langage assertionnel (appelé A-Box). Le langage assertionnel est dédié à la description de faits et le langage terminologique à la description de concepts et de rôles.

Dans le formalisme des LDs, un *concept* est une entité générique du domaine d'application représentant un ensemble d'individus. Un concept possède une description structurée qui se construit à l'aide d'un ensemble de constructeurs. L'intension d'un concept est l'ensemble des propriétés qui décrivent ce concept. L'extension d'un concept correspond à l'ensemble des individus du domaine d'interprétation qui possèdent en commun les propriétés décrites en intension.

Un concept peut être *primitif* ou *défini* :

- Les propriétés décrivant un concept *primitif* sont des conditions nécessaires mais pas suffisantes. Un individu associé à un concept primitif possède, ainsi, les caractéristiques de ce concept.
- Les propriétés décrivant un concept *défini* sont des conditions nécessaires et suffisantes. Autrement dit, ces propriétés définissent un concept. Un individu associé à un concept défini possède les caractéristiques de ce concept. Inversement, si un individu possède l'ensemble des caractéristiques d'un concept défini, cela suffit pour associer cet individu à ce concept.

La deuxième entité définie et manipulée dans une LD est le *rôle*. Un rôle introduit une relation binaire, entre concepts (quelques LDs autorisent cependant la définition de rôles n-aires [Calvanese *et al.*, 1997]). A l'instar d'un concept, un rôle peut être primitif ou défini et peut posséder une description structurée, où figurent les propriétés associées au rôle. Les rôles, et les restrictions attachées à ces rôles, sont introduits au niveau de la définition d'un concept, appelé *domaine*. Le concept avec lequel le rôle établit une relation correspond au *co-domaine* : les restrictions portent généralement sur ce concept. La cardinalité d'un rôle fixe le nombre

minimal et maximal de valeurs élémentaires, instances ou valeurs, que peut prendre le rôle pour une valeur donnée du domaine.

8.4.1 Sémantique

La signification formelle d'une description de concept est donnée au moyen d'une sémantique extensionnelle³ par une interprétation I qui est un couple $(\Delta^I; \cdot^I)$. Le domaine Δ^I est un ensemble arbitraire non vide d'individus et \cdot^I est une fonction d'interprétation faisant correspondre à chaque concept un sous-ensemble de Δ^I et à chaque rôle un sous-ensemble de $\Delta^I \times \Delta^I$. La sémantique d'un concept de la LD est donnée par le tableau 3.1.

Une interprétation I est un modèle pour le concept C si C^I est non-vide. C est *satisfiable* si et seulement s'il existe une interprétation I qui est un *modèle* de C . Selon cette sémantique, C subsume D (i.e. C est plus général que D au sens où l'ensemble des individus représenté par C contient l'ensemble des individus représenté par D), que l'on note $C \sqsubseteq D$, si et seulement si $C^I \subseteq D^I$ pour toute interprétation I .

Deux opérations, basées sur la relation de subsomption, sont à la base du raisonnement terminologique :

- la classification appliquée aux concepts (ou aux rôles) permet de déterminer la position d'un concept (ou d'un rôle) dans la hiérarchie.
- L'instanciation permet de retrouver les concepts auxquels un individu est susceptible d'appartenir.

8.4.2 Les constructeurs des LDs

Une LD est définie à partir d'un ensemble de concepts primitifs, un ensemble de rôles primitifs, des constantes \top et \perp et des règles de syntaxe et de sémantique représentés dans le tableau 3.1.

Les constructeurs utilisés dans cette syntaxe déterminent la puissance d'expression de la LD ainsi définie. Par exemple, la description du concept suivant :

$\text{Personne} \sqcap \exists \text{travailPour} . (\text{Projet} \sqcap \exists \text{finacéPar}.\text{Institut})$, décrit toutes les personnes qui travaillent dans un projet financé par un institut.

³ Qui satisfait à toutes les propriétés énoncées dans un domaine conceptuel

Selon les applications considérées, les constructeurs sont plus ou moins utiles. Ainsi pour décrire des objets, les constructeurs de cardinalité maximum et minimum peuvent être parfois très utiles. Pour décrire des actions et des processus, les constructeurs de composition, de conjonction et de disjonction de rôles semblent nécessaires.

A moins de posséder un algorithme de décision pour la logique contenant l'ensemble de tous les constructeurs (ce qui n'est actuellement pas le cas), il semble donc difficile de croire qu'une seule logique de description va permettre de couvrir les besoins de toutes les applications.

Constructeur	Syntaxe	Sémantique
CONCEPTS		
Plus général	T	Δ^I
Absurde	\perp	Φ
Concept primitif	P	$P^I \subseteq \Delta^I$
Négation	$\neg C, \neg P$	$\Delta^I \setminus C^I, \Delta^I \setminus A^I$
Conjonction de concepts	$C \cap D$	$C^I \cap D^I$
Disjonction de concepts	$C \cup D$	$C^I \cup D^I$
Restriction universelle	$\forall r.C$	$\{x \in \Delta^I / \forall y. (x, y) \in r^I \Rightarrow y \in C^I\}$
Restriction existentielle	$\exists r.C$	$\{x \in \Delta^I / \exists y. (x, y) \in r^I \wedge y \in C^I\}$
Cardinalité minimum	$\geq n r.C$	$\{x \in \Delta^I / \#(y / (x, y) \in r^I \wedge y \in C^I) \geq n\}$
Cardinalité maximum	$\leq n r.C$	$\{x \in \Delta^I / \#(y / (x, y) \in r^I \wedge y \in C^I) \leq n\}$
ROLES		
Rôle	r	$r^I \subseteq \Delta^I \times \Delta^I$
Rôle primitif	q	$q^I \subseteq \Delta^I \times \Delta^I$
Conjonction de rôles	$r_1 \cap r_2$	$\{(x, y) \in \Delta^I \times \Delta^I / \#(y / (x, y) \in r_1^I \wedge y \in r_2^I) \geq n\}$
Disjonction de concepts	$r_1 \cup r_2$	$\{(x, y) \in \Delta^I \times \Delta^I / (x, y) \in r_1^I \wedge (x, y) \in r_2^I\}$
Inverse de rôle	r^{-1}	$\{(x, y) \in \Delta^I \times \Delta^I / (x, y) \in r^{-1I} \Rightarrow (y, x) \in r^I\}$
Composition de rôles	$r_1 \circ r_2$	$\{(x, y) \in \Delta^I \times \Delta^I / \exists z \in \Delta^I : (x, z) \in r_1^I \wedge (z, y) \in r_2^I\}$

Tableau 3.1 – Quelques constructeurs d'une LD

Dans ce tableau C, D et P désignent des concepts, q et r, T désignent des rôles et signifie le concept universel, la source de tous les concepts c'est à dire le concept le plus général alors que \perp signifie la notion d'absurdité qui est le sous-concept de tous les concepts existants, c'est-à-dire le puits de la hiérarchie.

8.4.3 Problèmes et procédures de décision pour les LDs

Les deux principaux problèmes de décision pour les LDs sont de savoir si :

- **Un concept C est (in)satisfiable.** Le test de satisfiabilité d'un concept, permet de vérifier que ce concept peut admettre des instances quelle que soit l'interprétation des termes.
- **Un concept C est subsumé par un concept D** (noté $C \subseteq D$).

Pour les LDs incluant la négation, ces deux problèmes de décision sont équivalents (chacun peut être exprimé à partir de l'autre). En effet $C \subseteq D$ est équivalent à « $C \cap \neg D$ est insatisfiable ».

Pour tester les relations de subsumption entre les descriptions, deux voies ont été principalement exploitées jusqu'à présent : les algorithmes de type normalisation-comparaison, abrégés en *algorithme NC*, et une méthode dérivées de la méthode des tableaux sémantiques en logique classique dans le but d'essayer de construire un modèle de C [Baader et Sattler, 2000].

Pour une LD donnée, l'objectif est de trouver une procédure de décision correcte et complète, i.e. une procédure qui termine et telle que :

- Si la procédure trouve un modèle, alors C est satisfiable (Complétude pour le problème de l'indécidabilité),
- Si la procédure ne trouve pas de modèle, alors C est insatisfiable (Correction pour le problème de l'indécidabilité),

Ces caractéristiques majeures font des logiques de descriptions une classe importante des langages de représentation des connaissances.

Nous terminons cette présentation des LDs en signalant les références bibliographiques suivantes : le lecteur peut consulter [Napoli, 1997] pour une introduction générale sur les LDs, et [Baader et Sattler, 2000] pour une introduction sur les algorithmes de décision.

9. La représentation des connaissances sur le Web

Il existe plusieurs langages de spécification d'ontologies (ou langage d'ontologies) qui ont été développés pendant les dernières années, et ils deviendront sûrement des langages

d'ontologie dans le contexte du Web sémantique. Certains d'entre eux sont basés sur la syntaxe de XML (Extended Markup Language), tels que XOL (Ontology Exchange Language), SHOE (Simple HTML Ontology Extension - qui a été précédemment basé sur le HTML), OML (Ontology Markup Language), RDF (Resource Description Framework), RDF Schéma. Trois langages additionnels sont établis sur RDF(S) pour améliorer ses caractéristiques: OIL (Ontology Inference Layer), DAML+OIL (DARPA Agent Markup Language + Ontology Inference Layer) et OWL (Web Ontology Language).

La figure 3.4 représente les rapports principaux entre tous ces langages sous la forme d'une pyramide des langages du Web sémantique.

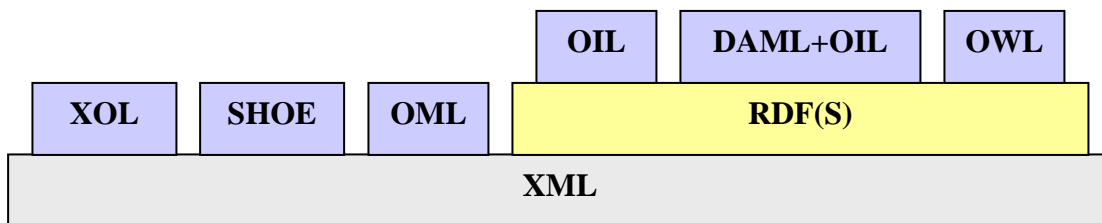


Figure 3.4 - La pyramide des langages basés Web.

9.1 RDF (Resource Description Framework)

Le W3C a adopté le langage RDF [RDF, 2004] comme formalisme standard de représentation. Utilisant la syntaxe XML, qui constitue déjà un standard, RDF permet de décrire des ressources Web en termes de ressources, propriétés et valeurs.

- Les **ressources** : Toutes choses décrites par des expressions RDF sont appelées des *ressources* (une page Web entière ou partie d'une page Web, un site Web, etc.). Elles sont identifiées par une URI⁴ (*Universal Resource Identifier*) et ne sont pas obligatoirement accessibles via le Web (ex., un livre imprimé);
- Les **propriétés** : il s'agit de caractéristiques, d'attributs ou de relations utilisées pour décrire des ressources particulières ;
- Les **valeurs** : la valeur de la propriété peut être une autre ressource ou elle peut être littérale ; c.-à-d., une ressource (spécifiée par une URI) ou une simple chaîne ou autre type de données primitif défini par XML.

⁴ Toute chose peut avoir une URI : L'extensibilité des URIs permet l'introduction d'identificateurs pour toute entité imaginable.

9.2 RDF Schema (RDF Vocabulary Description Language)

RDF, langage dédié à l'expression d'assertions sur les relations entre objets, s'est heurté à la nécessité de définir les propriétés des classes dont ces objets sont instances. Cependant, l'extension à RDFS [RDFS, 2004] permet de déclarer les propriétés et le type des ressources.

RDFS fournit les primitives de modélisation permettant de construire des ontologies simples. Beaucoup lui ont donc reproché son manque d'expressivité par comparaison avec ce que les langages de représentation des connaissances pouvaient faire. Dans l'optique d'une utilisation d'ontologies sur le Web, le langage RDF(S) a été enrichi par l'apport du langage OIL (Ontology Interchange Language [Fensel et al., 2000]) qui permet d'exprimer une sémantique à travers le modèle des frames tout en utilisant la syntaxe de RDF(S).

9.3 OIL (Ontology Inference Layer)

Le langage OIL, combine les primitives de modélisation utilisés dans le langage de *frames* et le raisonnement formel des logiques de description pour exprimer des ontologies sur le Web [Fensel et al., 2000]. Par rapport au langage RDFS, OIL reprend la distinction faite en logique de description entre concept *primitif* (décrit par des conditions nécessaires d'appartenance) et concept *défini* (décrit par des conditions nécessaires et suffisantes d'appartenance).

OIL offre de nouvelles primitives permettant de définir des classes à l'aide de mécanismes ensemblistes (intersection de classes, union de classes, complémentaire d'une classe).

9.4 DAML+OIL (DARPA Agent Markup Langage)

DAML+OIL est issu de l'union de deux langages: le langage OIL et DAML-ONT [Hendler et McGuinness, 2000]. L'intégration de OIL rend possible les inférences compatibles avec les logiques de description, essentiellement les calculs des liens de subsumption.

Ce nouveau langage supporte désormais les types de données primitifs, et la définition d'un certain nombre d'axiomes comme l'équivalence de classes ou de propriétés. Il est équivalent à une logique de descriptions très expressive, la logique *SHIQ*. Pour cette logique, il existe des procédures de décision correctes et complètes, et un certain nombre d'outils l'implémentant sont disponibles : OILed, OntoEdit, Protégé2000, et WebODE (voir section 10).

9.5 OWL (Ontology Web Language)

Le langage OWL (Ontology Web Language) est destiné à être utilisé lorsque les informations contenues dans les documents doivent être traitées par des applications logicielles, par opposition aux situations où le contenu est simplement présenté aux humains, cas traditionnel de l'utilisation du Web d'aujourd'hui [OWL, 2004].

Pour des raisons de modularité, OWL a été fractionné en trois langages distincts :

- **OWL Lite** ne contient qu'un sous-ensemble réduit des constructeurs disponibles, mais son utilisation assure que la comparaison de types pourra être calculée. [Horrocks et Patel-Schneider, 2003] montrent ainsi que ce langage est dans la même classe de complexité que la logique de description SHIF(D) et que le système Racer [Haarslev et Möller, 2001], par exemple l'implémente complètement (voir section 11.1). Le tableau 3.2 donne la liste des constructeurs disponibles dans ce langage ;

<p>RDF Schema</p> <ul style="list-style-type: none"> • Class • rdf : Property • rdfs : subclassOf • rdfs : subPropertyOf • rdfs : domain • rdfs : Range • individual 	<p>(In)égalité</p> <ul style="list-style-type: none"> • equivalentClass • equivalentProperty • sameAs • differentFrom • allDifferent 	<p>Propriétés algébriques</p> <ul style="list-style-type: none"> • inverseOf • TransitiveProperty • symmetricProperty • FunctionalProperty • InverseFunctionalProperty
<p>Intersection de classes</p> <ul style="list-style-type: none"> • intersectionOf 	<p>Version</p> <ul style="list-style-type: none"> • versionInfo • priorVersion • backwardCompatibleWith • incompatibleWith • DeprecatedClass • DeprecatedProperty 	<p>Annotation</p> <ul style="list-style-type: none"> • rdfs : label • rdfs : comment • rdfs : seeAlso • rdfs : isDefinedBy
<p>Types de données</p> <ul style="list-style-type: none"> • DatatypeProperty 		
<p>Restriction sur les rôles</p> <ul style="list-style-type: none"> • allValuesFrom • someValuesFrom 	<p>Cardinalité (restreinte)</p> <ul style="list-style-type: none"> • minCardinality (0-1) • maxCardinality (0-1) • cardinality (0-1) 	<p>En-tête</p> <ul style="list-style-type: none"> • ontology • imports

Tableau 3.2- Liste des constructeurs disponibles dans le langage OWL Lite

- **OWL DL** contient l'ensemble des constructeurs, mais avec des contraintes particulières sur leur utilisation qui assurent la décidabilité de la comparaison de types. Par contre, la grande complexité de ce langage semble rendre nécessaire une approche heuristique.

Ce langage a une correspondance avec la logique de descriptions SHOID(D) [Horrocks et Patel-Schneider, 2003];

- **OWL Full**, sans aucune contrainte est la fusion de RDFS et de OWL DL (c'est-à-dire qu'il abolit la distinction forcée entre classes et ressources); le problème de comparaison de types y est vraisemblablement indécidable, (cf. tableau 3.3).

Nous venons de voir dans cette section les différents langages qui ont conduit à la définition d'un standard permettant de représenter des ontologies sur le Web. Mais les ontologies sont des objets complexes comme nous l'avons vu dans la section 3. Nous étudions dans la section suivante les différentes méthodologies permettant de les construire.

Axiome de classes	Combinaison booléenne de classes
<ul style="list-style-type: none"> • oneOf, dataRange • disjointWith • equivalentClass (pour desexpression de calsse) • rdfs : subClassOf (pour desexpression de calsse) 	<ul style="list-style-type: none"> • unionOf • intersectionOf • complementOf
Cardinalité	Restriction sur les rôles
<ul style="list-style-type: none"> • minCardinality • maxCardinality • cardinality 	<ul style="list-style-type: none"> • hasValue

Tableau 3.3- Liste des constructeurs supplémentaires disponibles dans les langages OWL DL et OWL Full

10. Environnements et outils de modélisation

Après avoir étudié quelques méthodologies de conception d'ontologies et les langages de représentations et détaillé ceux que nous avons finalement adoptés, nous présentons maintenant les outils informatiques permettant de les créer afin de voir si l'un d'entre eux peut être utilisé pour suivre les principes fournis par la méthodologie choisie. Ces outils peuvent se regrouper grossièrement en deux catégories. Dans la premières, on trouve les plus anciens historiquement, qui permettent de spécifier les ontologies au niveau symbolique (voir par exemple, le serveur Ontolingua [Farquhar et al., 1995]). Dans la seconde catégorie, les outils prennent mieux en compte l'importance du niveau des connaissances : ils proposent à l'utilisateur de créer l'ontologie de manière relativement indépendante de tout langage implémenté et prennent ensuite automatiquement en charge l'opérationnalisation de

l'ontologie, en la transposant dans divers langages. Cette évolution tend à rapprocher les ontologies de leur but original : il semble en effet naturel de chercher à s'abstraire – dans un premier temps – du niveau symbolique.

Nous avons regroupé dans ce récapitulatif les outils de construction d'ontologies les plus connus :

- **Le serveur ONTOLINGUA**⁵ [Farquhar et al., 1995]. Le plus connu des environnements de construction d'ontologies en langage Ontolingua. Il consiste en un ensemble d'outils et de services qui supportent la construction en coopération d'ontologies, entre des groupes séparés géographiquement.
- **ONTOSAURUS** [Swartout et al., 1997]. Consiste en un serveur utilisant *LOOM* [Macgregor, 1991] comme langage de représentation des connaissances, et en un serveur de navigation créant dynamiquement des pages HTML qui affichent la hiérarchie de l'ontologie.
- **WEBONTO**⁶ [Domingue, 1998]. WebOnto et Tadzebao sont des outils complémentaires. Tadzebao permet aux ingénieurs des connaissances de tenir des discussions sur les ontologies, en mode synchrone et asynchrone. WebOnto supporte la navigation collaborative, la création et l'édition d'ontologies sur le Web.
- **ODE (Ontology Design Environment)** [Blazquez et al., 1998]. Un éditeur qui assure fidèlement le support de la méthodologie maison METHONTOLOGY (cf. section 7.2).
- **HOZO** [Mizoguchi et al., 2000]. Un environnement composé d'un éditeur et d'un serveur d'ontologies. L'éditeur est développé en applets Java afin de pouvoir fonctionner comme un client via l'Internet.
- **OILED**⁷ [Bechhofer et al., 2001]. Un éditeur freeware d'ontologies, destiné à supporter le développement d'ontologies de petites et moyennes tailles, dans le langage de représentation OIL.
- **OntoEdit**⁸ [Sure et al., 2002]. Il présente les fonctionnalités essentielles communes aux autres éditeurs (hiérarchie de concepts, expression d'axiomes, export de

⁵ <http://www-ksl-svc.stanford.edu:5915/>

⁶ <http://webonto.open.ac.uk/>

⁷ <http://oiled.man.ac.uk/>

⁸ Une version de démonstration est disponible sur le site d'Ontoprise, la société qui le développe en collaboration avec l'AIFB de Karlsruhe

l'ontologie dans des langages divers) et a le mérite de s'appuyer sur une réflexion méthodologique significative.

- **WebODE**⁹ [Arpirez et al., 2001]. D'un point de vue méthodologique, l'outil fait suite à ODE. L'accent a plus été mis sur la possibilité d'un travail collaboratif ou sur la mise à disposition d'outils complémentaires, comme un moteur d'inférences.
- **DOE** (*Differential Ontology Editor*)¹⁰ ([Troncy et Isaac, 2002] ; [Bachimont et al., 2002]). Cet Outil n'a pas pour ambition de concurrencer les grands environnements existants, mais plutôt de fournir un début d'implémentation à la méthodologie de structuration différentielle proposée par B. Bachimont [Bachimont, 2000].
- **KAON** (Karlsruhe Ontology and Semantic Web) [Maedche et al., 2003]. Un environnement *open source* modulaire, basé sur Java, destiné à la conception, au développement et à la gestion d'ontologies. L'environnement intègre les modules suivants : API¹¹, Query, Serveurs (d'ontologie et d'application), Générateur de portails Web (basés sur les ontologies), éditeur d'ontologie (construction et maintenance).

10.1 Protégé2000¹²

Protégé-2000 [Noy et al., 2000], successeur de *ProtégéWIN*, est un outil, une plate-forme et une librairie d'ontologies. Dans le modèle des connaissances de Protégé2000 les ontologies consistent en une hiérarchie de *classes* qui ont des attributs (*slots*), qui peuvent eux-mêmes avoir certaines propriétés (*facets*). L'édition des listes de ces trois types d'objets se fait par l'intermédiaire de l'interface graphique, sans avoir besoin d'exprimer ce que l'on a à spécifier dans un langage formel : il suffit juste de remplir les différents *formulaires* correspondant à ce que l'on veut spécifier. Ce modèle autorise d'ailleurs une liberté de conception assez importante puisque le contenu des formulaires à remplir peut être modifié suivant les besoins via un système de *méta-classes*, qui constituent des sortes de « patrons » de connaissance.

L'interface, très bien conçue, et l'architecture logicielle permettant l'insertion de *plugins* pouvant apporter de nouvelles fonctionnalités (par exemple, la possibilité d'importer et d'exporter les ontologies construites dans divers langages opérationnels de représentation ou

⁹ <http://delicias.dia.fi.upm.es/webODE/>

¹⁰ L'outil est disponible gratuitement à <http://opales.ina.fr/public/>

¹¹ API : Application Programming Interface

¹² <http://protege.stanford.edu/index.shtml>

encore la spécification d'axiomes) ont participé au succès de Protégé2000 qui regroupe une communauté d'utilisateurs assez importante et constitue une référence pour beaucoup d'autres outils.

Pour conclure, nous rappelons que dans l'ensemble, les outils de l'ingénierie ontologique ont mûri considérablement au cours des dernières années, mais qu'ils doivent encore s'enrichir de certaines caractéristiques propres à l'ontologie et à la collaboration. Pour une énumération exhaustive des outils permettant d'utiliser les ontologies d'une manière générale, nous renvoyons le lecteur au Déliverable OntoWeb 1.3 [Fensel et Gómez-Pérez, 2002].

11. L'exploitation des ontologies

Nous allons voir maintenant comment nous allons utiliser les ontologies. Celles-ci sont encodées dans un langage formel doté d'une sémantique, ce qui permet de conduire des inférences sur les connaissances représentées. Nous présentons donc quelques moteurs d'inférences opérationnels implémentant la sémantique des langages du Web Sémantique.

11.1 Les moteurs d'inférences

Parmi les moteurs d'inférences, nous distinguons les langages et API programmables qui permettent assez rapidement de construire ses propres applications, des raisonneurs « prêts à l'emploi » généralement développés autour des LDs.

Dans la première catégorie :

- **TRIPLE** est un langage modulaire de règle basé sur la logique de Horn et sur F-Logic qui permet à la fois d'interroger et d'effectuer des inférences sur des données RDF [Sintek et Decker, 2002]. Pour les LDs DAML+OIL et OWL, il fournit un accès à des classifieurs externes comme FaCT.
- **KAON** [Maedche et al., 2003] (Voir section 10) contient, entre autre, une API JAVA qui permet d'accéder aux ontologies en s'affranchissant de la manière dont celles-ci sont stockées. Les langages RDF et DAML+OIL sont implémentés.
- **Jena2**¹³ est également une API JAVA, qui fournit des raisonneurs pré-définis implémentant les langages RDFS et OWL Lite. Cette API permet aussi de construire un

¹³ <http://jena.sourceforge.net/index.html>

raisonneur à base de règles pour faire du chaînage avant, du chaînage arrière ou pour lancer une exécution hybride.

Dans la seconde catégorie, on peut trouver :

- **FaCT**, un classifieur de LDs, développé par I. Horrocks [Horrocks, 1998]. Celui-ci est aujourd'hui relié à différents éditeurs d'ontologies (OilEd, Protégé2000) afin de tester leur consistance tout au long du processus de construction. C'est avant tout un classifieur terminologique (T-Box).
- **Racer** est également un classifieur de LDs qui implémente la sémantique du langage OWL Lite [Haarslev et Möller, 2001]. Il est beaucoup plus performant que FaCT en ce qui concerne le raisonnement assertionnel (A-Box).

12. Synthèse

L'un des principaux problèmes du Web est de trouver l'information ou le service dont l'utilisateur a besoin dans un espace d'information vaste et peu ou pas structuré. Avec le Web Sémantique, les ontologies vont devenir les points centraux des systèmes d'information de demain. Même si leur principal défaut aujourd'hui est d'être élaborées pour des applications particulières, néanmoins il existe déjà des banques d'ontologies sur le Web. Cette révolution technologique offre de nouvelles opportunités au développement de systèmes et de services intelligents et facile d'utilisation fondés sur la connaissance des contenus du Web.

Une ontologie est le résultat d'un consensus social et définit un vocabulaire partagé par les utilisateurs et par les machines. Elle permet une interprétation automatisée grâce au fait qu'elle assure une sémantique unique pour chacun des concepts représentés. L'utilisation d'une ontologie dans un système à base de connaissance avec un moteur d'inférence permet de faire des requêtes sur la connaissance et de faire de la recherche d'information plus précise au moyen de métadonnées sémantiques.

Dans ce chapitre, nous avons examiné les développements récents survenus dans le domaine de l'ontologie. Dans la situation actuelle, l'on remarque une bonne compréhension globale de la nature et de la fonction des ontologies et il apparaît que les travaux réalisés se fondent désormais sur les acquis pour évoluer dans de nouvelles directions. Il est clair que les ontologies acquièrent une importance notable dans un grand nombre de domaines tels que

l'Internet, l'intégration de systèmes d'information, l'extraction de connaissances, le maintien de la cohérence dans les bases de données, la recherche d'information, etc. Les recherches en cours convergent sur le plan des motivations : permettre le partage des connaissances et des données au sein d'une ou plusieurs communautés mais aussi parmi les utilisateurs et entre les applications.

Nous avons montré comment construire des ontologies afin de capturer la connaissance d'un domaine particulier. Ces ontologies sont présentées dans un langage formel doté d'une sémantique et permettent de construire des systèmes à base de connaissances sur lesquels il est possible de raisonner.

Il est donc important de prendre en compte le Web Sémantique et donc les ontologies dans la conception des modèles utilisateurs et de systèmes de personnalisation comme nous les concevons. L'utilisation d'ontologies permet de définir un vocabulaire commun aux utilisateurs et sans ambiguïté. Quant au Web sémantique, il permet de séparer la structure de présentation, la structure logique d'une ressource et la sémantique de son contenu et ainsi, de faciliter les raisonnements sur ce qui a réellement du sens : *le contenu*.

Nous allons maintenant établir comment nous pouvons adapter le concept d'ontologie à la spécification des connaissances des utilisateurs. Nous allons donc voir dans le prochain chapitre pourquoi et comment nous proposons d'employer les ontologies pour décrire la structure et le contenu des profils des utilisateurs pour la mise en œuvre de la personnalisation.

Modèle de l'utilisateur

CHAPITRE 4 : Modèle de l'utilisateur	75
1. Introduction	76
2. Un modèle utilisateur, pourquoi faire?	78
3. Dégagement de la définition du modèle utilisateur	78
3.1 Les données personnelles	79
3.2 Profession	79
3.3 Rôle de l'utilisateur	81
3.4 Les préférences	83
3.5 Le centre d'intérêt	84
3.5.1 La modélisation d'intérêts	84
3.6 Le modèle de connaissance	85
3.6.1 La génération du modèle de recouvrement	86
3.6.2 L'expérience et les compétences	88
3.7 L'historique	88
3.7.1 Gestion de bookmarks	89
3.8 Modèle de groupe	89
3.9 La sécurité	90
3.9.1 P3P: Platform for Privacy Preferences Project	91
3.10 Gestion des conditions matérielles	91
4. Les propriétés des caractéristiques utilisateurs	94
4.1 Granularité, quel est le sujet modélisé ?	94
4.2 La donnée est-elle stable ?	95
4.3 Quelle est sa durée de vie ?	95
4.4 Dans quel cadre est utilisée la caractéristique ?	95
4.5 L'acquisition. Qui renseigne la caractéristique ?	95
4.6 Quelles sont les principales sources de données ?	96
4.6.1 Les formulaires (ou questionnaires)	96
4.6.2 Les fichiers Log et les clickstreams	96
4.6.3 Le retour de préférences ou Relevance Feed-back	97
4.7 Identification des utilisateurs	98
4.7.1 Adresse IP	98
4.7.2 Identifiants de session	98
4.7.3 Les Cookies	99
4.7.4 Combinaison d'un pseudonyme et d'un mot de passe	99
4.8 Implication de l'utilisateur dans la génération des profils	99
5. Critères d'évaluation	100
6. Conclusion	102

L'objectif de cette étape de modélisation est de permettre de décider comment les utilisateurs vont être représentés ? Quelles sont les caractéristiques des différents utilisateurs ? Dans le cadre applicatif des laboratoires de recherche. C'est sur ces quelques caractéristiques que la personnalisation reposera.

1. Introduction

Nous devons apporter des innovations supplémentaires au niveau des modélisations d'utilisateur existantes. Ceci contribue, selon nous, à tendre vers de meilleures capacités de personnalisation et de prise en compte des caractéristiques et des besoins des différents utilisateurs.

C'est dans cette optique que nous souhaitons maintenant définir un concept de modélisation de l'utilisateur. Après avoir alors exploré différentes techniques de conception et d'implémentation des modèles d'utilisateurs déjà existants, nous prolongerons nos travaux de recherche par une définition d'un modèle de l'utilisateur hiérarchisé, plutôt que des combinaisons de vecteurs comme le font la plupart des systèmes antérieurs (cf., chapitre 1), afin de pouvoir jouer sur le niveau de généralité et sur le nombre de termes utilisés. En effet, si nous pouvons espérer dans l'avenir disposer de nombreux services et/ou documents personnalisés, il sera nécessaire d'avoir accès à différentes catégories du modèle utilisateur qui ne seront sans doute pas sans intersection commune.

Nous nous sommes inspirés de la classification de [Koch, 2000] des modèles des utilisateurs en fonction de la connaissance qu'ils véhiculent (cf., Chapitre 1, section 2.1). D'où l'idée de concevoir des modèles indépendants du domaine d'application et de modèles dépendants du domaine.

En plus, l'approche adoptée par [Moukas, 1997], dans laquelle il combine jugement de l'utilisateur et détection automatique de ces centres d'intérêt nous semble très intéressante (cf. chapitre 1, section 3.1), sauf que dans notre modèle nous préconisons deux techniques de représentation des centres d'intérêts. La première, permettant la création de modèles qui sont dépendants du domaine, est déterminée en restreignant à une ontologie du domaine la possibilité offerte à un utilisateur de donner ses centres d'intérêts. Le modèle d'intérêt définit un ensemble de couples (Attribut, Valeur) où l'attribut correspond à un concept du domaine, et la valeur traduit l'intérêt de l'utilisateur pour ce concept. Nous considérons donc une ontologie du domaine pour compléter la définition du centre d'intérêt en explicitant la sémantique de certains termes (ex. JAVA : langage de programmation et danse).

La seconde technique, qui est indépendante du domaine, est accomplie à travers l'analyse des documents personnels de l'utilisateur. Dans ce cas, les centres d'intérêt sont les documents eux-mêmes. Ainsi, la création du modèle d'intérêt de l'utilisateur consiste à trouver les bonnes caractéristiques à partir des documents. Pour cela nous utilisons la mesure TFIDF (Term Frequency - Inverse Document Frequency) (cf. Chapitre 1, section 3.3.1) qui propose de pondérer chaque terme par sa fréquence dans le document.

Quant à la modélisation des connaissances de l'utilisateur, nous nous sommes inspirés essentiellement du modèle overlay (cf., chapitre 1, section 2.2.2) - qui associe pour un utilisateur particulier, une valeur à chaque concept du domaine. Chaque valeur correspond à une estimation du niveau de connaissance de l'utilisateur pour le concept. Nous pouvons noter que la modélisation des connaissances repose sur un modèle du domaine (ou ontologie du domaine), notre contribution ne portant pas sur la définition d'une nouvelle ontologie de domaine, nous avons choisi d'utiliser dans nos expériences une ontologie existante.

Et pour respecter certaines structures dans les laboratoires de recherche, nous intégrons, dans le modèle utilisateur, la profession ou la fonction de ce dernier, car les personnes ayant des fonctions différentes accéderont à des fonctionnalités différentes du système.

Un autre aspect à prendre en compte est l'activité professionnelle de l'utilisateur. Une approche formelle, issue de [Prat, 1997], est utilisée pour décrire les objectifs utilisateurs, nous l'avons donc adaptée pour le contexte de notre travail.

Nous avons intégré dans le modèle utilisateur proposé des caractéristiques démographiques, comme le préconise [Kobsa et al., 2001], en plus des données personnelles, des préférences des utilisateurs, de l'historique de leurs actions, des données environnementales et de sécurité.

Un autre point important consiste en la prise en compte des besoins de groupes d'utilisateurs. Pour cela, on utilise, pour la définition des stéréotypes d'utilisateurs (ou modèles de groupe) une seule caractéristique qu'est le domaine de recherche de ces derniers.

Donc, ce chapitre est organisé comme suit :

Nous justifions dans la première section la nécessité de disposer d'un modèle de l'utilisateur performant pour favoriser la personnalisation. La section 3 décrit les différentes informations retenues pour la description de ce modèle. Nous discuterons dans la section 3.5 comment peut-on modéliser les intérêts de l'utilisateur. La section 3.6 s'articule autour de la modélisation des connaissances de l'utilisateur. La section 3.7 concerne la prise en compte de

l'historique des actions de l'utilisateur pour la construction du modèle utilisateur. Nous définissons, dans la section 3.8, le modèle de groupe qui exploite la notion de stéréotype. Nous décrivons ensuite, dans la section 4, les propriétés des informations retenues et les critères d'évaluation dans la section 5.

2. Un modèle utilisateur, pourquoi faire?

Le modèle utilisateur est le cœur même de tout système de personnalisation. Il a pour but de représenter « le plus fidèlement » possible l'utilisateur d'un système et permet de personnaliser le contenu, la présentation ainsi que la navigation.

La prise en compte de l'utilisateur est donc un élément essentiel, non seulement pour modéliser le savoir faire et apporter une meilleure gestion du capital intellectuel, mais aussi pour fournir d'autres fonctions avancées. Il contient des caractéristiques sur les connaissances, les préférences, les objectifs, les centres d'intérêts, etc. d'un utilisateur. Chacune des caractéristiques peut avoir des propriétés particulières que nous développerons plus loin dans la section 4. Pour chacune de ces caractéristiques, une classe est décrite dans le modèle utilisateur ; les associations avec les autres classes des modèles (utilisateur et du domaine) sont spécifiées.

Il faut donc souligner l'importance d'établir une bonne connaissance des utilisateurs et de leurs aspirations avant de choisir les informations à inclure dans un profil.

3. Dégagement de la définition du modèle utilisateur

Nous voulons décrire l'ensemble des caractéristiques et énumérer les nombreuses « qualités » qui sont susceptibles de nous aider à distinguer, et par là même à spécifier, chacun de nos utilisateurs.

En conséquence à nos recueils d'informations construits jusqu'ici, nous conservons et travaillons plus avec des données que nous avons surtout cherchées à rendre synthétiques, orientées vers la pratique et les implémentations futures. Nous aboutissons à la définition ci-dessous faisant suite à maintes réflexions dans le but principal de créer un « modèle de l'utilisateur » à la fois ergonomique, efficace et exhaustif.

Le modèle utilisateur est défini selon:

- Les données « personnelles » permettant d'identifier et de catégoriser l'utilisateur.
- Les préférences regroupant l'ensemble des caractéristiques sur lesquelles l'utilisateur peut agir pour modifier le comportement du système (niveau de détails souhaités,

préférences de visualisation, préférences concernant la présentation des informations, etc.).

- L'historique ayant pour objectif de mémoriser l'activité de l'utilisateur.
- Données de sécurité. Restrictions sur l'accès au profil.
- Les données environnementales, regroupant les caractéristiques du dispositif d'accès, de la plate forme matérielle et logicielle.
- Les connaissances de l'utilisateur sur le domaine de l'application ou l'état de son savoir.
- Les centres d'intérêts de l'utilisateur.
- Groupe d'utilisateur. Prise en compte des besoins d'un groupe d'utilisateurs.
- Le rôle de l'utilisateur au sein de l'organisation (laboratoire de recherche, groupe de recherche, etc.), ou la tâche à réaliser.

3.1 Les données personnelles

Pour que l'utilisateur d'un système personnalisable puisse en bénéficier, il est nécessaire qu'il soit identifié dans le système. Les données personnelles ont un double objectif, d'une part la gestion de l'identification de l'utilisateur (Nom, prénom, qui permettent une personnalisation très simple par affichage de ces données dans la page d'accueil et donne le sentiment à l'utilisateur d'être un interlocuteur privilégié) et d'autre part, elles permettent de catégoriser l'utilisateur en fonction de caractéristiques telles que les attributs d'authentification (login, mot de passe, adresse IP, etc.), les facteurs démographiques (âge, sexe, première langue, lieu de naissance, particularités sociales et culturelles, etc.), les contacts personnels et professionnels de l'utilisateur et d'autres informations comme le groupe sanguin, le numéro du compte bancaire, etc. Ces données sont stables et renseignées par l'utilisateur.

3.2 Profession

La profession ou la fonction est un facteur important, surtout pour la conception d'un système qui sera employé dans certains établissements tels que les laboratoires de recherche. Les personnes ayant des fonctions différentes accéderont à des fonctionnalités différentes du système.

Dans un premier temps, nous caractérisons l'utilisateur en fonction de son niveau d'étude. Le tableau 4.1 regroupe une liste non exhaustive des différents niveaux éducationnels.

Niveau primaire ou secondaire
Aucun diplôme/ certificat d'étude primaire
BEF (Brevet d'Etudes Fondamentales)
Baccalauréat
Diplôme universitaire de premier cycle (BAC+2)
Diplôme universitaire de deuxième cycle (BAC+3 ou BAC+4) :
- Licence,
- Maîtrise,
- DEUA,
- Etc.
Diplôme universitaire de troisième cycle (BAC+5)
Diplôme d'ingénieur
Diplôme d'une grande école autre qu'une Ecole d'ingénieur
Doctorat, Thèse

Tableau 4.1 – Niveaux éducationnels

Nous avons procédé, ensuite, à une classification professionnelle de la communauté scientifique et technique d'un laboratoire de recherche. Ainsi, on trouve les niveaux suivants: Chercheur, Enseignant /chercheur, Etudiant et Staff administratif.

Chaque profession est structurée hiérarchiquement. Cette graduation est un indicateur d'expérience. Le dernier diplôme pour un étudiant, le titre officiel (maître de conférence, professeur) pour un enseignant/chercheur, maître de recherche et directeur de recherche pour un chercheur, secrétaire pour le staff administratif. Associé au domaine disciplinaire de la profession, nous pouvons en déduire un *niveau d'expertise* dans le domaine de l'individu.

Dans la catégorie des enseignants/chercheurs, nous distinguons les différents grades selon l'expertise acquise et la production scientifique fournie. La notion de production scientifique validée dépend des modes de publication (et donc de légitimation) dont disposent les différentes communautés scientifiques. Pour certaines cette validation passera par des publications dans des revues internationales à comité de lecture. Mais pour d'autres, la meilleure légitimation se concrétisera par des publications dans des communications dans certains colloques internationaux.

Pour les enseignants, on peut distinguer les grades présentés dans le tableau 4.2. Quant aux chercheurs, le conseil a examiné les textes concernant les promotions des grades de chercheurs associés (Décision N°4 du 25/03/1997 relative aux conditions de mise en poste, de travail et de rémunération des chercheurs associés). Les propositions retenues par le conseil sont présentées dans le tableau 4.3.

Grade	Conditions
Professeur (Prof)	Cinq années d'exercice en qualité de Maître de conférences avec, au moins, un article publié dans une revue internationale, une communication et un magistère soutenu.
Maître de Conférences (MC)	Doctorat d'état avec trois années d'exercices en qualité de Chargé de Cours et au moins un article publié dans une revue internationale.
Chargé de Cours (CC)	Maître assistant titulaire avec trois années d'exercices en qualité d'enseignant
Maître Assistant Titulaire (MAT)	Maître assistant stagiaire avec neuf mois d'exercices en qualité d'enseignant
Maître Assistant Stagiaire (MAS)	Magistère ou diplôme étranger équivalent.

Tableau 4.2- Grades des enseignants

Grade	Conditions
Directeur de Recherche (DR)	Professeur avec six années d'exercice en qualité de chercheur, deux articles publiés dans des revues internationales, deux communications et deux magistères soutenus.
Maître de Recherche (MR)	Professeur ou Maître de conférences avec trois années d'exercice en qualité de chercheur, un article publié dans une revue internationale, une communication et un magistère soutenu.
Chargé de Recherche (CR)	Maître de conférence ou Chargé de cours avec deux années d'exercice en qualité de chercheur, un article publié dans une revue internationale ou deux communications.
Attaché de Recherche (AR)	Maître assistant

Tableau 4.3- Grades des chercheurs

3.3 Rôle de l'utilisateur

Un *Rôle Utilisateur* possède un *nom* et est défini par un sous-ensemble d'*Objectifs Utilisateurs*. On peut voir le rôle comme la définition des différentes fonctionnalités qu'un utilisateur est autorisé à mener à bien en fonction de son statut.

Une approche formelle, issue de [Prat, 1997], est utilisée pour décrire les objectifs utilisateurs. Un Objectif Utilisateur est exprimé par une clause composée d'un *verbe* et d'un ensemble de *paramètres*, chacun étant associé à une fonction sémantique. Les paramètres qui entrent dans la description d'un objectif sont de quatre types, certains étant affinés en sous-types :

- la *Cible* qui désigne les entités affectées par l'objectif ; il peut s'agir d'un *Objet* ou d'un *Résultat* selon que la cible préexiste ou non à la réalisation de l'objectif.
- la *Façon* qui exprime la *Manière* et/ou les *Moyens* utilisés pour atteindre le but ;

- la *Direction* qui identifie la *Source* ou la *Destination* d'un objectif ;
- le *Bénéficiaire* qui correspond à la personne ou au groupe en faveur de qui l'objectif doit être atteint.

Donc cette partie du modèle utilisateur répond à la question suivante : Quel est le rôle de l'utilisateur dans l'organisation? Il est évident que le rôle est lié à une activité ponctuelle de l'individu. Cette activité s'inscrit dans le cadre de la profession. Intéressons-nous aux activités et établissons un tableau d'activité par profession (tableau 4.4).

Rôle fonctionnel	Activité (ensemble d'objectifs)	Moyens
Etudiant	<p>Rédiger</p> <ul style="list-style-type: none"> - Exposé (oral, écrit) - Mémoire - Rapport de synthèse - Etudes de cas <p>Réaliser un projet de fin d'étude</p>	<p>Pour la rédaction d'exposés, de mémoires :</p> <ul style="list-style-type: none"> - Autres exposés - Articles de synthèses - Articles primaires - Mémoires, thèses <p>Pour la réalisation d'un projet de Recherche ou de développement :</p> <ul style="list-style-type: none"> - articles scientifiques - rapports techniques - articles professionnels
Chercheur, enseignant /chercheur	<p>Rédiger</p> <ul style="list-style-type: none"> - article de recherche - article de synthèse - Contrat - PV - Etc. <p>Préparer</p> <ul style="list-style-type: none"> - Cours (général, approfondi) - Evénement éducationnel - Etc. <p>Veille technologique</p> <ul style="list-style-type: none"> - Mettre à jour des connaissances <p>Organiser</p> <ul style="list-style-type: none"> - Organisation d'événement (assistance, animation, organisation, etc.) <p>Encadrer</p> <ul style="list-style-type: none"> - Encadrement pédagogique (conseil, orientation, évaluation) <p>Synthèse bibliographique</p> <p>Administration</p>	<p>Pour la rédaction d'article de recherche</p> <ul style="list-style-type: none"> - Articles de recherches - Articles de synthèses - Résumés et rapports de conférence <p>Pour la préparation d'un cours.</p> <ul style="list-style-type: none"> - Ouvrages de référence - Autres cours (écrits à caractère didactique) - Eventuellement articles de vulgarisation. - CD - Internet - Etc. <p>Veille scientifique, mise à jour des connaissances</p> <ul style="list-style-type: none"> - Articles primaires récents - Résumés et rapports de conférence

Tableau 4.4 – Liste non exhaustive des activités par profession

Les chercheurs et thésards ont une activité centrée sur la rédaction d'article de recherche, de thèses, de mémoires ou de rapports. La veille technologique est une activité commune aux chercheurs (et tout spécialiste du domaine) sur l'information récente dans un domaine précis.

Une classe particulière, incluse dans la classe des chercheurs, est celle des enseignants chercheurs qui en plus de leur activité de recherche doivent assumer une activité d'enseignement. Cette activité implique la préparation ou la mise à jour d'un cours. L'enseignant/chercheur élabore en principe un cours dans son domaine de prédilection, mais il n'est pas rare que la diversité des enseignements l'amène à enseigner dans un domaine plus ou moins proche.

Notre dernière catégorie concerne la classe des étudiants (sauf les thésards que nous incluons dans les chercheurs). Leur activité est moins centrée sur des recherches de haut niveau mais plutôt sur des exposés de synthèse, des études de cas ou bien la rédaction de mémoire.

3.4 Les préférences

Tous les utilisateurs ne sont pas intéressés par les mêmes informations, ou par la même présentation de l'information.

Les préférences de l'utilisateur indiquent s'il est possible pour l'utilisateur d'exprimer ses préférences de visualisation, afin de permettre la personnalisation de la présentation (cf. section 3.10). Ceci concerne d'abord tout ce qui est lié aux modalités de présentation des résultats en fonction de la plateforme, de la nature et du volume des informations délivrées, des préférences esthétiques ou visuelles de l'utilisateur. A ces modalités de présentation, on peut ajouter les modalités d'exécution.

Les préférences d'un utilisateur peuvent donc concerner :

- le choix des méthodes de personnalisation et de services offerts,
- le choix des composants graphiques à ne pas afficher,
- les préférences concernant la présentation des informations,
- les modalités d'exécution, décrivant le moment d'exécution d'une requête (mode pull : l'utilisateur peut filtrer l'information conformément à ses intérêts et besoins ou push : l'utilisateur reçoit l'information susceptible d'être la plus pertinente pour lui en fonction de son profil),
- le niveau de détails souhaités,
- etc.

Mais bien souvent l'utilisateur a des idées floues sur ce qu'il veut réellement et par conséquent a du mal à exprimer clairement ses désirs, et il est très possible que cet utilisateur fausse lui-même les données le concernant.

Nous pouvons donc envisager des profils extraits à partir du contenu des documents le concernant, qui pourraient fournir des informations intéressantes; c'est ce que nous allons présenter dans la section suivante.

3.5 Le centre d'intérêt

Le centre d'intérêt d'un utilisateur exprime le domaine d'expertise de l'utilisateur ou son périmètre d'exploration. Il peut être défini par un ensemble de mots clés ou peut être vu comme une présélection virtuelle qui réduit la masse d'informations à prendre en compte. Ces centres d'intérêts seront ensuite utiles pour filtrer et organiser le résultat des recherches.

3.5.1 La modélisation d'intérêts

La modélisation des centres d'intérêts est utilisée pour raffiner et faciliter le travail de l'utilisateur dans certaines activités. Elle est accomplie à travers l'analyse des documents personnels de l'utilisateur, on suppose que chaque utilisateur maintient un espace dans lequel sont stockés tous les documents le concernant au sein de l'organisation (documents de travail, publications scientifiques, les listes de ses sites préférés (cf. section 3.7.1), une page Internet, etc.). Dans ce cas, les centres d'intérêt sont les documents eux-mêmes.

La première étape dans la création du modèle d'intérêt de l'utilisateur consiste à trouver les bonnes caractéristiques à partir des documents. Pour cela nous utilisons la mesure TFIDF (Term Frequency - Inverse Document Frequency) qui propose de pondérer chaque terme par sa fréquence dans le document (voir chapitre 1, section 3.3.1). Cette pondération reflète l'idée qu'un terme est important s'il est fréquent dans un document et peu fréquent dans les autres. Diverses méthodes de pondération utilisant cette idée ont été proposées dans la littérature. Notre contribution ne portant pas sur la définition d'une nouvelle méthode de pondération, nous avons choisi d'utiliser dans nos expériences la fonction TFIDF très utilisée en classification de documents ou en recherche d'informations.

Un nombre limité de termes est alors choisi pour représenter les documents sous forme d'un vecteur, où chaque valeur du vecteur est la valeur TFIDF d'un terme.

TF_i est la fréquence du terme i dans la collection de documents et IDF est la fréquence inverse par rapport le nombre de documents, calculé avec $\log(N/n_i)$ où N est le nombre total de documents dans la collection et n_i est le nombre de documents qui contiennent le terme i .

Ainsi, la valeur TFIDF d'un terme quelconque i par rapport à une collection de documents est donnée par :

$$TFIDF(i) = TF_i \times \log\left(\frac{N}{n_i}\right)$$

La représentation d'un certain document d prend en compte les fréquences des termes en d , et est donnée par le vecteur $d = \left\{ TF_1^d \times \log\left(\frac{N}{n_1}\right), \dots, TF_n^d \times \log\left(\frac{N}{n_n}\right) \right\}$

Où TF_i^d est la fréquence du terme i dans le document d . Ces vecteurs peuvent ensuite être utilisés pour déterminer le centre d'intérêt de l'utilisateur en retenant les mots les plus fréquents.

La mesure TFIDF pour la sélection de termes a été choisie, car celle-ci est souvent utilisée dans d'autres travaux. Par ailleurs, elle est très simple de calculer et sa sémantique semble robuste. Cette technique permet la création de modèles qui ne sont pas dépendants du domaine d'application et tendent à réduire le besoin de feedback.

Une autre technique permettant la création de modèles qui sont dépendants du domaine est déterminée en restreignant à une ontologie du domaine la possibilité offerte à un utilisateur de donner ses centres d'intérêts

Le modèle d'intérêt définit un ensemble de couples (Attribut, Valeur) définis comme suit:

- un attribut correspond à un concept du domaine,
- la valeur traduit l'intérêt de l'utilisateur pour ce concept (très bon, bon, intéressant, mauvais ou très mauvais).

Une fois les centres d'intérêts de l'utilisateur définis on complète le contenu du profil avec les connaissances de l'utilisateur qui sont spécifiques pour un domaine donné. Le modèle de connaissance est présenté dans la section suivante.

3.6 Le modèle de connaissance

La façon la plus simple de gérer les connaissances est de mémoriser ce que l'utilisateur connaît ou ne connaît pas. Pour cela, on utilise un modèle de recouvrement qui associe aux concepts d'un domaine un niveau de connaissance (cf. chapitre 1, section 2.2.2).

Nous considérant une ontologie du domaine pour la définition du niveau de connaissance de l'utilisateur pour les concept de ce domaine.

Cette information peut être obtenue explicitement par l'utilisateur ou implicitement par le système sur la base de règles d'inférences.

Les éléments qui constituent le modèle de connaissance sont des couples (concept, niveau) définis comme suit :

1. Concept du domaine
2. Niveau de connaissance. L'utilisateur associe aux concepts du domaine un niveau qu'il juge être le sien parmi 4 (Expert, Confirmé, Novice, Inconnu).

3.6.1 La génération du modèle de recouvrement

Le principe consiste à déduire le niveau de connaissance de l'utilisateur sur l'ensemble des concepts du domaine en fonction des niveaux qu'il a associé à un sous-ensemble des concepts.

La première règle est importante car elle consiste à ne pas réévaluer un concept estimé par l'utilisateur. En effet, nous prenons comme hypothèse que l'utilisateur est capable d'évaluer sa connaissance correctement. Cela permet l'expression de niveaux tels que : «Je suis expert en Programmation, même si je me considère confirmé en Pascal et novice en Java. »

Soit D l'ensemble des concepts du domaine organisés hiérarchiquement, N l'ensemble des niveaux possibles de connaissance et U l'ensemble des utilisateurs, Fils (d, c) signifie que le concept c est un sous-concept direct du concept d et Evaluation (d, u, n) signifie que le concept d est évalué au niveau n pour l'utilisateur u .

Soit $d \in D$, et $u \in U$:

La première règle concerne l'existence d'un niveau pour le concept d . Il s'agit donc de conserver le niveau associé par l'utilisateur. Les règles suivantes permettent d'associer des niveaux aux concepts non renseignés par l'utilisateur.

- R1: $\exists n \in N / \text{Evaluation}(d, u, n)$

Pour les règles suivantes, l'utilisateur n'a pas spécifié de niveau de connaissance pour le concept d . Elles consistent à évaluer un concept à partir de ses sous-concepts qui sont plus spécifiques.

- R2 : $\neg \exists c \in D / \text{Fils}(d, c) \wedge \neg \exists n \in N \text{ Evaluation}(d, u, n)$
 $\Rightarrow \text{Evaluation}(d; u; \text{Inconnu})$
- R3 : $\forall c \in D / (\text{Fils}(d, c) \wedge \text{Evaluation}(c, u, \text{Inconnu}))$
 $\Rightarrow \text{Evaluation}(d, u, \text{Inconnu})$
- R4 : $\forall c \in D / (\text{Fils}(d, c) \wedge \text{Evaluation}(c, u, \text{Expert}))$
 $\Rightarrow \text{Evaluation}(d, u, \text{Expert})$
- R5 : $\exists c \in D / (\text{Fils}(d, c) \wedge \text{Evaluation}(c, u, \text{Confirmé}))$

\Rightarrow Evaluation (d, u, Confirmé)

- R6 : $\exists c_1, c_2 \in D / (\text{Fils}(d, c_1) \wedge \text{Fils}(d, c_2) \wedge \text{Evaluation}(c_1, u, \text{Expert}) \wedge (\text{Evaluation}(c_2, u, \text{Novice}) \vee \text{Evaluation}(c_1, u, \text{Inconnu})))$

\Rightarrow Evaluation (d, u, Confirmé)

- R7 : $\forall c_1 \in D / (\text{Fils}(d, c_1) \wedge (\text{Evaluation}(c_1, u, \text{Novice}) \vee \text{Evaluation}(c_1, u, \text{Inconnu}))), \exists c_2 \in D / (\text{Fils}(d, c_2) \wedge \text{Evaluation}(c_2, u, \text{Novice}))$

\Rightarrow Evaluation (d, u, Novice)

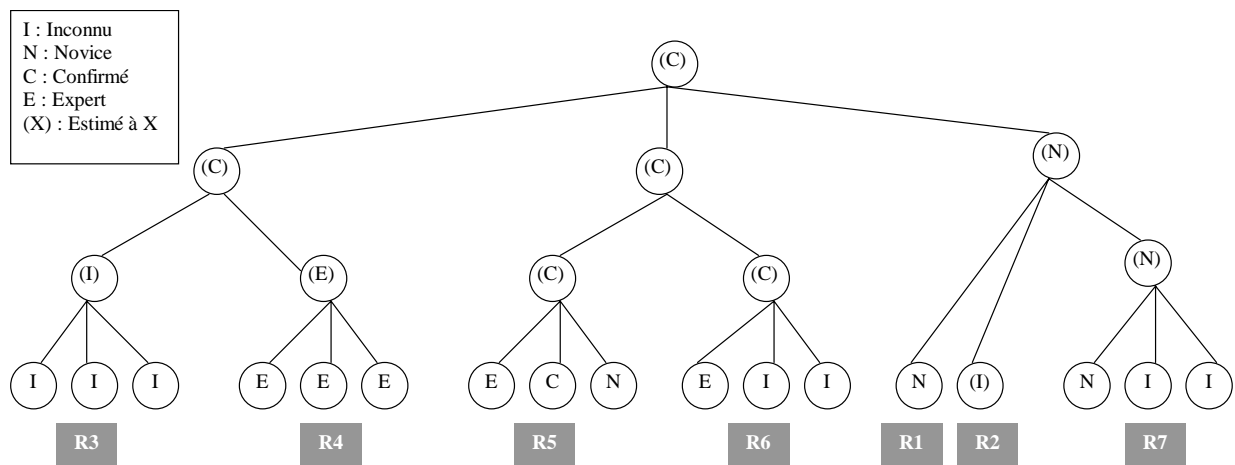


Figure 4.1 - Exemple graphique des règles de génération

Les règles R2 et R3 évaluent un concept à «Inconnu » soit parce qu'il n'a pas de fils soit parce que tous ses fils sont évalués à «Inconnu ». La règle R4 évalue un concept à «Expert » car tous ses fils sont évalués à «Expert ».

Les règles R5 et R6 évaluent un concept à «Confirmé » parce qu'au moins l'un de ses fils est au minimum évalué à «Confirmé » ou parce que parmi les évaluations de ses fils, il y a des niveaux «Expert » et des niveaux «Novice » ou «Inconnu ». La règle R7 évalue à «Novice » un concept dont les fils sont tous évalués «Novice » ou «Inconnu » avec au moins un fils évalué à «Novice ».

Cet ensemble de règles est facilement modifiable car il est défini dans un langage déclaratif associé aux ontologies.

3.6.2 L'expérience et les compétences

L'expérience et les compétences sont deux caractéristiques similaires à la connaissance définie dans le paragraphe précédent mais qui en diffèrent de par la nature même de l'information qu'elles représentent.

L'expérience de l'utilisateur représente son savoir-faire, la familiarité et l'aisance qu'il possède avec le type de système qui lui est présenté.

Les compétences possédées par l'utilisateur correspondent aux connaissances qui ne relèvent ni du domaine, ni de l'expérience mais qui sont néanmoins considérées comme pertinentes dans le fonctionnement du système. La compétence, en réalité, existe en relation avec autres concepts décrivant la capacité, l'habileté et le niveau d'expertise d'une personne.

Prenons l'exemple d'un étudiant en Informatique qui consulte un hypermédia sur l'Égypte Ancienne. Les connaissances évoquées dans le paragraphe précédent concernent l'Égypte Ancienne, alors que l'expérience porte sur les systèmes hypermédia et que les compétences peuvent regrouper ses connaissances en Informatique, en politique, etc.

3.7 L'historique

Lorsqu'il est fait mention de profils, il est généralement fait allusion à une dimension temporelle relativement longue. Un profil peut en effet contenir des données couvrant une période étendue, c'est-à-dire relevant de sessions différentes. C'est ce que nous avons appelé un profil basé sur des données « historiques ».

La solution que nous avons choisie a été de conserver les traces de navigation de l'utilisateur (fichiers Log¹ et clickstreams²). Les traces de l'utilisateur sont gérées automatiquement par le système. Ces informations pourront servir aussi à des fins statistiques mais nous ne nous y intéressons pas pour le moment.

Le système déduit des informations à partir d'une session d'utilisation: pour ce faire il mémorise l'historique de ses actions et des documents visités. Grâce à l'exploitation de ces informations, le système apprend certains centres d'intérêt de l'utilisateur (cf. section 3.5.1). Il les utilise ensuite pour la présentation de résultats.

¹ Les fichiers Log servent à garder une trace de l'activité d'un programme, il y inscrit toutes les actions effectuées

² Le clickstream provient de l'observation de la suite des actions de l'utilisateur sur le site et est disponible grâce aux fichiers Log

3.7.1 Gestion de bookmarks

Les documents jugés pertinents par l'utilisateur sont déduits des bookmarks³. En effet, lorsqu'il trouve un document intéressant, correspondant à ses besoins, un utilisateur peut sauvegarder son adresse (URL) au sein de ses bookmarks sous la forme d'un signet pour permettre un accès ultérieur rapide. Ces bookmarks représentent donc les pages web ayant un intérêt certain pour l'utilisateur ; intérêt pouvant être double du fait que ces pages web font l'objet :

- d'un accès fréquent (moteurs de recherches, sites d'actualités...),
- d'une grande pertinence concernant les centres d'intérêts de l'utilisateur (site web d'un chercheur dans le même domaine...).

De plus, ces bookmarks étant organisés sous la forme d'une arborescence, les besoins généraux se trouvent à la racine et les besoins spécifiques au niveau des feuilles. Cette arborescence peut donc être interprétée comme un ensemble hiérarchique de besoins.

Le format standard le plus évolué de *bookmarks* existant actuellement, *XBEL* (*The XML Bookmark Exchange Language*)⁴, ne fait que reprendre les caractéristiques communes des formats *Microsoft* (les « favoris ») et *Netscape* (les « signets »), dont les informations les plus utiles sont peut-être la dernière date de passage sur une page et la dernière date de modification de celle-ci.

3.8 Modèle de groupe

Un autre point important consiste en la prise en compte des besoins de groupes d'utilisateurs. Un laboratoire de recherche, regroupe des utilisateurs variés ne présentant pas les mêmes besoins, faisant preuve d'expériences ou de connaissances variées, exprimant des préférences diverses, etc. L'identification de différents groupes d'utilisateurs est un moyen de réduire la complexité sous-jacente à une telle hétérogénéité. En effet, les documents ayant un intérêt pour un groupe sont des documents qui sont pertinents pour une grande majorité d'individus.

Pour cela, on utilise un stéréotype (ou modèle de groupe). La notion de stéréotype, introduite par Rich [Rich, 1989], a pour but de grouper les utilisateurs en un nombre restreint de catégories, de manière à éviter une dispersion de l'effort de personnalisation due à une

³ Un bookmark (ou signet) est un raccourci vers une URL (adresse web). Les bookmarks correspondent aux liens « favoris » de l'utilisateur.

⁴ <http://pyxml.sourceforge.net/topics/xbel/>

trop grande variété de profils. Mais cette approche pose le problème des utilisateurs qui n'entrent dans aucun des stéréotypes définis, et nécessite de pouvoir créer ces stéréotypes.

Pour la définition des stéréotypes d'utilisateurs, on utilise des caractéristiques telles que l'âge ou l'activité professionnelle, ou encore des caractéristiques plus fines comme la connaissance ou le niveau de connaissance de tel ou tel concept du domaine.

Dans notre cas, le stéréotype d'utilisateur est décrit par une seule caractéristique pour le moment : le domaine de recherche. Le concept de *Groupe d'utilisateurs* permet de fédérer les informations relatives à plusieurs utilisateurs qui partagent certaines caractéristiques. Tous les membres du même groupe ont ainsi en commun le même domaine de recherche. Ainsi, toute personne qui n'est pas familiarisé avec cette classification académique trouve du mal à situer son domaine d'activité.

Néanmoins, bien qu'appartenant à un même groupe, les utilisateurs présentent des caractéristiques individuelles variant parfois dans une large mesure. En ce sens, que le système soit capable de s'adapter à différents groupes n'est pas suffisant. La prise en compte d'un utilisateur en tant qu'individu est également nécessaire. L'objectif est de proposer en priorité des personnalisations basées sur des spécifications personnelles ; en l'absence de celles-ci, les spécifications des groupes sont alors utilisées.

Un Utilisateur peut appartenir à un ou plusieurs groupe(s), et peut, en conséquence, remplir les différents rôles fonctionnels associés aux groupes dont il est membre.

3.9 La sécurité

La sécurité est une dimension fondamentale du profil. Elle peut concerner les données que l'on interroge ou modifie, les informations que l'on calcule, les requêtes utilisateurs elles-mêmes ou les autres dimensions du profil. La sécurité des données peut être exprimée par des niveaux de sécurité prédéfinis qui dépendent de la hiérarchie des vues autorisées. En effet, on doit mentionner clairement quels sont les droits de l'utilisateur à l'égard du modèle le concernant. Notamment la lecture et la modification (ou rectification) des caractéristiques et préférences par l'utilisateur lui-même devraient être autorisées même si de telles informations sont automatiquement déterminées par le système. Ceci permet de limiter le sentiment que peut avoir l'utilisateur de subir le système ou d'insatisfaction si les personnalisations ne lui conviennent pas.

Il existe aussi actuellement de nombreux travaux sur la description des politiques de

gestion des données utilisateur – publiques/privées, qui s'inscrivent dans le cadre des travaux du W3C autour de P3P⁵.

3.9.1 P3P: Platform for Privacy Preferences Project

Le projet P3P (Platform for Privacy Preferences) ou plate-forme d'expression de choix en matière de respect de la vie privée conçoit la protection de la vie privée et des données comme devant faire l'objet d'un accord entre l'internaute dont les données sont recueillies et le site Web qui collecte ces données.

- P3P est une politique de protection des données personnelles, notamment sur l'usage des informations générées par les cookies (cf. section 4.7.3). Il n'accepte donc que des cookies en provenance de sites certifiés P3P.
- Technologie de filtrage conçue pour favoriser la confidentialité et la confiance dans le cyberspace en permettant aux fournisseurs de contenu d'énoncer leurs pratiques d'utilisation des données personnelles et aux utilisateurs de contrôler les données concernant leur vie privée.

3.10 Gestion des conditions matérielles

Nous devons également tenir compte dans la représentation de l'utilisateur des conditions matérielles dans lesquelles il accède au système. Le dispositif utilisé constitue, en effet, le chaînon intermédiaire dans les interactions homme-machine.

Afin de considérer la diversité croissante des dispositifs d'accès exploités par les utilisateurs, il est nécessaire de proposer des solutions pour gérer les différences de capacités inhérentes aux caractéristiques techniques des machines utilisées. La taille d'affichage, capacité de la mémoire, vitesse du réseau, etc., varient ainsi considérablement en fonction des conditions matérielles d'utilisation : PC, ordinateur portable, télévision interactive, téléphone WAP⁶, PDA⁷, etc. La description des caractéristiques matérielles est utilisée dans le but de diffuser une information dans un format cohérent avec les capacités du dispositif de sortie.

Nous illustrons ce domaine à travers un format de données spécifiquement proposé par le W3C, appelée CC/PP⁸ (*Composite Capability / Preference Profile*) [Klyne et al., 2001]. Cette proposition est basée sur RDF (*Resource Description Framework* [RDF, 2004]) et les

⁵ <http://www.w3c.org/P3P/>

⁶ Wireless Access Protocol

⁷ Personal Digital Assistant

⁸ <http://www.ccpp.info/>

schémas associés. Elle offre un cadre générique pour la description des capacités des dispositifs d'accès. Ce qui est appelé *préférences des utilisateurs* dans CC/PP correspond en fait à la définition des caractéristiques particulières du dispositif utilisé par l'utilisateur. Le profil contient à la fois les caractéristiques du navigateur (comme sa capacité ou non à lire ActiveX ou Javascript), mais aussi les préférences de l'utilisateur (par exemple, même si Netscape est capable de lire des extraits sonores, il est inutile d'en envoyer si l'utilisateur a désactivé cette option). A ces dernières peuvent facilement s'ajouter des caractéristiques d'usage soit : des caractéristiques techniques (type de matériel, système d'exploitation, la bande passante ou le fait de déconnexion du profil pour certaines requêtes...).

La Figure 4.2 illustre un profil CC/PP défini en XML. Un tel profil est décrit par trois composants qui permettent de caractériser respectivement la plate-forme matérielle (*TerminalHardware*), la plate-forme logicielle (*TerminalSoftware*) et une application telle qu'un navigateur (*TerminalBrowser*). Chacun de ces composants est lui-même décrit à l'aide d'attribut (*e.g. display, name, version, etc.*).

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ccpp="http://www.w3.org/2000/07/04-ccpp#">
  <rdf:Description rdf:about="MyProfile">
    <ccpp:component>
      <rdf:Description rdf:about="TerminalHardware">
        <rdf:type rdf:resource="HardwarePlatform" />
        <display>320x200</display>
      </rdf:Description> </ccpp:component>
    <ccpp:component>
      <rdf:Description rdf:about="TerminalSoftware">
        <rdf:type rdf:resource="SoftwarePlatform" />
        <name>EPOC</name>
        <version>2.0</version>
        <vendor>Symbian</vendor>
      </rdf:Description>
    </ccpp:component>
    <ccpp:component>
      <rdf:Description rdf:about="TerminalBrowser">
        <rdf:type rdf:resource="BrowserUA" />
        <name>Mozilla</name>
        <version>5.0</version>
        <vendor>Symbian</vendor>
        <htmlVersionsSupported>
          <rdf:Bag>
            <rdf:li>3.0</rdf:li>
            <rdf:li>4.0</rdf:li>
          </rdf:Bag>
        </htmlVersionsSupported>
      </rdf:Description>
    </ccpp:component>
  </rdf:Description>
</rdf:RDF>

```

Figure 4.2 – Exemple d'un profil CC/PP décrit en XML

Le modèle utilisateur ainsi défini est décrit dans le tableau 4.5.

1	Données Personnelles	Données personnelles concernant l'utilisateur
1.1	Identité	Son identité
1.1.1	Nom	Son nom
1.1.2	Prénom	Son prénom
1.2	Authentification	Données d'authentification
1.2.1	Login	Son pseudonyme
1.2.2	Mot de passe	Le mot de passe correspondant
1.3	Facteurs démographiques	Données socio-démographiques
1.3.1	Age	Son âge
1.3.2	Sexe	Sexe de la personne
1.3.3	Langues	Langues maîtrisées...
1.3.4	Email	Son adresse électronique
1.3.5	Page Web	Sa page Web personnelle
1.3.6	N° téléphone	Son numéro de téléphone
1.3.7	Etat Civil	Célibataire, marié, etc.
1.3.8	Adresse	Où habite t-il ?
1.3.9	Handicap	Utilisateur mal voyant par exemple
2	Profession	Activité professionnelle ou qualification de l'utilisateur (chercheur, secrétaire, étudiant, etc.)
2.1	Discipline	Discipline ou spécialité couverte;
2.2	Niveau	Niveau éducationnel
3	Rôle	Activité Professionnelle ou rôle au sein du laboratoire de recherche. Quel type de travail ?
3.1	Activité	Un rôle est lié à une activité ponctuelle de l'individu, et il est défini par un sous-ensemble d'objectifs
3.2	Objectif	Un Objectif Utilisateur est exprimé par un verbe
3.2.1	Cible	Désigne les entités affectées par l'objectif ; il peut s'agir d'un Objet ou d'un Résultat selon que la cible pré existe ou non à la réalisation de l'objectif.
3.2.2	Façon	Exprime la Manière et/ou les Moyens utilisés pour atteindre le but
3.2.3	Direction	Identifie la Source ou la Destination d'un objectif
3.2.4	Bénéficiaire	Correspond à la personne ou au groupe en faveur de qui l'objectif doit être atteint.
4	Préférences	Données sur les préférences de l'utilisateur.
4.1	Détail	Niveau de détail (pointue ou généraliste).
4.2	Visualisation	Préférences de visualisation
4.3	Exécution	Les modalités d'exécution. le moment d'exécution d'une requête (mode pull ou push), la manière de notifier les résultats (différé, immédiat), etc.
4.4	Méthodes de personnalisation	le choix des méthodes de personnalisation et de services offerts

5	Centres d'intérêt	Ses centres d'intérêt
5.1	Domaine	Centres d'intérêt sur un domaine
5.1.1	Attribut	Un concept du domaine
5.1.2	Valeur	L'intérêt de l'utilisateur pour ce concept
5.2	Mots clés	Mots clés significatifs extraits des documents personnels de l'utilisateur
5.3	Poids	La mesure TFIDF (Term Frequency - Inverse Document Frequency) propose de pondérer chaque terme par sa fréquence dans le document
6	Connaissance	Données sur les connaissances de l'utilisateur
6.1	Domaine	Connaissances sur le domaine
6.1.1	Élément	Un concept du domaine
6.1.2	NiveauConnaissance	Un niveau de connaissance (Expert, Confirmé, Novice, Inconnu)
7	Historique	Données sur les accès aux dossiers
7.1	Bookmark	Données de bookmark
9	Données environnementales	Description des caractéristiques du dispositif d'accès exploité
9.1	Caractéristiques matérielles	Caractérise la plateforme matérielle. PC, Ordinateur portable, PDA, téléphone WAP
9.1.1	Taille d'affichage	Taille de l'écran
9.1.2	Capacité de la mémoire	Capacité mémoire du dispositif d'accès
9.2	Caractéristiques logicielles	Caractérise la plateforme logicielle
9.2.1	Système d'exploitation	Système d'exploitation de l'utilisateur (Windows, Linux, etc.)
9.2.2	Navigateur	Navigateur ou Browser (Internet Explorer, Netscape, etc.)
9.3	Connexion	Informations de connexion
9.3.1	Nom de connexion	Le nom de connexion
9.3.2	Bande passante	Vitesse du réseau
10	Groupe	Groupe d'utilisateurs. Communauté d'appartenance
10.1	Domaine	Domaine d'activité ou de recherche
11	Sécurité	Données de Sécurité. C'est une liste d'utilisateurs ayant les droits d'accès à ce profil.

Tableau 4.5 - Description du modèle utilisateur

4. Les propriétés des caractéristiques utilisateurs

Les caractéristiques modélisées possèdent des propriétés particulières soit en raison de leur acquisition, soit en raison de leur utilisation. Nous allons présenter ici les principales propriétés avec des questions qu'il est nécessaire de se poser :

4.1 Granularité, quel est le sujet modélisé ?

On peut modéliser un individu ou une classe d'individus (en utilisant des stéréotypes).

4.2 La donnée est-elle stable ?

Les caractéristiques considérées comme stables sont celles qui ne peuvent pas être remises en question ou qui sont rarement modifiées. Par exemple, le nom d'un utilisateur, son activité professionnelle, sa localisation géographique, etc.

Trois parties distinctes du modèle utilisateurs sont alors décrites :

- **La partie statique** est celle qui concerne les informations personnelles de l'utilisateur, comme par exemple, le nom, la date de naissance, l'adresse, la profession, etc. Ces informations, sont évidemment fournies par l'utilisateur;
- **La partie semi-dynamique** est celle qui est modifiée par l'utilisateur au fil du temps. Par exemple, les activités à faire;
- **La partie dynamique** est celle acquise au fur et à mesure que l'utilisateur interagit avec le système. Les données de navigation, par exemple.

4.3 Quelle est sa durée de vie ?

Un modèle utilisateur peut contenir des informations dont la validité se limite à la session courante, ou des informations qu'il est nécessaire de conserver pour une durée plus longue.

4.4 Dans quel cadre est utilisée la caractéristique ?

Toutes les informations contenues dans un modèle utilisateur ne sont pas utiles dans tous les cas. Certaines caractéristiques sont utilisées par le système pour des tâches particulières : recherche d'information, identification, personnalisation, etc.

Les caractéristiques utilisateurs mises en évidence peuvent contribuer, de façon complémentaire, à l'évaluation du besoin d'aide de l'utilisateur. Néanmoins, on peut à priori juger que ce besoin est fortement lié au niveau de connaissances du système et au niveau de spécialisation de l'utilisateur dans le domaine recherché.

4.5 L'acquisition. Qui renseigne la caractéristique ?

Certaines caractéristiques sont renseignées par l'utilisateur parce qu'il est le seul à les connaître comme les données personnelles, les préférences, mais d'autres peuvent être déduites par le système, comme par exemple la connaissance ou le centre d'intérêt qui peuvent être déduites de l'activité professionnelle de l'utilisateur ou acquises par des

techniques d'apprentissage ou de datamining exploitant les historiques des actions et des choix passés des utilisateurs, via les fichiers log et les clickstream (cf. section 3.7).

Le tableau 4.6 illustre la source de données pour chaque dimension du modèle utilisateur.

Caractéristique	Source
Données Personnelles	Fournies explicitement par l'utilisateur
Niveau d'études	Fourni par l'utilisateur. Certificats et diplômes
Profession	Fournie par l'utilisateur
Préférences	Exprimées par l'utilisateur
Centres d'intérêt	Fournis par l'utilisateur ou inféré par le système.
Connaissances	Fournies explicitement par l'utilisateur ou déduite par le système (modèle de recouvrement).
Rôle	Défini par l'utilisateur ou déduit de son activité professionnelle
Historique	Fichiers Log, clickstream et bookmark
Données environnementales	CC/PP (<i>Composite Capability / Preference Profile</i>)
Sécurité	Définie par l'utilisateur et assuré par le système

Tableau 4.6 – l'acquisition des données

4.6 Quelles sont les principales sources de données ?

Nous avons choisi de ne présenter que trois sources de données parmi les plus fréquemment employées. D'autres sources existent, telles que les logiciels de récolte automatique de données (*Automated Data Collection Software*), mais elles dépassent le cadre de ce mémoire. Le lecteur intéressé pourra se reporter à l'article de Cantor et Etgen [Cantor et Etgen, 1999] pour plus de précisions.

4.6.1 Les formulaires (ou questionnaires)

Le recueil d'informations à l'aide de questionnaires permet d'obtenir des informations sur les utilisateurs. Il n'est donc pas nécessaire de se lancer dans de complexes études sur les données pour déduire ces renseignements précieux. En plus les informations collectées sont précises (l'incertitude est minimale, les paramètres à estimer sont peu nombreux).

4.6.2 Les fichiers Log et les clickstreams

Cette source des données provient de l'observation de la suite des actions de l'utilisateur. Cette séquence porte le nom de « *clickstream* » et est disponible grâce aux fichiers Log (tableau 4.7).

Ce type de fichier sert à garder une trace de l'activité d'un programme, il y inscrit toutes les actions qu'il a effectuées. Dans le cas d'un site Internet, c'est le serveur HTTP (plus exactement le programme qui exploite le serveur) qui inscrit son activité dans un fichier Log.

Ce fichier est donc constitué d'une suite « brute » d'actions et il nécessite un traitement avant d'être en mesure de fournir les clickstreams.

Données	CLF	ECLF	Description
Hote	×	×	Nom de domaine pleinement qualifié du client ou adresse IP si le nom n'est pas disponible
Identifiant	×	×	Information d'identité fournie par le client si l'identification est active
Utilis. Authentifié	×	×	Si la requête concerne un document protégé par mot de passe, identifiant utilisateur utilisé dans la requête
Heure	×	×	Heure à laquelle la requête a atteint le serveur au format de date CLF {jj/Mmm/aaaa :hh :mm :ss zone}
Requête	×	×	Première ligne de la requête du client, en générale entre guillemets
Etat	×	×	Code d'état sur trois chiffres renvoyé au client
Octets	×	×	Nombre d'octets renvoyés au client, en excluant les en-têtes HTTP
Adresseur		×	URL du serveur adressant la requête
Agent utilisateur		×	Nom et version du client (navigateur)
Durée de traitement			Délai nécessaire pour servir la requête
Adresse IP			Adresse IP de l'hote distant (voir hote ci-dessus)
Port du serveur			Port canonique du serveur qui traite la requête
ID de processus			ID du processus enfant ayant traité la requête
Date formatée			Date et heure au format strftime(3) spécifié.
URL demandée			Chemin de l'URL demandée
Nom de serveur			Nom canonique du serveur traitant la requête
Cookie			Valeur du cookie extrait du fichier de cookie du client

Tableau 4.7 – Format d'un fichier log

CLF : Common Log Format (norme d'origine des fichiers Logs des serveurs Web).

ECLF : Extended Common Log Format (norme de base + extension)

4.6.3 Le retour de préférences ou Relevance Feed-back

Le retour de préférences est une alternative de collecte d'informations sur les préférences de l'utilisateur pour construire un profil utilisateur de manière dynamique. Au début, l'utilisateur fournit un ensemble de mots-clés décrivant ses préférences afin de permettre au système de construire un profil initial. A l'arrivée de nouveaux documents, le système utilise le profil utilisateur pour choisir les documents pouvant potentiellement l'intéresser. Par un processus de retour de pertinence (Relevance feedback), l'utilisateur n'indique pas seulement les documents pertinents mais aussi les documents non-pertinents.

Ces informations sont soit directement fournies par l'utilisateur (Explicites) ou soient récupérées ou dérivées à son insu (Implicites), il peut s'agir par exemple de l'exclusion des résultats qu'il n'aime pas, du nombre de clics qu'il a effectué sur le lien d'une page ou du nombre et de la nature des requêtes qu'il a émises, du temps passé sur une page, etc.

4.7 Identification des utilisateurs

Il est indispensable d'identifier clairement chaque utilisateur. Si le système ne peut différencier les requêtes qui lui parviennent, toute démarche de personnalisation est impossible. Nous décrivons plusieurs techniques permettant de différencier les utilisateurs. Il est bien entendu possible de les combiner afin de profiter de leurs différents avantages tout en réduisant leurs inconvénients respectifs.

4.7.1 Adresse IP

Sur Internet, les ordinateurs communiquent entre eux grâce au protocole TCP/IP⁹. Chaque ordinateur appartenant au réseau est identifié par une séquence unique de 32 bits (l'adresse IP) écrite à l'aide de quatre nombres compris entre 0 et 255.

Les adresses IP ont l'avantage d'être toujours disponibles et de ne nécessiter aucun traitement préalable. En revanche, elles possèdent principalement deux limites¹⁰ : 1) une adresse IP peut n'identifier qu'un groupe d'ordinateurs « cachés » derrière le serveur proxy d'un fournisseur d'accès à Internet ou d'un réseau local. 2) son inadéquation à la réattribution dynamique.

4.7.2 Identifiants de session

Les identifiants de session¹¹ permettent à un site entièrement dynamique d'identifier les utilisateurs individuellement. Cette technique permet d'attacher un identifiant à chacun des liens hypertextes présents sur une page. Lors de la première requête émise, le serveur attribue arbitrairement à cette requête un identifiant de session, la réponse du serveur sera une page préparée dynamiquement. Le serveur peut ainsi insérer l'identifiant de session dans tous les liens hypertextes de cette page. Lorsque l'utilisateur cliquera sur l'un de ces liens, sa requête

⁹ Transmission Control Protocol / Internet Protocol

¹⁰ Au niveau de la fiabilité de l'identification, il existe également un risque de spoofing, c'est-à-dire d'usurpation d'adresse. Cette technique est d'une telle complexité que nous ne l'avons pas jugée pertinente.

¹¹ Rappelons qu'une session se définit comme un intervalle de temps entre l'instant où l'utilisateur débute la consultation d'un site Web et le moment où il y met fin. L'internaute ne doit pas impérativement rester connecté durant la totalité de cet intervalle de temps.

contiendra automatiquement l'identifiant qui lui a été attribué au départ. Cette technique est très fiable mais limite l'identification de l'utilisateur à une seule session.

Par exemple, on peut utiliser l'adresse IP concaténé à la date et l'heure du début de la connexion : SessionID = « 189.156.339.39 29/12/04 18 :11 :19 »

4.7.3 Les Cookies

Les cookies sont de petits fichiers, d'un maximum de 4 Kbytes, constitués par une chaîne de caractères, gardés en mémoire par l'ordinateur jusqu'à la fermeture du navigateur ou stockés sur le disque dur même de l'utilisateur. En anglais, cette caractéristique est dénommée « *Client Side Storage* ». Ils peuvent contenir des informations telles que la date et l'heure de visite, un code d'identification de l'utilisateur, les réponses aux formulaires remplis, etc. Chaque fois que l'utilisateur introduit une URL (*Uniform Resource Locator*), le navigateur parcourt les cookies. Si l'un d'eux contient cette URL, la partie du cookie contenant les données associées est transférée conjointement à la requête.

Un cookie contient les données suivantes :

- **Nom** : chaîne arbitraire désignant le cookie.
- **Valeur** : valeur chaîne de caractères à enregistrer dans le cookie.
- **Expire le** : date et heure d'expiration du cookie.
- **Domaine** : nom de domaine du ou des serveurs pouvant lire le cookie.
- **Chemin** : nom du chemin dans le domaine sur lequel le cookie est valide

4.7.4 Combinaison d'un pseudonyme et d'un mot de passe

Pour identifier l'utilisateur de manière certaine, la seule possibilité est de collaborer avec celui-ci. Si l'utilisateur s'identifie lui-même à l'aide d'un pseudonyme (en anglais : Login) et un mot de passe (Password), le système ne devra se livrer à aucune approximation, il sera sûr de l'identité de son utilisateur. Cette technique permet donc d'identifier les utilisateurs de façon permanente et fiable.

4.8 Implication de l'utilisateur dans la génération des profils

Afin d'éviter le tout automatique, l'utilisateur, s'il le souhaite, a la possibilité de modifier les différents termes d'un profil. Il peut ainsi supprimer les termes qui lui semble non pertinents ou les réorganiser (augmenter ou diminuer l'importance d'un terme) voire même en ajouter.

5. Critères d'évaluation

Cette rubrique permet de définir comment est représenté l'utilisateur dans le modèle proposé. Les critères retenus sont présentés et décrits dans le Tableau 4.8.

Notre approche de détermination du profil d'un utilisateur consiste à assigner à chaque élément du « modèle » un ou des attributs bien déterminés. Ces attributs peuvent être des données que l'utilisateur fournit au système ou une évaluation à partir des attributs fournis.

Nous considérons un *modèle des utilisateurs* permettant la représentation aussi bien des groupes que des individus. L'idée de groupe d'utilisateurs (ou groupe de recherche) n'est pas complètement écartée, il s'est avéré que c'est un point important à prendre en considération. En effet, chaque utilisateur travaillant au sein d'un même groupe possède ses propres centres d'intérêts mais partage également une optique de travail commune au groupe. Les besoins d'un groupe passent nécessairement par une exploitation et un recoupement des besoins de chacun des individus composant ce groupe.

Le modèle utilisateur est à *long terme* car il est nécessaire de garder le profil de l'utilisateur d'une session à l'autre (ce que nous avons appris sur ses goûts et sur ses préférences). Il est en plus *Dynamique*, car il est pertinent d'adapter le modèle utilisateur en fonction de ses actions dans le système.

L'utilisateur est invité à demander des modifications de son profil, au fur et à mesure que ses centres d'intérêt évoluent.

Le Tableau 4.8 donne une vision synthétique de l'évaluation du modèle utilisateur.

Nom du critère	Description	Dimensions adoptées
Granularité	Décrit le niveau auquel sont représentés les utilisateurs, i.e. groupe, individu, les deux.	Individu / groupe
Caractéristiques :	Précise les informations qui permettent de représenter un groupe ou un utilisateur.	Domaine de recherche
Connaissances spécifiques au domaine d'application	Indique si les connaissances de l'utilisateur en la matière peuvent être représentées dans le modèle (i.e. niveau de compétence, d'expérience en informatique, etc.).	Modèle de recouvrement (cf., section 3.6)
Connaissances autres		
Conditions d'Exploitation	Indique si les conditions (matérielles ou logicielles) dans lesquelles l'utilisateur exploite l'application sont représentées.	Vocabulaire CC/PP (cf., section 3.10)

Fonctionnalités	Indique si une représentation des fonctionnalités (objectif, buts) exploitées par l'utilisateur est donnée.	Oui
Préférences :	Indique s'il est possible pour l'utilisateur d'exprimer ses préférences et à quel propos :	Oui
Contenu	Représentation d'un concept, granularité de l'information souhaitée	Oui
Navigation	Composition des nœuds en pages et gestion des liens	Oui
Présentation	Charte graphique personnalisée	Oui
Acquisition	Exprime comment sont collectées les informations à propos des utilisateurs.	Explicite / Implicite (cf. section 4.5)
Droits sur le modèle utilisateur	Exprime les droits de l'utilisateur à l'égard des informations le concernant	Aucun / Lecture / Écriture
Nom du critère	Description	Dimensions adoptées
Durée de vie	A long terme ou à court terme	Long terme
Identification des utilisateurs	Exprime les données d'identification d'un utilisateur.	Combinaison pseudonyme, mot de passe

Tableau 4.8 – Critères d'évaluation relatifs à la modélisation des utilisateurs

Le modèle utilisateur ainsi défini doit répondre aux exigences suivantes:

- Le modèle utilisateur et le mode de qualification des ressources reposent sur une même représentation de la connaissance.
- La construction et la mise à jour du modèle utilisateur utilisent une ontologie du domaine. Ceci d'une part permet à un utilisateur d'exprimer ses besoins en retenant des mots et des expressions qui lui sont présentées, plutôt que d'avoir à les saisir.
- Le modèle utilisateur doit permettre la représentation d'un utilisateur ou groupe d'utilisateurs à travers ses (leurs) préférence(s)
- Il ne doit pas être ennuyeux pour l'utilisateur : il doit poser le minimum de questions aux utilisateurs
- Le modèle utilisateur doit évoluer, pour tenir compte de la multiplicité des utilisateurs.

6. Conclusion

La modélisation des utilisateurs a pour but à la fois de déterminer les caractéristiques à retenir pour construire des profils utilisateur et d'établir les liens entre ces caractéristiques et leurs relations avec les éléments du modèle.

Si ces analyses achèvent l'élaboration du modèle utilisateur proposé, nous avons constaté que le rôle de l'homme ne cesse d'évoluer. En effet, nous ne le considérons plus comme un simple utilisateur manipulant une application, mais comme une personne possédant des connaissances qui lui sont propres et qu'elle utilise pour interagir. Le système se doit donc d'intégrer des mécanismes pour tenter de capturer et de comprendre ces connaissances, ce pour, en définitive, mieux maîtriser l'utilisateur. Ce dernier est alors impliqué dans le système dès la spécification initiale du modèle.

Le modèle utilisateur, ainsi défini, s'attachera tout d'abord à fournir une analyse complète ainsi que des renseignements riches résultant des diverses actions de l'utilisateur. Il servira de base pour la construction d'une ontologie des utilisateurs qui répondra à un besoin de modélisation et de représentation et qui constitue un élément clé pour une structuration sémantique du modèle utilisateur. Il deviendra par suite rapidement un maillon primordial du processus de personnalisation.

D'un autre côté, nous remarquons encore que l'utilisation conjointe d'ontologies liées à un domaine et de l'ontologie liée aux utilisateurs nous paraît très intéressante. Ainsi, une ontologie du domaine est prise en compte par notre modèle utilisateur, à travers l'application partielle de trois des caractéristiques de notre précédente définition, à savoir la profession (cadre de travail), le modèle de connaissances (niveau de connaissance par rapport à un concept du domaine), et le centre d'intérêt (domaine d'action, contexte).

Ce chapitre nous a permis de fixer la terminologie, c'est-à-dire les concepts fondamentaux qui seront à la base de la construction de l'ontologies des utilisateurs. Le but de l'ontologie est de permettre à une communauté d'utilisateurs de partager les connaissances de leur domaine spécifique et de capturer les caractéristiques des utilisateurs dans un langage permettant aux concepts et à leurs relations d'être interprétés par les ordinateurs. Le processus de développement de l'ontologie préconisé fait l'objet du prochain chapitre.

 Une approche fondée sur les ontologies

CHAPITRE 5	103
1. Introduction	104
2. La modélisation de la connaissance : pourquoi opter pour les ontologies ?	105
3. Une approche fondée sur les ontologies	105
3.1 Une ontologie du domaine	106
3.2 Ontologie des utilisateurs	107
3.2.1 Prise en compte des caractéristiques des ressources	107
3.3 L'ontologie des documents	107
3.3.1 Les publications scientifiques	108
3.3.2 Schéma Dublin Core	111
3.3.3 Un schéma de métadonnées	112
3.4 Une ontologie de l'application	113
3.4.1 A qui s'adresse la personnalisation ?	114
3.4.2 Choix du niveau de personnalisation	114
3.4.3 L'évaluation des documents	115
3.5 Liens entre les ontologies	117
4. Cycle de vie de l'ontologie	118
4.1 La spécification	119
4.1.1 L'évaluation des besoins	120
4.2 Acquisition des connaissances	122
4.2.1 Techniques de recueil de données	123
4.2.2 Envisager une éventuelle réutilisation des ontologies existantes (intégration)	124
4.3 La conceptualisation	125
4.3.1 Création de concepts	125
4.3.2 Création d'une taxonomie	126
4.3.3 Représentations intermédiaires	128
4.4 La formalisation de l'ontologie	130
4.4.1 Un formalisme pour représenter les ontologies : les Logiques de Description	131
4.4.2 Le langage de description <i>SHIQ</i>	131
4.4.3 La classification	132
4.5 L'implémentation	134
4.5.1 Le langage ontologique OWL	136
4.5.2 La structure d'une ontologie OWL	137
4.5.3 Visualisation de l'ontologie	145
4.5.4 Les requêtes	146
4.5.5 Le raisonnement	147
5. Conclusion	148

***L**e chapitre 4 nous a permis de définir nos besoins en terme de modélisation utilisateur. L'objectif de ce chapitre est de définir la structure conceptuelle que nous avons choisie pour la représentation et la manipulation de ce modèle dans le cadre applicatif des laboratoires de recherche. Nous précisons l'organisation de ces connaissances et montrons que la notion d'ontologie répond à nos attentes. Nous abordons alors plus formellement les fonctionnalités attendues de notre approche à base d'ontologies. Nous discutons ensuite du langage adéquat pour représenter les connaissances, en se basant sur l'état de l'art proposé dans le chapitre 3.*

1. Introduction

Les ontologies sont de bons candidats pour la représentation des connaissances relatives aux utilisateurs. Outre leur apport en matière de réutilisabilité, de modularité et de partage de connaissances, les ontologies permettent de définir un vocabulaire précis, sur lequel est basée la communication entre les différents utilisateurs. Nous pouvons donc les utiliser pour la représentation des caractéristiques des utilisateurs, de leurs activités, pour la gestion de la personnalisation et des documents scientifiques et techniques, la représentation de leur contenu et de leur manipulation. La représentation du contenu permet de le manipuler pour faire de la recherche par spécialisation, par similitude, par analogie, etc.

L'utilisation de ces outils de traitement de l'information permet le partage effectif des connaissances dans l'organisation (entreprise, équipe de recherche, etc.).

Notre approche prend en compte quatre types d'ontologies fédérées sous une seule et même structure. Une ontologie des utilisateurs, représentant le modèle utilisateur défini au chapitre 4, qui est en relation avec une ontologie des documents, car une normalisation minimale des documents exploités ou produits par les utilisateurs est indispensable, surtout que l'activité principale de la communauté scientifique d'un laboratoire de recherche est la production de documents ; avec une ontologie du domaine sur laquelle est basée la définition des centres d'intérêts des utilisateurs, leurs niveaux de connaissance, leurs domaine de recherche, etc. et enfin avec une ontologie d'application qui comporte des informations relatives à la gestion de la personnalisation, comme la liste des stratégies de personnalisation avec les paramètres nécessaires à leur fonctionnement. Il s'agit donc de représentations séparées, mais néanmoins interconnectées.

Dans ce chapitre, nous choisissons une définition des ontologies (section 2), ensuite, nous présentons les réalisations et les expérimentations menées autour de la modélisation d'une

ontologie dans le cadre applicatif des laboratoires de recherche. Nous commençons par détailler le contenu de cette ontologie compte tenu de la méthodologie de construction que nous avons suivie tout en justifiant nos choix concernant le formalisme et le langage utilisé pour les coder. Nous placerons le plus possible l'accent sur le côté pratique de la mise en œuvre en présentant les critères de choix et leurs implications.

2. La modélisation de la connaissance : pourquoi opter pour les ontologies ?

Dans le chapitre 3 nous avons détaillé les définitions des ontologies, leurs rôles et les différentes méthodes qui existent pour les concevoir et les représenter.

La littérature propose de nombreuses définitions d'une ontologie, nous avons retenue la définition de T. Gruber [Gruber, 1993] : « *une ontologie est une spécification formelle explicite d'une conceptualisation partagée* », c'est peut-être l'une de celles qui caractérisent le mieux l'essence d'une ontologie pour le Web et les raisons de développer des ontologies : avoir une compréhension partagée entre des personnes ou des agents logiciels du sens des termes et de leurs relations, un vocabulaire contrôlé et commun, ce qui implique une définition formelle des concepts.

3. Une approche fondée sur les ontologies

Les ontologies sont une solution pour répondre à un besoin de modélisation et de représentation. Elles constituent également l'élément central du système de personnalisation.

Les ontologies, dans notre approche, traitent de quatre sphères d'activités :

- Ontologie du domaine (Notre contribution ne portant pas sur la définition d'une nouvelle ontologie de domaine, nous avons choisi d'utiliser dans nos expériences une ontologie existante).
- Ontologie d'application (cette ontologie contient des informations relatives à la gestion de la personnalisation).
- Ontologie des utilisateurs. A la différence de nombreux systèmes, notre modèle utilisateur se focalise sur la création automatique de profils utilisateurs basés sur des ontologies. Cette ontologie de l'utilisateur agit en quelque sorte d'interface entre les utilisateurs et les autres ontologies.
- Ontologie des documents, se concentre sur la gestion des ressources documentaires du laboratoire (documents scientifique et techniques).

La séparation de ces différentes catégories de connaissance facilite leurs évolutions et leurs maintenances. Elle permet aussi d'améliorer leur compréhension et de définir des rôles

différents pour chacune des ontologies, c'est-à-dire de les spécialiser. Il s'agit donc de représentations séparées, mais néanmoins interconnectées. En effet, certaines métadonnées vont prendre leurs valeurs dans le modèle de domaine, d'autres dans le modèle utilisateur, et les sections connaissances et centres d'intérêts du modèle utilisateur sont fondées sur l'ontologie du domaine.

Notre approche est ainsi fondée sur des ontologies. Les quatre ontologies citées précédemment sont fédérées sous une seule et même structure (cf. Figure 5.1).

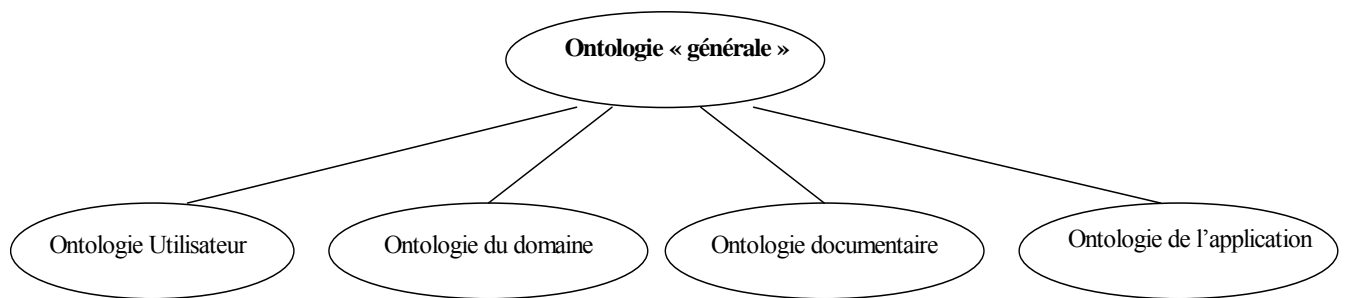


Figure 5.1 – Structure des ontologies

Le découpage en plusieurs ontologies que nous avons choisi est similaire aux découpages proposés dans les projets KnowMore [Liao et al. 1999] et O'CoMMA [Gandon et al. 2002]. Il s'agit de projets concernant les mémoires organisationnelles¹.

Dans chacun des cas, il existe une ontologie du domaine et une ontologie pour la gestion de documents (ou de l'information dans KnowMore). Les différences avec notre approche viennent des autres ontologies. En effet, aucun des deux projets ne possède de modèle utilisateur représenté par une ontologie. En plus, dans les deux projets, l'ontologie de l'entreprise est différente de notre ontologie de l'application car elle représente l'organisation en général de l'entreprise et non une activité particulière. D'ailleurs, dans KnowMore cette ontologie est générique et valable pour différentes entreprises.

3.1 Une ontologie du domaine

L'ontologie du domaine représente la connaissance sur un domaine particulier et offre des possibilités de raisonnements sur ces connaissances. Elle permet de décrire le contenu des

¹ Une mémoire organisationnelle est « une représentation persistante, explicite, désincarnée des connaissances et des informations dans une organisation, afin de faciliter leur accès, leur partage et leur réutilisation par les membres adéquats de l'organisation, dans le cadre de leurs tâches individuelles et collectives » [Dieng et al., 2001].

documents par le biais des métadonnées, ainsi que les connaissances de l'utilisateur au sein du modèle utilisateur. Elle permet ainsi aux utilisateurs d'exprimer leurs centres d'intérêts de façon à recevoir des informations ciblées.

3.2 Ontologie des utilisateurs

Le modèle utilisateur, défini au chapitre précédent (cf. chapitre 4), est présenté par une ontologie.

Par rapport à la sémantique du domaine, l'ontologie utilisateurs doit comporter des informations concernant l'utilisateur : quelle est sa connaissance du domaine (aspect relatif à la sémantique du domaine) et quels sont ses objectifs par rapport à ce domaine (aspect relatif à la tâche de l'utilisateur), en plus de ses centres d'intérêt, préférences, activités, etc.

3.2.1 Prise en compte des caractéristiques des ressources

L'objectif de l'utilisateur est de compléter son état de connaissance par l'acquisition d'information contenue par des documents pertinents. En plus, l'activité principale de la communauté scientifique d'un laboratoire de recherche est la production de documents scientifiques. Or, l'utilisateur et le document ne se limitent pas respectivement à une requête et à un contenu. Ils possèdent tous deux des caractéristiques propres : le domaine disciplinaire, la taille, le type de publication... pour le document; la profession, le niveau de connaissance, les domaines de compétences, l'expérience pour l'utilisateur. Les caractéristiques du document sont couramment appelées «métadonnées» tandis que nous appellerons «profil» l'ensemble des caractéristiques de l'utilisateur.

Nous avons donc jugé indispensable que les éléments de métadonnées rentrant dans la description de ressources comportent des éléments pertinents de profil d'utilisateur.

3.3 L'ontologie des documents

L'ontologie des documents se concentre sur la gestion des sources d'information parmi des groupes de personnes. Elle est conçue selon les concepts de métadonnées sur des documents, de l'ontologie Dublin Core², qui décrivent le document indépendamment du domaine.

² Dublin Core est un ensemble des éléments de méta-données prévu pour faciliter la découverte des ressources électroniques en décrivant le titre, l'auteur, le sujet, la description, l'éditeur, le collaborateur, la date, le type, le format, l'identificateur, la source, la langage, la couverture, les droits. <http://dublincore.org/>

Dans un premier temps, nous essayerons de présenter une description de la communication et des publications scientifiques et lister les documents traitant les notions à appréhender. Nous présenterons ensuite les éléments de métadonnées du Dublin Core qui serviront de base pour la définition de notre schéma de métadonnées.

3.3.1 Les publications scientifiques

L'information scientifique et technique est le résultat et le témoin de l'activité de la communauté scientifique. Parce que le développement de la science est une œuvre collective, la circulation et l'exploitation de cette information sont essentielles.

Notre but est de représenter la documentation d'un laboratoire de recherche, car dans ce domaine plus qu'ailleurs, la structuration des documents influe sur la quantité d'information transmise aux utilisateurs demandeur d'information.

« Communication scientifique », « information scientifique et technique », « publications scientifiques » et bien d'autres, sont toutes des expressions qu'on rencontre dans plusieurs documents et dont on parle souvent surtout dans le milieu universitaire. Les publications scientifiques sont celles issues des travaux de la recherche scientifique dans n'importe quelle discipline, sous n'importe quelle forme (revue, rapport, ouvrage,...) et sur n'importe quel support (papier, électronique,...).

Benichoux et ses collègues présentent la communication scientifique comme suit : « *la véritable communication scientifique de fin de recherche se présente sous plusieurs formes : la communication orale dans un congrès, un séminaire ou une réunion interne, l'article scientifique ou la revue générale et enfin le mémoire, la thèse, le livre etc.* » [Benichoux et al., 1985].

En se référant à cette définition, on peut déduire que les deux formes que peut prendre une communication scientifique sont la communication orale avec ses formes de réunions, séminaires,...; et la communication écrite ou encore les écrits scientifiques avec les revues, les ouvrages, les rapports de recherche,... qui est la catégorie la plus officielle dans l'évaluation et la validation des travaux scientifiques.

3.3.1.1 La communication scientifique orale

Cette communication s'établit généralement entre les scientifiques ayant des intérêts communs ou appartenant à la même discipline. Elle peut se présenter sous plusieurs formes :

- **Les conférences.** Ce type de communication peut prendre plusieurs formes à savoir : le congrès national ou international, le séminaire, le symposium, le colloque, etc.

- **Les réunions.** Ce sont les réunions qui s'établissent entre scientifiques ou chercheurs du même laboratoire ou du même groupe de recherche ou encore entre des groupes de recherche ayant des intérêts communs.
- **Les autres types.** D'autres types de communication orale peuvent se présenter telles que les discussions entre chercheurs, les communications téléphoniques, etc.

Il est intéressant de signaler que la plupart de ces formes de communication se transforment souvent en écrits scientifiques comme par exemple les rapports de conférences qui sont publiés sous forme d'actes de congrès (Proceeding) dans des livres ou journaux scientifiques, etc.

3.3.1.2 La communication scientifique écrite

La communication scientifique est basée sur l'écrit. En effet, même si on trouve des communications orales, l'écrit s'impose toujours dans la communication scientifique puisqu'il sert de preuve et c'est par son intermédiaire que le travail de recherche original est approuvé.

Ce système de validation des travaux scientifiques se traduit par la publication d'articles dans des revues scientifiques dotées d'un comité scientifique composé de chercheurs ayant leur place au sein d'une communauté donnée- pour l'évaluation des articles soumis.

Toutefois, l'écrit scientifique ne se limite pas à ces journaux scientifiques, il peut prendre plusieurs formes selon l'objectif. Ainsi on trouve l'écrit scientifique primaire (qui publie les résultats originaux de la recherche et qui est écrit par les chercheurs pour les chercheurs); l'écrit didactique (orienté vers les étudiants ou élèves) et enfin l'écrit scientifique de vulgarisation (orienté vers le grand public).

A. Les journaux scientifiques (primaires)

Appelés aussi les revues scientifiques ou encore les revues primaires. Une revue scientifique est une publication en série, à parution régulière, dotée d'un titre déposé et composée d'une suite d'articles évalués par un comité de lecture en fonction de critères scientifiques.

Dans ces revues, on peut trouver plusieurs types d'articles qui diffèrent du point de vue de leurs contenus, leurs structures et leurs rôles. Ainsi, on trouve l'article primaire appelé encore dans la littérature anglo-saxonne « scientific paper », l'article de synthèse, la revue générale ou encore « review paper », la lettre à l'éditeur, la controverse (réaction directe ou indirecte d'un spécialiste), etc.

B. La littérature grise

Cette littérature peut renfermer plusieurs types de documents. Les rapports de conférences, les brevets, les thèses... sont des documents non publiés, appelés aussi «la littérature souterraine » ou «littérature grise ».

Concernant les thèses de doctorat, elles prennent une place importante dans la littérature grise scientifique. En effet, même si les résultats de ces recherches sont publiés dans des articles primaires avant même la soutenance de la thèse, elles restent des documents scientifiques de référence puisque on y trouve les détails des recherches.

C. La littérature interne

Elle renferme les rapports préliminaires de fin de recherche, les correspondances entre les chercheurs, les rapports d'activités des laboratoires de recherche...

On trouve aussi dans cette littérature les propositions de recherche dans le but de trouver un financement pour les recherches accomplies par un laboratoire. Il ne faut pas sous-estimer cette littérature car elle prend énormément de temps pour un chercheur.

D. Les ouvrages scientifiques

Ce sont des publications scientifiques qui se différencient des autres par le volume, le contenu et la structure. Comme exemples les livres destinés aux étudiants dès le 1er cycle universitaire ou aux chercheurs.

E. Autres publications scientifiques

On peut trouver d'autres écrits scientifiques qui ne peuvent pas être inclus dans les types précédents. On pense notamment aux normes, aux bibliographies, aux écrits orientés aux techniciens et professionnels. Que ce soit ouvrages ou revues, ces écrits peuvent contenir des informations scientifiques intéressantes.

- Les documents scientifiques (primaires) décrivant les résultats originaux d'une recherche répondent généralement à une structure plus ou moins normalisée ou au moins connue par les membres de la communauté scientifique de chaque spécialité.
- Les travaux universitaires qui sont destinés à un public bien défini ont une structure généralement connue par les universitaires (étudiants, enseignants, chercheurs...)

- Les documents de vulgarisation destinés à un public large ou averti, qu'ils soient des ouvrages ou des articles, ont une structure qui répond à cet objectif de vulgarisation.

3.3.2 Schéma Dublin Core

La norme de métadonnées du Dublin Core est un ensemble d'éléments simples mais efficaces pour décrire une grande variété de ressources en réseau et faciliter l'interopérabilité des applications. La norme du Dublin Core comprend 15 éléments dont la sémantique a été établie par un consensus international de professionnels provenant de diverses disciplines telles que la bibliothéconomie, l'informatique, le balisage de textes et d'autres domaines connexes.

Les 15 éléments formant le Dublin Core sont demeurés inchangés depuis 1996. Leur utilisation n'est pas sujette à un ordre établi et chaque élément est optionnel, tout en pouvant être répété. Les noms des éléments, les identificateurs et les définitions de la version 1.1 de l'ensemble d'éléments sont listés dans le tableau 5.1 (cellules grisées).

Les éléments Dublin Core répondent à cinq objectifs:

1. simplicité et facilité de maintenance
2. sémantique facilement compréhensible
3. consensus international
4. extensibilité
5. modularité

Comme nous venons de le voir, les métadonnées Dublin Core ont pour objectif d'être simples et universelles dans le but de s'adapter au maximum de documents possibles. Néanmoins cette universalité de la sémantique est une barrière lorsque l'on souhaite déterminer plus précisément une propriété. Pour pallier ceci, Dublin Core met à notre disposition deux solutions:

1. la possibilité de définir des sous-catégories dans les 15 éléments. (ex Creator.age)
2. L'utilisation de « qualifieurs » ou « attributs » dont le rôle est de compléter la signification de chaque élément Dublin Core. Par exemple, appliqué à l'élément Date, l'attribut type sert à préciser si la date doit être interprétée comme une date de création ou de modification.

3.3.3 Un schéma de métadonnées

Pour construire notre schéma de métadonnées, nous avons pris comme éléments de description de ressources les éléments du Dublin Core dont nous avons complété la définition par l'ajout de sous-catégories (tableau 5.1).

	Nom	Description
1	Titre	Le titre d'une publication scientifique sert d'enseigne, c'est pourquoi il doit être soigneusement choisi.
2	Auteur	Auteur de la ressource
2.1	Champ disciplinaire	Champ disciplinaire de l'auteur ou affiliation institutionnelle dans lequel est menée la recherche objet de l'article. Valeur prise dans l'ontologie du domaine. Par contre le grade universitaire du/ou des auteurs n'apparaît pas sauf lorsqu'une notice biographique leur est consacrée.
3	Mots Clés	Ils constituent une spécificité des articles scientifiques. Ces mots clés sont généralement choisis par l'auteur de l'article.
4	Description	Brève description du thème général de l'étude et pas de l'article.
4.1	Résumé	Brève description de la ressource. Généralement placé au début de l'article, il constitue avec la conclusion la partie la plus lue des articles scientifiques.
4.2	Domaine	On peut décrire le contenu de la ressource en utilisant la terminologie présente dans l'ontologie de domaine. Ces informations permettent d'une part de sélectionner les ressources par rapport à leurs contenus et d'autre part d'effectuer un filtrage en fonction du modèle de connaissance de l'utilisateur.
4.2.1	Concept	Nom de concept. Valeur prise dans l'ontologie de domaine.
4.2.2	Niveau	Niveau de connaissance requis. A chaque concept du domaine, il est possible de définir un niveau de connaissance minimal requis pour comprendre la ressource. Valeur prise dans le modèle utilisateur sous le concept « NiveauConnaissance », il s'agit des niveaux Expert, Confirmé, Novice, Inconnu.
5	Editeur	L'entité responsable de la diffusion de la ressource tels, un rédacteur en chef, un département universitaire, une entreprise, etc. Typiquement, le nom d'une maison d'édition devrait être utilisé ici.
5.1	Nom	Nom de l'éditeur
6	Collaborateur	Liste d'instances d'auteurs. Généralement on trouve le nom du (ou des) auteur(s), car dans la majorité des cas il s'agit d'articles co-écrits (jusqu'à 5 auteurs).
7	Date	Il s'agit de la date (année) de publication de l'article et non celle de la soumission de ce dernier non plus de son acceptation.
8	Type Publication	Type de la publication scientifique : Organisationnel, recherche, Pédagogique, Promotionnel, etc.
9	Format	Format de la ressource. Ces caractéristiques sont utiles pour sélectionner un document en fonction des capacités du support de lecture.
9.1	Support	Electronique, papier, etc.
9.1.1	Papier	Documents imprimés
9.1.2	Electronique	Documents électroniques (ou numériques)

	Nom	Description
9.2	Type	Type de la ressource (HTML, Word, PowerPoint, etc.)
9.2.1	DOC	Documents Word
9.2.2	HTML	Documents HTML (HyperText Markup Language)
9.2.3	PPT	Document Power Point
9.2.4	PS	Document Post Script
9.2.5	XLS	Feuille de style
9.2.6	XML	Document XML
9.2.7	TXT	Document Texte

9.3	Taille	Taille de la ressource en Kilo-octets
9.4	Besoins	Ce qui est nécessaire à la lecture de la ressource. Instances de «Ressources » de l'ontologie d'application.
9.4.1	Logiciels	Noms de logiciels et versions (Word, Navigateur Internet, etc.) Liste d'instances de « ResourceLogicielle » de l'ontologie d'application.
9.4.2	Matériels	Besoins matériels (carte son) nécessaires à sa consultation. Liste d'instances de « ResourceMaterielle » de l'ontologie d'application.
10	Identifiant	Un numéro dans une série, numéro d'un rapport ou ISBN d'un livre.
10.1	Localisation	Lieu de stockage de la ressource. Adresse URL.
10.2	Série	Le nom d'une série d'ouvrages, d'une collection de bouquins, ...
10.3	Edition	Indique le numéro de l'édition, ou plutôt son ordinal
10.4	Volume	Le numéro de volume dans une série ou dans une collection
10.5	Version	Numéro de version de la ressource
11	Source	Source dont le document a été dérivé. Cette propriété doit contenir le titre de la revue de laquelle l'article est extrait ou le titre de la collection de l'ouvrage objet du document (les références du périodique dans lequel est publié l'article).
12	Langue	Langue de la ressource. Valeur prise dans l'ontologie de l'application sous le concept « Langues »
13	Relation	Rapports avec d'autres ressources : PartieDe, VersionDe,...
14	Couverture	couverture géographique ou temporelle de la ressource
15	Droits	Il s'agit d'informations concernant les restrictions d'utilisation de la ressource ainsi que les droits liés à cette ressource.
15.1	Accès	Restrictions d'accès
15.2	Copyrights	Copyrights ou licences
15.3	Coût	Coût d'utilisation

Tableau 5.1 – Schéma de métadonnées

3.4 Une ontologie de l'application

L'ontologie de l'application se compose des informations relatives à la gestion de la personnalisation, comme la liste des stratégies de personnalisation avec les paramètres nécessaires à leur fonctionnement, elles doivent être différentes selon les utilisateurs et peuvent s'opérer de différentes manières. L'utilisateur peut spécifier une stratégie, et l'ordonne selon ses préférences, de celle qu'il préfère le plus à celle qu'il préfère le moins. Le

système appliquera celle que l'utilisateur préfère le plus. Elle contient aussi les différentes classes disponibles pour l'évaluation des documents.

3.4.1 A qui s'adresse la personnalisation ?

On peut opter pour une solution « *one to one* », c'est-à-dire que la personnalisation se fait en fonction des caractéristiques de chaque individu. On peut également opter pour une solution « *one to many* » afin de réduire l'ampleur de l'effort. Dans ce dernier cas, la personnalisation s'adressera à des groupes d'utilisateurs. Les utilisateurs étant répartis selon plusieurs champs disciplinaires (ou domaine d'activité), nous avons choisi de les regrouper selon leur appartenance à une équipe de recherche (cf. chapitre 4, section 3.8).

3.4.2 Choix du niveau de personnalisation

Quel(s) type(s) de personnalisation on souhaite offrir aux utilisateurs ?

Nous avons abordé au deuxième chapitre les différents axes qui peuvent faire l'objet de personnalisations (cf. chapitre 2, section 2).

Le premier volet de cette partie consiste à choisir quels types de personnalisation seront offerts aux utilisateurs : celle du contenu informationnel, de la navigation ou de la présentation. Ces trois types de personnalisation sont indépendants, ils peuvent (ou non) être présents simultanément.

Ce choix est influencé par le degré d'implication des utilisateurs. Pour chaque type sélectionné, on choisira également quelle technique de personnalisation employer et quels éléments du profil utiliser.

Le tableau 5.3 illustre les différentes stratégies de personnalisation prises en compte.

3.4.2.1 Personnalisation de la présentation

Le système dispose de plusieurs présentations d'une information, chacune correspondant à un média différent. En fonction des préférences de l'utilisateur et du contexte de consultation (support, logiciels disponibles), le système propose l'information sur l'un ou l'autre des médias (cf. chapitre 2, section 2.3).

La personnalisation de la présentation exploite des informations relatives aux capacités du dispositif d'accès (taille d'écran, capacité de la mémoire, vitesse du réseau, etc.) et aux préférences de l'utilisateur (mise en page souhaitée, langage, etc.). Elle est basée sur la définition d'un *Profil* décrit en utilisant le vocabulaire CC/PP (cf. chapitre 4, section 3.10).

Outre les composants proposés par CC/PP pour la description des caractéristiques matérielles, logicielles et du navigateur, on peut modifier la présentation en fonction de certaines caractéristiques de l'utilisateur (e.g. handicaps : utilisateur malvoyant, etc.).

3.4.2.2 Personnalisation du contenu informationnel

C'est ce type de personnalisation qui connaît le plus de succès à l'heure actuelle, nombre de systèmes l'utilisent et les formes qu'elle peut adopter sont très variées (le filtrage, les recommandations et le re-ordonnancement des résultats de la requête sont les trois exemples que nous avons retenus au chapitre 2.

Rappelons que, dans le cadre de la personnalisation du contenu informationnel, l'objectif principal est d'améliorer l'information présentée aux utilisateurs et d'adapter le contenu d'une page ou d'un document accédé par un utilisateur particulier, à sa connaissance, ses objectifs ou encore à d'autres caractéristiques qui lui sont propres (cf. chapitre 2, section 2.1).

En plus, la personnalisation du contenu découle des stratégies de navigation adaptative et consiste donc à filtrer les fragments en fonction de leur pertinence pour l'utilisateur. C'est ce que nous présenterons dans la section suivante.

3.4.2.3 Personnalisation de la navigation

Nous prenons en compte les quatre stratégies de navigation adaptative les plus couramment utilisées, c'est-à-dire le tri, l'annotation, le masquage et le guidage direct (cf. chapitre 2, section 2.2).

Pour chacune des stratégies de personnalisation, il est nécessaire d'évaluer la pertinence de chaque lien ou du fragment pointé par ce lien afin de le classer et/ou de le filtrer (cf. section 3.4.3).

3.4.3 L'évaluation des documents

Nous supposons que les utilisateurs donnent leur opinion, que nous appelons *jugement*, sur les références documentaires. Les utilisateurs ne devraient pas rechigner à fournir un jugement des références reçues car, ces jugements ne sont pas destinés à « une machine » mais à des personnes qui les utilisent immédiatement. De plus, ils savent que ces jugements sont importants pour améliorer la qualité des résultats finaux et pour collaborer avec les autres (filtrage collaboratif), ce qui nécessite de disposer de critères d'évaluation généraux.

Nous avons choisi d'évaluer les documents en fonction des centres d'intérêts de l'utilisateur et en fonction de ses connaissances. La pertinence d'un document est évaluée dans l'ordre suivant :

1. Selon les centres d'intérêts : Le document est pertinent si au moins un des concepts des métadonnées figurent dans la liste des centres d'intérêts de l'utilisateur.
2. Selon le niveau de connaissance : Le document est pertinent si l'utilisateur possède le niveau minimum requis pour chacun des concepts décrivant le document.

Nous proposons l'utilisation de 5 classes pour lesquelles l'utilisateur définit la pertinence des documents. Elles sont représentées dans l'ontologie de l'application et sont nommées « Très mauvais », « Mauvais », « Intéressant », « Bon » et « Très bon ».

Le tableau 5.2 représente les différents concepts de l'ontologie de l'application

	Nom	Description
1	Evaluation	Evaluation de la pertinence d'un document
1.1	Très bon	Le document est jugé très bon
1.2	Bon	Le document est jugé bon
1.3	Intéressant	Le document est jugé intéressant
1.4	Mauvais	Le document est jugé mauvais
1.5	Très mauvais	Le document est jugé très mauvais
1.6	Centre d'intérêt	Valeurs prises dans l'ontologie des utilisateurs sous le concept «CentresInterets »
1.7	Niveau de connaissance	Valeur prise dans l'ontologie des utilisateurs sous le concept « NiveauConnaissance »
2	Ressources	Description du dispositif de l'utilisateur ou ressources nécessaires à la lecture d'un document.
2.1	RessourceMaterielle	Les ressources matérielles
2.1.1	Mémoire	La taille de la mémoire en Kilo-Octet
2.1.2	Carte Son	La carte son
2.1.3	Carte Vidéo 3D	La carte vidéo trois dimensions

2.2	RessourceLogicielle	Les ressources logicielles
2.2.1	Système d'exploitation	Le système d'exploitation (Windows, Unix, etc.)
2.2.1	Navigateur	Le navigateur

2.3	Langage	Le langage utilisé
2.3.1	Français	
2.3.2	Anglais	

3	Methodes Personnalisation	Les méthodes de personnalisations
3.1	Presentation	Personnalisation de la présentation. Listes d'instance du concept Ressources.

	Nom	Description
3.2	Contenu	Personnalisation du contenu informationnel
3.2.1	Texte conditionnel	Toute l'information disponible sur un concept est divisée en plusieurs fragments. Une condition sur le niveau de connaissance de l'utilisateur est associée à chaque fragment.
3.2.2	Stretchtext	Certains mots clés d'un texte peuvent être remplacés par une description plus longue
3.2.3	Réorganisation	Le système peut réorganiser les fragments présentés à l'utilisateur en fonction de ses préférences ou de ses objectifs
3.2.4	FragmentsVariants	Présenter la même information de différentes façons
3.3	Navigation	Personnalisation de la navigation
3.3.1	Guidage Direct	Sélectionner parmi les meilleurs liens celui qu'il faut proposer au lecteur. Les autres sont ignorés
3.3.2	Tri des liens	Le système trie les différents liens du meilleur au plus mauvais
3.3.3	Masquage des liens	Les liens estimés les plus pertinents sont retenus pour l'utilisateur
3.3.3.1	Masquage	Masquage du lien
3.3.3.2	Désactivation	Désactivation du lien
3.3.3.3	Suppression	Suppression du lien
3.3.4	Annotation	Chaque lien est présenté différemment en fonction du niveau de pertinence qui lui est associé.
4	Execution	Modalité d'exécution
4.1	Pull	Les nouvelles informations sont diffusées automatiquement vers les usagers
4.2	Push	Les informations sont recherchées et fournies aux usagers en fonction de leurs requêtes.
5	Portée	La portée de la personnalisation
5.1	Individu	La personnalisation se fait en fonction des caractéristiques de chaque individu.
5.2	Groupe	Personnalisation s'adresse à des groupes d'utilisateurs

Tableau 5.2 – Ontologie d'application

3.5 Liens entre les ontologies

Les ontologies citées ci-dessus ont des liens entre elles, ainsi un élément décrit dans l'ontologie du domaine est cité dans une instance de l'ontologie des documents. Ou bien une tâche nécessite des connaissances décrites dans l'ontologie des utilisateurs.

Le lien s'établit par des propriétés dont l'objet est défini dans une autre ontologie du groupe. Ainsi la propriété « NiveauConnaissance » dans l'ontologie des utilisateurs indique que l'utilisateur a un niveau de connaissance relatif à un concept (instance) décrit dans l'ontologie du domaine.

4. Cycle de vie de l'ontologie

Le processus de développement ontologique selon *Methontology* (cf. chapitre 3, section 7.2) identifie les tâches à accomplir (planification, supervision, assurance qualité, spécification, acquisition de connaissances, conceptualisation, intégration, formalisation, implémentation, évaluation, maintenance, documentation et gestion de configuration).

Methontology commence par identifier les activités suivantes impliquées dans le développement d'une ontologie.

- *Spécifications* : identifier le but de l'ontologie, les utilisateurs, les scénarios d'usage, le degré de formalisme requis, et la portée de l'ontologie comprenant l'ensemble de termes à représenter, leurs caractéristiques et leur granularité. Le résultat de cette phase est un document en langage naturel de spécifications de l'ontologie.
- *Acquisition de connaissance* : ceci se produit en grande partie parallèlement à l'étape de spécification.
- *Conceptualisation* : les termes du domaine sont identifiés comme concepts, instances, relations ou propriétés et représentés en utilisant une représentation intermédiaire (semi-formelle) par le biais de tableaux et de graphes.
- *Intégration* : afin d'obtenir une certaine uniformité à travers des ontologies, des définitions d'autres ontologies devraient être incorporées.
- *Formalisation* : l'ontologie est formellement représentée en logique de descriptions.
- *Implémentation* : l'ontologie est implémentée dans un langage, dans notre cas, OWL.
- *Evaluation* : Les techniques utilisées sont en grande partie basées sur celles utilisées dans la validation et la vérification de SBCs (Systèmes à base de connaissances). Dans [Gómez-Pérez, 1997] un ensemble de directives est donné sur la façon de constater l'incomplétude, les contradictions et les redondances.
- *Documentation* : classement des documents qui résultent d'autres activités.

Le cycle de vie de l'ontologie est illustré dans la figure 5.2. Il est à noter que ce processus n'est pas linéaire et que de nombreux allers-retours sont a priori nécessaires pour bâtir une ontologie opérationnelle adaptée aux besoins.

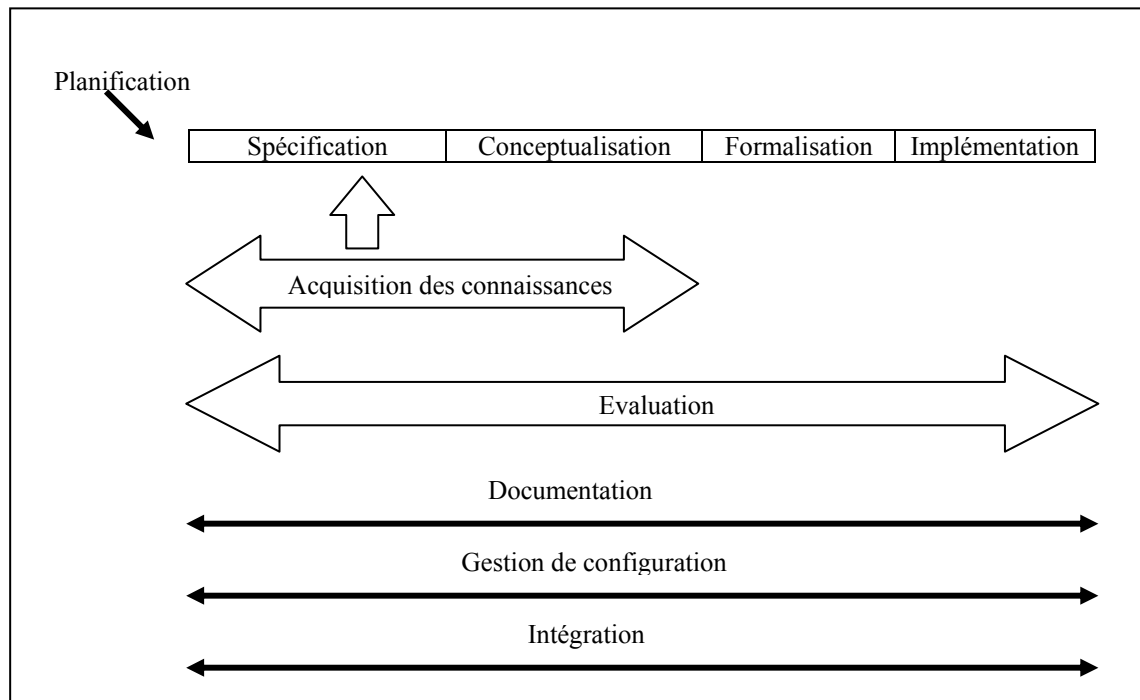


Figure 5.2 – Cycle de vie de l'ontologie

4.1 La spécification

Cette phase, permet d'identifier le but de l'ontologie, les utilisateurs, les scénarios d'usage, le degré de formalisme requis, et la portée de l'ontologie comprenant l'ensemble de termes à représenter, leurs caractéristiques et leur granularité. Le résultat de cette phase est un document de spécifications en langage naturel de l'ontologie (figure 5.3).

<p>Domaine : utilisateurs et ressources techniques d'un laboratoire de recherche</p> <p>Développée par : Ilhem Boussaid</p> <p>L'objectif opérationnel (But) : décrire les utilisateurs du laboratoire de recherche, le contenu des ressources documentaires ainsi que le contexte d'utilisation de ces ressources.</p> <p>Les utilisateurs : Chercheurs, enseignants, étudiants et fonctionnaires d'un laboratoire de recherche.</p> <p>Le degré de formalisme : Formel</p> <p>Capacité : (nombre de concepts)</p> <p>Liste de concepts : données personnelles, objectif, préférence, connaissances, titre, auteur, etc.</p> <p>Granularité (niveau de détail) : Granularité large</p>

Figure 5.3 - Le document de spécification

4.1.1 L'évaluation des besoins

Nous commençons le développement de l'ontologie en répondant à quelques questions de base :

- Dans quel but utiliserons-nous l'ontologie, utilisations prévues, finalités ? Il est important de clairement identifier *pourquoi* l'ontologie est construite? *Ce que* sont ses utilisations prévues ? Celles-ci peuvent être un point de départ pour l'identification du but d'une ontologie. Il sera également utile d'identifier et de caractériser la portée des utilisations prévues de l'ontologie.
- Quel est le domaine que va couvrir l'ontologie ? *Le domaine de connaissance* doit être délimité aussi précisément que possible, et découpé, si besoin est, en termes de connaissances du domaine, connaissances de raisonnement, connaissances de haut niveau (communes à plusieurs domaines) ;
- A quels types de questions l'ontologie devra-t-elle fournir des réponses ? Il est indispensable de bien préciser l'objectif opérationnel de l'ontologie, en particulier à travers des scénarios d'usage ;
- Utilisateurs potentiels de l'ontologie. Qui va utiliser et maintenir l'ontologie ? *Les utilisateurs* doivent être identifiés autant que faire se peut, ce qui permet de choisir le degré de formalisme de l'ontologie et sa granularité.

Les réponses à ces questions peuvent varier au cours du processus de la conception de l'ontologie, mais à chaque moment donné, elles aident à limiter la portée du modèle.

4.1.1.1 Une ontologie pourquoi faire ?

Outre leur apport en matière de réutilisabilité, de modularité et de partage de connaissances, les ontologies permettent de définir un vocabulaire précis, sur lequel est basée la communication entre les différents utilisateurs.

L'ontologie permet :

- La modélisation des connaissances partagées de la communauté scientifique et technique d'un laboratoire de recherche.
- L'identification et description des utilisateurs potentiels et de leurs buts.
- La représentation de la connaissance, des compétences, des rôles, des positions, des préférences, des intérêts des individus, des groupes requis dans le lieu de travail pour exécuter certaines fonctions de l'organisation.

- La gestion des ressources humaines (la communauté scientifique et technique du laboratoire de recherche).
- La gestion documentaire : la définition d'une représentation explicite du contenu sémantique des documents du laboratoire. Ceci afin de pouvoir déterminer le degré de pertinence pour un utilisateur des informations contenues dans les documents.
- Le référencement des ressources à l'aide de spécifications standard, dans le but de les utiliser par la suite à des fins de recherche et d'enseignement.
- L'accès à une description non ambiguë du contenu des documents.
- L'accès rapide et pour tous les utilisateurs aux informations particulièrement intéressantes.
- L'indexation/annotation sémantique de la base documentaire à des fins d'amélioration de la recherche de documents ou d'informations contenues dans les documents

4.1.1.2 Scénarios à étudier

Il peut être intéressant d'exploiter les techniques actuelles visant au Web sémantique pour permettre plusieurs scénarios d'utilisation, parmi lesquels :

- La recherche d'une personne qui remplit un certain nombre de critères ou travaillant sur un thème de recherche donné.
- Bénéficier d'un accès personnalisé au Web sur la base du profil utilisateur établi.
- Un utilisateur enseignant cherche des ressources pédagogiques à consulter ou à utiliser comme base d'un cours qu'il souhaite construire.
- Un utilisateur étudiant cherche un cours sur un thème donné pour un niveau donné.
- Un chercheur cherche un cours sur un thème donné pour s'instruire sur ce sujet et l'exploiter éventuellement dans ses recherches.
- Un groupe de chercheurs, ayant traité différents sous-thèmes d'une problématique de recherche, dans la même approche épistémologique, veut mettre à disposition d'autres chercheurs, les recherches faites et publiées.
- L'étudiant qui doit préparer un exposé.
- L'ingénieur qui souhaite « se tenir au courant » des évolutions techniques dans son domaine.
- L'organisation qui souhaite détecter les avancées de la recherche ou de la technologie dans un secteur très précis.
- Le chercheur qui élabore une synthèse ou un état de l'art.

- L'expert d'un domaine qui cherche des travaux originaux et novateurs,
- Etc.

Après avoir passé en revue les principaux scénarios, nous serons amenés à en privilégier certains. On abordera sûrement de manière différente, la satisfaction du besoin d'un étudiant et celle d'un chercheur.

Il est remarquable de noter que nous pouvons tous nous reconnaître dans différents rôles, selon les circonstances et les multiples facettes de nos vies professionnelles et privées.

Nous souhaitons obtenir des volumes d'informations exploitables (et donc réduits), mais il est indispensable que la réduction se fasse par une meilleure prise en compte du type d'objectif que nous poursuivons.

4.1.1.3 Le niveau de détail de l'ontologie

Nous avons envisagé une granularité fine, correspondant à une ontologie très détaillée, possédant ainsi un vocabulaire plus riche capable d'assurer une description détaillée des concepts pertinents d'un domaine ou d'une tâche. Ce niveau de granularité peut s'avérer utile lorsqu'il s'agit d'établir un consensus entre les agents qui l'utiliseront.

4.1.1.4 Le degré de formalisme

Par rapport au niveau du formalisme de représentation du langage utilisé pour rendre l'ontologie opérationnelle, on a opté pour l'utilisation d'un langage formel contenant une sémantique formelle, ainsi que des théorèmes et des preuves des propriétés telles la robustesse et l'exhaustivité.

A l'issue de l'étape de spécification, la modélisation conceptuelle est engagée à partir des spécifications obtenues jusqu'ici.

4.2 Acquisition des connaissances

L'acquisition de connaissance se fait tout au long du cycle de vie du processus de développement de l'ontologie, la plupart au début, moins vers la fin.

Des sources très diverses sont utilisées dans le but d'établir une liste de termes et de définitions de termes afin de les utiliser ensuite dans le modèle conceptuel et dans la construction de l'ontologie. Parmi elles on peut citer : experts, livres, manuels, figures, tables, d'autres ontologies, les données de navigation, documents, diplômes, ...

4.2.1 Techniques de recueil de données

Plusieurs techniques existent pour le recueil de données, héritées des recherches effectuées en acquisition des connaissances. Ces techniques alimentent le processus de modélisation incluant la construction de l'ontologie. On peut identifier les techniques suivantes :

- Séance de réflexion
- Entretiens avec les futurs utilisateurs du système. Il s'agit le plus souvent de discussions avec des utilisateurs ou avec personnes susceptibles d'être ou de devenir des utilisateurs du système.
- Questionnaires
- L'analyse de texte informelle, pour étudier les principaux concepts et pouvoir ébaucher une première représentation des connaissances.
- Analyse formelle de textes afin d'identifier les différentes structures (définitions, affirmations, ...) et le type de connaissances contenu dans chacune d'entre elles (concepts, attributs, valeurs, relations).
- Outils d'acquisition de connaissances, etc.

Nous avons essentiellement utilisé trois de ces techniques : les entretiens (ou interviews), les questionnaires et l'analyse de documents.

4.2.1.1 L'interview

L'interview est utile au tout début du processus, elle doit être structurée et le plus souvent associée à d'autres techniques d'acquisition de connaissances.

L'interview préalable et l'obtention « en direct » d'informations concernant l'utilisateur courant se doivent alors d'être aussi succinctes et implicites ou automatiques que possible. Il nous faut effectivement éviter de demander trop de perspicacité et d'efforts à l'utilisateur par des requêtes trop fréquentes, longues et fastidieuses. En conséquence, l'interview se doit de rester toujours très discret, en évitant de submerger son utilisateur.

4.2.1.2. Le questionnaire

Le questionnaire est peu coûteux et complémentaire à l'interview. Il nécessite toutefois une formulation non ambiguë des questions.

4.2.1.2 L'analyse de documents

Cette analyse suppose une collecte systématique des documents pertinents pour les scénarios considérés. Cela inclut les documents textuels (rapports, thèses, etc.). Ces documents indiquent les personnes impliquées dans leur production et utilisation et précise le vocabulaire requis pour décrire les différents documents impliqués.

Ces activités doivent être raffinées au fur et à mesure que la conceptualisation émerge. Les interviews très ouvertes doivent alors laisser place à des questionnaires permettant, par exemple, de préciser la sémantique d'un concept mis en évidence lors d'une interview. De même, l'analyse informelle des textes doit être doublée par une analyse formelle qui permet de détecter les termes et structures sémantiques (définitions, règles) présentes dans le corpus.

Lors de cette phase il faut également faire bien attention à repérer les différentes terminologies ou ontologie existantes afin de les intégrer dans le modèle en cours de création.

4.2.2 Envisager une éventuelle réutilisation des ontologies existantes (intégration)

Le recueil de données inclut également la réutilisation, la fusion et l'intégration d'ontologies existantes. Il est toujours utile de prendre en considération ce que d'autres personnes ont fait et d'examiner si nous pouvons élargir des sources existantes et les affiner pour répondre aux besoins de notre tâche particulière.

La réutilisation d'ontologies est à la fois séduisante (elle permet une économie de temps et d'efforts et favorise la normalisation) et difficile (il faut réajuster les engagements et les conceptualisations de l'ontologie réutilisée pour la nouvelle ontologie). Plutôt que de partir de rien, nous avons basé notre travail sur les résultats publiés par la communauté de l'acquisition de connaissances (KA)² (*Knowledge Annotation Initiative of the Knowledge Acquisition Community*) [Blazquez et al., 1998]. Cette ontologie n'a pas été importée directement ni traduites automatiquement. La meilleure façon de la réutiliser était d'analyser sa version informelle comme un document recueilli. Les différences entre les objectifs et les contextes de modélisation et d'utilisation de cette ontologie et ceux de notre ontologie, ainsi que les divergences dans la façon d'envisager une ontologie, nous ont amenés à penser qu'aucune importation automatique n'était possible et que la supervision humaine était obligatoire.

Cependant nous avons réalisé beaucoup de simplifications et d'adaptations de ce modèle au nôtre, en particulier pour la mise en place des activités professionnelles et pour respecter certaines structures déjà en place dans le laboratoire de recherche, etc.

4.3 La conceptualisation

Durant la phase de spécification et d'acquisition de connaissances, termes et définitions ont été recueillis. La phase suivante est de conceptualiser les notions sous-jacentes au vocabulaire utilisé.

Gruber [Gruber, 1993] formule la définition suivante, à savoir « *une ontologie est une spécification explicite d'une conceptualisation* », La construction d'une ontologie n'intervient donc qu'après que le travail de conceptualisation ait été mené à bien. Ce travail consiste à identifier les connaissances à représenter.

Tout comme pour le processus de développement d'un logiciel, plus les problèmes de conception sont rencontrés tard dans le cycle de développement, plus ils seront coûteux en termes d'efforts et de budget. Dans ce contexte, il est impératif d'approfondir la phase d'analyse autant que possible et d'y inclure, dès le début, les spécifications liées à la personnalisation.

4.3.1 Création de concepts

Durant la collecte des données et la définition des objectifs différents termes sont identifiés, parfois plusieurs pour un même concept. L'étude de la terminologie est au cœur de l'ingénierie des connaissances et est en charge de sélectionner des termes candidats et de fournir une définition consensuelle.

Voici quelques règles à suivre pour la création de définitions :

- Identification des termes dénotant les concepts et les relations,
- Définitions précises et non ambiguës pour les concepts et relations,
- Ecrire une définition en langage naturel aussi clair que possible
- S'assurer de la consistance avec les termes déjà existants
- Indiquer les relations avec les autres termes
- Eviter les définitions circulaires
- La définition d'un terme doit être nécessaire et suffisante autant que possible
- Placer les termes dans une catégorie
- Conserver toutes les traces des décisions prises afin de pouvoir procéder à d'éventuelles modifications
- Grouper les termes similaires dans les mêmes catégories

4.3.2 Création d'une taxonomie

La création d'une taxonomie est une étape importante dans la réalisation d'une ontologie. Différentes approches peuvent être envisagées :

- **Approche ascendante (Bottom-up)** : L'ontologie est construite par généralisation en partant des concepts des basses couches taxinomiques (les concepts les plus spécifiques). Cette approche encourage la création d'ontologies spécifiques et adaptées (figure 5.4).
- **Approche descendante (Top-down)** : Partir d'un nombre réduit de concepts que l'on spécialise. L'ontologie est construite par spécialisation en partant de concepts des hautes couches taxinomiques. Cette approche encourage la réutilisation d'ontologies (figure 5.5).
- **Approche centrifuge (Middle-down)** : La priorité est donnée à l'identification de concepts centraux à l'application que l'on va ensuite généraliser et spécialiser pour compléter l'ontologie. Cette approche encourage l'émergence de domaines thématiques dans l'ontologie et favorise la modularité (figure 5.6).

Après expérience, nous ne sommes pas convaincus qu'il existe réellement une approche purement ascendante, descendante ou centrifuge. Pour nous, il s'agit de trois perspectives complémentaires d'une méthodologie complète. Lors de la structuration taxinomique, la spécialisation d'un concept générique, ou la généralisation d'un concept spécifique sont présentes de façon concurrente à différents niveaux de granularité (au niveau des concepts, au niveau de groupes de concepts).

Cependant nous ne pensons pas qu'il faille opposer les différentes approches, mais plutôt considérer qu'elles correspondent à trois perspectives combinées en ingénierie ontologique.

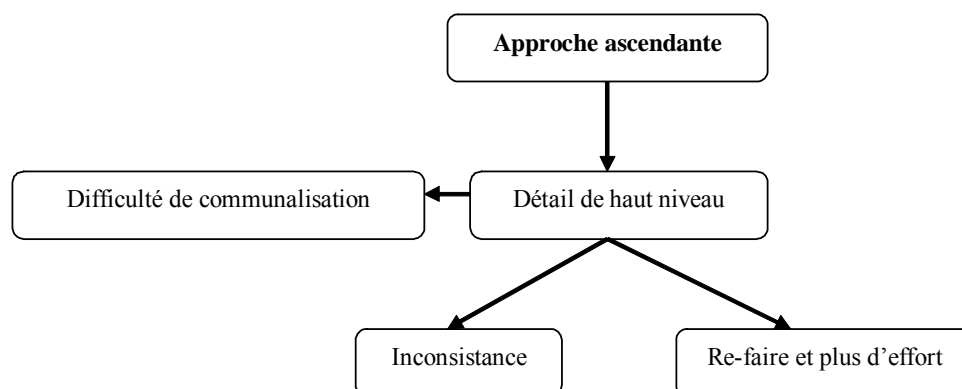


Figure 5.4 - L'approche ascendante

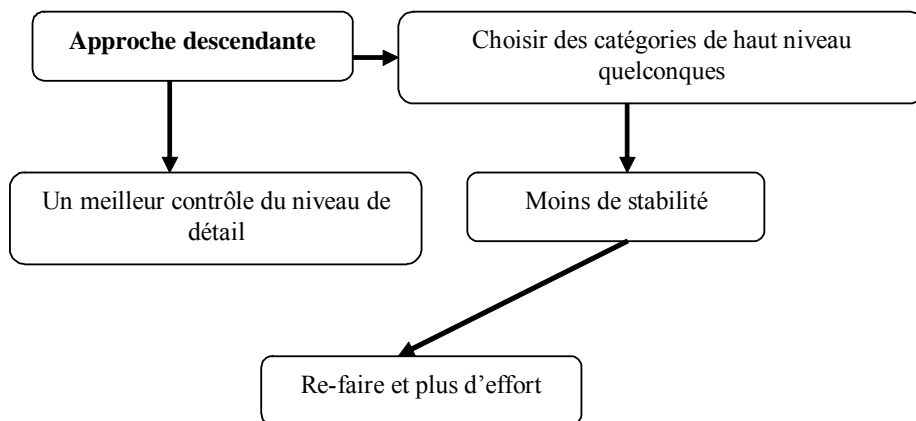


Figure 5.5 - l'approche descendante

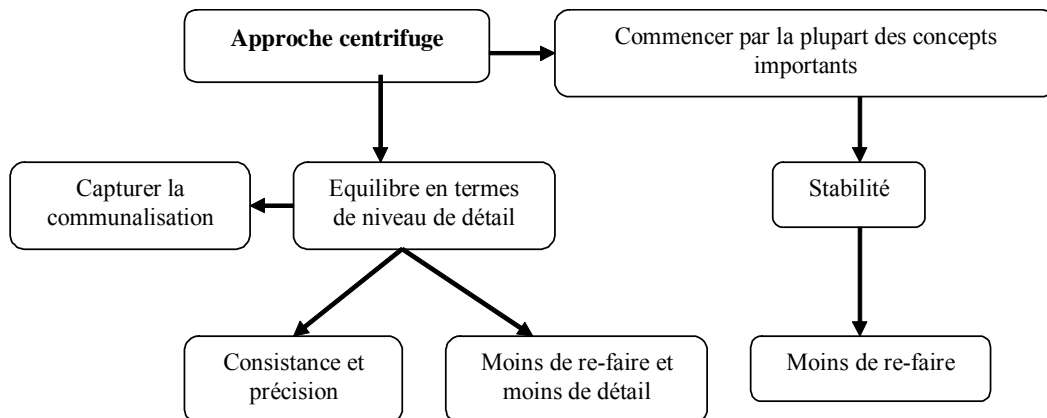


Figure 5.6 - L'approche centrifuge

Pour notre travail, nous avons effectivement effectué certaines tâches en parallèle:

- En perspective descendante, nous avons étudié des ontologies de haut niveau et les couches hautes d'autres ontologies pour structurer la couche supérieure.
- En perspective centrifuge, nous avons étudié différentes branches d'autres ontologies ainsi que les principaux thèmes discernés pendant le recueil, pour circonscrire les principaux domaines nécessaires et regrouper leurs concepts.
- En perspective ascendante, nous avons exploité les rapports de la spécification pour identifier les concepts spécifiques et les regrouper par généralisation.

Chaque approche influence les autres, un événement dans une perspective déclenche des vérifications et des tâches dans les autres.

Une fois les concepts et relations identifiés par leurs termes, il faut en décrire la sémantique en indiquant, *à priori* en langage naturel, leurs instances connues, les liens qu'ils entretiennent entre eux et leurs propriétés, à travers des représentations intermédiaires. Des choix liés aux contextes d'usage de l'ontologie doivent donc être effectués dès cette étape.

4.3.3 Représentations intermédiaires

La conceptualisation se situe au niveau de la connaissance en utilisant un ensemble de représentations intermédiaires basé sur des tableaux ou des graphiques. Methontology préconise un ensemble de représentations intermédiaires : arbres de classification des concepts, diagramme des relations binaires, arbres de classification des attributs, glossaire des termes, dictionnaire des concepts, tables des relations binaires, tables des attributs d'instances, table des instances, etc.

On a adopté une autre présentation intermédiaire. A partir de la terminologie informelle, on a donc commencé à séparer les attributs et les relations des concepts et on a obtenu trois tableaux, dont la structure finale était: le libellé des concepts, relations ou attributs ; les concepts liés par les relations ou le concept et le type élémentaire de l'attribut ; les liens d'héritage du concept ou de la relation ; les termes synonymes du libellé ; une expression de l'intension en langue naturelle. En plus des tables de relations binaires.

4.3.3.1 Table des Concepts

Pour identifier et recueillir tous les concepts du domaine utiles et potentiellement utilisables, leurs significations, attributs, instances, etc. (cf. tableau 5.3)

Nom du Concept (Concept ID)	Parent IDs	Autres termes (synonyme)	Acronyme	Description
Thing	-	Top Level		Le concept le plus général
Personne	Thing	Humain, Person, Human		
Employe	Personne	Employee, salarié, salaried		
EnseignantChercheur	Employe			
ChargeCours	MaitreAssistantTitulaire	Assistant Professor	CC	Maître assistant titulaire avec 3 années d'exercices en qualité d'enseignant
Chercheur	EnseignantChercheur			
AttacheRecherche	Chercheur	Research attaché	AR	
Étudiant	Personne	Student		
Gradue	Etudiant, Chercheur	Thésard, Phd-Student		
Evenement	Thing	Event		Un événement peut prendre plusieurs formes à savoir : la conférence nationale ou internationale, etc.
Workshop	Événement			
Document	Thing			

Publications	Document	Publications scientifiques		Un lien vers les publications de cette personne
Proceedings	Publication			
PublicationEvenement	Publication, Evenement			
Organisation	Thing			
LaboratoireRecherche	Organisation	Research institute, laboratory		Cadre dans lequel travaillent chercheurs, staff administratifs, étudiant post-gradués, etc.
Domaine	Thing	Domain topic		Concept père regroupant les concepts d'un domaine donné
Diplome	Thing			Diplôme d'étude
CentresInterets	Thing	Interest		Centres d'intérêts
MethodePersonnalisation	Thing			Les méthodes de personnalisations
...

Tableau 5.3 –Table des concepts

4.3.3.2 Table des relations

La table des relations (cf. tableau 5.4), contient le nom de la relation, son domaine (ou concept source), son rang (ou concept cible), sa cardinalité et une description en langage naturel.

Relation	Domaine	Rang	Cardinalité	Description
ACentreInteret	Personne	CentreInteret	(0, n)	
ParticipeEvenement	Personne	Evenement	(0, n)	
Affiliation	Personne	Organisation	(0, n)	Le fait d'adhérer à une organisation, rattachement, admission.
DateRecrutement	Employe	Date	(0, 1)	date à partir de laquelle un employé a commencé à exercer son activité
Communication	EnseignantChercheur	Presentation	(0, n)	
Editer	EnseignantChercheur	Document	(0, n)	
Membre	Chercheur	GroupeRecherche	(1, 1)	
DomaineRecherche	Chercheur	Domaine	(1, n)	
Finance	Organisation	Projet	(0, n)	
MembresProjet	Projet	Personne	(1, n)	
FinancePar	Projet	Organisation	(0, n)	Un projet est financé par une ou plusieurs organismes
PresidentJury	These	EnseignantChercheur	(1, 1)	
LieuEvenement	Evenement	Lieu	(1, 1)	Lieu dans lequel s'organise l'événement
DateObtention	Diplome	Date	(1, 1)	
...				...

Tableau 5.4 - Tables des relations

4.3.3.4 Tables des attributs

La table des attributs fournit des informations au sujet des attributs et de leurs type (cf. tableau 5.5)

Attribut	Domaine	Rang (type de données)	Cardinalité	Autres termes	Description
Nom	Personne	String	(1, 1)	First Name, Family Name	Le nom de famille d'une personne
Sexe	Personne	{Femelle, Male}	(1, 1)	Gender, Mal, Femal, , Man	Une personne peut être une femme ou un homme
NumInscription	Etudiant	String	(1, 1)		Numéro d'inscription d'un étudiant
TitreCour	Cours	String	(1, 1)		Un cour porte sur un thème précis (Système d'exploitation, génie logiciel, etc.)
URL	Document	String	(1, 1)		L'adresse URL du document publié
ISBN	Livre	Integer	(1, 1)		L'ISBN d'un livre
NiveauConnaissance	Domaine	{Expert, Confirmé, Novice, Inconnu}	(0, 1)		Niveau de connaissance relatif à un concept du domaine
...					...

Tableau 5.5 - Tables des attributs

4.3.3.5 Tables des relations binaires

(cf. tableaux 5.6, et 5.7).

Nom de la relation	OrganiseEvenement
Concept source	EnseignantChercheur
Cardinalité source	(0, n)
Concept destination	Evenement
Relation inverse	ComiteOrganisation

Tableau 5.6 - La relation binaire OrganiseEvenement

Nom de la relation	ComiteDeProgramme
Concept source	Evenement
Cardinalité source	(1, n)
Concept destination	EnseignantChercheur
Relation inverse	MembreComiteProgramme

Tableau 5.7 - La relation binaire ComiteDeProgramme

L'étape suivante, selon *Methontology*, consiste à formaliser l'ontologie. Les concepts sont définis dans un langage formel.

4.4 La formalisation de l'ontologie

Une fois les ressources cognitives passées au travers du tamis de la conceptualisation, il convient de formaliser le modèle conceptuel obtenu.

Il est à retenir que les ontologies ont à être compréhensible à la fois par les humains et les ordinateurs. Pour obtenir un bon équilibre il est important pour chaque définition de conserver une description informelle. Enfin, comme l'ontologie devra être exploitée par un ordinateur, il est nécessaire qu'elle soit calculable. Et pour cela il est nécessaire de l'implémenter dans un langage formel. Une représentation formelle d'ontologie permet un traitement automatique du

développement. Elle soutient également l'automatisation du processus de vérification de fiabilité de systèmes.

Le travail de formalisation ne consiste pas à substituer une version formelle à une version informelle. Il s'agit d'augmenter une version informelle avec la correspondance formelle des aspects sémantiques intéressants et pertinents de l'ontologie afin d'obtenir une ontologie documentée et opérationnelle (description formelle des aspects sémantiques appropriés nécessaires aux opérations du système envisagé).

4.4.1 Un formalisme pour représenter les ontologies : les Logiques de Description

Les logiques de descriptions (LD) se sont spécialisées dans la description des connaissances ontologiques. On peut noter également que les LDs s'adaptent bien au Web, c'est-à-dire à un monde ouvert, puisque tous les faits non présents ne sont pas forcément faux. C'est pourquoi le formalisme des LD nous a semblé le plus adapté à nos besoins.

Le formalisme de LD effectue une séparation claire entre les connaissances terminologiques (T-box) permettant de décrire les concepts et les rôles et les connaissances assertionnelles (A-box) permettant de décrire les règles de contraintes qui s'appliquent aux concepts ainsi que les instances de concepts. Il est équipé d'une sémantique formelle qui permet d'effectuer des raisonnements sur les bases de connaissances produites.

L'ontologie utilise une logique de description comme langage de représentation des connaissances, il s'agit de SHIQ.

4.4.2 Le langage de description SHIQ

La grammaire de SHIQ est donnée au tableau 5.8. Les expressions construites grâce à une telle grammaire sont aussi appelées expressions conceptuelles.

Le langage SHIQ est représenté par l'ensemble des constructeurs qu'il utilise $SHOIQ = \{C \cap D, C \cup D, \neg C, \forall R.C, \exists R.C, \geq n R.C, \leq n R.C, R^-\}$.

Par exemple, la description suivant décrit le concept MaîtreDeConférence :
 Personne \cap \exists ADiplôme.Doctorat \cap ≥ 3 .AnnéesEnseignement.Integer \cap
 \exists A-Publication.Revue

Constructeur	Syntaxe	Sémantique
CONCEPTS		
TOP	T	Δ^I
BOTTOM	\perp	Φ
Concept primitif	A	$A^I \subseteq \Delta^I$
Négation	$\neg C$	$\Delta^I \setminus C^I$
Conjonction de concepts	$C \cap D$	$C^I \cap D^I$
Disjonction de concepts	$C \cup D$	$C^I \cup D^I$
Restriction universelle	$\forall R.C$	$\{x \in \Delta^I / \forall y. (x, y) \in R^I \Rightarrow y \in C^I\}$
Restriction existentielle	$\exists R.C$	$\{x \in \Delta^I / \exists y. (x, y) \in R^I \wedge y \in C^I\}$
Cardinalité minimum	$\geq n R.C$	$\{x \in \Delta^I / \#(y / (x, y) \in R^I \wedge y \in C^I) \geq n\}$
Cardinalité maximum	$\leq n R.C$	$\{x \in \Delta^I / \#(y / (x, y) \in R^I \wedge y \in C^I) \leq n\}$
ROLES		
Rôle	R	$R^I \subseteq \Delta^I \times \Delta^I$
Rôle inverse	R^-	$\{(x, y) \in \Delta^I \times \Delta^I \mid (y, x) \in R^I\}$

Tableau 5.8 - La grammaire du langage de description de concepts SHIQ.
C, D et A désignent des concepts, R désigne un rôle.

4.4.3 La classification

Les systèmes classificatoires s'appuient sur les développements théoriques réalisés dans le cadre des logiques de descriptions. Les opérations principales qui sont à la base du raisonnement sont :

- le test de subsomption vérifie qu'une classe C est plus générale qu'une classe D,
- la classification de classes qui consiste à placer une nouvelle classe X dans une hiérarchie H, la classification d'instances qui consiste à déterminer les classes dont un objet x donné peut être une instance (en particulier, une classe C n'est satisfiable que si elle peut avoir effectivement des instances),
- la recherche de propriétés qui consiste à retrouver les propriétés détenues par une classe ou une instance.

Le raisonnement par classification s'appréhende comme une procédure de déduction opérant sur une hiérarchie. Sa mise en oeuvre repose sur un cycle comprenant trois étapes :

- 1) La recherche des **subsumants les plus spécifiques SPS** (concepts qui subsument le concept à classer et dont les sous-concepts ne le subsument pas),
- 2) la recherche des **subsumés les plus généraux SPG** (concepts subsumés par le concept à classer et dont les sur-concepts ne sont pas subsumés par lui)
- 3) et puis l'insertion du concept dans la hiérarchie (Figure 5.7).

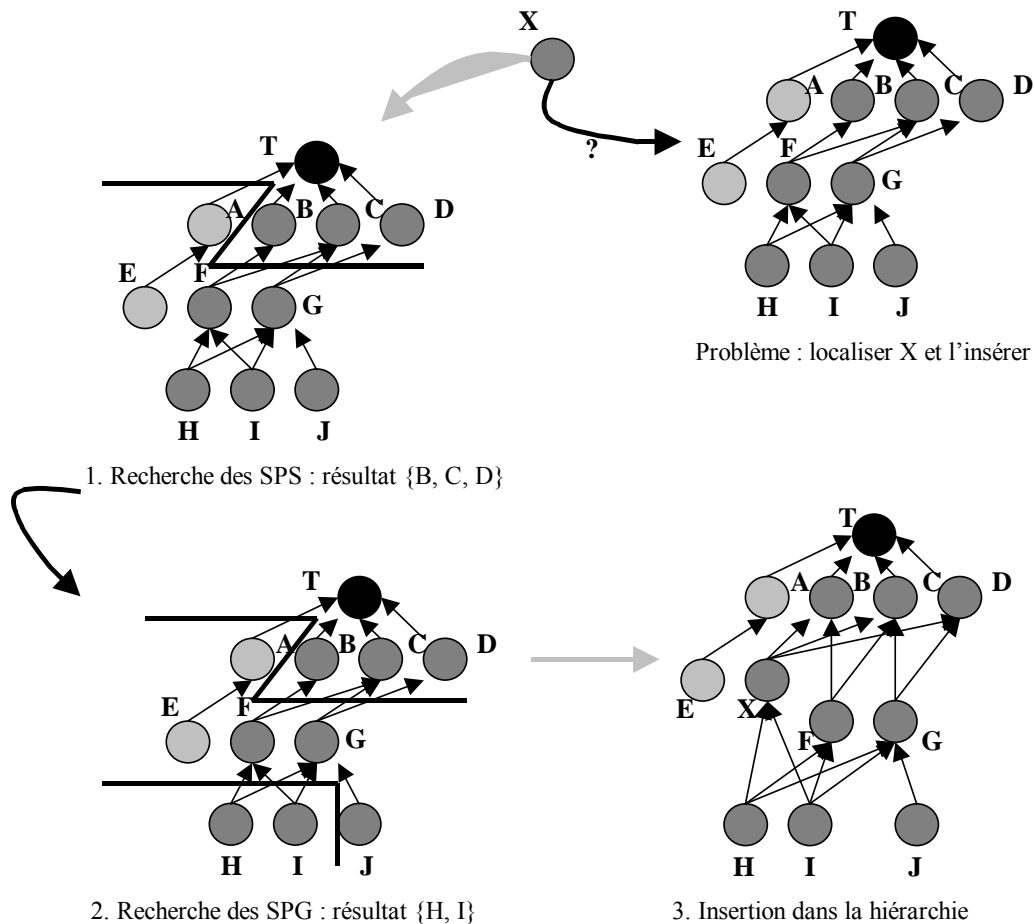


Figure 5.7- La classification d'un nouveau concept comporte trois étapes : la recherche des subsumants les plus spécifiques SPS, la recherche des subsumés les plus généraux SPG, et l'insertion du concept.

4.4.3.1 Appariement entre deux concepts

L'appariement dans la classification de concepts est le calcul qui permet d'établir la relation de subsomption existante entre deux concepts. L'algorithme de subsomption dépend des descripteurs offerts par le langage. Cependant tous les algorithmes comportent des tâches générales comme la vérification des relations de subsomption entre concepts primaires et la vérification de subsomption au niveau des rôles et des descripteurs.

Dans la plupart des systèmes, l'appariement comporte une première étape de **normalisation** des concepts à comparer³ [Nebel, 1990]. La normalisation d'un concept C se fait à deux niveaux, celui des concepts subsumant explicitement C et celui des rôles de C et de leurs descripteurs.

³ Lorsque la subsomption est utilisée à l'intérieur de l'algorithme de classification, comme c'est le cas qui nous occupe, la normalisation du concept à classer est faite une seule fois au début de la classification.

Normaliser les subsumants explicites consiste à « aplatir » la description du concept C en recopiant toute l'information de ses catégories primaires et en ne laissant comme catégories primaires que des concepts primitifs. Par exemple, si la base de connaissances a le concept primitif *Personne* et les concepts définis *Etudiant* et *Etudiant_étranger* décrits par :

$$\text{Personne} = (\exists \text{Prénom.String} \cap \exists \text{Nom.String} \cap \exists \text{Nationalité.String})$$
$$\text{Etudiant} = \text{Personne} \cap (\geq 1.\text{Numero_Inscription.Integer} \cap \leq 1.\text{Numero_Inscription.Integer})$$
$$\text{Etudiant_étranger} = \text{Etudiant} \cap (\exists \text{Nationalité.String} \cap \forall \text{Nationalité.} \neg (\text{Algérienne}))$$

Alors la normalisation du concept *Etudiant_étranger* donne :

$$\begin{aligned} \text{Etudiant_étranger} = & \exists \text{Prénom.String} \cap \exists \text{Nom.String} \cap \exists \text{Nationalité.String} \cap \\ & \geq 1.\text{Numero_Inscription.Integer} \cap \leq 1.\text{Numero_Inscription.Integer} \\ & \cap \forall \text{Nationalité.} \neg (\text{Algérienne}) \end{aligned}$$

La normalisation enlève les descripteurs redondants et génère des versions équivalentes d'une description donnée. Le processus de normalisation produit les formes normales des descriptions, qui sont ensuite effectivement comparées.

Il existe une certaine dualité entre normalisation et comparaison : plus il y a de travail fait lors de la normalisation, moins il y en aura à faire lors de la comparaison.

4.5 L'implémentation

Pour l'implémentation (ou l'opérationnalisation) de notre ontologie, on a exploité l'environnement de développement Protégé 2000⁴, dont les éléments sont structurés en différents items (cf. Chapitre 3, section 10.1). Originellement Protégé 2000, dans sa version 2.1.1, peut interpréter des ontologies écrites en OWL.

⁴ <http://protege.stanford.edu/index.shtml>

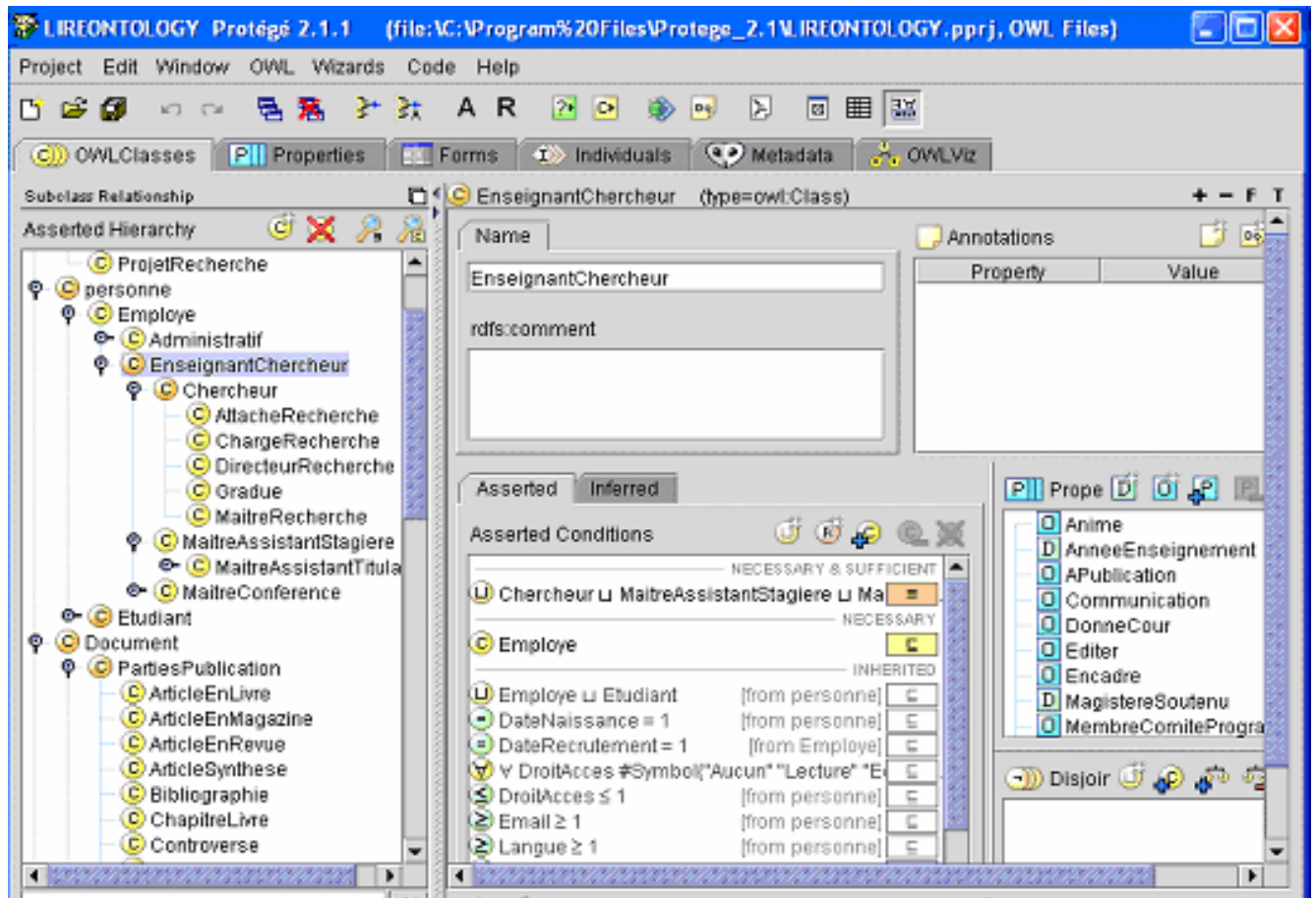


Figure 5.8 - Illustre Protégé 2000 avec une ontologie d'exemple.

Comme illustré dans la figure 5.8, l'interface utilisateur du plugin OWL contient divers onglets associés à des tâches spécifiques : un onglet pour éditer et visionner les classes OWL, un autre pour les propriétés (les « slots »), un autre pour les instances et un onglet pour les métadonnées. Lorsque l'on choisit un élément particulier comme une classe ou une instance, des fenêtres s'ouvrent pour afficher les informations concernant cet élément. En plus protégé2000 offre un éditeur d'expressions arbitraire en OWL (figure 5.9).



Figure 5.9 – Editeur d'expressions OWL

Le tableau 5.9 illustre la relation entre la conceptualisation et l'implémentation.

Implémentation	Conceptualisation
Class	« Nom du concept » mentionné dans la table des concepts
Documentation	« Description » dans la table des concepts
Superclass	Nom des superclasses auxquelles est reliée la classe dans la hiérarchie des classes
Subclass	Nom des sous-classes auxquelles est reliée la classe dans la hiérarchie des classes
Instance	Instances des concepts
Cardinalité Max	Cardinalité maximale exprimée dans le champ « Cardinalité » de la table des attributs, ou du champ « Cardinalité source » de la table de la relations binaire.
Cardinalité Min	Cardinalité minimale exprimée dans le champ « Cardinalité » de la table des attributs, ou du champ « Cardinalité source » de la table de la relations binaire.

Tableau 5.9 – Relation entre la conceptualisation et l'implémentation

4.5.1 Le langage ontologique OWL

Inspiré des logiques de descriptions (et successeur de DAML+OIL), OWL fournit un grand nombre de constructeurs permettant d'exprimer de façon très fine les propriétés des classes définies. On est alors en droit de se demander ce que ce langage apportera de plus aux langages déjà existant à savoir, RDF, RDFS et DAML+OIL.

OWL se base sur XML et RDF(S) tout en empruntant beaucoup de DAML+OIL. Il offre également une version allégée OWL Lite conçue pour être plus facile à exploiter par un plus grand nombre d'utilisateurs.

Il est relativement facile de convertir DAML+OIL en OWL, car ce dernier est une version expurgée des éléments redondants et moins utiles de DAML+OIL. Les principales différences entre DAML+OIL et OWL sont [Petel-Schneider et al., 2002]:

- La suppression de quantificateur de restriction
 - `daml:cardinalityQ`
 - `daml:hasClassQ`
 - `daml:maxCardinalityQ`
 - `daml:minCardinalityQ`
- Élimination de terme redondant de DAML+OIL ayant des synonymes dans RDF(S) comme `daml:comment`, `daml:domain`, `daml:label`, `daml:isDefinedBy`, `daml:Literal`, `daml:Property`, `daml:range`, `daml:seeAlso`, `daml:subClassOf`, `daml:subPropertyOf`, `daml:type`, `daml:value`

- Changement de nom pour certaines propriétés (tableau 5.10):

DAML+OIL	OWL
Daml :differentIndividualFrom	owl :differentFrom
Daml:hasClass	owl:someValuesFrom
Daml:toClass	owl:allValuesFrom
Daml:UnambiguousProperty	owl:InverseFunctionalProperty
Daml:uniqueProperty	owl:FunctionalProperty

Tableau 5.10- Quelques différences entre DAML+OIL et OWL

4.5.2 La structure d'une ontologie OWL

Une ontologie écrite en langage OWL comporte tout d'abord une déclaration des espaces de noms. En effet, avant de pouvoir employer un jeu de termes, nous avons besoin d'une indication précise de l'endroit (URL) où est défini le vocabulaire employé.

Par exemple :

```
<rdf:RDF
  xmlns      = "http://www.LIRE.org/lireontology#"
  xmlns:owl  = "http://www.w3.org/2002/07/owl#"
  xmlns:rdf  = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs = "http://www.w3.org/2000/01/rdf-schema#"
  xmlns:daml = "http://www.daml.org/2001/03/daml+oil#"
  xmlns:dc   = "http://purl.org/dc/elements/1.1/"
  xmlns:xsd  = "http://www.w3.org/2000/10/XMLSchema#"
```

Une ontologie OWL comporte une entête :

```
<owl:Ontology rdf:about="http://www.LIRE.org/lireontology">
```

L'attribut *about* aura normalement toujours pour valeur l'URL du fichier courant indiquant que le sujet de ce document est ce document lui même.

L'entête peut être complété par les informations suivantes :

```
<owl:Ontology rdf:about="http://www.LIRE.org/lireontology">
<rdfs:comment>Laboratoire de recherche LIRE</rdfs:comment>
<owl:versionInfo>
$Id: Overview.html,v 1.0 2005/01/02 16:42:25 connolly Exp $
</owl:versionInfo>
<owl:imports rdf:resource="http://www.w3.org/exemple.owl"/>
</owl:Ontology>
```

- `<rdfs:comment>` permet de commenter l'ontologie.
- `<owl:versionInfo>` permet de contrôler les versions.
- `<owl:imports>` permet d'importer une ontologie et d'inclure ses définitions dans l'ontologie en cours de création.

Les préfixes `rdf:` ou `rdfs:` sont employés lorsque les termes sont déjà présents dans RDF ou RDF Schéma. Sinon les termes sont introduits par OWL. OWL reprends tous les constructeurs de RDF (c'est-à-dire fournit des mécanismes permettant de définir un individu comme instance d'une classe et de mettre des individus en relation), utilise les mots-clés de RDFS (`rdfs:subClassOf`, `rdfs:Property`, `rdfs:subPropertyOf`, `rdfs:range`, `rdfs:domain`), avec la même sémantique.

Nous allons, dans ce qui suit, présenter quelques caractéristiques définies pour ce langage. Nous donnons ci-dessous un panorama des constructeurs utilisés dans OWL, dans une syntaxe simplifiée (les mots-clés réservés de OWL sont préfixés de OWL).

4.5.2.1 Définition de classes et de hiérarchies entre ces classes

Les classes fournissent un mécanisme d'abstraction pour regrouper des ressources ayant des caractéristiques communes. Comme les classes RDF, toute classe OWL (`owl:Class`) est associée à un ensemble d'individus appelés l'extension de la classe.

Exemple :

On définit ici trois classes `Personne`, `Lieu` et `Organisation`.

```
<owl:Class rdf:ID="Personne" />
<owl:Class rdf:ID="Lieu" />
<owl:Class rdf:ID="Organisation" />
```

On crée ensuite une sous-classe `Etudiant` de la classe `Personne`.

```
<owl:Class rdf:ID="Etudiant">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#Personne" />
  </rdfs:subClassOf>
</owl:Class>
```

4.5.2.2 Instanciation

OWL ne parle pas d'objet ou d'instances mais d'individus. Nous allons créer ici une instance (ou individu) de la classe `Personne` :

```
<personne rdf:ID="ilhem">
  <Sexe rdf:datatype=http://www.w3.org/2001/XMLSchema#string
    >Femelle</Sexe>
  <Age
rdf:datatype=http://www.w3.org/2001/XMLSchema#positiveInteger
    >25</Age>
  <Langue>
    <Langues rdf:ID="arabe" />
  </Langue>
```

```
<Nom rdf:datatype=http://www.w3.org/2001/XMLSchema#string
  >boussaid</Nom>
<Email rdf:datatype=http://www.w3.org/2001/XMLSchema#string
  >ilhemboussaid@hotmail.com</Email>
<Nationalite
rdf:datatype=http://www.w3.org/2001/XMLSchema#string
  >algérienne</Nationalite>
  ...
</personne>
```

OWL fournit trois constructeurs concernant l'identité des individus :

- `owl:sameAs` est utilisé pour énoncer que deux références se rapportent à un même individu. Le constructeur `owl:sameIndividualAs` est un synonyme de `owl:sameAs`.
- `owl:differentFrom` est utilisé pour énoncer que deux références se rapportent à des individus différents.
- `owl:AllDifferent` est utilisé pour énoncer qu'une liste d'individus sont tous différents.

4.5.2.3 Propriétés

Une propriété peut être considérée comme un attribut ou bien comme une relation binaire entre deux classes. OWL distingue deux types de propriétés :

- `owl:ObjectProperty`. Les propriétés d'objet ont un « range » de valeur d'individus de classe et lie ainsi les individus aux individus.
- `owl:DatatypeProperty`. Les propriétés de type de donnée ont un range de valeur de données, et lie ainsi les individus aux valeurs de données.

Exemples :

```
<owl:ObjectProperty rdf:ID="PresentePar">
  <rdfs:domain rdf:resource="#Presentation"/>
  <rdfs:range rdf:resource="#EnseignantChercheur"/>
</owl:ObjectProperty>
```

On précise donc ici que la propriété `PresentePar` s'applique à un objet de la classe `Presentation` et prend ses valeurs dans la classe `EnseignantChercheur`.

```
<owl:onProperty>
  <owl:DatatypeProperty rdf:about="#Prenom"/>
</owl:onProperty>
```

Types de données

OWL utilise les mécanismes de RDF pour les valeurs de données. Les valeurs de données sont des instances de la classe RDF-Schéma `rdfs:Literal`. Les littéraux peuvent être typés ou non. Les types de données sont des instances de la classe `rdfs:Datatype`. Les types de données XML Schéma recommandés pour les ontologies de OWL sont les suivants :

- Le type de données primitif `xsd:string` et ses dérivés suivants : `xsd:normalizedString`, `xsd:token`, `xsd:language`, `xsd:NMTOKEN`, `xsd:Name` et `xsd:NCName`.
- Le type de donnée primitif `xsd:boolean`.
- Le type de donnée numérique primitif `xsd:decimal`, `xsd:float` et `xsd:double`, et les dérivés du type `xsd:decimal` (`xsd:integer`, `xsd:positiveInteger`, `xsd:nonPositiveInteger`, `xsd:negativeInteger`, `xsd:nonNegativeInteger`, `xsd:long`, `xsd:int`, `xsd:short`, `xsd:byte`, `xsd:unsignedLong`, `xsd:unsignedInt`, `xsd:unsignedShort`, `xsd:unsignedByte`).
- Les types de données primitifs relatifs au temps : `xsd:dateTime`, `xsd:time`, `xsd:date`, `xsd:gYearMonth`, `xsd:gYear`, `xsd:gMonthDay`, `xsd:gDay`, et `xsd:gMonth`.
- Les types de données primitifs `xsd:hexBinary`, `xsd:base64Binary`, et `xsd:anyURI`.

4.5.2.4 Caractéristiques des propriétés

Souvent, les axiomes de propriété définissent les caractéristiques additionnelles de propriétés. OWL supporte les constructeurs suivants :

- Les constructeurs RDF-S : `rdfs:subPropertyOf`, `rdfs:domain`, et `rdfs:range`
- Les relations à d'autres propriétés : `owl:equivalentProperty` et `owl:inverseOf`

- Les contraintes globales de cardinalité : `owl:FunctionalProperty` et `owl:InverseFunctionalProperty`
- Les caractéristiques logiques de propriété : `owl:SymmetricProperty` et `owl:TransitiveProperty`

Propriété `inverseOf` (`owl:inverseOf`)

Une propriété `R1` marquée comme `inverseOf` de la propriété `R2` satisfait l'axiome suivant : $R1(x,y)$ si $R2(y,x)$

Exemple:

```
</owl:ObjectProperty>
  <owl:ObjectProperty rdf:ID="DonneCour">
    <rdfs:range rdf:resource="#Cours"/>
    <owl:inverseOf rdf:resource="#EnseignePar"/>
    <rdfs:domain rdf:resource="#EnseignantChercheur"/>
  </owl:ObjectProperty>
```

Un cours est enseigné par un EnseignantChercheur et un EnseignantChercheur Donne un cours. Ainsi les deux propriétés `DonneCour` et `EnseignePar` sont inverses.

Propriété `functional` (`owl:functionalProperty`)

Cette propriété précise que chaque instance de la classe ne peut avoir qu'une seule valeur pour cette propriété (équivalent à donner une cardinalité de 1).

Une propriété `R`, marquée comme `functional` satisfait la règle suivante :

$$R(x,y) \text{ et } R(x,z) \rightarrow y = z$$

Dans notre ontologie, la propriété `ISBN` est du type `functional`. En effet, un livre n'a qu'un seul `ISBN`.

```
<owl:DatatypeProperty rdf:ID="ISBN">
  <rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
  <rdfs:domain rdf:resource="#Livre"/>
  <rdf:type
rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty
"/>
</owl:DatatypeProperty>
```

Propriété `InverseFunctionalProperty` (`owl:InverseFunctionalProperty`)

La propriété `Inversefunctional` implique que les éléments de l'intervalle (range) donnent un identifiant unique à chaque élément du domaine.

Une propriété R, marquée comme `InverseFunctional` satisfait

$$R(y,x) \text{ et } R(z,x) \rightarrow y = z$$

```
<owl:InverseFunctionalProperty rdf:ID="DroitAcces">
  <rdfs:domain rdf:resource="#personne"/>
  <rdfs:comment
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >droits d'accès aux données du profil</rdfs:comment>
  <rdfs:range rdf:resource="#NiveauSecurite"/>
  <rdf:type
rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"
  />
  <rdf:type
rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProp
erty"/>
</owl:InverseFunctionalProperty>
```

Propriétés symétriques (`owl:SymmetricProperty`)

Une propriété R marquée comme symétrique doit satisfaire la règle suivante :

$$R(x,y) \text{ si } R(y,x)$$

Exemple :

La propriété `CoopereAvec` est symétrique.

```
<owl:ObjectProperty rdf:ID="CoopereAvec">
  <rdf:type rdf:resource="&owl;SymmetricProperty" />
  <rdfs:domain rdf:resource="#Chercheur" />
  <rdfs:range rdf:resource="#Chercheur" />
</owl:ObjectProperty>
```

Propriétés transitives (`owl:TransitiveProperty`)

Une propriété R marquée comme transitive satisfait la règle suivante :

$$R(x,y) \text{ et } R(y,z) \rightarrow R(x,z)$$

4.5.2.5 Restrictions sur les propriétés

Une restriction de propriété est une sorte de description de classe. Elle définit une classe anonyme, une classe dont tous les individus satisfont la restriction. OWL distingue deux sortes de restrictions : les contraintes de valeur et les contraintes de cardinalité.

Les contraintes de valeur

La contrainte de valeur est une propriété qui lie une classe de restriction à une classe de description ou à un range de donnée.

allValuesFrom (owl:allValuesFrom)

La classe contient toutes les valeurs possibles de la propriété R (c-à-d que toutes instances rencontrant les valeurs associées à la propriété appartiennent à la classe).

someValuesFrom (owl:someValuesFrom)

La classe contient la valeur de la propriété R (c.à.d. spécifie qu'au moins une seule valeur de la propriété doit être rencontrée pour définir la classe).

La différence entre les deux formulations est en fait la différence entre une quantification universelle et existentielle.

Exemple de la restriction someValueFrom :

```
<owl:Restriction>
  <owl:someValuesFrom>
    <owl:Class rdf:about="#Revue"/>
  </owl:someValuesFrom>
  <owl:onProperty>
    <owl:ObjectProperty rdf:about="#APublication"/>
  </owl:onProperty>
</owl:Restriction>
```

Les contraintes de cardinalité

- Un élément owl:cardinality contient le nombre exact d'éléments différents pour la propriété R.
- Un élément owl:maxCardinality contient le nombre maximum d'éléments différents pour la propriété R.
- Un élément owl:minCardinality contient le nombre minimum d'éléments différents pour la propriété R.

Exemple:

On restreint la propriété Nom avec une cardinalité égale à 1. Une personne n'accepte qu'un et un seul nom.

```
<owl:Restriction>
  <owl:Cardinality
rdf:datatype="http://www.w3.org/2001/XMLSchema#int "
  >1</owl:Cardinality>
  <owl:onProperty>
    <owl:DatatypeProperty rdf:about="#Nom"/>
  </owl:onProperty>
</owl:Restriction>
```

L'élément énumération (`owl:oneOf`)

Une énumération est un élément `owl:oneOf` contenant une liste d'objets (qui sont des instances). La classe définie par un `oneOf` contient exactement les instances contenues dans la liste, pas une de plus, ni une de moins.

Exemple :

```
<owl:Restriction>
...
<owl:oneOf rdf:parseType="Resource">
  <rdf:first
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Femelle</rdf:first>
  <rdf:rest rdf:parseType="Resource">
    rdf:first
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Male</rdf:rest>
  <rdf:rest
rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-
ns#nil"/>
...
<owl:onProperty>
<owl:DatatypeProperty rdf:about="#Sexe"/>
</owl:onProperty>
</owl:Restriction>
```

4.5.2.6 Opérateurs ensemblistes : `intersectionOf`, `unionOf`, `complementOf`

Il s'agit de combinaisons de différentes classes qui peuvent être construites grâce à :

- l'élément `owl:intersectionOf` contenant une liste de noms de classes et définissant l'intersection exacte, c'est-à-dire les objets en communs à toutes ces classes. L'intersection est un constructeur de classe employé dans la logique de description. Il peut être vu comme l'opérateur AND sur les classes.
- l'élément `owl:unionOf` contenant une liste de noms de classes et définissant l'union exacte de tous les objets de chaque classe.
- l'élément `owl:complementOf` contenant un nom de classe uniquement et définissant le complémentaire.

Exemple :

L'exemple suivant montre l'utilisation du constructeur `unionOf` :

```
<owl:ObjectProperty rdf:ID="PapierPresente">
  <rdfs:domain rdf:resource="#Presentation"/>
  <rdfs:range>
    <owl:Class>
      <owl:unionOf rdf:parseType="Collection">
        <owl:Class rdf:about="#PapierConference"/>
        <owl:Class rdf:about="#PapierWorkshop"/>
      </owl:unionOf>
    </owl:Class>
  </rdfs:range>
</owl:ObjectProperty>
```

Classes disjointes

Un élément `disjointWith` `owl:disjointWith` signifie que deux classes ne peuvent avoir d'instances en commun.

Nous n'avons pas cité ici certains constructeurs, qui peuvent être trivialement implémentés grâce à ceux que nous avons évoqués (par exemple `owl:sameClassAs`, servant à affirmer que deux classes sont identiques, peut être écrit grâce à deux `rdfs:subClassOf`).

4.5.3 Visualisation de l'ontologie

OntoViz⁵ est un plugiciel pour permettre à Protégé 2000 de visualiser des ontologies. Cette application nécessite le logiciel de graphisme Graphviz⁶.

Graphviz est un logiciel de code source libre permettant à d'autres applications de créer des graphiques pour visualiser des graphes. L'installation d'OntoViz nécessite des étapes supplémentaires : en plus de copier les fichiers « jar » nécessaires dans le répertoire « plugin » de Protégé 2000, il faut installer Graphviz et ensuite aller placer quelques paramètres dans un fichier de configuration de Protégé 2000 pour lier Graphviz à Protégé. Un fichier d'instruction accompagnant OntoViz explique les étapes d'installation. Une fois installé, il s'agit d'activer OntoViz dans Protégé 2000 en cochant une case dans l'item « configuration » du menu « Project » pour ajouter l'onglet OntoViz. Les relations de classes et sous classes sont représentées par la relation « is a ». En choisissant l'option « slot » et « slot edge » nous

⁵ <http://protege.stanford.edu/plugins/owl/download.html>

⁶ <http://www.research.att.com/sw/tools/graphviz/>

faisons apparaître d'autre relation comme « DonneCour » ainsi que les propriétés des classes. Voir la figure 5.10.

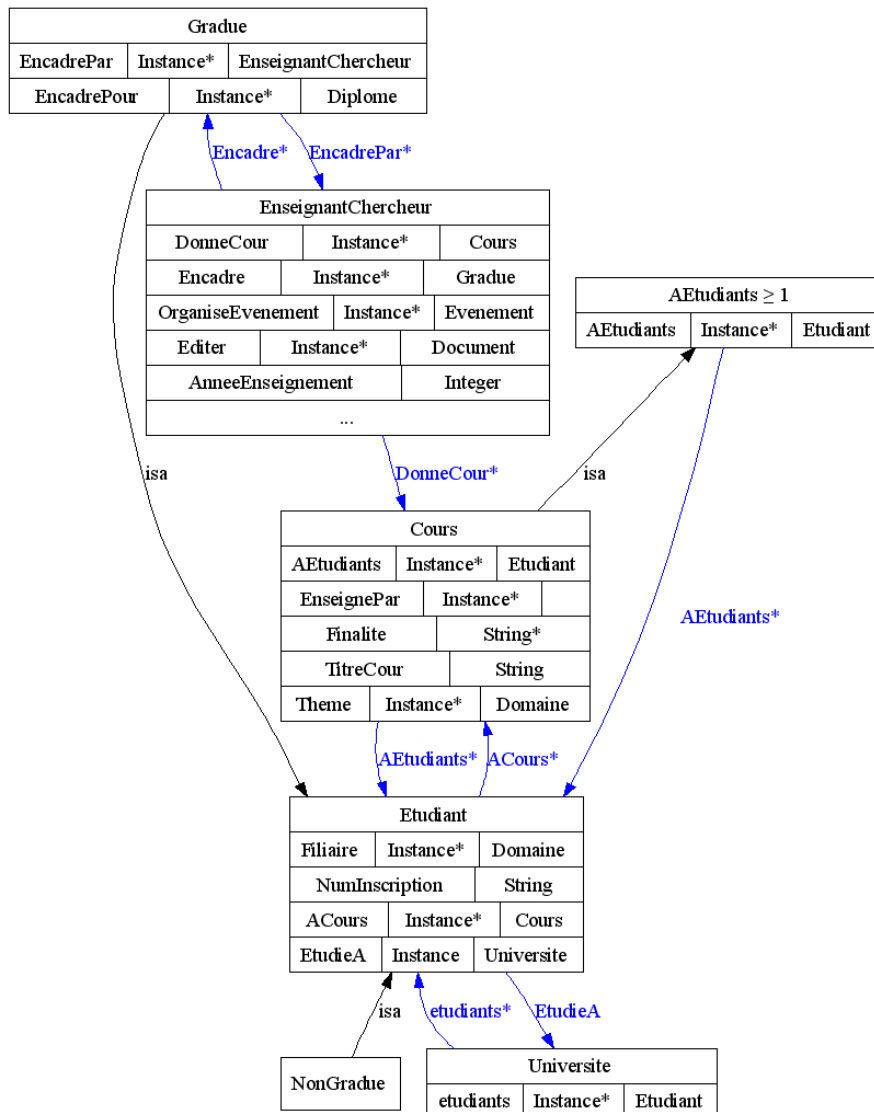


Figure 5.10 : OntoViz -Visualisation des attributs des classes.

4.5.4 Les requêtes

La formation des requêtes, dans protégé 2000, est guidée par une interface graphique (figure 5.11).

Voici quelques exemples de requêtes qu'on peut formulé sur l'ontologie construite :

- Quel est le centre d'intérêt pour un utilisateur X?
- Trouver les cours dont Z. Boufaïda est enseignante ?
- Qui enseigne le cours SIA ?
- Y a-t-il un enseignant-chercheur ?

- Y a-t-il une Personne de LIRE ?
- Trouver des documents écrits par X
- Trouver un institut qui *travaille* sur le sujet S.

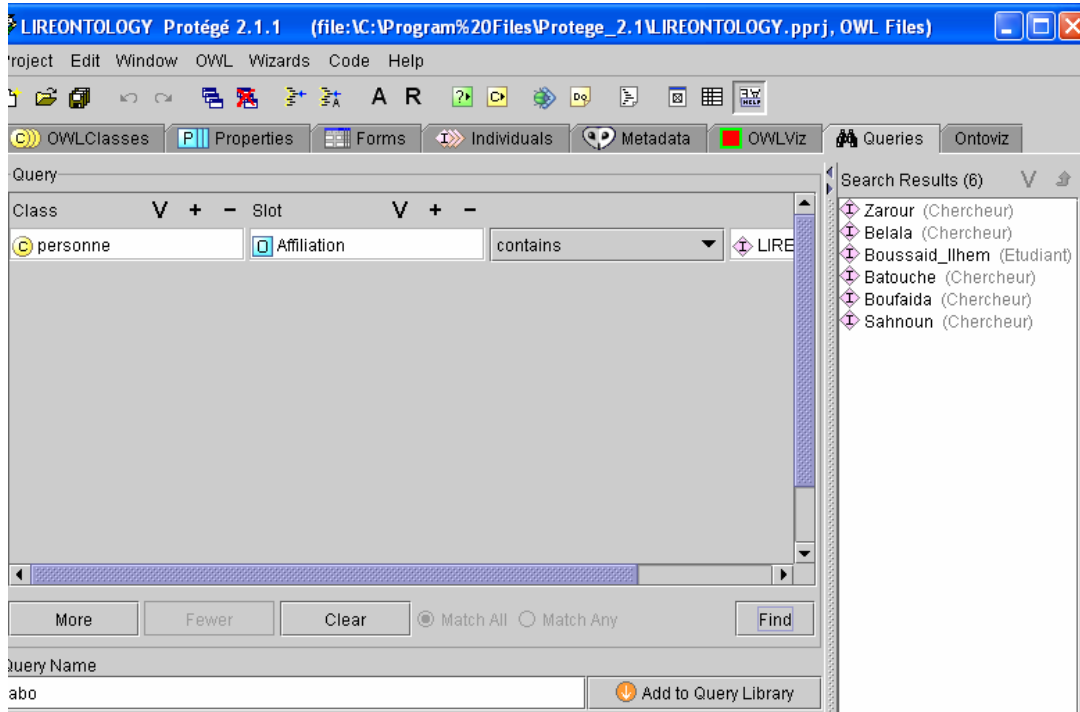


Figure 5.11 – Formulation de la requête : Y a-t-il une Personne de LIRE ?

4.5.5 Le raisonnement

La représentation des connaissances avec OWL a été illustrée, nous présentons dans cette section le raisonnement sur cette représentation à l'aide d'un moteur d'inférence. En se limitant à un sous-ensemble du langage OWL (OWL-DL), un type de raisonnement particulier, la classification automatique, devient faisable. Ce type de raisonnement peut être utilisé pour construire des taxonomies automatiquement sur base des propriétés, et classer automatiquement des instances de classes sur base de leurs propriétés. Cela peut être réalisé avec Protégé, le « plugin » OWL, et l'outil d'inférence RACER.

La fidélité de OWL à la logique de description permet de supporter des inférences au niveau de l'ontologie grâce au mécanisme de subsomption. Le raisonnement peut être utile entre autres :

- Pour valider la relation entre les classes lors de la conception de l'ontologie. Ceci est d'autant plus important pour les ontologies comportant plusieurs auteurs.
- Pour faciliter l'intégration d'information entre ontologies et éviter entre autres des inconsistances au niveau des hiérarchies entre les classes.

- Pour son déploiement : vérifier si un individu est une instance d'une classe, déterminer si un ensemble de faits peut être consistant avec l'ontologie.
- Pour permettre aux agents intelligents de mieux exploiter le contenu de l'ontologie.

5. Conclusion

Les connaissances sont souvent dispersées sous des formes diverses, explicites comme les documents ou tacites comme les savoir-faire, ce qui rend leur réutilisation difficile. Pour réaliser cette gestion de connaissances, nous nous sommes appuyés sur les notions d'ontologies et de Web sémantique.

Ce chapitre a présenté la modélisation de la connaissance que nous avons adoptée. Nous avons justifié notre choix d'utiliser les ontologies et l'avons appliqué à la modélisation de la connaissance relative aux utilisateurs, documents, domaine dans le cadre applicatif d'un laboratoire de recherche.

Ces travaux ont permis d'apporter d'une part un principe de personnalisation pour un utilisateur ou groupe d'utilisateurs, et d'autre part un principe de modélisation des utilisateurs pour la mise en œuvre de ces personnalisations, pour un utilisateur particulier (ou groupe d'utilisateurs) sur Internet. Ces principes sont fondés sur des ontologies, ou plutôt sur une ontologie de niveau plus abstrait qui fédère sous une seule et même structure les quatre ontologies dont nous avons besoin : ontologie du domaine, ontologie de l'application, schéma de métadonnées (ou ontologie des documents) et l'ontologie des utilisateurs. Le processus de personnalisation présuppose l'existence de cette structure comme base pour son fonctionnement, c'est pourquoi les ontologies sont des paramètres du processus de personnalisation.

L'ontologie de domaine représente l'ensemble des concepts du domaine de l'application ainsi que les relations entre ces concepts. Elle est utilisée comme un vocabulaire commun pour décrire le contenu des documents et les connaissances des utilisateurs. L'ontologie de l'application représente la gestion de la personnalisation. Le schéma de métadonnées est nécessaire pour améliorer la recherche et la gestion d'informations hétérogènes. L'ontologie utilisateur permet de représenter sous une même entité l'ensemble des caractéristiques d'un utilisateur. Il est nécessaire au filtrage et à la sélection de l'information pertinente pour l'utilisateur, mais aussi pour personnaliser son utilisation du système.

La séparation de ces différentes catégories de connaissance facilite leurs évolutions et leur maintenance. Elle permet aussi d'améliorer leur compréhension et de définir des rôles différents pour chacune des ontologies, c'est-à-dire de les spécialiser. Il s'agit donc de

représentations séparées, mais néanmoins interconnectées comme nous venons de le voir. Ces ontologies sont des paramètres du système, ce qui permet de les modifier sans remettre en cause les principes de personnalisation.

Nous avons ensuite expliqué comment ces ontologies sont mises en relation avec les données réelles. L'utilisation d'ontologie comme base à la qualification revêt celle-ci d'une dimension sémantique. Cette modélisation de la connaissance sert de base au raisonnement appliqué par le système lors de la personnalisation de l'information.

Ce travail nous a montré que l'utilisation des ontologies et du Web sémantique se révèle effectivement très intéressante pour la gestion des connaissances dans un laboratoire de recherche.

Conclusion Générale

Le besoin de personnalisation semble évident si nous envisageons dans le futur des machines capables de remplacer les êtres humains dans certaines activités. Pour cela, le système doit avoir une représentation suffisamment étoffée et complexe de l'utilisateur.

Après avoir effectué une revue des propositions actuelles de modélisation des utilisateurs, nous avons apporté des innovations supplémentaires au niveau des modèles déjà existants.

Le but de notre modélisation des utilisateurs est de déterminer à la fois les caractéristiques à retenir pour construire des profils utilisateur et d'établir les liens entre ces caractéristiques et leurs relations avec les éléments du modèle.

Notre choix pour la modélisation des utilisateurs s'est porté sur les *ontologies*. L'état de l'art que nous avons présenté dans le chapitre 3 permet d'en cerner la définition ainsi que les langages et les techniques qui permettent de les concevoir. Au-delà de leur qualité pour la modélisation de la connaissance, les propriétés des ontologies favorisent le raisonnement sur les connaissances représentées.

Quatre ontologies particulières ont été présentées dans le cadre de ce mémoire :

1. **Ontologie des utilisateurs.** A la différence de nombreux systèmes, notre modèle utilisateur se focalise sur la création automatique de profils utilisateurs basés sur des ontologies. L'ontologie des utilisateurs agit en quelque sorte d'interface entre les utilisateurs et les autres ontologies. Elle nous renseigne sur les préférences et les connaissances de ce dernier exprimées en fonction de l'ontologie du domaine. Les informations qui caractérisent un utilisateur sont représentées dans un profil. Il contient des informations administratives et les préférences explicites qui vont de la personnalisation d'interface aux centres d'intérêt. Il positionne également l'utilisateur dans l'organisation (rôle, position, connaissance, etc.) permettant au système de cibler ses propres actions.
2. **Ontologie du Domaine** par laquelle le système accède à une connaissance du domaine. Elle représente la connaissance sur un domaine particulier et offre des possibilités de raisonnements sur ces connaissances. Elle permet ainsi aux utilisateurs d'exprimer leurs centres d'intérêts de façon à recevoir des informations ciblées.

- 3. Ontologie d'application** qui contient des informations relatives à la gestion de la personnalisation. Dans le chapitre 2, nous avons mis en évidence comment peut s'exprimer la personnalisation. Nous avons ainsi observé les différents types de personnalisation pouvant être mis en œuvre, décrit les dimensions sur lesquelles peuvent porter les personnalisations (contenu, navigation, présentation) et souligné l'importance d'une représentation de l'utilisateur pour le processus de personnalisation. Ainsi, le premier pas vers la construction d'un nouveau système capable de traiter les données en fonction des besoins et les préférences d'un utilisateur est le choix du service de personnalisation que ce système va fournir.
- **Ontologie des documents**, se concentre sur la gestion des ressources documentaires d'un laboratoire de recherche (documents scientifique et techniques). Cette étude préliminaire s'est révélée extrêmement instructive et intéressante, car elle a soulevé de multiples questions à la fois théoriques et pratiques. Elle permet de recenser et de structurer l'ensemble de la production scientifique. En effet, ce type de démarche permet d'obtenir des comparaisons entre établissements, à un niveau national ou international.

Il est clair que les ontologies acquièrent une importance notoire dans un grand nombre de domaines. Mais la construction d'ontologies pour un Web Sémantique, réactualise inévitablement un certain nombre de questions classiques mais récurrentes de représentation des connaissances : les « multiples points de vue ou multiples hiérarchies », la notion de contexte, etc. Mais elle soulève aussi d'autres questions nouvelles plus spécifiques au Web Sémantique :

1. Fusionner, connecter, ou importer les ontologies ?

Le Web étant fondamentalement distribué, contiendra vraisemblablement plusieurs ontologies *modulaires*, produites indépendamment par divers utilisateurs mais qui auront inévitablement besoin *d'assembler* ces multiples ontologies. Développées indépendamment et pour des usages différents, ces ontologies

- 1) utiliseront un vocabulaire et des langages de représentation différents
- 2) contiendront inévitablement des connaissances qui se recoupent.

Une question essentielle alors est : comment réconcilier ces différences ? Les ontologies devront elles être assemblées par une « *fusion* » produisant une nouvelle ontologie où les concepts similaires seraient fusionnés en un unique concept (C de l'ontologie $A \equiv D$ de

l'ontologie B), de nouvelles relations ajoutées etc. ?, ou par une « *interconnexion* » définissant les règles sur les relations entre les différentes ontologies, leurs concepts, propriétés et relations par exemple, « C de l'ontologie A implique D de l'ontologie B » ou encore par « *des mécanismes de modularité* » adaptés qui permettent l'importation d'ontologies modulaires (ontologie A utilise ontologie B où le concept D de B est renommé en C de A). En garantissant la cohérence sémantique de l'ontologie résultat.

2. Quels mécanismes de modularité et quel langage d'assemblage ?

Si cette voie semble prometteuse, elle pose deux questions fondamentales :

- 1) définir les mécanismes de modularité et
- 2) les langages à utiliser pour permettre l'intégration sémantique d'ontologies modulaires réutilisables. En particulier il faudra déterminer si OWL et ses sous-langages permettent de répondre à ce besoin. Cette question rejoint des questions fondamentales sur la modularité, bien étudiées en génie logiciel.

Toutefois, ces limites peuvent être atténuées pour deux raisons :

1. Compte tenu des efforts actuels, il est raisonnable de penser qu'un nombre croissant d'ontologies de domaines seront disponibles et pourront donc être réutilisées moyennant quelques adaptations. Dans l'avenir, il existera très certainement plusieurs ontologies sur le même domaine ou ayant le même objectif. Elles vont devoir coexister et collaborer. En effet, il est difficile de mettre tous d'accord avec une unique ontologie dans certains domaines. Il sera donc nécessaire d'aider les utilisateurs à poser une requête par sélection d'une ou plusieurs ontologies pour un domaine donné ou par proposition de différents vocabulaires. Il s'agit ici de développer des systèmes médiateurs capables d'établir les correspondances entre différentes modélisations d'un domaine.
2. De nombreux travaux tentent également de proposer des méthodes et des outils pour construire ces ontologies semi-automatiquement à partir de ressources textuelles.

Ces efforts permettent de réduire fortement le coût de construction des ontologies.

Évolutions possibles

Nous avons pensé à quelques évolutions concernant l'approche proposée. Dans un premier temps, il serait intéressant d'extraire de l'ontologie de l'application une ontologie de la personnalisation de façon à rendre cette dernière réutilisable dans différentes applications. Nous envisageons également d'associer des stéréotypes d'utilisateurs à chacune des stratégies de personnalisation prévues dans le système (déclarée dans l'ontologie de l'application) pour réduire l'effort de personnalisation et de préciser les éléments du modèle utilisateurs pris en compte pour chaque stratégie.

Il serait aussi intéressant d'étudier les possibilités de réutilisation de l'approche pour des portails de gestion de connaissance, des systèmes d'information en ligne, des bibliothèques numériques, des intranets d'entreprise, ou pour la gestion des mémoires organisationnelles.

En plus, les caractéristiques utilisateurs sont susceptibles d'évoluer. A cet égard, on ne peut se satisfaire d'une représentation figée dans le temps de l'utilisateur pour mettre en œuvre la personnalisation. Il est possible de faire évoluer le modèle utilisateur à l'aide d'une fonction d'apprentissage, à l'aide d'un modèle probabiliste ou à l'aide d'agents logiciels. Par conséquent, on peut utiliser la plupart des méthodes d'apprentissage vues précédemment (cf., chapitre 1) pour faire évoluer ces modèles et optimiser sa performance au fur et à mesure que le temps passe.

Enfin, avec cette première version de l'ontologie, se posent maintenant les problèmes de son cycle de vie et de la maintenance, non seulement de l'ontologie mais aussi de ce qui a été construit au-dessus (modèles, inférences...). Il nous reste à étudier comment gérer ce cycle de vie global.

BIBLIOGRAPHIE

- [Aamodt, 1994] A. Aamodt et E. Plaza. Case-Based Reasoning: Foundational Issues, Methodological Variations and Systems. *AI Communications*, vol. 7, n°1, p 36–59, 1994.
- [Allen et al., 2001] C. Allen, D. Kania et B. Yaeckel. *One-to-One Web Marketing, Second Edition* - John Wiley & Sons, 2001
- [Amato et Straccia, 1999] G. Amato et U. Straccia. User Profile Modeling and Applications to Digital Libraries, Proc. 3rd European Conf. Research and Advanced Technology for Digital Libraries, ECDL, 1999.
- [Amazon, 2004] <http://www.amazon.com>
- [Armstrong et al., 1995] R. Armstrong, D. Freitag, T. Joachims et T. Mitchell. WebWatcher: A Learning Apprentice for the World Wide Web. In *1995 AAAI Spring Symposium on Information Gathering from Heterogeneous Distributed Environments, 1995*.
- [Arpirez et al., 1998] J. Arpirez, A. Gómez-Perez, A. Lozano et S. Pinto. (ONTO)2Agent: An ontology- based WWW broker to select ontologies. Paper presented at the Workshop on Applications of Ontologies and PSMs, Brighton, England, 1998.
- [Arpirez et al., 2001] J. Arpirez, O. Corcho, M. Fernandez-Lopez et A. Gómez-Pérez. WebODE : a Workbench for Ontological Engineering. In *First international Conference on Knowledge Capture (K-CAP'01)*, p. 6–13, Victoria, Canada: ACM, 2001.
- [Asnicar et Tasso, 1997] F. Asnicar et C. Tasso. IfWeb: A Prototype of User Models Based Intelligent Agent for Document Filtering and Navigation in the WorldWideWeb. In *Proceedings of UM'97*. Sardinia, Italy: Chia Laguna, 1997.
- [Baader et Sattler, 2000] F. Baader et U. Sattler. Tableau algorithms for description logics. *In Proceedings of Tableaux 2000, University of St Andrews, Scotland*. Springer-Verlag, 2000.
- [Baader et al., 2003] F. Baader, D. Calvanese, D. MacGuinness, D. Nardi et P.F. Patel-Schneider. *The Description Logique Handbook: Theory, Implementation and applications*. Cambridge University Press, 2003.
- [Bachimont, 2000] B. Bachimont. Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances. In J. Charlet, M. Zacklad, G. Kassel et D. Bourigault, Eds., *Ingénierie des connaissances : évolutions récentes et nouveaux défi*, chapitre 19. 305-323. Paris: Eyrolles, 2000.

- [**Bachimont et al., 2002**] B. Bachimont, A. Isaac et R. Troncy. Semantic Commitment for Designing Ontologies: A Proposal. In A. Gómez-Pérez et V. Benjamins, Eds., 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW'02), volume (2473) of Lecture Notes in Artificial Intelligence, p. 114–121, Sigüenza, Espagne: Springer Verlag, 2002.
- [**Basu et al., 1998**] C. Basu, H. Hirsh et W. Cohen. Recommendation as Classification: Using Social and Content-Based Information in Recommendation. In *Proceedings of AAAI'98*, 714–720, 1998.
- [**Bazsalicza et Naim, 2001**] M. Bazsalicza et P. Naim. Data Mining pour le Web. Edition Eyrolles, Paris, 2001
- [**Bechhofer et al., 2001**] S. Bechhofer, I. Horrocks, C. Goble et R. Stevens. OilEd: a Reasonable Ontology Editor for the Semantic Web. In Joint German/Austrian conference on Artificial Intelligence (KI'01), volume (2174) of Lecture Notes in Artificial Intelligence, p. 396–408, Vienne, Autriche: Springer Verlag, 2001.
- [**Belkin et Croft 1992**] N. J. Belkin, W. B. Croft. Information filtering and information retrieval: two sides of the same coin?, *Communications of the ACM*, vol. 35, n° 12, p. 29-38, décembre 1992.
- [**Benichoux et al., 1985**] R. Benichoux, J. Michel, D. Pajaud. Guide pratique de la communication pratique : comment écrire- comment dire ?, Paris : Gaston Lachurié. 268p, 1985.
- [**Bernaras et al., 1996**] A. Bernaras, I. Laresgoiti et J. Corera. Building and Reusing Ontologies for Electrical Network Applications. In: *Proceedings European Conference on Artificial Intelligence (ECAI96)*. pp. 298-302. Budapest, Hungary, 1996.
- [**Berners-Lee et al., 2001**] T. Berners-Lee, J. Hendler et O. Lassila. The Semantic Web. *Scientific American*, 284(5):34-43, Mai c2001.
- [**Billsus et Pazzani, 1999**] D. Billsus et M. J. Pazzani. A Hybrid User Model for News Classification. In *Proceedings of UM'99*, 99–108. Wien, New York: Springer-Verlag, 1999.
- [**Blazquez et al., 1998**] M. Blazquez, M. Fernández, J.M. Garcia-Pinar et A. Gomez-Perez. Building Ontologies at the Knowledge Level using the Ontology Design Environment. Paper presented at the Proceedings of the Banff Workshop on Knowledge Acquisition for Knowledge-based Systems, Canada.1998.
- [**Bollacker et al., 1999**] K. D. Bollacker, S. Lawrence, C. L. Giles. A system for Automatic Personalized Tracking of Scientific Literature on the Web, in *Proceeding of the 4th ACM Conference on Digital Library*, Berkeley, CA, August 11-14, 1999.

- [Boone, 1998] G. Boone. Concept Features in RE:Agent, an Intelligent Email Agent. In *The Second International Conference on Autonomous Agents (Agents '98)*. Minneapolis/St. Paul, 1998.
- [Borzsonyi et al., 2001] S. Borzsonyi, D. Kossmann, K. Stocker. The Skyline Operator, IEEE Conf. On Data Engineering, 2001
- [Brachman et Schmolze, 1985] R.J. Brachman et J.G. Schmolze, An overview of the KL-One knowledge representation system, in *Cognitive Science* 9(2), pages 171-216, 1985.
- [Bradley et al., 2000] K. Bradley, R Rafter, B. Smyth, Case-Based User Profiling for Content Personalisation, *Lecture Notes in Computer Science*, 2000.
- [Brusilovsky, 1997] P. Brusilovsky. Efficient techniques for adaptive hypermedia, In Nicholas C. & Mayfield J. (Eds.), *Intelligent hypertext: Advanced techniques for the World Wide Web*, LNCS 1326, pp. 12-30, Berlin, Springer-Verlag, 1997.
- [Brusilovsky et Eklund, 1998] P. Brusilovsky, L. Eklund et al. Web-based Education for all: A tool for development adaptive courseware. *Seventh International World Wide Web Conference. Computer Networks and ISDN Systems*, (30): 291-300, 1998.
- [Brusilovsky, 2001] P. Brusilovsky. Adaptive Hypermedia. In *User Modeling and User-Adapted Interaction 11*- Kluwer Academic Publishers - pp87-110, 2001.
- [Calvanese et al., 1997] D. Calvanese, G. De Giacomo, M. Lenzerini. Conjunctive query containment in Description Logics with n-ary relations. *International Workshop on Description Logics*, 1997.
- [Cannataro et al., 2001] M. Cannataro, A. Cuzzocrea et A. Pugliese. A probabilistic adaptive hypermedia system, 1st International Workshop on Web Dynamics, In conjunction with the 8th International Conference on Database Theory, London, UK, 3 January 2001
- [Cantor et Etgen, 1999] J. Cantor et M. Etgen. - What does getting WET (Web Event-logging Tool) Mean for Web Usability? - AT&T Labs -, pp. 1-3, 06/1999
- [Carby et al., 2000] O. Corby, R. Dieng et C. Hébert. A Conceptual Graph Model for W3C Ressource Description Framework. In 8th International Conference on Conceptual Structures (ICCS'00), pages 468-482, Darmstadt, Allemagne, 14-18 Août 2000.
- [CDNow, 2001] <http://www.cdnw.com>
- [Chan, 1999] P.K. Chan. A non-invasive approach to building Web user profiles, *Workshop on Web Usage Analysis and User Profiling*, 5th International Conference on Knowledge Discovery & Data Mining, San Diego, CA, August 15-18, 1999.

- [Charlet, 2002] J. Charlet. L'Ingénierie des connaissances : développements, résultats et perspectives pour la gestion des connaissances médicales. Habilitation à diriger des recherches, Université Paris 6, 2002.
- [Chomicki, 2002] J. Chomicki, Querying with Intrinsic Preferences, EDBT, 2002
- [Cooley et al., 1999] R. Cooley, P. N. Tan, et J. Srivastava. WebSift: The Web Site Information Filter System. In *Proceedings of the 1999 KDD Workshop on Web Mining*. San Diego, CA: Springer-Verlag, 1999.
- [Cunningham et al., 2001] P. Cunningham, R. Bergmann, S. Schmitt, R. Traphoner, S. Breen et B. Smyth. WebSell: Intelligent Sales Assistants for the World Wide Web. In *E-2001*, 2001.
- [Cutting et al., 1992] D. R. Cutting, D. R. Karger, J. O. Pederson, J. W. Tukey. Scatter/Gather: A cluster-based approach to browsing large document collections. *Actes de la 15ième conférence Annuelle ACM/SIGIR Research and Development in Information Retrieval*, Copenhagen, Danemark, 1992, p.318-329.
- [Devedzic, 1999] V. Devedzic. ITS Ontology Engineering: Borrowing from Design Patterns. Workshop on Ontologies for Intelligent Educational Systems, Ninth International Conference on Artificial Intelligence in Education, AI-ED'99, Le Mans, France, July 19-23th, 1999. <http://www.ei.sanken.osaka-u.ac.jp/aied99/aied99-onto.html>
- [Chomicki, 2002] J. Chomicki, Querying with Intrinsic Preferences, EDBT, 2002.
- [Croft et al., 2001] W. B. Croft, St. Cronen-Townsend, V. Lavrenko, Relevance Feedback and Personalization: A Language Modeling Perspective, Computer Science, DELOS Workshop, 2001
- [Daniels, 1986] J.P. Daniels. Cognitive Models in Information Retrieval – An Evaluation Review, *Journal of Documentation*, 42(4), pp. 272-304, December 1986.
- [Dieng et al., 2001] R. Dieng-Kuntz, O. Corby, F. Gandon, A. Giboin, J. Golebiowska, N. Matta, M. Ribière. *Methodes Et Outils Pour La Gestion Des Connaissances: Une Approche pluridisciplinaire du Knowledge Management (2nd Edition)*, Dunod Edition, INFORMATIQUES Série Systèmes d'information - ISBN 2 10 006300 6, 2001.
- [Domingue , 1998] J. Domingue. Tadzebao and WebOnto: Discussing, Browsing, and Editing Ontologies on the Web. Paper presented at the 11th Workshop on Knowledge Acquisition, Modeling and Management, KAW'98, Banff, Canada, 1998.
- [Euzenat, 1999] J. Euzenat. Sémantique des représentations de connaissance. Notes de cours de DEA d'Informatique, Université Joseph Fourier, Grenoble, 1999.

- [Euzenat et al., 2000] J. Euzenat, A. Napoli et R. Ducournau. Les représentations des connaissances par objets. *Techniques et science informatiques*, 19(1), 2000.
- [Farquhar et al., 1995] A. Farquhar, R. Fikes, W. Pratt, J. Rice. Collaborative Ontology Construction for Information Integration Technical Report KSL-95-10. Knowledge Systems Laboratory. Stanford University. CA. 1995
- [Faure et Nedellec, 1999] D. Faure et C. Nedellec. Knowledge acquisition of predicate argument structures from technical texts using machine learning: The system ASIUM. In Proc. of the 11th European Workshop, Knowledge Acquisition, Modelling and Management, number 1937 in LNAI, p.329–334, Juan-les-Pins, France: Springer-Verlag. 1999.
- [Fensel et al., 1998] D. Fensel, S. Decker, M. Erdmann, et R. Studer. Ontobroker: How to make the WWW Intelligent. Research report, Institute AIFB, in Proceedings KAW98, the 11th Banff Knowledge Acquisition for Knowledge-Based System Workshop research report, Banff, Kanada, April 1998.
<http://www.aifb.uni-karlsruhe.de/WBS/broker/ontobroker.html>
- [Fensel et al., 2000] D. Fensel, I. Horrocks, F. Van Harmelen, S. Decker, M. Erdmann et M. Klein. OIL in a nutshell. In Proceedings of European Knowledge Acquisition Workshop (EKAW'2000), Springer-Verlag LNAI 1937, pages 1-16, Juan-les-Pins, France, 2000.
- [Fensel et Gómez-Pérez, 2002] D. Fensel et A. Gómez-Pérez. A survey on ontology tools. *OntoWeb Deliverable 1.3*, 31 Mai 2002.
<http://ontoweb.aifb.uni-karlsruhe.de/About/Deliverables/>.
- [Fernández et al., 1997] M. Fernández, A. Gómez-Perez et N. Juristo. Methontology: From Ontological Art Toward Ontological Engineering. Paper presented at the Spring Symposium Series on Ontological Engineering. AAAI97, Stanford, USA, 1997.
- [Fernández et al., 1999] M. Fernández, A. Gómez-Perez, A. Pazos et J. Pazos. Building a chemical ontology using methontology and the ontology design environment. *IEEE Intelligent System and their Applications*, 14(1), 37–46, 1999.
- [Fink et Kobsa, 2000] J. Fink et A. Kobsa. A Review and Analysis of Commercial User Modeling Servers for Personalization on the World Wide Web, *International Journal of User Modeling and User-Adapted Interaction*, 10, Kluwer Academic Publishers, pp. 209-249, 2000.
- [Gandon et al., 2002] F. Gandon, R. Dieng-Kuntz, O. Corby et A. Giboin. Web sémantique et approche multi-agents pour la gestion d'une mémoire organisationnelle distribuée. Dans 13^{èmes} Journées Francophones d'Ingénierie des Connaissances, pages 15-26. IC2002, 2002.

- [Gavray, 2002] G. Gavray. Personnalisation des sites web : élaboration d'une méthodologie de mise en œuvre et application au cas DGTRE. Mémoire Projet Présenté par en vue de l'obtention du titre de : ingénieur de gestion année 2001-2002.
- [Goldberg et al., 1992] D. Goldberg, D. Nichols, B.M. Oki et D.Terry. Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM*, **35**: 61– 70, 1992.
- [Gòmez-Pérez, 1997] A. Gòmez-Pérez. A Framework to Verify Knowledge Sharing Technology. *Expert Systems with Applications*, **11**(4), 519-529. 1997.
- [Gòmez-Pérez, 1999] A. Gòmez-Pérez. Ontological Engineering: A state of the art. *Expert Update*, 2(3), 33-43. Madrid,1999.
- [Gonzales et Wullemin, 1998] C. Gonzales, P.H. Wullemin. Réseaux baysiens en modélisation d'utilisateurs. Dans *Modélisation de l'Apprenant*, revue Science et Techniques Educatives Vol. 5, N°2, Hermès, pp. 123-140, juin, 1998.
- [Good et al., 1999] N. Good, J. Schafer, J. Konstan, A. Borchers, B. Sarwar, J. Herlocker et J. Riedl. Combining Collaborative Filtering with Personal Agents for Better Recommendations. In *Proceedings of AAAI*, Vol. 35, 439–446. AAAI Press, 1999.
- [Gruber, 1991] Thomas R. Gruber. The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases. In *Proceedings of the Second International Conference Principles of Knowledge Representation and Reasoning, (KR & R-91)*, J. Allen, R. Fikes, et E. Sandewall (Eds.), Morgan Kaufmann Publishers: San Mateo, CA, pp.601-602, 1991.
- [Gruber, 1993] R. T. Gruber, A Translation Approach to Portable Ontology Specification, *Knowledge Acquisition* (5), pp 199-220, 1993
- [Gruber, 1994] GRUBER T. & OLSEN G., An ontology for engineering mathematics, in J. DOYLE F. S.&TORANO P., eds., *Proceedings of the 4th International Conference on Principles of Knowledge Representation and Reasoning*, Morgan- Kauffmann, 1994.
- [Gruber, 1995] T.R. Gruber. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. Revision of paper presented at the international workshop on Formal Ontology, Padova, Italy, March 1993, in Special issue of the *International Journal of Human-Computer Studies*, Vol. 43, N° 5-6, Nicola Guarino et Roberto Poli (Eds.), 1995.
- [Grüninger et Fox, 1995] M. Grüninger et M.S. Fox. Methodology for the Design and Evaluation of Ontologies. Paper presented at the *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing*, Montreal, 1995.

- [Guarino,1997] N. Guarino. *Understanding, building, and using ontologies: A commentary to Using Explicit Ontologies in KBS Development, by van Heijst, Schreiber, and Wielinga*. International Journal of Human and Computer Studies 46: 293- 310.,1997
- [Haarslev et Möller, 2001] V. Haarslev et R. Möller. RACER System Description. *Proceedings of the First International Joint Conference on Automated Reasoning table of contents*. 701-706, 2001.
- [Hayes et Cunningham, 2000] C. Hayes et P.Cunningham. Smart Radio: Building Music Radio on the Fly. In *Proceedings of Expert Systems 2000 (ES2000)*. Cambridge, UK, 2000.
- [Hendler et McGuinness, 2000] J. Hendler et D.L. McGuinness. The DARPA Agent Markup Language. IEEE Intelligent Systems- Trends and Controversies, 15(6): 67-73, Janvier 2000.
- [Hill et al., 1997] W. Hill, L. Stead, M. Rosenstein et G. Furnas. Recommending and Evaluating Choices in a Virtual Community of Use. In *Proceedings of CHI'95*, 194-201. Denver, 1997.
- [Horrocks, 1998] I. Horrocks.The FaCT System. *Proceedings of Automated Reasoning with Analytic Tableaux and Related Methods: International Conference (TABLEAUX'98)*. 307-312, 1998.
- [Horrocks et Patel-Schneider, 2003] I. Horrocks et P. F. Patel-Schneider. Reducing OWL Entailment to Description Logic Satisfiability. In D. Fensel, K. Sycara et J. Mylopoulos, editors, 2nd International Semantic Web Conference (ISWC'03), volume (2870) of Lecture Notes in Computer Science, pages 17-29, Sanibel Island, Floride, USA, 20-23 Octobre 2003Springer Verlag.
- [Huberman et Kaminsky, 1996] B. Huberman et M. Kaminsky. Beehive: A System for Cooperative Filtering and Sharing of Information. *Technical Report, Dynamics of Computation Group*. Palo Alto, CA: Xerox, Palo Alto Research Center, 1996.
- [IEEE, 2002] Draft standard for learning object metadata. Technical report, IEEE, <http://ltsc.ieee.org/wg12/> .
- [Ikeda et al., 1999] M. Ikeda, Y. Hayashi, J. Lai, W. Chen, J. Bourdeau, K. Seta et R. Mizoguchi. *An ontology more than a shared vocabulary*. Workshop on Ontologies for Intelligent Educational Systems, Ninth International Conference on Artificial Intelligence in Education, AI- ED'99, Le Mans, France, July 19-23, 1999. <http://www.ei.sanken.osaka-u.ac.jp/aied99/aied99-onto.html>
- [Janowski et Sarner, 2001] W. Janowski, A. Sarner. Five Opportunities for Personalization. Gartner Group, pp. 1, 05/2001.

- [Joachims et al., 1997] T. Joachims, D. Freitag et T. Mitchell. Webwatcher: A tour guide for the World Wide Web. In Proc. IJCAI-97, 1997.
- [Jung et al., 2002] S. Y. Jung, J-H. Hong, T-S. Kim. A Formal Model for User Preference, IEEE International Conference on Data Mining, 2002.
- [Kappel et al., 2000] G. Kappel, W. Retschitzegger et W. Schwinger. Modeling Customizable Web Applications - A Requirement's Perspective, in Proceedings of the International Conference on Digital Libraries, (ICDL 2000), Kyoto, Japan, November 13-16, 2000.
- [Kayser, 1997] D. Kayser. La représentation des connaissances, Hermès, 1997.
- [Kifer et al., 1995] M. Kifer, G. Lausen et J. Wu. Logical Foundations of Object-Oriented and Frame-Based Languages. Journal of the Association for Computing Machinery, 1995.
- [Klyne et al., 2001] G. Klyne, F. Reynolds, C. Woodrow et H. Ohto. Composite Capability/Preference Profile (CC/PP): Structure and Vocabularies, W3C, 2001, <http://www.w3.org/TR/CCPP-struct-vocab>
- [Kobsa et Wahlster, 1989] A. Kobsa et W. Wahlster. User Models in Dialog Systems. Springer-Verlag. Koivunen M. et Miller E. (2001). W3C semantic web activity. Dans Proceedings of the Semantic Web Kick-off Seminar. W3C, 1989. <http://www.w3.org/2001/12/semweb-fin/w3csw> .
- [Kobsa, 2001] A. Kobsa. Generic User Modeling Systems, International Journal of User Modeling and User-Adapted Interaction, 11, Kluwer Academic Publishers, pp. 49-63, 2001.
- [Koch, 2000] N. Koch. Software Engineering for Adaptative Hypermedia Systems – Reference Model, Modelling Techniques and Development Process, Ph.D Thesis, Fakultät der Mathematik und Informatik, Ludwig-Maximilians- Universität München, December 2000.
- [Kok, 1991] A. J. Kok. A review and synthesis of user modelling in intelligent systems. The Knowledge Engineering Review, 6(1) : 21-47, 1991.
- [Kostadinov, 2003] D. Kostadinov. Personnalisation de l'information et gestion des profils utilisateurs, Rapport de DEA, Université de Versailles, France, 2003.
- [Krulwich, 1997] B. Krulwich. LifeStyle Finder: Intelligent User Profiling Using Large-Scale Demographic Data. *AI Magazine* 18(2): 37-45, 1997.
- [Krulwich et Burkey, 1995] B. Krulwich et C. Burkey. ContactFinder: Extracting Indications of Expertise and Answering Questions with Referrals. *Working Notes of the 1995 Fall*

- Symposium on Intelligent Knowledge Navigation and Retrieval*, 85–91. Technical Report FS-95-03, The AAAI Press, 1995.
- [Lang, 1995] K. Lang. NewsWeeder: Learning to Filter News. In *Proceedings of the 12th International Conference on Machine Learning*, 331–339. Lake Tahoe, CA, 1995.
- [Liano et al., 1999] M. Liao, A. Abecker, A. Bernardi, K. Hinkelmann et M. Sintek. Ontologies for knowledge retrieval in organizational memories. Dans F. Bomarius, éditeur, *Proceedings of the Learning Software Organizations (LSO'99) Workshop*, pages 19-26, Kaiserslauten, Germany. <http://citeseer.nj.nec.com/liao99ontologies.html>.
- [Lieberman, 1995] H. Lieberman. Letizia: An agent that assists Web Browsing. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-95)*, pp. 924- 929, Montreal, Canada, August 20-25, 1995.
- [Lieberman et al., 1999] H. Lieberman, N. W. Van Dyke et A. S. Vivacqua. Let's Browse: A Collaborative Web Browsing Agent. In *Proceedings of International Conference on Intelligent User Interfaces*, 924–929, 1999.
- [Luke et al., 1997] S. Luke, L. Spector, D. Rager, J. Hendler. Ontology-based Web Agents. *Proceedings of First International Conference on Autonomous Agents*, AA-97, 1997. <http://www.cs.umd.edu/project/plus/SHOE/spec.html>
- [Macgregor , 1991] R. Macgregor. Inside the LOOM classifier, in *SIGART bulletin* (2)3, pages 70-76, 1991.
- [Maedche et Staab, 2000] A. Maedche et S. Staab. Mining ontologies from text. In R. Dieng et O. Corby, Eds., *Proc. of the 12th European Workshop, Knowledge Engineering and Knowledge management: methods, models and Tools*. 2000.
- [Maedche et al., 2003] A. Maedche, B. Motik, L. Stojanovic, R. Studer et R. Volz. Ontologies for Enterprise Knowledge Management. *IEEE Intelligent Systems*, 18(2): 26-33, Mars-Avril 2003.
- [Mctear, 1993] M. F. Mctear. User modelling for adaptive computer systems: a survey of recent developments.” *Artificial Intelligence Review* 7: 157-184, 1993.
- [Melville et al., 2001] P. Melville, R. J. Mooney, and R. Nagarajan, Content-boosted collaborative filtering, In *Proceedings of the 2001 Workshop on Recommender Systems*, New Orleans, LA, September 2001.
- [Miller et al., 1997] D. Miller, J. L. Maltz, L. R. Herlocker, A. Gordan, J. A. Riedl, B. N. Konstan. GroupLens: applying collaborative filtering to Usenet News. *Communications of the ACM*, vol. 40, n° 3, p. 77-87, mars 1997.

- [**Minio et Tasso, 1996**] M. Minio et C. Tasso. User Modeling for Information Filtering on Internet Services: Exploiting an Extended Version of the UMT Shell. In *UM96 Workshop on User Modeling for Information Filtering on the WWW*. Kailua-Kona, Hawaii, 1996.
- [**Minsky, 1975**] M. Minsky. A Framework for Representing Knowledge, in P. Henry Winston, editor, *The Psychology of Computer vision*, McGraw- Hill, pages 211-277. New York, USA, 1975.
- [**Mizoguchi, 1998**] R. Mizoguchi. A step Towards Ontological Engineering. Translation of the paper presented at the 12th National Conference on AI of JSAI, pp.24-31, June 1998.
<http://www.ei.sanken.osaka-u.ac.jp/english/step-onteng.html>
- [**Mizoguchi et al., 2000**] R. Mizoguchi, K. Kozaki, T. Sano et Y. Kitamura. Construction and Deployment of a Plant Ontology. The 12th International Conference, EKAW2000, (Lecture Notes in Artificial Intelligence 1937), 113-128, 2000.
- [**Mladenic, 1996**] D. Mladenic. Personal WebWatcher: Implementation and Design. *Technical Report IJS-DP-7472, Department of Intelligent Systems*. Slovenia: J. Stefan Institute, 1996.
- [**Mobasher et al., 2000**] B. Mobasher, R. Cooley et J. Srivastava. Automatic Personalization Based on Web Usage Mining, *Communications of the ACM*, 43(8), pp. 142-151, August 2000.
- [**Moukas, 1997**] A. Moukas. Amalthea: Information Discovery and Filtering using a Multiagent Evolving Ecosystem', in *Applied Artificial Intelligence: An International Journal*, 11(5), pp. 437-457, 1997.
- [**Napoli, 1997**] A. Napoli. Une introduction aux logiques de description. Rapport de recherche 3314, INRIA Lorraine, Nancy, France, 1997.
- [**Nebel, 1990**] B. Nebel. Reasoning and Revision in Hybrid Representation Systems, V(422). Springer-Verlag, Lecture notes in Artificial Intelligence edition, Juin 90. PhD thesis, Universitat des Saarlandes, Saarbrucken, West Germany, 1990.
- [**Noy at Hafner, 1997**] N.F. Noy et C.D. Hafner. The State of the Art in Ontology Design: A Survey and Comparative Review. *AI Magazine*, 18(3): 53-74, 1997.
- [**Noy et al., 2000**] N.F. Noy, R.W. Ferguson et M.A. Musen. The knowledge model of Protege-2000: Combining interoperability and flexibility. Paper presented at the 2th International Conference on Knowledge Engineering and Knowledge Management (EKAW'2000), volume (1937) of Lecture Notes in Computer Science, p. 17– 32, Juan-les-Pins, France: Springer Verlag, 2000.
- [**OWL, 2004**] OWL. Web Ontology Language Reference. W3C Recommendation, 2004.
<http://www.w3c.org/TR/owl-ref/>.

- [Pazzani et al., 1996] M. Pazzani, J. Muramatsu et D. Billsus. Syskill & Webert : Identifying Interesting Web Sites. In Proceedings of the National Conference on Artificial Intelligence, pp. 54-61, Portland, OR, 1996.
- [Petel-Schneider et al., 2002] P. F. Petel-Schneider, I. Horrocks et F. Van Harmelen. *OWL Web Ontology Language 1.0 Abstract Syntax*, W3C working draft 29 July 2002; <http://www.w3.org/TR/2002/WD-owl-absyn-20020729/>
- [Porter, 1980] M. F. Porter. An Algorithm For Suffix Stripping. *Program*, vol 14, n° 3, p. 130-137. 1980.
- [Potter et Trueblood, 1998] G. Potter et R. Trueblood. Traditional, Semantic, and Hyper-Semantic Approaches to Data Modeling. *IEEE Computer* **21**(6): 53–63, 1998.
- [Powell, 1998] A. Powell. Dublin core in RDF. Technical report, 1998.
- [Prat, 1997] N. Prat. Une approche linguistique pour la formalisation et la classification des buts en ingénierie des processus, 1st International Workshop on the Many Facets of Process Engineering (MFPE'97), Gammarth, Tunisie, Septembre 1997.
- [Pretschner et Gauch, 1999] A. Pretschner et S. Gauch. Ontology Based Personalized Search, in Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence, pp. 391-398, Chicago, November 1999.
- [Py, 1998] D. Py. Quelques méthodes d'intelligence artificielle pour la modélisation de l'élève. Dans Modélisation de l'Apprenant, revue Science et Techniques Educatives Vol. 5, N°2, Ed. Hermès, pp. 123-140, juin, 1998.
- [Quillian, 1968] M.R. Quillian. Semantic memory. In Semantic Information Processing, pages 227-270. MIT Press, 1968.
- [Ranwez, 2000] Sylvie Chabert-Ranwez, Composition Automatique de Documents Hypermédia Adaptatifs à partir d'Ontologies et de Requêtes Intentionnelles de l'Utilisateur. Thèse Pour obtenir le grade de Docteur de l'Université Montpellier II Soutenue le 21 décembre 2000
- [Resnick et al., 1994] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom et J. Riedl. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In *Proceedings of ACM CSCW'94*, 175–186, 1994.
- [RDF, 2004] RDF. Resource Description Framework Primer. W3C Recommendation, 2004. <http://www.w3.org/TR/rdf-primer/> .
- [RDFS, 2004] RDFS. RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation, 2004. <http://www.w3.org/TR/rdf-schema/> .

- [Rich, 1989] E. Rich. Stereotypes and User Modeling, chapitre 2, pages 35-51. Springer Verlag, 1989.
- [Rich, 1999] E. Rich. Users are individuals: individualizing user models, *Int Journal on Human-Computer Studies*, 51, (Ré-édition d'un article de 1981), pp. 323-338, 1999.
- [Riordan et Sorensen, 1995] A. Riordan et H. Sorensen. An Intelligent Agent for High-Precision Information Filtering. In *Proceedings of the CIKM-95 Conference, 1995*.
- [Roberston et Spark-Jones, 1976] Roberston S.E, Spark-Jones K., « Relevance Weighting of Search Terms », *Journal of the American Society of Information Science*, 27, 1976, p.129-146.
- [Rossi et al., 2001] G. Rossi, D. Schwabe et R. Guimarães. Designing Personalized Web Applications, 10th International World Wide Web Conference (WWW10), May 1-5, Hong Kong, 2001.
- [Salton, 1971] G. Salton. The SMART Retrieval System – Experiments in Automatic Document Processing, Prentice Hall, Inc., Englewood Cliffs, NJ, 1971.
- [Salton et Buckley, 1988] G. Salton, C. Buckley. Term weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5), 1988, p. 513-523.
- [Schwab et al., 2001] I. Schwab, A. Kobsa et I. Koychev. Learning User's Interests Through Positive Examples Using Content Analysis and Collaborative Filtering. *Submitted*, 2001.
- [Seo et Zhang, 2000] Y.W. Seo, B.T. Zhang. A Reinforcement Learning Agent for Personalized Information Filtering, *Proceedings of the 2000 International Conference on Intelligent User Interfaces*, New-Orleans, USA, ACM, pp248-251, Jan 9-12, 2000.
- [Shearin et Lieberman, 2001] S. Shearin, H. Lieberman, *Intelligent Profiling by Example*, MIT Media Lab, Cambridge, 2001.
- [Sheth et Maes, 1993] B. Sheth et P. Maes. Evolving Agents for Personalized Information Filtering. In *Proceedings of the Ninth Conference on Artificial Intelligence for Applications*. IEEE Computer Society Press. 1993.
- [Sintek et Decker, 2002] M. Sintek et S. Decker. TRIPLE - A Query, Inference and transformation Language for the Semantic Web. In I. Horrocks and J. Hendler, editors, first international Semantic Web Conference (ISWC'02), volume (2342) of *Lecture Notes in Computer Science*, pages 364-378, Chia, Sardaigne, Italie, 9-12 Juin 2002. springer Verlag.
- [Smith et Welty, 2001] B. Smith et C. Welty. Ontology: Towards a new synthesis. In N. Guarino, editor, *Formal Ontology in Information Systems*, *Proceedings of the 2nd International Conference (FOIS-01)*, Ogunquit, Maine, 17-19 October 2001. ACM.

- [Soltysiak et Crabtree, 1998] S. J. Soltysiak, I. B. Crabtree., Automatic learning of user profiles-towards the personalisation of agent services. *BT Technol J.* Vol 16 No 3, pp 110-117, July 1998.
- [Sorensen et McElligott, 1995] H. Sorensen, M. Mc Elligott, PSUN : A Profiling System for Usenet NewsCIKM'95 Intelligent Information Agents Workshop, Baltimore, December 1995.
- [Sorensen et al., 1997] H. Sorensen, A. O. Riordan et C. O. Riordan. Profiling with the INFORMER Text Filtering Agent. *Journal of Universal Computer Science* 3(8): 988–1006, 1997.
- [Sowa, 1984] J. F. Sowa.. *Conceptual Structures : Information Processing in Mind and Machine.* Addison-Wesley, Massachusset, USA, 1984.
- [Stefani et Strappavara, 1998] A. Stefani et C. Strappavara. Personalizing Access to Web Wites: The SiteIF Project. In *Proceedings of HYPERTEXT'98.*
- [Studer et al., 1998] Studer R., Benjamins R. et Fensel D. (1998). *Knowledge Engineering: Principles and Methods.* Data Knowledge Engineering.
- [Sure et al., 2002] Y. Sure, M. Erdmann, J. Angele, S. Staab, R. Studer et D. Wenke . *OntoEdit: Collaborative Ontology Engineering for the Semantic Web.* In I. Horrocks et J. Hendler, Eds., *First International Semantic Web Conference (ISWC'02)*, volume (2342) of *Lecture Notes in Computer Science*, p. 221–235, Chia, Sardaigne, Italie: Springer Verlag, 2002.
- [Swartout et al., 1997] B. Swartout, R. Patil, K. Knight et T. Russ. *Towards Distributed Use of Large-Scale Ontologies.* *Spring Symposium Series on Ontological Engineering*, pp.138-148, 1997.
- [Troncy et Isaac, 2002] R. Troncy et A. Isaac. *DOE : une mise en œuvre d'une méthode de structuration différentielle pour les ontologies.* In *13es Journées Francophones d'Ingénierie des Connaissances (IC'02)* , p. 63–74, Rouen, France, 2002.
- [Trousse, 2001] B. Trousse. *Recommandations personnalisées pour l'aide à la recherche d'informations Web basées sur l'analyse et l'utilisation du comportement des utilisateurs,* ActionAxis, INRIA Sophia Antipolis, 2001
- [Uschold et King, 1995] M. Uschold et M. King. *Towards a methodology for Building Ontologies.* Presented at the *Workshop on Basic Ontological Issues in Knowledge Sharing*, International Joint Conference on A.I. Montreal, 1995.
<http://www.aii.ed.ac.uk/aii/pubtrs95.html>
- [Villa et Chalmers, 2001] R. Villa, M. Chalmers, *A Framework for implicitly tracking data,* DELOS Workshop: *Personalisation and Recommender Systems in Digital Libraries* 2001.

[Villanova, 2002] M. Villanova-Oliver. Adaptabilité dans les systèmes d'information sur le Web : Modélisation et mise en œuvre de l'accès progressif. Thèse de doctorat, institut national polytechnique de Grenoble, 2002.

[Yan et Garcia-Molina, 1995] T. W. Yan et H. Garcia-Molina. Sift – a Tool for Wide-Area Information Dissemination. In *Proceedings of the 1195 USENIX Technical Conference*, 177– 186, 1995.

[Zweigenbaum, 1999] ZWEIGENBAUM P., Encoder l'information médicale : des terminologies aux systèmes de représentation des connaissances, in *Innovation stratégique en information de santé (ISIS)* (2-3), pages 27-47, 1999.

Résumé

Nos recherches se résument à la tentative constante de simplification automatique de la vie de l'utilisateur et de la réduction de l'effort cognitif fourni par celui-ci devant une activité de recherche ou d'exploration d'un domaine. Le besoin de personnalisation semble donc évident si nous envisageons dans le futur des machines capables de remplacer les êtres humains dans certaines activités. Pour cela, le système doit avoir une représentation suffisamment étoffée de l'utilisateur.

L'ontologie peut jouer un rôle important pour la représentation et la gestion des connaissances relatives aux utilisateurs. Elle fournit un cadre unificateur pour réduire et éliminer les confusions conceptuelles et terminologiques et assurer une compréhension partagée par la communauté visée. Elle peut donc permettre la communication entre des utilisateurs, même s'ils travaillent dans des contextes différents, avec des besoins et des points de vue différents.

Notre approche préconise un modèle utilisateur présenté par une ontologie qui est en relation avec d'autres ontologies : ontologie des documents, ontologie du domaine et ontologie d'application, dans le cadre applicatif des laboratoires de recherche.

Ces quatre types d'ontologies sont fédérés sous une seule et même structure. L'ontologie des utilisateurs agit en quelque sorte d'interface entre les utilisateurs et les autres ontologies ; l'ontologie des documents permet de structurer et d'exploiter les documents ; l'ontologie du domaine sur laquelle est basée la définition des centres d'intérêts des utilisateurs, leurs niveaux de connaissance, leurs domaines de recherche, etc. et enfin l'ontologie d'application comporte des informations relatives à la gestion de la personnalisation, comme la liste des stratégies de personnalisation avec les paramètres nécessaires à leur fonctionnement. Il s'agit donc de représentations séparées, mais néanmoins interconnectées.