



HAL
open science

Propositions de méthodologies pour la valorisation de la médecine traditionnelle fondées sur une ontologie

Behou Gérard N'Guessan

► To cite this version:

Behou Gérard N'Guessan. Propositions de méthodologies pour la valorisation de la médecine traditionnelle fondées sur une ontologie. Informatique [cs]. UNIVERSITE NANGUI ABROGOUA, 2016. Français. NNT: . tel-01376966

HAL Id: tel-01376966

<https://hal.science/tel-01376966>

Submitted on 6 Oct 2016

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

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



Unité de Formation et de Recherche des
Sciences Fondamentales et Appliquées
(UFR-SFA)

N° d'ordre : 293

THESE

Pour obtenir le grade de

Docteur de l'Université Nangui Abrogoua

Mention : Informatique

Spécialité : Génie Informatique

Présentée par :

N'GUESSAN Behou Gérard

Sujet :

Propositions de méthodologies pour la valorisation de la médecine traditionnelle fondées sur une ontologie.

Soutenue le 29/02/2016 devant le Jury composé de :

<i>M. OUMTANAGA Souleymane</i>	<i>Professeur Titulaire</i>	<i>INP-HB, Yamoussoukro - Côte d'Ivoire</i>	<i>Président</i>
<i>M. AKA Boko</i>	<i>Professeur Titulaire</i>	<i>Université Nangui Abrogoua, Abidjan - Côte d'Ivoire</i>	<i>Co-Directeur</i>
<i>M. BABRI Michel</i>	<i>Maître de Conférences</i>	<i>INP-HB, Yamoussoukro - Côte d'Ivoire</i>	<i>Co-Directeur</i>
<i>M. LISHOU Claude</i>	<i>Professeur Titulaire</i>	<i>Ecole Supérieure Polytechnique de Dakar-Sénégal</i>	<i>Rapporteur</i>
<i>M. LOUM Georges Laussane</i>	<i>Maître de Conférences</i>	<i>INP-HB, Yamoussoukro - Côte d'Ivoire</i>	<i>Rapporteur</i>
<i>M. KONE Tiémoman</i>	<i>Maître de Recherches</i>	<i>Université Félix Houphouët Boigny, Abidjan-Côte d'Ivoire</i>	<i>Examineur</i>

A mes Parents

Remerciements

Nos premiers remerciements vont à l'endroit du TOUT PUISSANT pour tout ce qu'il fait pour nous. Que son nom soit béni.

Qu'il nous plaise de remercier très sincèrement tous les membres du jury pour l'honneur qu'ils nous font de juger notre modeste travail ; nous pensons en particulier à nos maîtres :

M. OUMTANAGA Souleymane, Professeur Titulaire à l'INP-HB de Yamoussoukro ;

M. BABRI Michel, Maitre de Conférences à l'INP-HB de Yamoussoukro ;

M. LOUM Georges, Maitre de Conférences à l'INP-HB de Yamoussoukro ;

M. BROU Marcellin, Maitre-Assistant à l'INP-HB de Yamoussoukro ;

pour avoir bien voulu accepté d'encadrer nos travaux de recherche dans leur laboratoire. Ils ont été, pour nous, à la fois des éducateurs émérites et des guides rigoureux mais bienveillants. Ils ont accepté de nous suivre dans cette recherche ; leurs disponibilités et leurs conseils ont efficacement relevé le niveau scientifique de cette étude. Tout au long de ce travail de recherche, il nous a été donné de noter leurs qualités humaines de tolérance et de compréhension, conséquences d'une solide expérience scientifique. Nous ne cesserons de les remercier pour toute la confiance qu'ils nous ont témoignée et des efforts qu'ils ont faits pour l'aboutissement de cette thèse.

Nous avons eu également la très grande chance de travailler avec M. AKA Boko, Professeur Titulaire à l'Université Nangui Abrogoua, notre Directeur de thèse. Ses conseils et sa promptitude à faire face à nos préoccupations nous ont été d'une très grande utilité ; nous saluons ici tous les efforts qu'il fait pour rendre vivante la coopération entre l'INP-HB de Yamoussoukro et l'Université Nangui Abrogoua d'Abidjan.

Nous remercions également tous ceux qui se sont donnés la peine de relire notre travail et de qui nous avons pu bénéficier de pertinentes remarques et conseils techniques, académiques et professionnels; il s'agit de :

- Docteur N'GUESSAN Kotchi Rémi, enseignant à l'INPHB
- Docteur GOORE Bi Tra, enseignant à l'INP-HB
- Docteur TANOHI Lambert KADJO, enseignant à l'INPHB
- Docteur KIMOU Prosper, enseignant à l'INP-HB
- Docteur TRAORE Issa, enseignant chercheur à l'Université Félix Houphouët Boigny
- M. ANDERSON Robert, sociologue à la retraite.

Nos remerciements vont également à l'endroit de toutes ces personnes qui, d'une manière ou d'une autre, nous ont encouragé ou aidé à parfaire notre travail. Nous pensons notamment à M. ATIAMPO Kodjo Armand et M. ACHIEPO Yapo Odilon pour leur appui.

Nous remercions tous les membres du Laboratoire de Recherche en Informatique et Télécommunications (LARIT), du Laboratoire des Techniques de l'Information et de la Communication (LABTIC) et du Laboratoire d'Energie Solaire et Nanotechnologie (LESN).

Nous tenons enfin à manifester notre profonde gratitude à l'endroit des membres de notre famille pour le soutien moral qu'ils n'ont cessé de nous apporter tout au long de ces mois de recherches ; nous pensons en particulier à notre mère biologique Ayé Bakou Jacqueline, à notre mère nourricière Ovo Adèle et à notre compagne de tous les jours Sidonie Yapi qui nous ont été d'un secours inestimable dans la poursuite et l'achèvement de ce travail.

Résumé : Le travail présenté dans cette thèse porte sur la problématique de la valorisation de la médecine traditionnelle. La médecine traditionnelle est d'une diversité biologique et culturelle très riche. Sa pratique est très répandue et se fait sous des formes variées. La valorisation de cette médecine constitue aujourd'hui un enjeu très important ; elle permettra de capitaliser ce savoir, de le vulgariser, et donc d'améliorer ses prestations en termes de diagnostic, de traitement et de coût. L'Organisation Mondiale de la Santé (OMS) propose, du reste, de l'intégrer dans le système national de santé. Mais la pratique et l'exercice de cette médecine rencontrent de nombreux problèmes qui rendent son application difficile. Parmi ces problèmes, nous pouvons relever le caractère informel de sa pratique, son contenu non formalisé, son mode d'accès non déterminé, etc. Nous proposons, dans cette thèse, des techniques de pratiques plus efficaces puisque basées sur les nouvelles technologies de l'information et de la communication ; celles-ci reposent plus spécifiquement sur la ressource sémantique telle que l'ontologie qui est la structuration formelle d'une connaissance. Les méthodes proposées dans cette thèse permettent de formaliser le contenu de cette médecine, pour en faciliter l'exercice et en définitive, pour aboutir à sa revalorisation.

Mots-clés : Médecine traditionnelle, ontologie visuelle, conflit, apprentissage, Réseaux Bayésiens, Raisonnement médical, Raisonnement à Partir de Cas, Mesures de similarité, Diagnostic médical.

Abstract: The work presented in this thesis focuses on the problematic of the valorization of traditional medicine. Traditional medicine is a very rich biological and cultural diversity. His practise is widespread and occurs in various forms. The valorization of this medicine is now a very important issue; it will capitalize this knowledge to popularize, and thus improve its performance in terms of diagnosis, treatment and cost. World Health Organization (WHO) proposes its integration into the national health system. But the practice and exercise of this medicine face many problems which make its implementation difficult. Among these problems, we can list the informal nature of its practice, its content is not formalized, its access mode is not determined, etc. We propose in this thesis, practices more efficient based on the new technology of information and communications; they based specifically on semantic resource such as the ontology which is the formal structure of an acquaintance. The methods proposed in this thesis allow to formalize the contents of this medicine, to facilitate its exercise and ultimately to succeed its revalorization.

Keywords: Traditional Medicine, visual ontology conflict, learning, Bayesian Networks, Medical Reasoning, Case-Based Reasoning, similarity measures, medical diagnosis.

Table des matières

Liste des Figures	xii
Liste des abréviations	xiv
Liste des publications	xvi
INTRODUCTION GENERALE.....	1
I. Position du problème.....	1
II. Objectif de la recherche	2
III. Contribution de la thèse.....	3
IV. Organisation du mémoire	3
Chapitre 1.....	5
MEDECINE TRADITIONNELLE	5
1.1. Fondements historiques de la médecine traditionnelle.....	5
1.2. Concepts de la médecine traditionnelle africaine	6
1.2.1. Guérisseur.....	6
1.2.2. Féticheur	6
1.2.3. Herboriste	6
1.2.4. Réducteur de fractures.....	7
1.3. Plantes médicinales	7
1.4. Autres éléments de la médecine traditionnelle (minéraux, parties d’animaux et pratiques magico-spirituelles).....	7
1.5. Quelques découvertes de plantes dans la médecine traditionnelle	7
1.6. Prévention dans la médecine traditionnelle	8
1.7. Diagnostic dans la médecine traditionnelle	8
1.8. Travaux existants.....	8
1.9. Logique de description	10
1.9.1. TBox (Terminological Box)	11
1.9.2. Assertional Box (ABox)	11
1.10. Connaissances	11
1.11. Base de connaissances.....	12
1.12. Connaissances empiriques.....	12
1.13. Connaissances théoriques.....	12
1.14. Moteur d’inférence	13
1.15. Langage naturel	13
1.15.1. Système de compréhension du langage naturel.....	14

1.15.2.	Langages de communication entre agents	14
1.16.	Concept.....	14
1.17.	Base de règles.....	15
1.18.	Conclusion.....	16
Chapitre 2.....	17	
WEB SEMANTIQUE	17	
2.1.	Introduction	17
2.2.	Web	18
2.3.	Outils de création de données sur le web	18
2.4.	Web et la médecine	19
2.5.	Représentation des connaissances	20
2.6.	Avantages du web sémantique	20
2.7.	Quelques inconvénients.....	21
2.8.	Langages du web sémantique.....	23
2.8.1.	Langage XML	23
2.8.2.	Langage RDF (Resource Description Framework)	24
2.8.3.	Langage RDF-Schéma.....	25
2.8.4.	Langage OWL (Web Ontology Language)	26
2.9.	Différentes couches du web sémantique	27
2.9.1.	URI (Universal Ressource Identifier).....	27
2.9.2.	XML	28
2.9.3.	RDF et le RDF SCHEMA	29
2.9.4.	XML-Schéma	30
2.9.5.	Ontologies	30
2.9.5.1.	Thésaurus.....	32
2.9.5.2.	Graphes conceptuels.....	32
2.9.5.3.	Logiques de description.....	33
2.9.5.4.	Ontologie et web-sémantique	34
2.9.5.5.	Composantes d'une ontologie	35
2.9.5.6.	Logique.....	36
2.9.6.	Preuve.....	37
2.9.7.	Couche Confiance	37
2.9.8.	Langage de requête SPARQL	38
2.9.9.	Inter-opérabilité d'ontologie (Co-ontologie).....	39

2.9.10.	Conclusion.....	39
Chapitre 3.....		41
CONSTRUCTION D'UNE ONTOLOGIE.....		41
3.1.	Introduction	41
3.2.	Origine de la construction d'ontologie.....	42
3.3.	Ontologies biomédicales	42
3.4.	Domaine visuel.....	43
3.5.	Etapes de la construction d'ontologie.....	43
3.5.1.	Acquisition de connaissances	44
3.5.2.	Evaluation.....	45
3.5.3.	Intégration	46
3.5.4.	Formalisation.....	46
3.5.5.	Spécification.....	46
3.6.	Méthodologie de conception d'une ontologie.....	47
3.7.	Classification des méthodes de construction d'ontologies	48
3.8.	Méthode d'élicitation	48
3.9.	Création d'un Domaine	49
3.10.	Construction d'ontologie.....	50
3.11.	Algorithme de diagnostic d'une maladie.....	52
3.11.1.	Fonctionnement de l'algorithme proposé.....	52
3.11.2.	Extensions Notationnelles	53
3.12.	Conclusion.....	54
Chapitre 4.....		55
APPRENTISSAGE A PARTIR D'UNE ONTOLOGIE.....		55
4.1.	Introduction	55
4.2.	Définition de l'Apprentissage.....	56
4.3.	Etapas historiques de l'apprentissage.....	56
4.4.	Conditions d'apprentissage	57
4.5.	Travaux existants.....	58
4.6.	Enseignement Assisté par Ordinateur (EAO).....	60
4.7.	Enseignement Intelligemment Assisté par Ordinateur (EIAO).....	61
4.8.	Module tutorial	63
4.9.	Apprentissage classique de la MT	63
4.10.	Apprentissage organisationnel.....	64

4.11.	Algorithme d'apprentissage.....	65
4.12.	Taxonomies des connaissances	65
4.13.	Dimension ontologique	66
4.13.1.	Modes de transformation des connaissances	66
4.13.2.	Modes d'apprentissage	67
4.14.	Ontologie et EIAO.....	68
4.15.	Architecture du modèle EIAO.....	69
4.15.1.	Description de l'architecture	71
4.15.2.	Description des flux.....	71
4.16.	Conclusion.....	72
	Chapitre 5.....	74
	PROCESSUS DE RECHERCHE D'INFORMATIONS DANS UNE BASE MEDICALE	74
5.1.	Introduction	74
5.2.	Recherche d'information (RI)	75
5.3.	Indexation.....	76
5.3.1.	Indexation par une liste de descripteurs	76
5.3.1.1.	Indexation libre.....	76
5.3.1.2.	Indexation par un langage documentaire.....	77
5.3.2.	Indexation dynamique d'ontologie.....	77
5.3.3.	Indexation sémantique d'ontologie	78
5.4.	Arbre conceptuel	79
5.5.	Concepts de base de la recherche d'information	79
5.6.	Recherche sémantique	80
5.7.	Moteurs de recherche	80
5.8.	Requêtes	80
5.9.	Outil de recherche sémantique	82
5.10.	Comparaison SQL / SPARQL.....	82
5.10.1.	Modèle.....	83
5.10.2.	Quelques aspects de l'outil.....	83
5.11.	Conclusion.....	84
	Chapitre 6.....	85
	RAISONNEMENT MEDICAL : APPLICATION DES RESEAUX BAYESIENS AU DIAGNOSTIC D'UNE MALADIE DANS LA MEDECINE TRADITIONNELLE.	85
6.1.	Introduction	85

6.2.	Raisonnement Bayésien.....	86
6.3.	Ontologie.....	86
6.4.	Réseaux Bayésiens	87
6.5.	Modélisation.....	88
6.5.1.	Structure du Réseau.....	89
6.5.2.	Tables de probabilités.....	90
6.5.3.	Factorisation des probabilités	91
6.6.	Algorithme de la méthode	92
6.7.	Exploitation du modèle	94
6.8.	Comparaison au niveau des Réseaux bayésiens	95
6.9.	Conclusion.....	96
	Chapitre 7.....	98
	GESTION DE CONFLITS DANS L'EXERCICE DE LA MEDECINE TRADITIONNELLE A BASE D'ONTOLOGIE MEDICALE.....	98
7.1.	Introduction	98
7.2.	Ontologie médicale : exemple de cas de conflit.....	99
7.3.	Modélisation.....	100
7.3.1.	Filtrage collaboratif	100
7.3.1.1.	Hypothèses du problème	100
7.3.1.2.	Détection du conflit	102
7.3.1.3.	Gestion des conflits	102
7.3.1.4.	Traitement des données manquantes	103
7.3.1.5.	Classification de la nouvelle donnée	105
7.4.	Illustration du modèle.....	106
7.5.	Comparaison au niveau de la gestion de conflit.....	106
7.6.	Conclusion.....	106
	Chapitre 8.....	107
	RAISONNEMENT MEDICAL : APPLICATION DU RAISONNEMENT A PARTIR DE CAS EN MEDECINE TRADITIONNELLE.....	107
8.1.	Introduction	107
8.2.	Etat des lieux	108
8.2.1.	Approche du raisonnement à partir de cas.....	108
8.2.2.	Principes de la mesure de similarité	110
8.3.	Modélisation de la similarité du diagnostic.....	111

8.4.	Différentes fonctions de la mesure de similarité	112
8.4.1.	Fonction α	112
8.4.2.	Fonction β	113
8.4.3.	Fonction γ	114
8.5.	Mesure de similarité globale	115
8.5.1.	Fonction Ψ	115
8.5.2.	Application générale de la mesure	115
8.6.	Simulation	116
8.7.	Comparaison au niveau de la mesure de similarité	118
8.8.	Conclusion.....	118
	Chapitre 9	120
	METHODOLOGIE D'APPROCHE VIA LE LANGAGE SPARQL	120
9.1.	Introduction	120
9.2.	Architecture Globale du Système Proposé	121
9.2.1.	Module de gestion	122
9.2.1.1.	Gestionnaire de la BC.....	122
9.2.1.2.	Module d'interrogation.....	122
9.2.1.3.	Gestionnaire de requêtes SPARQL	122
9.2.1.4.	Moteur SPARQL.....	123
9.3.	Accès aux données contenues dans l'entrepôt sémantique	123
9.4.	Interface graphique de recherche d'informations	123
9.5.	Ontologie du paludisme.....	124
9.6.	Résultats d'exécution	125
9.7.	Evaluation des différents modèles.....	126
9.8.	Evaluation de notre approche	126
9.9.	Conclusion.....	127
	CONCLUSIONS ET PERSPECTIVES	130
	Bibliographie	132

Liste des Figures

Figure 2.1 : Chronologie : avantages et inconvénients du web dans le temps	22
Figure 2.2 : Exemple de document XLM représenté sous forme textuelle	24
Figure 2.3 : Exemple de graphe RDF	25
Figure 2.4 : Document RDF simple	25
Figure 2.5 : Exemple de Description du modèle OWL	27
Figure 2.6 : Architecture du web sémantique.....	27
Figure 2.7 : Relation entre le SGML et XML.	29
Figure 2.8 : Description de concept à partir d'une ontologie	31
Figure 2.9 : Exemple d'une TBox.....	34
Figure 2.10 : Exemple d'une ABox.....	34
Figure 2.11 : Exemple de connexion.....	36
Figure 2.12 : Exemple de requête SPARQL	39
Figure 3 1 : Le cycle de vie d'une ontologie	44
Figure 3 2 : Exemple de représentation très informelle.....	49
Figure 3 3 : Exemple d'une représentation semi-informelle.....	49
Figure 3 4 : Exemple de représentation semi-formelle.....	50
Figure 3 5 : Exemple de représentation formelle d'une ontologie	50
Figure 3 6 : Architecture de concepts d'une ontologie.....	51
Figure 3 7 : Algorithme de diagnostic d'une maladie	52
Figure 4.1 : Principaux courants en informatique et éducation	57
Figure 4.2 : Composantes d'un système d'EIAO	62
Figure 4.3 : La hiérarchie des connaissances	66
Figure 4.4 : Architecture d'apprentissage de la médecine traditionnelle par le mécanisme d'EIAO	70
Figure 5.1 : Architecture du moteur de recherche Google	78
Figure 5.2 : Architecture d'un moteur de recherche.....	81
Figure 5.3 : Système d'interaction du SPARQL	83
Figure 5.4 : Requête SPARQL retournant les noms de plantes présentes dans un fichier RDF.	84
Figure 6.1 : Ontologie du paludisme	87
Figure 6.2 : Exemple de Réseaux Bayésiens evenement.....	88
Figure 6.3 : Structure du Réseau Bayésien.....	89
Figure 6.4 : Diagnostic et traitement du paludisme.....	94
Figure 7.1 : Ontologie de la maladie_Z et de la maladie_P	99
Figure 8 1 : Architecture d'un système de RàPC	109
Figure 8 2 : Exemple d'adaptation de cas	110
Figure 8 3 : Classification hiérarchique avec notre métrique	117
Figure 8 4 : Positionnement multidimensionnel avec notre métrique	118
Figure 9.1 : Architecture globale du système.....	121
Figure 9.2 : Requête du langage SPARQL.....	123
Figure 9.3 : Interface d'utilisation.....	124
Figure 9.4 : Chargement d'une base ontologique.....	124
Figure 9.5 : Ontologie du paludisme	125
Figure 9.6 : Exécution de la requête.....	125
Figure 9.7 : Le résultat de l'exécution de la requête SPARQL	126

Liste des Tableaux

Tableau 2.1 : Sémantique des concepts et des rôles en logique de description.....	33
Tableau 4.1 : Exemple d'éléments contenus dans : les bases de faits, les bases de règles et les bases de traitements.....	71
Tableau 4.2 : Description des flux.....	71
Tableau 5.1 : Comparaison des langages SQL et SPARQL.....	82
Tableau 7.1 : Présentation de cas de conflit	100
Tableau 7.2 : Exemple de matrice Y représentant le degré d'importance des symptômes dans les maladies de notre base d'apprentissage.....	104

Liste des Annexes

Annexe A.1 : Présentation du code source de l'implémentation.....	I
Annexe A.2 : Le code de la vérification des symptômes	I
Annexe A.3 : Les valeurs attendues pour les données manquantes	II
Annexe A.4 : Indication de la maladie	II
Annexe C.1 : Algorithme de la fonction alpha.....	V
Annexe C.2 : Algorithme de la fonction beta.....	V
Annexe C.3 : Algorithme de la fonction gamma.....	VI
Annexe C.4 : Algorithme de la fonction phi	VII

Liste des abréviations

AAC :	Augmentative and Alternative Communication
ABox :	Assertional Box
APO :	Applications Pédagogiques de l'Ordinateur
ATM :	Médecine Traditionnelle Africaine
BC :	Base de Connaissance
CaBIG :	Cancer du Biomedical Informatics Grid
CERN :	Centre Européen de Recherche Nucléaire
CISMef :	Catalogue et Index des Sites Médicaux de langue Française
DAML :	DARPA Agent Markup Language
DARPA :	Defense Advanced Research Projects Agency
DTD :	Document Type Définition
EAO :	Enseignement Assisté par Ordinateur
EIAH :	Environnements Informatiques pour l'Apprentissage Humain
EIAO :	Enseignements Intelligemment Assistés par Ordinateur
FC :	Filtrage Collaboratif
GLS :	Grid Learning Services
HTML :	Hypertext Markup Language
HTTP :	Hypertext Transfer Protocol
IA :	Intelligence Artificielle
IBS :	Intelligent Building Systems
ICN :	Institut du Cancer National
LD :	Logique de Descriptions
LOM :	Learning Object Metadata
MEDIGATE :	Médical Examination Direct Iconic and Graphic Augmented Text Entry
MeSH :	Medical Subject Headings
MOT :	Modèle-Organisationnel-des-Traitements
MT :	Médecine Traditionnelle
NTIC :	Nouvelles Technologies de l'Information et de la Communication
OGSA :	Open Grid Services Architecture
OIL :	Ontology Inference Layer
OMS :	Organisation Mondiale de la Santé
OWL :	Web Ontology Language
OWL-DL :	Ontology Web Language Description Logic
PC :	Personnal Computer
PHP :	Hypertext Preprocessor
RàPC :	Raisonnement à Partir de Cas
RDF :	Resource Description Framework
RDFS :	Resource Description Framework Schéma
RPC :	Remote Procedure Call
SAM :	Système Multi-Agents
SBC :	Système à Base de Connaissance
SGML :	Standard Generalized Markup Language

SQL :	Structured Query Language
STI :	Systèmes Tutoriels Intelligents
TBox :	Terminological Box
TCP\IP :	Transmission Control Protocol\Internet Protocol
TELEOS :	Technology Enhanced Learning Environment for Orthopaedic Surgery
TIC :	Technologies de l'Information et de la Communication
UDDI :	Universal Description Discovery and Integration
UDP :	User Datagram Protocol
UML :	Unified Modeling Language
URI :	Universal Ressource Identifier
URL :	Uniform Resource Locator
URN :	Uniform Resource Name
UTILIS :	Updating Through Interaction in Logical Information Systems
VAL :	Virtual Access to the Lexicon
VCM :	Visualisation des Connaissances Médicales
WWW :	World Wide Web
XML :	eXtensible Markup Language

Liste des publications

Durant cette thèse, notre travail a porté sur la valorisation de la médecine traditionnelle. Nous avons étudié la gestion des conflits lors de l'apprentissage et la pratique de la médecine traditionnelle, l'indice de similarité pour le traitement d'un nouveau cas, l'utilisation des réseaux bayésiens en cas d'incertitude lors de la pratique de la médecine traditionnelle et une autre méthode via le langage de requête SPARQL. Certains de ces travaux ont fait l'objet de publication dans des journaux.

Articles publiés dans les journaux internationaux

[Gues 15a] Behou Gérard NGUESSAN, Konan Marcellin BROU, Odilon Yapo M. ACHIEPO, Michel BABRI et Boko AKA. «Raisonnement médical : Application des Réseaux Bayésiens au diagnostic d'une maladie dans la médecine traditionnelle». *Afrique Science*, Vol.11, N°5 (2015), 1 septembre 2015, <http://www.afriquescience.info/document.php?id=5159>. ISSN 1813-548X

[Gues 15b] G. B. N'GUESSAN, K. M. BROU, M. BABRI and S. OUMTANAGA “ Effective Management of Conflict in Learning the Traditional Medicine-Based Medical Ontology ”, *IJCSI International Journal of Computer Science Issues*, Volume 12, Issue 2, March 2015 ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784

[Gues 15c] O Y M. ACHIEPO, B. G. N'GUESSAN, K. M. BROU, “Similarity Measure in the Case-Based Reasoning Systems for Medical Diagnostics in Traditional Medicine,” *International Journal of Computer Science and Telecommunications* Volume 6, Issue 4, April 2015, Pages: 1-5

INTRODUCTION

I. Position du problème

Le projet de thèse relatif à notre sujet concernant la médecine traditionnelle (MT) et le web sémantique est né du constat selon lequel certains éléments de la médecine traditionnelle ont tendance à disparaître. L'absence d'un système formalisé de stockage et de transmission des connaissances, la déforestation abusive des forêts sont, parmi tant d'autres, les facteurs qui mettent en péril la pratique de la médecine traditionnelle. Cette médecine a pourtant été pratiquée presque exclusivement à l'époque par une forte proportion de populations vivant en Afrique, en Asie et en Amérique latine. Nous assistons, depuis un certain temps, à un jeu de pouvoir très important de la part de nombreuses personnes (tradipraticiens, universitaires, chercheurs, etc.) qui œuvrent pour la sauvegarde et la valorisation de cette médecine. Celle-ci réunit l'ensemble des connaissances, des compétences et des pratiques basées sur des théories, des croyances et des expériences auxquelles différentes cultures ont recours pour entretenir la santé. Autrement dit, la médecine traditionnelle permet de prévenir, de diagnostiquer, de soulager ou de soigner des maladies physiques et mentales [Ra 93]. L'Organisation Mondiale de la Santé (OMS) estime que la plupart des maladies en Afrique, sont traitées par la médecine traditionnelle [Caune 00]. Cette médecine a fait ses preuves là où la médecine moderne a parfois montré des insuffisances [Arm 12]. C'est du reste, l'une des raisons pour lesquelles l'OMS propose de l'intégrer dans les systèmes sanitaires nationaux [Arm 12] [Marta 12]. L'ensemble des connaissances et des données que manipule le tradipraticien, lorsqu'il est versé dans le domaine de la connaissance publique, pourrait promouvoir l'intégration du praticien de la MT dans le système sanitaire national. Le problème reste que le profil d'intégration des données du tradithérapeute reste difficilement appréhendable. Il n'existe pas de standard qui intègre les connaissances irrationnelles, mystico-religieuses et ethno-socio culturelles qui entourent la pratique de cette science même si des recherches sont de plus en plus entreprises dans cette direction par les socio-anthropologues. Le fait demeure qu'il reste absolument nécessaire que soient sécurisés les flux d'informations et l'ensemble des connaissances dont disposent les praticiens de la médecine traditionnelle.

Pour résoudre ce problème, une première étape serait déjà d'utiliser une approche scientifique pour structurer les éléments contenus dans la MT (plantes médicinales, maladies, individus, etc.) à l'aide des sciences informatiques (ontologie, etc.) ; ainsi, ces éléments du contenu seraient plus formalisés, plus accessibles et les données qu'ils renferment seraient plus stables et plus cohérentes.

La deuxième étape serait d'arriver à la vulgarisation des méthodes d'apprentissage et de pratique de cette médecine au moyen d'un outil informatique tel que l'ontologie. Il est vrai que le secteur de la MT accuse un certain retard au regard de l'usage des dispositifs que présentent les nouvelles technologies telles que le web sémantique qui, à vrai dire, s'avère être un outil adéquat pour le tradipraticien.

Le Web sémantique est d'abord une infrastructure permettant l'utilisation de connaissances formalisées en plus du contenu informel actuel du Web. En médecine traditionnelle, l'utilisation d'une « sémantique partagée » via les ontologies, terminologies ou thésaurus facilite l'intégration de sources multiples lorsque celles-ci sont associées à un même concept. Le web sémantique a récemment normalisé un ensemble de méthodes et d'outils permettant d'intégrer des données en les associant à des concepts d'une ontologie suivant différentes approches : les bases de données à base ontologique. Utiliser une ressource sémantique telle qu'une ontologie pourrait être un moyen d'enrichir les données cliniques en vue de répondre plus précisément à des questions d'ordre médical complexes [Aria 10].

Ensuite, le web sémantique est un vaste espace d'échange de ressources entre êtres humains et machines permettant une exploitation qualitativement supérieure de grands volumes d'informations et de services variés [Bern 01]. Enfin, c'est aussi, de nos jours, un ensemble de technologies visant à rendre le contenu des ressources accessible et utilisable par des agents logiciels qui ont pour ambition de lever cette difficulté liée à la maîtrise de la nouvelle technologie par les tradipraticiens. Les ressources du Web seront plus aisément accessibles aussi bien par l'homme que par la machine grâce à la représentation sémantique de leurs contenus.

II. Objectif de la recherche

Le principal objectif de cette thèse est d'abord de contribuer à la valorisation de la MT puis, d'aider à la mise en place d'une certaine collaboration entre le pratiquant de cette science et celui de la médecine moderne afin de rehausser le niveau des prestations de santé offertes aux populations. Cependant, l'absence de courage des uns et des autres, pour orienter la MT vers les sciences informatiques, reste un des défis majeurs à relever. Nous nous sommes rendus compte en parcourant la littérature et en observant la réalité, que l'intérêt et les efforts de valorisation de cette science se sont généralement orientés vers l'analyse exclusive des moyens de traitement et de leur transformation en produits pharmacologiques sans que la recherche ne porte véritablement sur les modes de stockage, de préservation, de transmission et de pratique du contenu de cette science.

En d'autres termes, notre travail consiste à montrer, entre autres, comment l'informatique peut aider à l'émancipation du pratiquant de la MT.

III. Contribution de la thèse

Notre première contribution consiste en l'utilisation des réseaux bayésiens comme outils de recherche en santé dans le cadre de la MT. A cet effet, nous proposons un modèle probabiliste pour le diagnostic et le traitement de la maladie dans la pratique de la médecine traditionnelle.

Notre deuxième contribution consiste à proposer un système d'algorithme dont l'objectif est d'aider le praticien à résoudre les problèmes auxquels il peut être confronté devant un patient. Nous montrons, à travers une modélisation basée sur une approche collaborative, comment on pourrait éviter d'éventuels cas de conflit en termes de symptomatologie et/ou de thérapeutique de sorte à rendre la pratique de la médecine traditionnelle comme une activité conviviale.

Notre troisième contribution vise à proposer une modélisation fondée sur le raisonnement à partir de cas concrets de maladies guéries par la médecine traditionnelle. Ce raisonnement permet au praticien de la MT de pouvoir acquérir facilement les connaissances de cette science par la maîtrise des cas proposés qui deviennent alors des nœuds focaux sur lesquels se greffent d'éventuels autres cas.

La quatrième contribution porte sur l'utilisation d'un langage d'interrogation de requêtes appelé langage « SPARQL».

Comme application, nous avons mis en place une plateforme de navigation chargée de gérer les ressources ontologiques de la médecine traditionnelle à travers un langage de requête SPARQL. Nous avons également présenté une application permettant de gérer les cas de conflit lors de la pratique de la médecine traditionnelle. Toutes ces approches ont pour seul but de répondre à la problématique de la formalisation des connaissances de la MT, et ce, dans un souci de valorisation de cette science.

IV. Organisation du mémoire

Ce mémoire est organisé en neuf (9) chapitres :

Le chapitre 1 présente les généralités et l'état des lieux en matière de médecine traditionnelle.

Le chapitre 2 établit l'intérêt de l'utilisation du web sémantique dans le domaine médical traditionnel.

Au niveau du chapitre 3 nous présentons, dans un premier temps les différentes étapes de la construction d'une ontologie. Ensuite, nous élaborons les différentes techniques de sa mise en place. Enfin, nous proposons un algorithme qui permet de suivre les étapes successives dans la pratique médicale en médecine traditionnelle en nous basant sur le prototype ontologique proposé.

Le chapitre 4 présente les travaux de recherche réalisés sur les différentes méthodes d'apprentissage ; l'élaboration d'une approche novatrice pour l'acquisition des connaissances en sciences médicales traditionnelles procède du souci de rationaliser l'apprentissage et la pratique de la MT. Et ce, en vue de valoriser cette médecine. Nous avons, à ce propos, élaboré une architecture qui montre les différentes articulations dans un processus d'apprentissage.

Dans le chapitre 5 nous montrons comment utiliser les techniques de recherches proposées pour obtenir une information dans le domaine de la médecine traditionnelle.

Le chapitre 6 décrit un modèle de rationalisation de la pratique de la science médicale traditionnelle basé sur la méthode des réseaux bayésiens.

Le chapitre 7 parle d'une autre méthode qui entre dans la pratique de la médecine traditionnelle et qu'on dénomme méthode collaborative pour la gestion des conflits en cas d'ambiguïtés dans les résultats fournis par l'outil informatique.

Le chapitre 8 se focalise sur cette autre méthode qu'est le raisonnement à partir de cas concrets de maladies guéries par la médecine traditionnelle.

Le chapitre 9 présente le langage « SPARQL » qui est un mode de requête et d'interrogation d'une connaissance. « SPARQL » se présente comme étant une approche nouvelle dans l'exercice et la pratique de la médecine traditionnelle.

Nous terminons notre mémoire par une conclusion générale et dégageons quelques perspectives de recherche.

Dans les annexes, nous présentons des codes sources et des solutions de notre application.

Chapitre 1

MEDECINE TRADITIONNELLE

Résumé. La médecine traditionnelle est utilisée par bon nombre de personnes pour guérir de nombreux cas de maladies. Cette médecine est accessible et efficace. Les personnes de différents milieux culturels l'utilisent sous plusieurs formes. Dans ce chapitre, nous allons présenter succinctement l'histoire de la médecine traditionnelle, ses acteurs, ses potentialités, ses perspectives.

1.1. Fondements historiques de la médecine traditionnelle

Depuis l'aube des temps, l'homme a toujours recherché et trouvé des moyens pour soigner les nombreuses maladies qui menaçaient son existence. C'est ainsi, qu'au fil des années, l'homme a appris à connaître les plantes et autres éléments de la nature et en a cerné les vertus thérapeutiques. En même temps qu'il s'efforçait de s'adapter aux différents milieux de vie, il avait pu voir et développer des connaissances et des pratiques pour se soigner en utilisant des produits naturels d'origine végétale, minérale, animale etc.

L'ensemble de ces pratiques appelées médecine traditionnelle, serait né de l'observation et d'une démarche fondée sur la démonstration ; elle regroupe une immense variété de méthodes, d'approches, de savoirs, de croyances et de savoir-faire ; elle implique l'usage à des fins médicales de plantes, de parties d'animaux, de minéraux, de moyens spirituels, utilisés séparément ou en association pour prévenir, diagnostiquer, soigner les maladies. [Moh 05].

L'OMS définit la médecine traditionnelle comme étant « l'ensemble de toutes les connaissances et pratiques explicables ou non à l'état actuel de la science, pour diagnostiquer, prévenir ou éliminer un déséquilibre physique, mental ou social, en s'appuyant exclusivement sur l'expérience vécue et l'observation transmise de génération en génération oralement ou par écrit »¹. Plusieurs théoriciens, à diverses époques, ont entrepris de recenser les plantes médicinales et d'expliquer leur action sur l'organisme. Les connaissances et techniques de cette médecine traditionnelle sont transmises et enrichies de génération en génération.

Vers les années 500 av. J-C, dans les civilisations les plus avancées, cette médecine se sépare progressivement de l'univers magico-spirituel dans lequel elle avait évolué donnant ainsi naissance à la médecine Moderne [Caun 00].

¹ <http://www.medicine-montaigne.com/medecine-traditionnelle.ws>

1.2. Concepts de la médecine traditionnelle africaine

La médecine traditionnelle africaine est un système complexe de traitement dans lequel la maladie est considérée comme « un déséquilibre physique mental ou social » [Ghis 07]. Cette médecine regroupe plusieurs acteurs dans différents domaines ; ce qui rend les échanges de connaissances assez complexes et les traitements multiformes. Les connaissances, dans ce domaine, sont transmises par voie orale et par conséquent ne sont pas structurées. On peut identifier dans la médecine traditionnelle (MT) plusieurs acteurs qui ont des rôles spécifiques ou des fonctions différentes [Ghis 07].

1.2.1. Guérisseur

Le guérisseur est une personne célèbre et respectée dans son milieu. Il peut être « psychologue », « botaniste », « pharmacologiste » et/ou « médecin ». Il connaît la cosmogonie, les noms des plantes, des animaux, des pierres, etc. Dans le processus de traitement, un guérisseur examine le malade dans toute son intégralité, tant sur le plan physique que métaphysique ; il utilise, pendant ses interventions, les éléments de la nature et s'appuie sur certains aspects de la vie du malade. Les guérisseurs traditionnels sont proches culturellement de leurs patients ; ce qui facilite la communication au sujet du malade et le traitement des questions sociales inhérentes [Om 13].

1.2.2. Féticheur

Le féticheur est un guérisseur qui prédit des événements importants (malchance ou bonheur) et est souvent consulté par des patients en vue de pouvoir trouver la cause d'une maladie ou pour se protéger contre certaines malchances. Il joue à peu près le même rôle que le devin qui prédit et se trouve être comme l'intermédiaire entre les divinités et le patient. Il diagnostique généralement le mal dont souffre le patient mais, en cas de nécessité, peut lui conseiller un guérisseur.

1.2.3. Herboriste

Ce terme décrit un guérisseur traditionnel spécialisé dans l'utilisation exclusive des plantes médicinales pour traiter diverses maladies. On attend de lui une grande connaissance des plantes du point de vue de leur efficacité, de leur toxicité, de leur dosage et de leur mode de préparation.

1.2.4. Réducteur de fractures

Un réducteur de fracture est un guérisseur traditionnel qui réduit les fractures. Il procède en attachant des morceaux de bois ou des tiges (comme des attelles) avec une plante fibreuse (comme corde) autour de l'os fracturé et en appliquant de manière appropriée des médicaments traditionnels au point de la fracture. Il accomplit son métier sans l'aide de rayon X et procède parfois aussi à des manipulations d'articulations et à des massages.

1.3. Plantes médicinales

Les plantes médicinales sont utilisées en médecine traditionnelle sous différentes formes, en association ou seules pour le traitement des maladies. Pour qu'une plante médicinale puisse guérir, certaines règles strictes doivent être respectées lors de sa transformation en remède : incantation, sacrifices, choix des jours bénéfiques, situation et orientation de l'échantillon végétal prélevé, façon de cueillir, nombre de feuilles, de racines et tiges nécessaires à la préparation [Abdu 11].

Les plantes médicinales jouissent non seulement d'une grande importance dans le traitement et la prévention des maladies mais également dans la préservation de l'hygiène générale de la société [Nge 15]. Les plantes sont utilisées comme nourriture, moyen de guérison et également utilisées pour la fabrication des produits cosmétiques ou des huiles essentielles en aromathérapie. Elles sont également la base de tous les systèmes médicaux dans le monde [Mag 13].

1.4. Autres éléments de la médecine traditionnelle (minéraux, parties d'animaux et pratiques magico-spirituelles).

La plante reste l'un des ingrédients essentiels dans la pratique de la médecine traditionnelle. Cependant, l'usage des autres éléments (minéraux, parties d'animaux et pratiques magico-spirituelles) participent de la recherche de la santé c'est-à-dire « ...un état de complet bien-être physique mental et social... » [Caun 00]. Tous les éléments précités peuvent être utilisés soit en association, soit de façon spécifique.

La médecine traditionnelle intervient pratiquement dans le traitement de toutes les maladies auxquelles l'homme peut être confronté.

1.5. Quelques découvertes de plantes dans la médecine traditionnelle

Ces découvertes s'avèrent être un moyen très efficace de traitements de certaines maladies à caractère épidémique et/ou endémique dans nos régions. Par exemple,

l'artémisinine est une substance extraite des feuilles d'un arbuste appelé l'artemisia annua ou absinthe chinoise et est présentement l'un des remèdes les plus efficaces contre le paludisme [Priv 10]. De même, la pervenche de Madagascar (*Catharanthus roseus* (Apocynaceae) Syn.: *Vinca rosea*) qui est une plante cultivée au départ comme plante ornementale est, depuis quelques années, considérée comme une plante très efficace contre certains cancers dont le cancer du sang (la leucémie) [Des 05]. Cette plante contient jusqu'à 55 types différents d'alcaloïdes qui servent au traitement de ces maladies [Nge 15].

1.6. Prévention dans la médecine traditionnelle

En médecine traditionnelle, la plupart des maladies relèvent plus ou moins de causes surnaturelles ; aussi la prévention nécessite-t-elle souvent le respect de certaines règles et pratiques magico- religieuses, à savoir :

- ✓ Respecter les us et coutumes en se soumettant à leurs principes.
- ✓ Honorer les lois des ancêtres,
- ✓ Obéir aux aînés,
- ✓ Respecter les interdits....

1.7. Diagnostic dans la médecine traditionnelle

Dans toute approche thérapeutique (qu'il s'agisse de la médecine traditionnelle ou de la médecine moderne) le diagnostic est important dans le processus de guérison d'une maladie. En effet, le diagnostic permet de déterminer la cause de la maladie et la nature de l'affection. Celui-ci s'obtient à partir de certains éléments comme l'interrogatoire, la palpation, les techniques de divinations etc.

1.8. Travaux existants

Dans le souci de préserver les connaissances accumulées dans le domaine de la médecine traditionnelle, des travaux ont été effectués qui ont donné lieu à l'élaboration des systèmes spécifiques dans le cadre d'une assistance technologique. Nous pouvons citer entre autres B. R. Baker [Bak 02], qui présente les atouts du système IBS (Irritable Bowel Syndrome). En effet, le système IBS est un langage de requête graphique utilisant des icônes. Ce système permet aux médecins de pouvoir interroger une base de données relationnelles sans que ces derniers aient une idée de la manière dont les questions c'est-à-dire les requêtes SQL sont écrites. Ce même Auteur décrit les systèmes Minspeak et VAL (Virtual Access to the Lexicon) comme étant des prothèses de langue. Ces systèmes ont été mis en place en vue d'aider les personnes ne sachant

pas comment exprimer leur désir par la parole de pouvoir le faire. Cependant, l'auteur ne porte pas une attention particulière sur la signification des icônes. Par contre, d'autres auteurs [Ghis 07] [Fay 01] ont mis au point les systèmes Edith et HandiAS, spécialement conçus pour aider certains malades à communiquer.

Les systèmes Sibylle, VITIPI et Sibylle AAC (Augmentative and Alternative Communication) conçus par [Kob 10] [Ton 06, Bet 99] dans le but de permettre aux personnes lourdement handicapées (tétraplégiques, locked-in syndromes, infirmes moteurs cérébraux...) de pouvoir prédire les mots. En effet, ces systèmes donnent des moyens pictographiques adéquats aux utilisateurs pour contrôler l'énoncé au travers de l'ordinateur, quel que soit leur état psychique ; ceux-ci, cependant ne prévoient aucune alternative en cas d'oubli ou lorsque les mots prédits ne désignent pas exactement la maladie. De plus, ces systèmes ne tiennent pas compte des sujets à contenu complexe.

Les systèmes SEIBOGA, MEDITRA et MedTrad de [Leg 10, Fras 86] ont pour fonction d'identifier de façon iconique les plantes médicinales et de retrouver ensuite leurs vertus curatives. Ces systèmes permettent aussi de capitaliser la connaissance ancestrale africaine sur la phytothérapie et l'entreposage de la connaissance sur les maladies sans toutefois tenir compte de la ressemblance des feuilles du point de vue de leur morphologie et de leurs couleurs ; sans tenir compte également de la réaction du praticien vis-à-vis d'un nouveau produit.

Le système MEDIGATE (Médical Examination Direct Iconic and Graphic Augmented Text Entry) [Kof 09, Jos 92] est un système basé sur l'approche orienté-objet. Il a été conçu dans le but d'étudier certains problèmes liés à l'interface lors de la conception d'une application afin qu'elle soit le plus naturel possible pour l'utilisateur. Aussi permet-il d'aider les praticiens dans le développement et le maintien de l'enregistrement des résultats. Il autorise la manipulation directe des objets graphiques. Cependant, il ne prend pas en compte la cible c'est-à-dire les personnes autorisées à utiliser ce système.

Le système SADMEDTra [Ben 13], par contre, se base sur les agents intelligents pour diagnostiquer et traiter les maladies. Il offre aux tradipraticiens une meilleure compréhension de leurs traitements sans tenir compte de leur niveau intellectuel.

De même, VCM (Visualisation des Connaissances Médicales) [Den 98, Jean 06] est un système de langage iconique permettant de représenter un ensemble de concepts médicaux comme des maladies, des médicaments et des examens complémentaires. Il permet également d'accéder très rapidement au contenu d'un dossier médical et aux sections de documents textuels contenant des connaissances à priori utiles à la prise de décision.

Les systèmes décrits plus haut ont, malgré leurs pertinences, montré leurs limites en ce qui concerne :

- la signification des icônes,
- les cas d'oubli ou de non-conformité des mots suggérés par certains systèmes,
- les sujets à contenus complexes,
- la ressemblance des feuilles du point de vue de leur morphologie,
- la population cible (s'agissant des tranches d'âges) n'est pas déterminée dans certaines applications.

D'où l'intérêt de l'utilisation des langages de description, objet de la prochaine section.

1.9. Logique de description

La logique de description, connue sous le nom de « logique descriptive (LD) », regroupe une famille de langages de représentation de connaissances. Elle peut être utilisée pour représenter la connaissance terminologique d'un domaine d'application de manière formelle et structurée [Chan 04]. Les logiques de description ont été développées pour être utilisées dans des situations pratiques. Les opérateurs logiques permis sont donc soigneusement choisis pour s'assurer que les méthodes de preuve associées donnent des résultats dans un temps raisonnable pour des problèmes typiques [Berg 13]. Un système de représentation de connaissances à base de logique de description fournit des moyens pour construire des bases de connaissances, effectuer des raisonnements sur leur contenu et les manipuler [Fran 03]. Le développement d'une base de connaissances en utilisant la logique de description revient d'abord :

1- à définir les entités du domaine à modéliser en utilisant la terminologie (c'est-à-dire le vocabulaire) du domaine, ce qui représente la TBox (Terminological Box).

2- à introduire les assertions, l'ABox (Assertional Box) qui représente les individus et les concepts définis dans la TBox. Une des caractéristiques des logiques de description qui les distingue de leurs prédécesseurs, est leur sémantique formelle basée sur la logique [Rad 09]. Une autre caractéristique est l'accent mis sur le raisonnement qui est considéré comme un service central pour ce type de formalisme. Le raisonnement permet de déduire les connaissances implicites à partir de connaissances explicitement présentes dans la base de connaissances [Fran 03]. Les logiques de description exploitent, en général, des sous-ensembles décidables de la logique du premier ordre. Ces logiques ont été créées dans le but de surmonter les ambiguïtés des anciens systèmes à base de réseaux sémantiques et de frames.

1.9.1. TBox (Terminological Box)

La TBox introduit le vocabulaire du domaine de l'application et décrit les connaissances générales de ce domaine. Aussi, il contient les déclarations des primitives conceptuelles organisées en concepts et rôles (relations). La TBox est constituée de définitions de concepts. Elle décrit de nouveaux concepts à partir de concepts primitifs et de concepts déjà définis. Ces définitions utilisent des opérateurs logiques qui caractérisent chacune des logiques de description. Par exemple, l'opérateur \cap représente l'intersection de concepts et permet de définir : $Femme \equiv Humain \cap Femme$. L'opérateur universel de restriction de valeur $\forall RC$ permet de définir le concept Hsf , un homme avec seulement des fils comme : $Hsf \equiv Homme \cap \forall a Enfant.Homme$. Cet opérateur $\forall RC$ est une forme restreinte de l'opérateur universel de la logique des prédicats du premier ordre. Cette restriction permet une bonne expressivité tout en simplifiant la méthode de preuve. De la même façon, on a un opérateur de négation \neg , un opérateur existentiel de restriction de valeur $\exists RC$, un opérateur d'union \cup , un opérateur de décompte de valeurs sur un rôle ($> nR.C$) [Berg 13].

1.9.2. Assertional Box (ABox)

L'Assertional Box (ABox) contient des assertions sur les individus et les instances de concepts. Plusieurs ABox peuvent être associées à une même TBox ; chacune représente une configuration constituée d'individus, et utilise les concepts et les rôles de la TBox pour l'exprimer. Les ABox sont des restrictions sur les valeurs des prédicats et représentent les propriétés d'un objet pour un domaine données [Berg 13].

1.10. Connaissances

La connaissance est une notion qui a plusieurs sens. Elle définit tout aussi bien «l'état de celui qui sait» que «les choses sues ». Il n'y a pas aujourd'hui de définition plus substantielle de la connaissance qui soit largement admise. Les connaissances, leur nature et leur variété, leur processus d'acquisition, leur valeur, et leur rôle dans les sociétés humaines, sont étudiés par une diversité de disciplines, notamment la philosophie, l'épistémologie, la psychologie, les sciences cognitives, l'anthropologie et la sociologie [Gir 95]

Selon Beckman en 1998 [Rol 03], la ressource «connaissance» fait partie d'une chaîne de transformation ou encore d'un processus comportant un certain nombre d'étapes où les données (data) constituent la matière brute de la chaîne. Les connaissances permettent non seulement l'exécution des tâches et la prise de décision ; mais aussi la résolution des problèmes.

Les cas, les règles, les modèles sont trois types de connaissances. Les connaissances s'acquièrent par accumulation d'informations qui s'organisent progressivement par rubriques dans l'ensemble des moyens de stockage : ouvrage, bases de données, supports magnétiques [Ben 12]. Une information n'est rien d'autre qu'une donnée dont la forme et le contenu sont utiles pour une tâche particulière.

1.11. Base de connaissances

Une base de connaissances regroupe des connaissances spécifiques à un domaine spécialisé sous une forme exploitable par un ordinateur. Elle peut contenir des règles (dans ce cas, on parle de base de règles) permettant d'établir de nouveaux faits ou d'autres représentations. Elle contient aussi à la fois des règles et un moteur d'inférence. Le moteur d'inférence est le programme qui réalise les déductions logiques d'un système expert à partir d'une base de connaissances (faits) et d'une base de règles [Coul 08]. Les moteurs d'inférence simulent les raisonnements logiques et peuvent être utilisés pour déduire de nouveaux faits. En effet, une base de connaissance sert à rassembler de manière centralisée l'expertise d'un domaine généralement formalisée de manière déclarative [Chan 04]. On peut dire tout simplement que la base de connaissances est un ensemble de connaissances fournies par un spécialiste humain et représenté par des règles.

1.12. Connaissances empiriques

Les connaissances empiriques, c'est-à-dire les connaissances expérimentales, sont représentées par l'ensemble des cas pratiques observés sur un sujet (ensemble d'exemples). Ce sont des connaissances "ancestrales" qui n'ont pas été traitées, analysées ou modifiées. Ces connaissances représentent les résultats d'expériences ou les exemples de cas pratiques ; elles n'ont pas encore subi de transformations en vue d'obtenir une théorie plus générale sur le domaine. On peut dire que ce sont des connaissances de bas niveau.

1.13. Connaissances théoriques

Les connaissances théoriques modélisent les connaissances sur un sujet à l'aide d'une théorie correspondant au problème posé. Elles sont des connaissances "traitées" qui ont été obtenues à partir de l'analyse des connaissances de base. Ce type de connaissances représente une généralisation du savoir. Ce sont des connaissances dites de haut niveau. Elles sont habituellement représentées par des structures symboliques telles que les règles de production, les modèles mathématiques, les réseaux sémantiques, les objets structurés.

1.14. Moteur d'inférence

Le moteur d'inférence a pour fonction d'exploiter la base de connaissances afin de résoudre le problème qui lui est posé en s'appuyant sur des algorithmes. Il existe en effet, deux types d'algorithmes :

- les algorithmes à chaînage avant : les règles sont appliquées sur une base de faits initiaux et vont modifier cette base (ajouts, suppressions, modifications). Ces modifications impliquent le déclenchement d'autres règles (ce qui constitue le chaînage avant) jusqu'à épuisement des déductions possibles. Ce type d'algorithme a été utilisé dans plusieurs secteurs d'activités notamment dans les systèmes de diagnostics médicaux pour déduire des pathologies probables à partir de symptômes constatés. Exemple : en médecine moderne on utilise le système Mycin qui est un système expert de référence destiné à conseiller le médecin sur le choix d'un traitement adéquat à une infection bactérienne donnée.

- les algorithmes à chaînage arrière : dans ce cas, le moteur d'inférence s'appuie sur des faits à établir, recherche et vérifie les règles contenues dans sa base de règles puis déclenche de nouveaux faits ; ce même mécanisme s'effectue sur les autres faits existants dans sa base des faits et ce, de manière récursive.

Ce type d'algorithme est notamment utilisé pour le traitement de langage naturel [Leg 10].

1.15. Langage naturel

Par langage naturel, on entend tout simplement les langues qui sont parlées (parfois aussi écrites) par les humains [Fran 10]. Le langage naturel permet de rendre l'interaction Homme-Machine plus réaliste. Le langage naturel a d'abord été étudié dans le but de faire de la traduction automatique. Devant les difficultés, on s'est plutôt intéressé aux sous-problèmes de la compréhension et de la génération du langage naturel, domaines qui se rattachent également au dialogue et aux interfaces. Il est donc nécessaire de disposer de représentations internes, de connaissances syntaxiques, sémantiques, et même pragmatiques [Dom 00] en se basant sur les langages de communication d'agents intelligents. Un agent intelligent est « ... un système informatique, situé dans un environnement, et qui agit d'une façon autonome pour atteindre les objectifs pour lesquels il a été conçu » [Harb 07]. Les agents intelligents sont apparus dans le but de rendre accessible la tâche de la représentation du comportement humain par la machine. L'observation des conversations humaines est la base de l'élaboration de protocoles sophistiqués car, l'utilisateur s'exprime le mieux et le plus spontanément possible par ce langage appelé théorie des actes de langage [Fran 10]. Cette théorie est issue de la philosophie

du langage, initialement pensée pour le langage naturel. Sa nature formelle la rend utilisable pour les modèles de calcul automatique. Elle est reconnue depuis longtemps dans la communauté scientifique pour sa capacité à bien modéliser les communications. Il faut aussi garder à l'esprit que le langage naturel a le plus grand pouvoir d'expression et permet de tout dire sous certaines conditions et aussi d'être compris de tous. Cela est un aspect majeur du langage naturel [Ben 13].

1.15.1. Système de compréhension du langage naturel

Un système de compréhension du langage naturel nécessite successivement une analyse morphologique pour déterminer les morphèmes, une analyse lexicale pour identifier les mots, une analyse syntaxique pour avoir la structure syntaxique de la phrase, une analyse sémantique pour aboutir à une représentation sémantique, une analyse pragmatique pour avoir la représentation pragmatique et enfin un raisonnement à l'aide de règles d'inférences et de déduction pour aboutir au sens de la phrase [Sal 09].

1.15.2. Langages de communication entre agents

Les langages de communication entre agents se distinguent des approches plus traditionnelles de communication au niveau des programmes informatiques telles que UDP, TCP/IP, RPC, JAVARMI ou CORBA du fait qu'ils cherchent à échanger des propositions et des tâches ayant une sémantique de haut niveau comparativement à de simples données ou objets. Dès lors, les agents sont généralement impliqués dans une conversation ou un dialogue semblable à ce que font les humains et non dans un simple processus d'échange d'informations

1.16. Concept

Un concept n'a de valeur que dans un domaine. En effet, un concept peut être défini dans plusieurs domaines mais avec des objets différents. Le concept peut représenter une réalité abstraite, c'est-à-dire une idée qui représente un aspect de la réalité isolée par l'esprit ou encore une unité de pensée constituée d'un ensemble de caractères attribués à un objet ou à une classe d'objets [Brou 97]. Du point de vue traditionnel et purement linguistique, un concept est donné par une définition qui précise les conditions nécessaires et suffisantes pour son application [Thag 97]. Le concept est perçu comme une unité de pensée constituée par abstraction à partir des propriétés communes à un ensemble d'objets. Un concept est également vu comme un sous-ensemble d'un domaine.

- **L'Objet** est comme une instanciation d'un concept. L'instanciation est l'action d'instancier, de créer un objet à partir d'un modèle. Par définition, un objet est un élément de la réalité qui peut être perçu ou conçu. Il peut être matériel (exemple plante) ou immatériel (exemple maladie). Au vu de ces aspects, on peut aussi dire qu'un objet est la représentation concrète d'un concept. Par analogie au modèle objet, il peut être comparé à un objet d'une classe.

- **Un domaine** est un ensemble de concepts reliés les uns aux autres par des relations sémantiques. Une langue de spécialité est choisie dans un domaine et est utilisée par les experts du domaine afin d'avoir une interprétation commune des concepts qu'ils utilisent. En botanique, par exemple, le latin est la langue véhiculaire (c'est-à-dire langue de communication entre des communautés d'une même région ayant des langues maternelles différentes) pour nommer les plantes.

1.17. Base de règles

Une règle est une description de haut niveau, permettant de contrôler et/ou de prendre une décision, en utilisant des concepts propres à une entreprise ou une organisation. Ainsi, les règles décrivent ce qu'un expert doit faire pour prendre une décision, c'est-à-dire la connaissance des experts. La modélisation interne doit donc permettre [Leg 10] non seulement d'exprimer les concepts objets et de décomposer en règles simples des éléments de raisonnement, mais aussi d'organiser (ordonner, affecter des priorités, brancher) l'exécution des règles ou de paquets de règles. En d'autres termes, la modélisation interne est l'ensemble des connaissances du spécialiste qui lui permet de résoudre des problèmes spécifiques.

Exemple1 :

Si nous avons une eau dont la température est supérieure à 120°C alors cette eau bout ; elle est donc en ébullition.

$Eau \wedge Temperature > 100 \rightarrow EauBout$ (Règle en logique des propositions, ou logique d'ordre 0) $\forall x(Temperature(x) > 100 \wedge x = Eau \rightarrow Etat(x, Ebullition))$ (Règle en logique du premier ordre)

$Temperature > 100 \wedge liquide = Eau \rightarrow EtatLiquide = Ebullition$ (Règle dite d'ordre 0+ : les symboles propositionnels peuvent avoir des valeurs non booléennes, et on dispose des comparateurs booléens =, >, ..., < ; mais il n'y a ni variables, ni quantificateurs).

1.18. Conclusion

Dans ce chapitre, nous avons présenté les fondements historiques de la médecine traditionnelle. Nous avons pu comprendre quelle était son origine et comment elle fonctionnait. Nous en avons décrit les concepts et présenté quelques acteurs qui l'animent. Par la suite, nous avons décrit les éléments essentiels qui constituent les moyens par lesquels cette médecine opère. Nous avons pu nous rendre compte que plusieurs travaux de recherche avaient été entrepris dans le domaine de la médecine traditionnelle permettant ainsi de plus ou moins la valoriser. Nous avons pu constater que la médecine traditionnelle regorgeait d'atouts importants. Cependant, elle présente encore des limites majeures en ce qui concerne son mode de transmission et de fonctionnement.

Dans le chapitre suivant nous allons essayer de montrer l'intérêt de la nouvelle technologie : « le web sémantique » dans la pratique de la médecine traditionnelle afin qu'elle puisse être d'une meilleure accessibilité et d'une utilisation plus efficiente.

Chapitre 2

WEB SEMANTIQUE

***Résumé.** L'intérêt croissant porté à la recherche d'informations sur le Web a donné lieu à l'initiative du Web sémantique. L'objectif est de rendre les ressources du web facilement accessibles non seulement aux êtres humains mais et surtout à la machine grâce à la représentation sémantique de leur contenu. Pour rendre possible cette représentation, il est important de mettre l'accent sur des langages de structuration sémantique des connaissances. Celle-ci permettra également de fournir le vocabulaire et la structure des métadonnées associées aux différentes ressources.*

2.1. Introduction

Le Web sémantique désigne un ensemble de technologies visant à rendre le contenu des ressources du World Wide Web accessible et utilisable par les programmes et agents logiciels grâce à un système de métadonnées formelles, utilisant notamment la famille de langages développés par le W3C [Aria 10]. Comme l'écrit en substance Tim Berners-Lee, « le Web sémantique est ce que nous obtiendrons si nous réalisons le même processus de globalisation sur la représentation des connaissances que celui que le Web fit initialement sur l'hypertexte » [Phil 02]. Partant de ce fait, le Web sémantique peut se définir comme la mise en relation de l'information pour permettre aux ordinateurs de comprendre la sémantique (le sens) de l'information [Trus 12]. C'est aussi une extension du World Wide Web qui intègre le web actuel dans le but d'agréger plusieurs données liées entre elles, soit sémantiquement, soit par des attributs qui les déterminent [Hab 04]. La différence philosophique majeure entre ces deux technologies est que le web actuel est un ensemble de documents basés essentiellement sur l'identifiant URI du protocole HTTP et sur le langage HTML, dont la recherche se fait à travers des mots clés et utilisables par l'homme tandis que le web sémantique est un ensemble de connaissances basées sur le document XML (eXtensible Markup Language) et RDF (Resource Description Framework). D'après Tim Berners-Lee [Bern 01], aujourd'hui, les humains sont les seuls êtres à avoir la capacité de comprendre ce qu'ils trouvent et les seuls à pouvoir décider de ce qu'ils veulent chercher réellement. Cependant, par quels moyens les moteurs de recherche nous aident-ils ? En d'autres termes, comment les ressources textuelles ou multimédias doivent-elles être sémantiquement étiquetées par des métadonnées afin que les agents logiciels puissent les exploiter [Pha 05] ? La représentation explicite des contenus des ressources documentaires du Web est rendue possible grâce aux ontologies. Ces ontologies représentent une technologie clef pour la mise en œuvre de ce Web Sémantique [Flor 07]. Cependant, cultiver le web sémantique exige un équilibre délicat entre l'interopérabilité, l'homogénéité, la compréhension partagée, le malentendu et les besoins divergents. Le terme « Web Sémantique » est largement

utilisé souvent sans discernement, tant du point de vue de ses origines que de sa signification.

Il y a trois perspectives principales de la technologie web sémantique à savoir :

- la vision qui permet aux humains et aux machines de lire le programme
- lequel fournit les détails du déroulement de l'activité
- et la technologie qui sont les outils.

Ces termes sont généralement utilisés dans des situations pratiques pour résoudre de vrais problèmes [Am 08].

2.2. Web

Le World Wide Web est un système client/serveur qui intègre différents types d'informations sur le réseau internet. Il permet aux utilisateurs de naviguer sur différents sites à travers le monde et devient de plus en plus populaire. La technologie Web implique la combinaison des clients web, appelés navigateurs et des serveurs. Mis ensemble, ces deux éléments forment un outil graphique permettant facilement de naviguer sur le réseau internet [Fon 09]. Ainsi donc, le World Wide Web, devient la « toile d'araignée mondiale », en abrégé WWW ; en fait, le W3 ou le Web a été défini par une équipe dirigée par Tim Berners-Lee du CERN (Centre Européen de Recherche Nucléaire) de Genève dès 1989.

Au départ, l'objectif était de faciliter les échanges de documents (textes et images) par les réseaux entre chercheurs en sciences, entre particuliers et les différents centres de recherche dans le monde. Un nouveau type de logiciels appelés navigateurs (butineur au Canada, browser en Angleterre) a ensuite été créé pour automatiser ces échanges sur le Web. Ces logiciels, très faciles à utiliser et quasiment gratuits, ont permis de populariser l'accès aux documents du Web et aux divers services offerts sur internet [Jos 00].

2.3. Outils de création de données sur le web

Le web renferme une multitude d'outils de création de données. Nous allons en présenter quelques-uns. Pour la création d'un site web, les utilisateurs ont recours à des outils spécifiques. Par exemple, une simple feuille de bloc-notes permet de créer une page web ; en effet, pour avoir un site web ayant plusieurs pages, il faut des outils plus évolués tels que : « sublime texte, Notepad, HTML édit, Dreamweaver etc. ». Pour des pages web dynamiques, il faut le langage PHP. En effet, le PHP est un langage dont les atouts peuvent permettre de gérer une base de données avec le serveur MySQL. D'autres outils tels que java, Perl, permettent également de créer des pages web. Il est également bon de signifier que le stockage des données peut se faire

à travers des SGBD (Système de Gestion de Base de Données) de grandes capacités telles que MySQL que l'on peut obtenir grâce à EasyPHP, WanpServer, SQLyog etc.

2.4. Web et médecine

Le Web est l'une des ressources principales utilisées par les usagers pour rechercher de l'information médicale [Gab 11]. Les technologies de l'information et de la communication irriguent l'ensemble de la société et contribuent à faire émerger de nouvelles formes d'échanges et de relations sociales. Elles permettent ainsi de transformer les attitudes et les pratiques en matière de recherche et de production de l'information. Ces mutations sont particulièrement importantes dans le champ de la santé.

Notons qu'avec l'explosion du Web et la prolifération des connaissances biomédicales, les utilisateurs du monde médical ont potentiellement accès à des informations de plus en plus nombreuses [Golb 08]. En réalité, ces utilisateurs sont obligés de naviguer dans un véritable labyrinthe de pages car il leur est encore difficile d'obtenir de façon satisfaisante, c'est-à-dire rapidement, avec le minimum de bruit et de silence possible des informations médicales les plus récentes.

La recherche d'informations médicales est l'une des tâches les plus courantes des utilisateurs d'internet. Ceux-ci ont tendance à rechercher l'information médicale sur des moteurs de recherche populaires plutôt qu'au travers des sites spécialisés [Rad 09]. C'est dans ce contexte que le catalogue CISMef (Catalogue et Index des Sites Médicaux de langue Française) a été développé en 1995 pour assister les professionnels de santé, les étudiants et le grand public dans leur recherche d'informations de santé sur le Web [Golb 08]. Dans cette mouvance, les applications du Web sémantique tentent d'automatiser l'organisation et la présentation des données à l'aide d'algorithmes imaginés par l'homme. Ces innovations permettent aux utilisateurs de mieux comprendre la sémantique (le sens) de l'information [Trus 12]. Aussi, le web oblige les utilisateurs à mieux reconnaître leurs limites et les conduit quelques fois à refuser de se fier à une seule source d'informations médicales. Par ailleurs, de plus en plus de patients choisissent leur spécialiste à qui demander conseil après avoir fait des recherches sur internet ; ils fouillent le net, postent des messages sur des forums, discutent entre eux de leurs traitements. L'internet est ainsi devenu omniprésent dans le domaine médical car la quantité d'informations disponibles ne cesse de croître.

2.5. Représentation des connaissances

Une représentation est une structure de symboles permettant de décrire un modèle (une approximation) du monde dans le cadre d'une tâche particulière [Jer 10]. Pour que le Web sémantique fonctionne, les machines doivent avoir accès à des collections structurées d'informations et d'ensemble de règles d'inférence [Bern 01]. Ces machines les utilisent pour parvenir à un raisonnement automatisé. En effet, le défi du Web sémantique est de fournir un langage qui exprime à la fois des données et des règles. Reasonner sur ces données est ensuite nécessaire. Ainsi, les règles de n'importe quel système de représentation de la connaissance peuvent être exportées sur le Web sémantique [Pha 05].

Il existe différentes manières d'envisager la représentation des connaissances. Une première approche consiste à utiliser des représentations à base de faits et de règles. Cette approche, appelée « ingénierie des connaissances », a été employée dans la mise au point des premiers systèmes experts. C'est notamment ce que le système Mycin a fait en utilisant des outils inspirés des logiques multi-évaluées et est parvenu à de bons résultats prouvant ainsi l'intérêt des logiques pour la représentation des connaissances.

Les systèmes à base de connaissances qui encodent des faits et des règles ont été conçus dans une perspective de résolution de problèmes relativement spécifiques ; leurs fonctionnements sont éloignés des capacités du cerveau humain (inférences très rapides, mais connaissances peu expressives). De ce fait, pour des tâches plus proches des préoccupations humaines pouvant nécessiter des opérations de traitement automatique de la langue, il a été envisagé d'autres approches. Ces approches de représentation de connaissances sont basées sur des objets et des catégories ; elles sont plus aptes à simuler les modèles de représentations internes du savoir humain (ou représentations mentales) issues des sciences cognitives. Elles sont parfois qualifiées d'ingénierie ontologique.

2.6. Avantages du web sémantique

Les outils du Web sémantique présentent plusieurs avantages. Au niveau de son extension de base, ces outils font davantage appel à la signification des mots. Les outils deviennent plus flexibles et offrent de nouveaux moyens aux internautes en fournissant ainsi, des services de meilleures qualités qui permettent aux hommes et aux machines de coopérer. Ce mécanisme améliore la qualité des échanges de connaissances [Adam 07].

Le web sémantique en lui-même, fournit des outils nécessaires pour pouvoir identifier et décrire les données accessibles sur le web. Il facilite la construction des vocabulaires servant à

décrire les données de l'utilisateur et leurs relations, ainsi que leur exploitation des opérations telles que l'agrégation, la recherche et la présentation [ALJC 04]. L'utilisateur a un accès instantané à un niveau de base factuel de connaissances et à une multitude de sources crédibles. Cette accessibilité réduit le temps de recherche et d'organisation de l'information, ce qui permet à l'utilisateur de passer plus rapidement à l'analyse et à l'interprétation des données.

2.7. Quelques inconvénients

Malgré les atouts considérables du Web sémantique pour une gestion plus intelligente du contenu des données, ainsi qu'une meilleure collaboration entre les hommes et les machines à des fins de nouvelles possibilités d'automatisation, cette technologie fait l'objet de nombreuses critiques de la part de certains chercheurs qui lui reprochent principalement son manque de pragmatisme. En effet, Clay Shirky, dans son ouvrage [Mest 07], soutient que la gestion d'ontologies n'est pas une sinécure ; il propose d'autres alternatives plus adaptées à l'internet, telles que le «folksonomies» [Mest 07]. Cette alternative permet à l'utilisateur d'indexer des documents afin qu'il puisse plus aisément les retrouver grâce à un système de mots-clés. L'on estime également que le web sémantique est un outil peu révolutionnaire qui ne comprend que les documents et les données sémantiques qui nécessitent que les ordinateurs aient accès à des collections structurées d'informations. Il ne propose que des effets théoriques et non pratiques, ce qui paraît complexe, rigide et inutile, comparé au web 2.0.

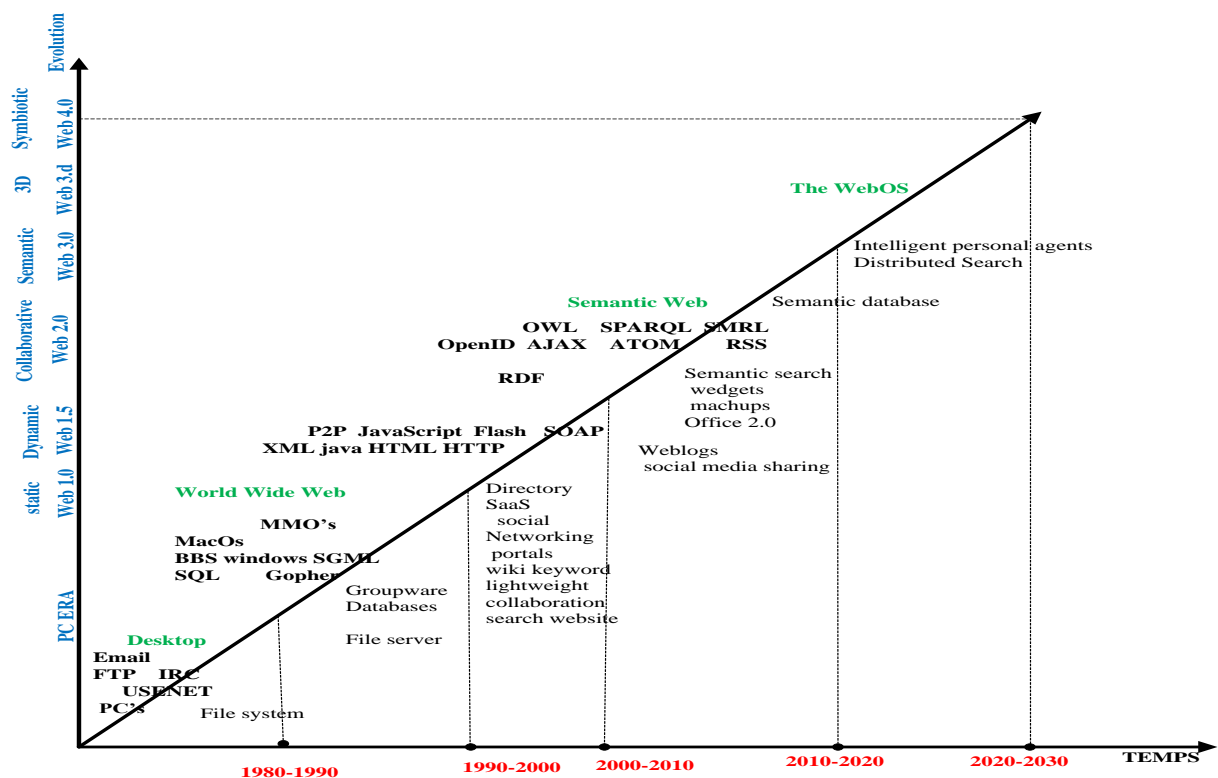


Figure 2.1 : Chronologie : avantages et inconvénients du web dans le temps.

Commentaire :

Au début des années 90, le World Wide Web était perçu comme un système d'information publique composé d'un ensemble de documents multimédias inter-reliés par des hyperliens et constitue de ce fait, une toile d'araignée mondiale connue sous le nom de Web 1.0. Une nouvelle version du Web appelée web 2.0 a vu le jour dans les années 2005 ; il consiste à considérer l'internaute comme un acteur primaire qui contribue à la production, à la consultation et à la diffusion des contenus des sites. Ce changement permet de regrouper les internautes en des communautés, et facilite ainsi, aussi bien la diffusion que le partage de l'information. Une telle vision a eu beaucoup de succès grâce aux communautés grandissantes de ce type de web. Précisons à ce niveau que les technologies du web 2.0 sont nombreuses ; nous citerons par exemple : les wikis, les blogs, les flux RSS, les réseaux sociaux, etc. Grâce à l'évolution technologique permanente, on parle déjà de la nouvelle vision du web du futur qui porte déjà les noms de Web 3.0, Web 3D ou encore Web 4.0 [Sbi 09]. La figure 2.1, établie sur inspiration des travaux de Sbihi [Sbi 09], montre l'évolution du web dans le temps depuis sa création jusqu'à nos jours, et de son évolution envisagée pour le futur.

2.8. Langages du web sémantique

Le Web sémantique est une infrastructure liée à l'utilisation de connaissances formalisées, en plus du contenu informel actuel du Web. Cette infrastructure doit permettre de localiser, d'identifier et de transformer des ressources du web de manière robuste et saine tout en renforçant l'esprit d'ouverture de celui-ci, avec sa diversité d'utilisateurs. Elle doit également s'appuyer sur un certain niveau de consensus portant sur les langages de représentation ou sur les ontologies utilisées [Jean 03]. Le Web sémantique devrait pouvoir être manipulé par les machines, d'où la nécessité de disposer de langages appropriés pour :

- exprimer les données et les métadonnées ;
- exprimer les ontologies ;
- décrire les services [Bag 04].

A l'état actuel de la technologie, il existe trois sortes de langages :

- les langages d'assertions (RDF et cartes topiques) ;
- les langages de définition d'ontologies pour le Web (OWL, XML, RDF etc.) ;
- les langages de description et de composition de Services (UDDI et autres).

2.8.1. Langage XML

Les développeurs du « world wild web consortium » (W3C) ont recommandé depuis le 10 février 1998 l'utilisation du langage « eXtensible Markup Language » (XML) qui connaît depuis ce temps, un succès indéniable.

Depuis longtemps le « Hypertext Markup Language » (HTML) a été considéré comme le seul langage facilement applicable pour définir la représentation des informations disponibles sur le web. Il fallait donc mettre en place une structure permettant d'automatiser le traitement du contenu de ces informations ; c'est à partir de cet instant qu'a vu le jour le langage XML [Sin 04].

Le XML (eXtensible Markup Language) est donc un langage à base de balises comme le HTML ; il apparaît comme une ressource intéressante pour constituer de véritables réservoirs de documents numériques pour le développement des systèmes de recherche d'informations ; ce qui permet d'évoluer notamment d'un Web traditionnel à un Web sémantique [Abde 06]. Le XML permet également de décrire les structures d'arbres avec une syntaxe linéaire ; celui-ci offre des installations aux utilisateurs pour définir leurs propres balises nécessaires à la description de la structure des documents.

Le contenu d'un document Web peut être facilement traité parce que le XML fournit un moyen d'échanger des informations sur le web. Le XML permet de séparer la présentation du contenu des documents. C'est un langage de base pour le Web sémantique, conçu essentiellement pour décrire des données en se concentrant sur leur structure et aussi pour assurer leur interopérabilité. Pour rendre possible cette méthode, XML utilise un «Document Type Définition» (DTD) et un XML Schéma pour un « modèle » de données, comme illustré par la figure 2.2.

```

1      <?xml version="1.0" encoding="UTF-8"?>
2      <!DOCTYPE etudiant SYSTEM "malade.dtd">
3      <etudiant>
4          <identite>
5              <nom>N'guessan</nom>
6              <prenoms>Gerard</prenoms>
7              <matricule attr="N°_unique">10AAA23308</matricule>
8              <age>45</age>
9              <maladie>
10                 <diagnostic>Paludisme</diagnostic>
11                 <temperature>40°C</temperature>
12            </maladie>
9          </identite>
6      </etudiant>

```

Figure 2.2 : Exemple de document XML représenté sous forme textuelle

2.8.2. Langage RDF (Resource Description Framework)

Bien qu'XML fournisse un langage pour décrire la structure de l'information, il ne parvient pas à définir de manière compréhensible et traitable la sémantique formelle des documents dans la machine. Ce sont les humains qui donnent un sens, une sémantique aux balises et leur contenu pour les machines. Le RDF (Resource Description Framework) est un langage formel développé par le W3C destiné à décrire les ressources Web et leurs métadonnées. Il permet le traitement automatique de telles descriptions et assure l'interopérabilité entre applications échangeant de l'information non formalisée et non structurée sur le Web. Le langage RDF vient donc remplir ce vide puisque c'est un langage qui permet de représenter des informations sur des ressources sous la forme de graphes orientés avec des nœuds et des arcs étiquetés (figure 2.3).

Un document RDF est un ensemble de triplets de la forme < sujet, prédicat, objet >. Les éléments de ces triplets peuvent être des URIs (Universal Resource Identifiers), des littéraux ou des variables. Cet ensemble de triplets peut être représenté de façon naturelle par un graphe, plus précisément un multi-graphe orienté et étiqueté où les éléments apparaissant comme sujet ou objet sont des sommets ; chaque triplet est représenté par un arc dont l'origine est son sujet

et la destination, son objet. Ce document sera codé en machine par un document RDF/XML (figure 2.4). Mais il est souvent représenté sous forme graphique. L'objectif de ce langage est précisément d'explicitier autant que possible le sens de l'information publiée sur le Web pour faciliter l'interopérabilité, l'agrégation des données dispersées et le raisonnement à partir de ces données. Etant un standard du W3C depuis 1999, le langage RDF est capable de décrire des données de toute nature. L'élément de base du RDF est la déclaration (ou « triplet »). Ce triplet RDF est composé, comme son nom l'indique, de trois (3) éléments à savoir (un sujet c'est-à-dire la ressource décrite, un prédicat qui attribue une propriété à la ressource décrite et un objet qui attribue une valeur à la propriété) et possède plusieurs syntaxes dont une syntaxe XML [Raim 13].

✓ **Exemple1 :**

Les maladies sont sur la page Web <http://www.inapg.fr/omip/maladie.htm>

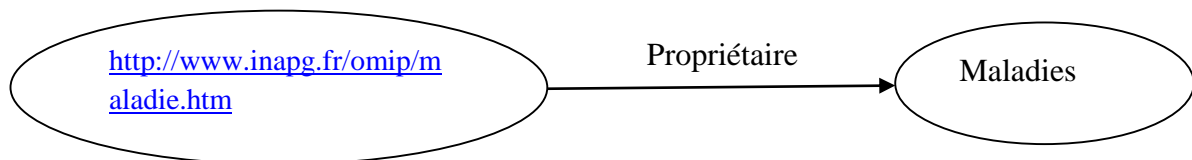


Figure 2.3 : Exemple de graphe RDF

✓ **Exemple2 :**

```

1  <? Xml version="1.0" encoding="UTF-8"?>
2  <rdf:RDF
3      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4      xmlns:monDom="http://www.inapg.fr/mon-rdf-ns#">
5      <rdf:Description rdf:about="http://www.inapg.fr/omip/maladie.htm">
6          <monDom:proprietaire>
7              Maladies
8          </monDom:proprietaire>
9      </rdf:Description>
10 </rdf:RDF>

```

Figure 2.4 : Document RDF simple

2.8.3. Langage RDF-Schéma

Le langage RDFS ajoute à RDF la possibilité de définir des hiérarchies de classes et de propriétés dont l'applicabilité et le domaine de valeurs peuvent être spécifiés à l'aide des attributs rdfs : domain et rdfs : range. Chaque domaine applicatif peut être ainsi associé à un schéma identifié par un préfixe particulier et correspondant à une URI. Les ressources instances sont ensuite décrites en utilisant le vocabulaire donné par les classes définies dans ce schéma. Les applications peuvent alors leur donner une interprétation opérationnelle. Cependant, on

peut noter que RDFS n'intègre pas en tant que tel des capacités de raisonnement mais plutôt des solutions de base de données dédiées à RDFS apparaissant comme l'architecture Sésame à laquelle est associé le langage de requête RQL [Phil 02].

2.8.4. Langage OWL (Web Ontology Language)

Le langage OWL (Web Ontology Language) a été conçu pour être utilisé par les applications qui traitent le contenu de l'information au lieu de la présenter seulement aux êtres humains. OWL facilite beaucoup plus l'interopérabilité au niveau machine du contenu du Web que ce qui est déjà supporté par XML, RDF et RDF Schéma (RDF-S). Il fournit un vocabulaire supplémentaire avec des sémantiques formelles, possédant ainsi des sous-langages de plus en plus expressifs, comme OWL Lite, OWL DL et OWL Full (W3C 2004). [Sab 04]. OWL, tout comme RDF est un langage XML profitant de l'universalité syntaxique de XML fondée sur la syntaxe de RDF/XML figure 2.5. En plus de posséder les fonctionnalités RDF, OWL est un langage d'ontologie qui intègre les outils de comparaison, des propriétés et des classes. Toutes ces fonctionnalités se résument en identité, équivalence, contraire, cardinalité, symétrie, transitivité, et disjonction, etc. [Xav 05]. Ainsi, si RDF et RDFS apportent à l'utilisateur la capacité de décrire des classes avec des constructeurs et des propriétés, le langage OWL profite de l'universalité syntaxique du XML en offrant la possibilité d'écrire des ontologies Web et permet de définir un vocabulaire extrêmement riche. Ce qui donne aux machines une plus grande capacité de manipulation du contenu Web. Le W3C a doté OWL de trois sous langages offrant des capacités d'expression croissantes:

- **OWL Lite** est un sous langage simple destiné à représenter des hiérarchies de concepts simples. Parmi les sous langages d'OWL, il est particulièrement bien adapté à la transformation des thésaurus comme WordNet en ontologie car il permet essentiellement de définir une hiérarchie de concepts. Toute ontologie valide en OWL Lite est une ontologie valide en OWL DL [Sab 04].

- **OWL DL**, ce langage a un fonctionnement plus complexe que le précédent. Il est fondé sur la logique descriptive, d'où son nom OWL Description Logics. Ce sous langage est adapté pour faire des raisonnements en garantissant la complétude et la décidabilité de ceux-ci. Toute ontologie valide en OWL DL est également valide en OWL Full. L'inverse n'est pas vrai : il faut y appliquer des extensions ou des restrictions [Thom 12].

- **L'OWL Full**, version la plus complexe de l'OWL, est destinée aux situations où il est important d'avoir un haut niveau de capacité de description, quitte à ne pas pouvoir garantir la complétude et la décidabilité des calculs liés à l'ontologie. C'est la version qui

présente le plus haut niveau d'expressivité. Comme RDF, il permet de redéfinir les primitives du langage. OWL Full est parfaitement compatible syntaxiquement et sémantiquement avec RDF : toute ontologie valide en OWL Full est également valide en RDFS. L'avantage de ce langage est sa forte expressivité. Néanmoins, OWL Full ne garantit ni la complétude des raisonnements car toutes les inférences ne sont pas calculables, ni leur décidabilité, le calcul ne se faisant pas nécessairement en un temps fini.

```

1  <owl:NamedIndividual rdf:about="&medtrad;VomissementsBilieux">
2    <rdf:type rdf:resource="&medtrad;Symptome"/>
3    <definition>Rejet par la bouche d&apos;un vomis amer.</definition>
4    <icone>Vomissement </icone>
5    <langue>Français</langue>
6    <icone>iconeVomissementsBilieux</icone>
7    <contexte>Médecine traditionnelle</contexte>
8    <terme>Vomissement</terme>
9    <estSymptomeDe rdf:resource="&medtrad;Paludisme"/>
10 </owl:NamedIndividual>

```

Figure 2.5 : Exemple de Description du modèle OWL

2.9. Différentes couches du web sémantique

Le Web sémantique étant une extension du Web classique, il repose sur une architecture en couches (figure 2.6) permettant à chacun de s'approprier les couches dont il a besoin. En d'autres termes, rien n'est imposé, tout est proposé : [Fab 06].

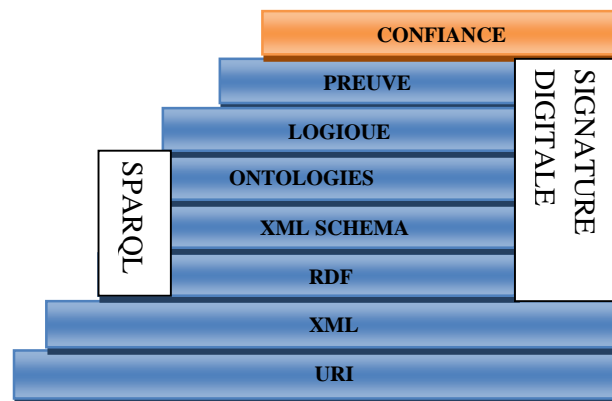


Figure 2.6 : Architecture du web sémantique

2.9.1. URI (Universal Resource Identifier)

L'URI (Universal Resource Identifier) permet d'identifier une ressource (physique ou abstraite) sur le web [Tan 04]. C'est un protocole simple et extensible pour identifier, de façon unique et uniforme, toute ressource sur le web. Il s'agit d'un aspect central de l'infrastructure. C'est pour cette raison que cet élément se trouve à la base de l'architecture en couches proposée

[Mest 07]. L'URI (Universal Resource Identifier) permet d'attribuer un identifiant unique non seulement à un ensemble de ressources sur le Web, mais aussi à d'autres éléments comme les documents, les téléphones, les portables, les personnes, etc. Cette notion connaît aujourd'hui de nombreuses extensions en cours de standardisation. Un URI peut être classé dans l'une des trois (3) catégories, selon qu'il est destiné à la localisation, au nommage ou aux deux (L'URL et l'URN).

Le terme URL (*Uniform Resource Locator*) désigne un sous-ensemble d'URI qui identifie les ressources via une représentation de leur mécanisme d'accès, plutôt que par le nom ou autre attribut de cette dernière, comme il en est le cas pour l'URN (*Uniform Resource Name*). L'URL et l'URN sont donc des cas particuliers d'URI. Elles sont à la base même des langages du W3C. Les URL sont des URI qui identifient la ressource par le moyen d'y accéder. Les URL peuvent être pérennes [Bern 11]. L'interconnexion des documents à l'échelle mondiale implique un moyen unique de les identifier sur le réseau Internet. Une page dont l'URL n'a pas été adressé peut de même être visitée si d'autres pages déjà référencées font un lien vers celle-ci [Tra 12].

2.9.2. XML

Le XML (eXtensible Markup Language) est un langage de formatage balisé de textes basé sur le standard de marquage de documents SGML (Standard Generalized Markup Language), dont est également issu HTML. XML est toutefois nettement plus puissant que HTML [Hair 00]. Sa structure est arborescente et très similaire au principe de l'orienté objet, chaque type de documents (article scientifique, bon de commande, document administratif, graphique, etc.) est représenté par plusieurs éléments (figure 2.7). Un article scientifique par exemple est représenté par son titre, l'auteur, le domaine de recherche, sa date de publication, etc. Chaque élément est lui-même décrit par d'autres propriétés (l'auteur est structuré par son nom, prénom, son grade, son adresse, etc.) et ainsi de suite. D'autres informations peuvent être ajoutées à l'intérieur même des balises. Cette façon de structurer un type quelconque de document porte le nom de DTD [Moh 02] (Document Type Définition) qui est une forme de grammaire du XML relativement ancienne car issue de l'univers SGML. Le DTD décrit l'ensemble des règles et des propriétés que doit suivre le document XML. Ces règles définissent généralement le nom et le contenu de chaque balise et le contexte dans lequel elles doivent exister [Tra 12] : "Définition de Types de Documents" ou de schémas XML [Moh 02]. Le XML décrit la structure stockée, transportée et échangée des données URLs. Dans la littérature relative au XML, on distingue les documents bien formés et les documents valides. Un

document est dit valide lorsque celui-ci respecte les règles spécifiques de son DTD. Un document bien formé est un document qui respecte les règles générales de la syntaxe XML. C'est-à-dire :

- La syntaxe des balises est respectée
- Les éléments qu'il contient sont bien imbriqués les uns les autres sous forme d'arbre.

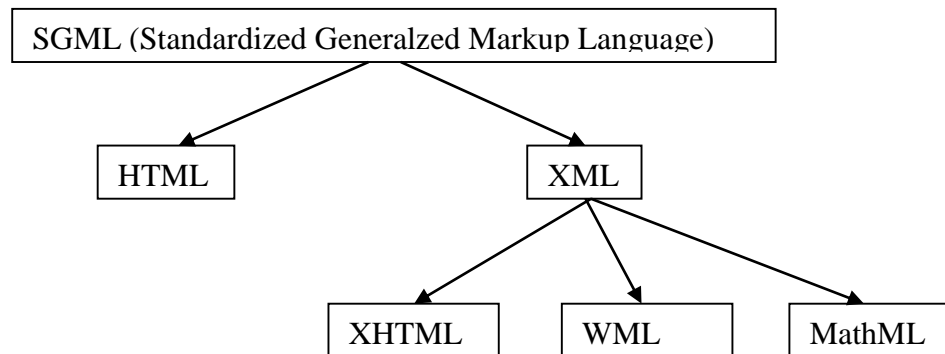


Figure 2.7 : Relation entre le SGML et XML.

2.9.3. RDF et le RDF SCHEMA

De manière générale RDF permet de voir le web comme un ensemble de ressources reliées par les liens étiquetés. Il est par ailleurs un des modèles de base et de syntaxe sur lequel le web sémantique se construit avec l'ajout de couches («layers»); caché au-dessus de RDF, on peut noter OIL (Ontology Inference Layer) et DAML (DARPA Agent Markup Language); OIL est utilisé pour définir des ontologies et DAML ajoute un petit nombre de caractéristiques au schéma RDF afin de rendre plus facile la définition de nouveaux langages permettant la communication entre agents intelligents. Initialement développées par la DARPA (pour DAML) et la Communauté Européenne (pour OIL), ces deux spécifications sont maintenant regroupées au W3C sous l'acronyme OWL [Blon 04].

Le langage OWL est une couche supérieure à RDF. Cela signifie qu'il réemploie le langage RDF avec certaines fonctions. Celles-ci permettent de structurer l'information selon des classes, des entités et des relations pour faciliter la modélisation d'un domaine. Ces fonctions offrent ainsi la possibilité de structurer l'information en ontologies, sur lesquelles vont pouvoir être plus facilement appliqués des raisonnements [Benj 09]. OWL permet d'exprimer des vocabulaires étendus, surtout quand il les complète avec RDFS qui offre un niveau supérieur de structuration. RDFS ajoute à RDF la possibilité de définir des hiérarchies de classes et de propriétés. En effet, l'application RDFS a pour but d'étendre le langage en décrivant plus précisément les ressources utilisées pour étiqueter les graphes. Pour cela, il fournit un mécanisme permettant de spécifier les classes dont les ressources seront des

instances, comme les propriétés [Berm 11]. Une propriété est spécifiée dans un schéma par une URI qui identifie la ressource, représente la propriété et/ou le type RDF. Le type de la ressource est RDF : property [Blon 04]. RDFS s'écrit toujours à l'aide de triplets RDF, en définissant la sémantique de nouveaux mots-clés.

Remarque : XML peut être vu comme la couche de transport syntaxique, RDF comme un langage relationnel de base. RDFS offre des primitives de représentation de structures ou primitives ontologiques.

2.9.4. XML-Schéma

XML-schéma tout comme le DTD est un langage de description de format de document XML. Ainsi donc, avant de commencer à organiser les informations dans un document XML, il est impératif de définir la structure de ce dernier, afin de permettre notamment de vérifier sa validité. En effet, il y a une légère différence entre le DTD et le XML-S. XML-S permet par exemple de définir des domaines de validité pour la valeur d'un champ, alors que cela n'est pas possible dans une DTD ; en revanche, il ne permet pas de définir des entités. Par ailleurs, il est à noter qu'un fichier XML-S est lui-même un fichier XML.

2.9.5. Ontologies

L'humain est capable, par la force de son langage, de référencer deux contextes différents en utilisant le même terme ou en employant plusieurs termes pour référencer la même idée. Or, l'automatisation ne peut s'en accommoder. Cela serait bien si la machine pouvait comprendre le sens des termes, même s'ils émanent de plusieurs bases de données. Comment régler le cas où un terme prend deux significations différentes ? Comment lever toute ambiguïté ? Ces questions sont traitées par le troisième pilier du web sémantique appelé ontologie [Moh 02].

Le terme « ontologie », construit à partir des racines grecques *ontos* (ce qui existe, l'existant) et *logos* (le discours, l'étude), est apparu pour la première fois dans un texte en 1613. Le mot anglais ontology est apparu pour la première fois dans « Universal Etymological English Dictionary » [Rad 09] en 1721. Il est apparu pour la première fois dans un titre de livre en 1733 [Jean 03]. C'est une description formelle des connaissances [Coul 08] qui est aussi un mot que l'informatique a emprunté à la philosophie au début des années 80 : on le trouve dans l'article de McCarthy et dans le livre de John Sowa avant que le terme « ontologie » ne devienne célèbre avec l'article de Thomas Gruber [Gand 10]. Une ontologie définit un ensemble de concepts, leurs caractéristiques et les relations qu'ils entretiennent entre eux. L'utilisation des ontologies est devenue de plus en plus courante dans les systèmes d'informations [Seb 08]. C'est une

théorie logique qui traduit clairement les notions mobilisées par la description d'une réalité des règles qui contraignent la structure de ces descriptions. L'ontologie permet donc de regrouper des connaissances relatives à un domaine donné. Une ontologie est une spécification formelle des concepts et des relations entre ces concepts" [Gand 10]. La figure 2.8 illustre l'exemple d'un cas de maladie tirée de la médecine traditionnelle.

Exemple : Un concept peut être une maladie, un symptôme, un remède et une plante. Une maladie a un nom ; un symptôme est un ensemble de signes cliniques, un remède a une description et une plante a un nom scientifique, etc.

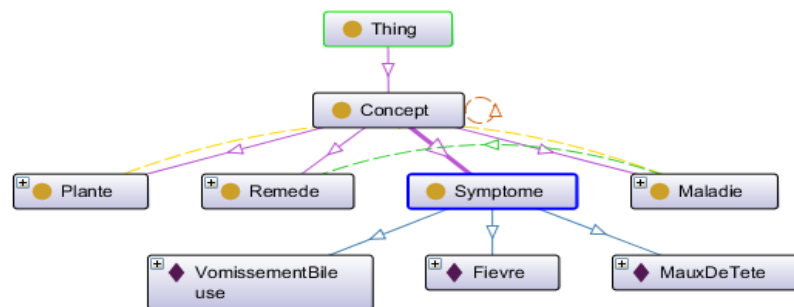


Figure 2.8 : Description de concept à partir d'une ontologie

En d'autres termes, une ontologie est une spécification formelle d'une conceptualisation partagée. Elle se compose donc d'un système symbolique d'opérations permises sur ce système (ex. règles de réécriture, règles d'inférence) et d'une interprétation. C'est aussi l'ensemble structuré des termes et concepts représentant le sens d'un champ d'informations tel que la représentation des concepts (principes, idées, catégories d'objets, notions potentiellement abstraites) et les relations entre ces concepts. Elle est employée pour raisonner à propos des concepts du domaine concerné [Seb 08]. Elle inclut généralement une organisation hiérarchique des concepts pertinents et des relations qui existent entre ces concepts, ainsi que des règles et axiomes qui les contraignent. Cela signifie qu'elle permet de donner un sens aux informations [Rok 07, Grig 08]. *Le langage de description d'ontologie du W3C est OWL : Web Ontology Language (après DAML+OIL)* [Tan 04]. Dans le cadre des mémoires collectives, une ontologie sert à capturer les notions qui interviennent dans les scénarios de recherche et de gestion de la connaissance. Elle fournit alors des primitives pour annoter des ressources et décrire des acteurs (groupes et individus). Les modèles à base d'ontologies permettent la conception d'inférences améliorant les mécanismes de gestion et d'exploitation de la mémoire [Fab 08]. Utiliser une ressource sémantique telle qu'une ontologie pourrait être un moyen d'enrichir les données cliniques en vue de répondre plus précisément à des questions d'ordre médical complexes. Car ceci apporte une compréhension partagée d'un domaine : interopérabilité sémantique :

- vaincre les différentes terminologies
- établir la cartographie entre ontologies.

En effet les ontologies sont utiles pour l'organisation et la navigation dans les sites web en se basant sur quelques langages.

2.9.5.1. Thésaurus

Nés dans les années 50, selon la norme internationale ISO 2788 depuis 1986, les thésaurus sont les « *vocabulaires d'un langage d'indexation contrôlés organisés formellement de façon à expliciter les relations a priori entre les notions (par exemple relation générique spécifique)* ». Selon la même norme, un langage d'indexation est un « *ensemble contrôlé de termes choisis dans une langue naturelle et utilisés pour représenter sous forme condensée le contenu des documents* » [Tric 00]. Les thésaurus sont également comme des vocabulaires contrôlés de termes représentant généralement un domaine particulier gérant des relations hiérarchiques, associatives et d'équivalence. On peut citer NLM's Medical Subject Headings (MeSH) dans le domaine médical pour indexer et rechercher des articles, le célèbre Wordnet, plus général, utilisé dans des travaux d'analyse sémantique [Noel 10]. L'usage d'un thésaurus autorise donc l'indexeur à utiliser autant de descripteurs que possible.

2.9.5.2. Graphes conceptuels

Le modèle des graphes conceptuels est un modèle général de représentation des connaissances fondé sur la description de concepts et sur la relation entre ces concepts [Tric 00]. Les connaissances exprimées dans ce modèle se structurent à deux niveaux. Dans un premier temps, au niveau terminologique où le vocabulaire conceptuel du modèle des graphes conceptuels est composé d'un ensemble ordonné de types de concepts principalement composés d'un treillis de concepts et d'un ensemble ordonné de relations conceptuelles. Dans un deuxième temps, nous avons le niveau assertionnel où les faits sont décrits par des graphes construits à partir des éléments du niveau terminologique [Mich 98].

Un graphe conceptuel est un graphe fini, connexe, biparti composé de sommets concepts représentant des entités, des attributs, des états ou des événements et de sommets relations décrivant la nature et les propriétés des liens entre les sommets concepts. Chaque sommet d'un graphe conceptuel est étiqueté :

- l'étiquette d'un sommet concept est un couple composé d'un type de concept spécifiant la nature de l'entité représentée et d'un référent permettant d'identifier cette entité. Ce référent

peut être un marqueur individuel si l'identité de l'individu est connue ou un symbole, appelé marqueur générique si cette identité n'est pas connue ;

– l'étiquette d'un sommet relation est un type de relation spécifiant la nature du lien qui existe entre les sommets concepts connectés par ce sommet relation.

2.9.5.3. Logiques de description

Les logiques de descriptions facilitent le développement des modèles ontologiques associés à des outils d'inférences. De même, elles offrent des techniques d'inférence et de la cohérence d'un ensemble d'axiomes [Nad 03]. En outre, les logiques de description sont une famille de langage de représentation de connaissances. En général, les connaissances sont divisées en deux parties. Les connaissances liées à la terminologie sont stockées dans une TBox, et celles liées aux individus sont stockées dans une ABox. Les logiques de description s'appuient sur les langages RDF et OWL pour manipuler des concepts, des individus (instances de concepts) et des rôles (relations entre concepts) [Thom 12]. Les concepts sont formés à partir des éléments suivants : le concept universel \top , le concept impossible \perp , les concepts atomiques (a et b) et (c ou d), les rôles (R et S), les constructeurs (\sqcup et \sqcap) et les quantificateurs (\forall , \exists). Un concept est défini comme un sous-ensemble d'un domaine ; et un rôle est interprété comme une relation binaire sur le domaine. Une interprétation est définie par un couple (Δ^I, \cdot^I) où Δ^I est le domaine d'interprétation et \cdot^I est la fonction d'interprétation. Le tableau 2.1 présente l'interprétation des différents concepts.

Tableau 2.1 : Sémantique des concepts et des rôles en logique de description..

Notation	Interprétation
\top	Δ^I
\perp	\emptyset
c	$c^I \subseteq \Delta^I$
R	$R^I \subseteq \Delta^I * \Delta^I$
$\neg c$	$\Delta^I \setminus c^I$
$d \sqcap c$	$d^I \cap c^I$
$d \sqcup c$	$d^I \cup c^I$
$\exists R.c$	$\{e_i \in \Delta^I \mid \exists e_j \in \Delta^I : (e_i, e_j) \in R^I \text{ et } e_j \in c^I\}$
$\forall R.c$	$\{e_i \in \Delta^I \mid \forall e_j \in \Delta^I : (e_i, e_j) \in R^I \Rightarrow e_j \in c^I\}$

En effet, une TBox est définie formellement par un ensemble fini d'axiomes, de rôles d'inclusion et de GCI (General Concept Inclusion). Une interprétation I est un modèle d'une TBox T s'il satisfait tous les axiomes de I . C'est-à-dire que :

– $c^I \subseteq d^I$ est vérifié pour chaque $c \sqsubseteq d \in T$, et

– $R^I \subseteq S^I$ est vérifié pour chaque $R \sqsubseteq S \in T$.

Soient $a, b \in I$ des noms d'individus, C un concept, et R un rôle. L'expression $c(a)$ est appelée une assertion de concept et $R(a, b)$ est appelée une assertion de rôle. La figure 2.9 présente le contenu d'une TBox. Elle spécifie que le symptôme du paludisme est soit une fièvre, soit un mal de tête, et que la dysenterie est une maladie ayant la diarrhée comme symptôme.

$\text{Symp_Paludisme} \sqsubseteq \text{Fievre} \sqcup \text{Mal_Tete}$ $\text{Dysenterie} \equiv \text{Maladie} \wedge \exists \text{Symp_Diarrhe}$
--

Figure 2.9 : Exemple d'une TBox

Une ABox est définie par un ensemble fini d'assertions de rôles et de concepts. Une fonction d'interprétation permet de lier chaque nom d'individu a à un élément a^I du domaine d'interprétation Δ^I . Une interprétation I satisfait :

- une assertion de concept $c(a)$ si $a^I \in c^I$, et
- un rôle d'assertion $R(a, b)$ si $\langle a^I, b^I \rangle \in R^I$.

Une interprétation qui satisfait toutes les assertions de concept et de rôle dans une ABox A est appelée un modèle de A . La figure 2.10 présente le contenu d'une ABox. Celle-ci spécifie qu'une fièvre est un symptôme, qu'un paludisme est une sensation, et que le paludisme est une maladie caractérisée entre autres, par la fièvre.

$\text{Symptome}(\text{Fievre})$ $\text{Sensation}(\text{Paludisme})$ $\text{Maladie}(\text{Paludisme}, \text{Fievre})$

Figure 2.10 : Exemple d'une ABox

2.9.5.4. Ontologie et web-sémantique

J. Charlet, dans les années 2002, déclare que, l'apparition des ontologies entendues comme une nouvelle approche de la modélisation des connaissances, s'est faite clairement dans le contexte de l'ingénierie informatique [Bour 02]. Les ontologies s'inscrivent dans la continuité de nombreux travaux sur la représentation des connaissances, des réseaux sémantiques, des cartes conceptuelles, des graphes conceptuels. Leur popularité a principalement bénéficié du développement du Web Sémantique, une vision prospective et normative du web proposée par Tim Berners Lee en 1994 mais qui n'a pas connu à ce jour le

succès escompté [Man 07]. Comme les principaux langages de représentation des connaissances issues des sciences cognitives et de l'Intelligence Artificielle, les ontologies visent également à décrire des concepts qui sont appréhendés comme des représentations mentales plus ou moins universelles ou comme des catégories a priori largement partagées dans la droite ligne de la philosophie de la connaissance. On distingue des ontologies de différents niveaux de généralité : des ontologies dites de haut niveau qui contiennent « des concepts très généraux comme l'espace, le temps, la matière, les objets, les événements, les actions, etc. qui ne dépendent pas d'un problème ou d'un domaine particulier » ; des ontologies de domaine (médecine, architecture, mécanique...) ; des ontologies de tâche (diagnostiquer, enseigner...) ; voire des ontologies d'application dans lesquelles les concepts appartiennent à un domaine et à une tâche particulière (enseigner la médecine).

2.9.5.5. Composantes d'une ontologie

Les connaissances traduites par une ontologie sont à véhiculer à l'aide des éléments suivants [Psy 03] :

- **Les concepts** : aussi appelés termes ou classes de l'ontologie ; ils correspondent aux abstractions pertinentes d'un segment de la réalité (le domaine du problème) retenues en fonction des objectifs qu'on se donne et de l'application envisagée pour l'ontologie. Ces concepts peuvent être classifiés selon plusieurs dimensions :

- le niveau d'abstraction (concret ou abstrait) ; l'atomicité (élémentaire ou composée) ;
- le niveau de réalité (réel ou fictif).

- **Les relations** : elles traduisent les associations pertinentes existant entre les concepts présents dans le segment analysé et la réalité. Ces relations permettent d'apercevoir la structuration et l'inter-relation des concepts les uns par rapport aux autres. Ces relations incluent également les associations suivantes :

- Sous-classe-de (généralisation –spécialisation ou « is-a ») ;
- Partie-de (agrégation ou composition ou « part-of ») ;
- Associée-à ;
- Instance-de, etc.

Les relations représentent un type d'interaction entre les notions d'un domaine. Elles sont formellement définies comme tout sous-ensemble d'un produit de n ensembles, c'est-à-dire $R = C_1 * C_2 * \dots * C_n$

Ex.1. Sous-classe-de : *concept1* * *concept2*

Ex.2. La *Maladie_1* est connectée-à la *Maladie_2* comme le montre la figure 2.11 suivant.

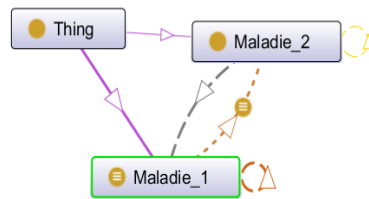


Figure 2.11 : Exemple de connexion

• **Les fonctions** : constituent des cas particuliers de relations dans lesquelles un élément de la relation, est défini en fonction des $n - 1$ éléments précédents. Formellement, les fonctions sont définies ainsi :

$$F : C_1 * C_2 * \dots * C_{n-1} \rightarrow C_n$$

Ex. Le prix d'une voiture usagée est fonction de son modèle, de son âge et de son kilométrage ; ce qui peut être représenté comme suit :

$$Modele * Annees * Kilometres \rightarrow Prix$$

• **Les axiomes** constituent des assertions acceptées comme vraies à propos des abstractions du domaine traduites par l'ontologie.

• **Les instances** constituent la définition en extension de l'ontologie. Ces objets véhiculent les connaissances (statiques, factuelles) à propos du domaine du problème.

2.9.5.6. Logique

La couche logique repose sur les langages ontologiques dans l'architecture recommandée par le W3C. En général, cette couche est utilisée pour exprimer les règles d'inférences. On dispose aujourd'hui d'une large variété de logiques. Étant le formalisme le mieux apprécié dans la représentation de la connaissance, la logique descriptive est celle qui est généralement la plus adoptée pour la représentation des règles d'inférences. La logique descriptive est définie comme « *une famille de formalismes de représentation de la connaissance basée sur la logique. Elle est conçue pour représenter et raisonner sur la connaissance d'un domaine d'application d'une manière structurée et bien comprise. Elle dérive des réseaux sémantiques* » [Mest 07].

Il existe deux concepts de base dans les logiques descriptives à savoir, les concepts (prédicats) et les rôles (relations binaires). Les notions de **Satisfiabilité** et de **subsumption** sont fondamentales dans la logique descriptive. Un concept « C » est dit satisfaisable quand son expression n'aboutit pas à un ensemble vide. Il existe donc une interprétation possible de

ce concept. La Satisfiabilité est un cas particulier de la subsumption qui n'est autre que la relation hiérarchique entre deux concepts.

Il existe un autre type de logique nommé logique normative ayant pour objet la manière dont l'homme doit raisonner. La logique, en tant que discipline, était à l'abri de toute espèce d'enquête empirique sur la manière dont les humains raisonnent ou sur la manière dont ils apprennent à le faire [Herm 10]. Elle permet d'exprimer des « règles » de raisonnement, des règles qui permettent de déduire de nouveaux faits à partir de faits existants [Guy 07]. Elle représente l'activité mentale par laquelle on produit un argument complet, ou par laquelle on produit ou on évalue la conclusion d'un argument ou de sa preuve [Pier 11].

2.9.6. Preuve

Une preuve est l'application successive de règles qui permettent de déduire un nouveau fait [Laub 04]. La couche Preuve a pour but de prouver la pertinence de l'information retournée par les couches de plus bas niveau et des déductions obtenues à partir des inférences. Une des façons de le faire est de garder la trace des sources d'information et des raisonnements effectués. Malheureusement, il n'existe pas encore de langage de preuve standardisé par le W3C². Un langage de preuve constitue un moyen simple pour prouver si une déclaration est juste ou pas. Une instance de ce dernier consiste en général en une liste de toutes les étapes d'inférence par lesquelles transite l'information en question [Mest 07].

2.9.7. Couche Confiance

Dans l'architecture proposée par Tim Bernes-Lee [Bern 01], la couche Confiance a pour objectif d'évaluer la fiabilité de l'information et des raisonnements. Cette couche repose sur les signatures numériques, le cryptage des données et sur la fiabilité des sources d'information (agents de confiances, certifications, etc.). Le web ne pourra atteindre son plein potentiel que si les utilisateurs ont confiance dans les transactions et la qualité de l'information fournie [Fab 06]. Si on utilise des agents pour prendre des décisions à notre place, il faudrait qu'on puisse avoir « confiance » aux résultats. Pour ce faire, il faut que l'agent utilisé puisse : expliquer clairement comment il arrive à ses conclusions (preuve), garantir la fiabilité et l'origine des informations utilisées (signature digitale) [Laub 04].

² <http://theses.ulaval.ca/archimede/fichiers/24629/ch02.html>

2.9.8. Langage de requête SPARQL

Le Web sémantique définit un protocole et un langage de requête pour interroger les graphes de connaissances exprimés en RDF (par extension RDFS et OWL) nommé SPARQL Protocol and RDF Query Language (SPARQL). C'est une recommandation de la W3C depuis 2008 capable d'exprimer une requête et de retourner des résultats en utilisant la formalisation d'une ontologie. SPARQL est décrit par le W3C comme le langage de requête et le protocole de transport des données sur le Web de Données [Pier 11]. Le protocole permet à un client Web de consulter, en exécutant une requête SPARQL, un service ou point d'accès SPARQL qui traitera la requête pour retourner la réponse sous différents formats (HTML, XML, RDF/XML, N3, JSON etc.). Ce langage est fondé sur l'union de motifs de graphes. C'est-à-dire que la requête est formée [Benj 09] d'un graphe RDF où certaines parties sont remplacées par des variables. Ce langage répond à la requête en fournissant tous les graphes RDF correspondants. SPARQL permet d'interroger des descriptions RDF en utilisant des clauses (similaires dans certains cas à celles du langage SQL) telles que PREFIX (qui spécifie l'adresse exploitée dans la construction de la requête), SELECT [FROM] WHERE (requête interrogative), CONSTRUCT (requête constructive), UNION, OPTIONAL (jointures, conditions optionnelles), FILTER (conditions obligatoires) et DESCRIBE, ASK (description d'une ressource, évaluation d'une requête) [Arian 10]. Cet opérateur permet de considérer SPARQL comme un langage de règles de transformation pour des métas modèles exprimés en RDF. Plusieurs travaux ont démontré cette capacité parmi lesquels nous pouvons citer :

- Le projet caBIG (cancer du Biomedical Informatics Grid) qui utilise le langage SPARQL pour formuler des requêtes au cours de la représentation sémantique des services de données liées au cancer.
- Le projet européen DebugIT qui permet d'interroger des données hétérogènes pour en extraire une connaissance susceptible d'alimenter des systèmes experts d'aide à la prescription d'antibiotiques [Pier 11].

La syntaxe de SPARQL est proche de celle de SQL dont voici un exemple d'utilisation de la clause SELECT :

```

1 PREFIX ex: <http://ics.upmc.fr/schema#>
2 SELECT ?name
3 WHERE {
4     ?person ex:name ?name .
5     {
6         { ?person rdf:type ex:Adult . }
7         UNION
8         { ?person ex:age ?age .
9         FILTER (?age > 17) }
10    }
11 }

```

Figure 2.12 : Exemple de requête SPARQL

2.9.9. Inter-opérabilité d'ontologie (Co-ontologie)

La plupart des systèmes ont été conçus pour le suivi d'une maladie chronique particulière et utilisent un ensemble de matériels assez important, notamment en termes de capteurs. En effet, la plupart des personnes âgées souffrent à la fois de plusieurs troubles chroniques liés à l'âge. Il est donc envisageable de multiplier les systèmes de suivi, de s'intéresser à l'interopérabilité de ces systèmes de manière à mutualiser le matériel commun pour réduire les différents coûts. Il s'agit d'éviter la prolifération de systèmes qui présentent peu d'intérêt puisque la plupart utilise des équipements similaires pour effectuer leurs mesures. Dans la pratique, cette mutualisation et cette prise en compte pluri-pathologique seront traduites par des plateformes de suivi fortement évolutives pouvant intégrer les connaissances relatives à différentes pathologies. Ces connaissances et les ressources utilisées sont donc naturellement hétérogènes et rendent leur compréhension et leur analyse très difficiles. La préservation du sens de l'information échangée est alors un problème important. C'est ce que l'on appelle « l'interopérabilité sémantique ». L'interopérabilité est la capacité que possède un produit ou un système dont les interfaces sont intégralement connues, à fonctionner avec d'autres produits ou systèmes existants ou futurs et ce, sans restriction d'accès ou de mise en œuvre. Une définition est communément admise pour l'interopérabilité sémantique : *"elle donne un sens aux informations échangées et s'assure que ce sens est commun dans tous les systèmes entre lesquels des échanges doivent être mis en œuvre"*[Haj 12].

2.9.10. Conclusion

Dans ce chapitre nous avons non seulement présenté les atouts de la technologie web sémantique ; mais également son importance, ses avantages et ses inconvénients.

Ce chapitre a aussi traité de manière générale quelques langages du web sémantique ; il a présenté l'architecture du web sémantique et décrit le rôle de chaque couche. Par ailleurs, nous avons relevé l'importance des ontologies dans le web sémantique et quelques composants de

cette technologie. De plus, cette étude a permis de comprendre que les ontologies donnent une meilleure structuration des données du web sémantique.

Comme les ontologies sont importantes, nécessaires et indispensables dans le fonctionnement du web sémantique, nous allons, dans le prochain chapitre, essayer de montrer le mode de construction d'une ontologie qui est un outil de représentation des connaissances.

Chapitre 3

CONSTRUCTION D'UNE ONTOLOGIE

***Résumé.** L'intérêt croissant que portent les ontologies en tant que technologies de base pour les systèmes fondés sur les connaissances, implique le raffinement des méthodes de développement des ontologies. Dans ce chapitre, nous proposons une nouvelle approche pour développer une ontologie de la médecine traditionnelle. Le but est d'élaborer un algorithme d'apprentissage utilisant les éléments fondamentaux de la science médicale traditionnelle.*

3.1. Introduction

Les ontologies sont des moyens indispensables pour représenter, exploiter les données, les connaissances d'un domaine, et plus particulièrement celles du domaine médical. Ces données et ces connaissances médicales ont été structurées au moyen de taxonomies et de classifications longtemps avant que les ontologies ne deviennent indispensables pour le Web sémantique [Rad 09]. De manière générale, l'utilisation de connaissances en informatique a pour but de permettre un dialogue et une coopération entre les utilisateurs et les systèmes d'enseignement assisté.

La présence des ordinateurs et des réseaux « internet » dans le monde, permet de faciliter, d'une manière générale, la pratique de la médecine moderne. Les praticiens de cette science peuvent prendre avantage sur la connaissance de la médecine traditionnelle dès lors que celle-ci aura complètement intégré le système des TICs ; ainsi, il sera construit un mode de gestion des connaissances [Ghis 08]. Pour ce faire, le système doit avoir accès non seulement aux termes utilisés par l'être humain mais également à la sémantique qui est associée à ces termes afin qu'une communication efficace soit possible. Les ontologies visent en effet, à représenter cette connaissance qui soit à la fois interprétable par l'homme et par la machine. Il faudrait un document formel dont le contenu et la sémantique soient traitables par des systèmes informatiques ; la connaissance qui y est exprimée doit être obtenue de manière consensuelle [Mich 09]. Il faut aussi construire des ontologies qui soient des processus déjà suffisamment complexes ; cette complexité est d'autant plus importante que les développeurs implémentent directement cette technologie dans un langage formel sans avoir à leur disposition un outil d'aide à la construction. Pour faciliter cette tâche, les premiers environnements de construction d'ontologies tels que l'outil « proteger » sont apparus dans les années 90. Ces outils aident l'utilisateur à effectuer certaines des tâches principales du processus de développement d'ontologies, telles que la conceptualisation, l'implémentation, le contrôle de cohérence et la documentation. Au cours de ces dernières années, le nombre d'outils pour le développement

d'ontologies a considérablement augmenté et s'est diversifié [Abd 06]. Le processus de conception d'une ontologie de domaine requiert une longue phase de collecte de connaissances auprès d'experts (éventuellement aidés par des traitements d'analyse lexicale de corpus du domaine) associée à une définition claire et sans ambiguïtés des différents concepts et de leur formalisation. Cette phase est d'autant plus longue qu'elle nécessite que ces experts s'accordent entre eux sur la définition des concepts qu'ils utilisent. Or, dans certaines situations, la connaissance peut être répartie entre différentes disciplines dont les membres n'ont pas forcément la même vision de leur domaine et n'utilisent pas les mêmes terminologies ; ceci implique de longs débats avant d'aboutir à un consensus. C'est particulièrement vrai dans le domaine des sciences de la vie où se côtoient des phénomènes biologiques, chimiques, physiques, etc [Syl 13]. L'utilisation d'ontologies dans un modèle de recherche d'informations a pour finalité de spécifier des connaissances qui seront interprétables d'une part par l'utilisateur du système et d'autre part par le système lui-même.

Dans ce chapitre nous proposons de construire un algorithme d'apprentissage utilisant les éléments fondamentaux de la science médicale traditionnelle à savoir l'ensemble des sémantiques sous-jacentes de concepts propres à cette science.

3.2. Origine de la construction d'ontologie

Les catégories d'Aristote peuvent être considérées comme la première tentative de construction d'une ontologie de ce qui existe [Rad 09]. En effet, Aristote a identifié et nommé dix catégories qui servent à classer n'importe quel objet ou être vivant. Ces catégories peuvent nous sembler curieuses aujourd'hui. Mais à l'époque, ce travail était original et, avec le recul, on peut le juger remarquable. Elles étaient établies sous forme de liste plate. Mais cinq siècles plus tard, Porphyre les a organisées en une structure d'arbre et a fourni les principes de base pour différencier les nœuds pères des nœuds fils ainsi que les nœuds du même niveau (les nœuds frères), chaque nœud ayant un père et un seul. Ces principes consistent à identifier des ensembles de caractéristiques qui distinguent deux nœuds proches. Ces principes sont connus sous le nom de principes différentiels et sont à la base de plusieurs approches contemporaines de construction d'ontologies [Rad 09].

3.3. Ontologies biomédicales

Dans le domaine biomédical, l'usage de l'ontologie a rapidement évolué. Dans certains pays comme les Etat Unis, il existe des Centres Nationaux d'Ontologies Biomédicales et même des laboratoires d'ontologies biomédicales contenant plus de 300 ontologies et terminologies

contrôlées. Les biologistes utilisent des ontologies biomédicales pour gérer de grands volumes de données. Les hôpitaux et entités apparentées les utilisent dans le processus d'enregistrement de l'information médicale. Ceci constitue un support puissant pour la prise de décision en matière de diagnostic. Ainsi, Rector et *al.* ont publié un modèle dans « SNOMED CT », qui est une des ontologies biomédicales les plus largement utilisées [Jon 12].

3.4. Domaine visuel

Certains domaines de l'information comme la médecine ou la géologie et la plupart des sciences naturelles nécessitent une connaissance visuelle en tant qu'un élément crucial du problème des processus de résolution. Le processus d'interprétation se fait par un motif visuel correspondant au domaine pour capturer les objets qui peuvent soutenir le chemin d'inférence. Certains de ces objets visuels peinent même à faire la traduction de la description propositionnelle. Par conséquent, l'ontologie construite dans les domaines visuels nécessite plus de descriptions symboliques pour expliquer les concepts. La visibilité est incluse car il est important de représenter l'ontologie dans un mode visuel. La présentation visuelle d'ontologie facilite la présentation de l'ontologie et permet aux utilisateurs de comprendre facilement le contenu de l'ontologie [Ahs 05]. C'est ainsi que de nombreux développeurs d'ontologies trouvent qu'il est plus facile de fournir des descriptions de leurs concepts de domaines et propriétés à l'aide de représentations visuelles plutôt que de faire seulement des descriptions formelles (ou des descriptions pures en langage naturel) [Gab 11].

3.5. Etapes de la construction d'ontologie

La construction d'une ontologie nécessite une étape d'élicitation des connaissances du domaine visé (c'est à dire une étape de formalisation dans un moteur d'inférence donnant la possibilité de reproduire artificiellement l'analyse d'une situation et la prise de décision de l'expert), suivie d'une étape de formalisation et d'une étape de validation. La nature des connaissances à représenter, l'hétérogénéité du support des connaissances à formaliser (documents plus ou moins formels, communication orale ou écrite, représentations graphiques, connaissances tacites, etc.) et le manque d'assistance automatisée rendent laborieuse la démarche de conception d'une ontologie [Mich 09]. Les ontologies dégagent et modélisent les primitives c'est à dire les aspects essentiels de l'ensemble des connaissances du domaine en question et ce, dans un langage formel de représentation qui sera utilisé pour exploiter ces mêmes connaissances dans des applications informatiques. La construction d'ontologies est certes un problème difficile mais la mise à disposition de méthodes peut en faciliter la

résolution. Cependant, de telles méthodes doivent tenir compte du contexte dans lequel elles sont élaborées. Le processus de construction d'une ontologie se fait en plusieurs étapes telles que décrites dans la figure 3.1.

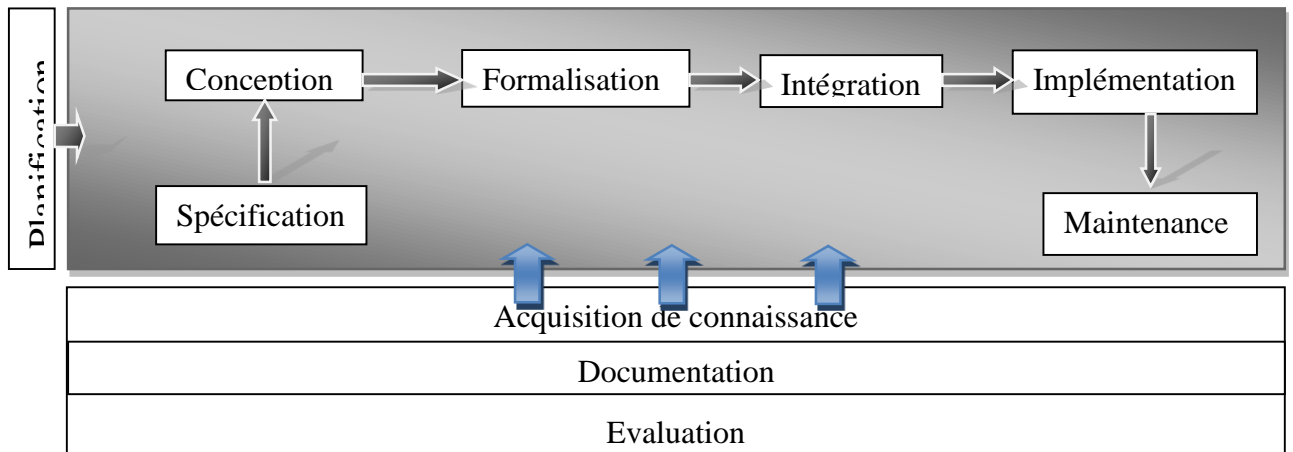


Figure 3 1 : Le cycle de vie d'une ontologie

3.5.1. Acquisition de connaissances

Toutes connaissances ou toutes méthodes servant à déclencher une analyse et ce, dans quelque domaine que ce soit, suppose qu'au préalable soient disponibles des informations existantes ou relevant d'analyse antérieure d'experts. L'acquisition de ces connaissances répond de ce fait, à un certain nombre de principes que sont:

- L'analyse approfondie des lexiques spécialisés comme la source de la connaissance.
- Le contrôle scientifique du glossaire conceptuel d'une discipline est une question importante, surtout dans le traitement naturel et artificiel d'une langue.
- La combinaison du vocabulaire précommandé et du vocabulaire post-contrôle ; ceci, pour avoir une description à plusieurs niveaux du glossaire conceptuel tels que la morphologie, le lexique, la sémantique, la pragmatique etc. [Xuez 04]

Dans le cadre de la médecine traditionnelle il importe de disposer de :

- Une bonne construction du cadre conceptuel de cette science pour répondre à tous les problèmes d'ingénierie de la connaissance des disciplines connexes.
- Un système complet de relations avec d'autres disciplines telles que la médecine moderne, la nature, l'agriculture, les sciences humaines et biologiques, etc.
- Un système de connaissances de cette médecine lié à des structures conceptuelles sémantiques complexes et à des types de relations qui se réfèrent au contenu multi-disciplinaire.

- Une analyse complète et de recherche sur les terminologies de la MT avant la conception de l'ontologie.
- Un contrôle de la combinaison des vocabulaires et des lexiques spécialisés.
- Une infrastructure de concept unifié sur la base des vocabulaires contrôlés et des lexiques spécialisés.

Remarque : Le vocabulaire contrôlé peut être considéré comme une ontologie qui n'a pas de cas. Le lexique de spécialiste peut être considéré comme les instances de l'ontologie. L'ensemble constitue une base de connaissances.

3.5.2. Evaluation

L'accent est mis ici sur l'étape de la "methontology" qui consiste à construire une ontologie à partir des termes extraits des ressources qui ne sont pas spécifiées. Ces ressources ont pour objectif de donner une définition à chaque concept et à chaque relation de l'ontologie dans un langage de Logique de Descriptions (LDs). La "methontology" nous spécifie les étapes à suivre pour construire une ontologie [Bend 07]. Les techniques utilisées dans cette partie sont essentiellement basées sur celles utilisées dans la validation et la vérification de Système à Base de Connaissances (SBC). Un ensemble de directives est donné sur la façon de chercher des imperfections, des incohérences et des redondances [Jon 98]. L'ontologie construite doit être évaluée en prenant des critères d'évaluation qui peuvent être classés en deux types à savoir : les critères génériques et les critères spécifiques. Les critères génériques traitent des facteurs tels que la clarté, la cohérence, la réutilisabilité. Quant aux critères spécifiques, ils vérifient l'ontologie générée contre les exigences d'usage de l'utilisateur. Sur papier, l'évaluation s'effectue de deux manières :

- l'évaluation interne faite par l'équipe de construction de l'ontologie,
- l'évaluation externe réalisée par l'utilisateur.

Les questions de compétence doivent également être évaluées pour le bien-fondé de l'ontologie.

En d'autres termes, l'évaluation de l'ontologie doit être effectuée à deux niveaux :

- le premier niveau traite de la pertinence du modèle sémantique de l'ontologie et
- le second niveau concerne la méthode utilisée pour l'extraction de l'ontologie.

Pour une bonne évaluation, il faut prendre également en compte deux types de paramètres : les paramètres liés à la précision de la relation taxonomique et ceux liés à la relation non taxonomique. La taxonomie étant définie comme la science des lois de la classification qui peut être une suite d'éléments formant des listes qui concernent un domaine, une science etc [Akil 11].

3.5.3. Intégration

L'intégration d'une ontologie donne une implémentation appropriée pour les problèmes de cycle et de séquence [Bati 06]. C'est aussi une étape permettant d'obtenir une certaine uniformité dans les ontologies ou dans les définitions d'autres ontologies telles que Ontolingua qui est une unité d'ontologie standard, intégrée [Akil 11]. Les ontologies sont maintenant en cours de construction dans les plates-formes de collaboration en utilisant des méthodologies et des cadres de prototypage rapide comme par exemple OntoEdit [Bati 06]. Cet outil permet l'inspection, la navigation, la codification et la modification des ontologies. L'utilisation de ces fonctionnalités supporte le développement de l'ontologie et l'entretien d'une tâche [Wach 01]. C'est un outil qui permet également d'éditer une hiérarchie de concepts qui peuvent être abstraits ou concrets et qui indiquent si oui ou non il est permis de faire des instances directes du concept.

L'intégration d'une information est importante pour les systèmes à base d'ontologies médicales. Cependant, l'intégration d'une ontologie de la médecine traditionnelle n'est possible que si les modèles destinés à la conceptualisation originale se chevauchent, ou bien les deux ontologies associées se chevauchent dans la procédure de développement. En effet, intégrer une ontologie de la médecine traditionnelle n'est pas une chose facile puis qu'il faut prendre en compte les différentes ontologies de chaque secteur du domaine.

3.5.4. Formalisation

La question clé dans la phase de formalisation est de décider si les concepts doivent être formalisés comme des classes ou des instances. Plusieurs discussions et expériences ont été nécessaires pour comprendre la meilleure façon de formaliser la conceptualisation définie précédemment. Certains concepts ont été formalisés sous forme de classes et leurs instances utilisent les hiérarchies définies comme des taxonomies où les valeurs des attributs sont de type classe.

3.5.5. Spécification

La Spécification est une étape qui permet d'identifier d'une part le but de l'ontologie, y compris les utilisateurs visés, les scénarios à utiliser, etc., et d'autre part sa portée, avec l'ensemble des conditions de représentations, de caractéristiques et de granularité nécessaire. La spécification des ontologies est partielle car une conceptualisation ne peut pas toujours être entièrement formalisée dans un cadre logique ; en effet, des ambiguïtés ou l'absence de représentation de leur sémantique dans le langage de représentation d'ontologies choisies

peuvent la contrarier. Cette phase se solde par un document de spécification de l'ontologie du langage naturel [Jon 98]. La spécification rend explicite le sens associé à ces concepts en leur donnant une définition [Nath 05].

3.6. Méthodologie de conception d'une ontologie

La conceptualisation permet d'identifier par un processus d'abstraction les concepts essentiels référencés par les termes du domaine. Chaque modalité de domaine est identifiée comme des concepts, des instances, des relations avec les verbes ou les propriétés et est représentée en utilisant une représentation informelle applicable [Jon 98]. La construction d'une ontologie est un travail important et précieux pour les données d'un domaine, en particulier pour le domaine complexe et complet tel qu'est la médecine traditionnelle (MT). Ces définitions peuvent être formelles ou non dans le cadre d'une manipulation de l'ontologie par des agents humains dès lors que les définitions de concepts en langage naturel sont nécessaires. En revanche, si l'ontologie est utilisée par des systèmes à base de connaissances, il faut associer à chaque concept une définition formelle écrite dans un modèle de représentation des connaissances. Ces représentations de connaissances correspondent à une spécification explicite et formelle d'une conceptualisation partagée ; étant au cœur du Web sémantique pour ajouter une couche sémantique au Web actuel, celles-ci font l'objet de nombreux travaux de recherche. Ces travaux s'attachent, d'une part, à définir des méthodologies et des techniques permettant leur élaboration à partir de textes et, d'autre part, à les utiliser dans les systèmes d'information. Une ontologie fournit une référence pour la communication entre les machines mais aussi entre humains et machines en définissant le sens des objets. Ceci est fait tout d'abord à travers les symboles (mots ou expressions) qui les désignent et les caractérisent et ensuite à travers une représentation structurée ou formelle de leur rôle dans le domaine [Nath 05].

Dans un système à base de connaissances, les expressions sont contrôlées par le recours à un langage de représentation. Celui-ci fournit le vocabulaire, les primitives et la grammaire pour créer des expressions valides ainsi que les éléments permettant de leur associer une interprétation formelle puis opérationnelle. Il faut alors construire une spécification de ce langage qui soit correcte vis-à-vis de la conceptualisation. Pour cela, on utilise des artefacts élaborés, des ontologies dont l'objectif est d'explicitier, autant que faire se peut, le sens du vocabulaire proposé par cette conceptualisation. Adosser un langage de représentation de connaissances à une telle ontologie s'appelle engagement d'ontologie. Cet engagement d'ontologie permet de donner une signification aux index qui seront construits par son moyen,

aussi bien pour le système chargé de leur manipulation que pour les utilisateurs humains qui vont les créer et les visualiser [Ant 05].

3.7. Classification des méthodes de construction d'ontologies

La construction des ontologies est classiquement divisée en une phase de repérage des concepts pertinents et une phase d'organisation de ces concepts, à l'aide de relations plus ou moins nombreuses et complexes. Les méthodes utilisables pour repérer les concepts d'un domaine et leurs relations appartiennent à deux grandes classes. La première, orientée du haut vers le bas (top-down), repose sur l'exploitation des conceptualisations théoriques du domaine faites par les experts. Pour la médecine moderne, il s'agit par exemple des formulaires structurés d'observation, des guides pour la pratique clinique, des ouvrages académiques. L'identification des concepts est plus ou moins directe selon le degré de formalisation de la ressource et la connaissance du domaine par l'ingénieur des connaissances. Lorsque la ressource est très structurée, avec une bonne connaissance du domaine, il est possible d'identifier directement les concepts sous-jacents. Aussi, lorsque la ressource est en texte libre et peu structurée, les outils de traitement automatique du langage naturel peuvent fournir une aide précieuse pour l'identification des concepts.

La seconde classe, orientée du bas vers le haut (bottom-up), s'appuie sur les traces textuelles laissées par les experts lors de la réalisation de leur activité. Pour la médecine moderne, il s'agit par exemple des observations médicales, des comptes rendus d'hospitalisation, des courriers etc. En effet, cette production textuelle contient les concepts de bas niveau dont les experts ont quotidiennement besoin. En revanche, ces textes ne contiennent pas les concepts de haut niveau qui structurent la représentation cognitive du domaine.

Néanmoins, les concepts structurants peuvent, dans une certaine mesure, être inférés des rapprochements entre les concepts de bas niveau opérés par une analyse terminologique. L'extraction des concepts et l'analyse de leurs interrelations nécessitent le recours à de gros corpus représentatifs de l'activité, et donc à des outils de traitement automatique du langage naturel pour leur analyse [Oliv 13].

3.8. Méthode d'élicitation

L'élicitation des connaissances est un domaine de recherche vaste et complexe qui a fait l'objet de nombreux travaux depuis plus d'une vingtaine d'années. De récentes recherches ont montré que des activités de transfert d'expertise impliquant la co-modélisation entre experts et novices permettent de produire des modèles de connaissances dont le contenu des propriétés

est consensuel et riche en connaissances du domaine. Afin de supporter et de stimuler la démarche intellectuelle des acteurs dans l'activité de co-modélisation, il est important de mettre à leur disposition un langage de représentation qui, par sa sémantique, impose une certaine structure à la pensée de manière à favoriser son déploiement, et ce, sans en entraver la créativité [Mich 09].

3.9. Création d'un Domaine

Une Ontologie est composée de quatre principaux éléments pour représenter un domaine. Ce sont :

i. le concept : il permet de représenter un ensemble d'entités dans un domaine [Akil 11]. Un concept peut avoir plusieurs noms qui sont essentiellement un moyen de faire la part des choses. De plus, cela nous permet de différencier les synonymes fins d'un concept donné [Wach 01]. Il peut participer à la mise en place d'une bonne relation entre plusieurs entités.

ii. la relation : elle spécifie l'interaction entre les concepts [Wach 01]. Les relations peuvent être commandées dans une hiérarchie qui permet l'héritage des relations et qui sont raffinées en imposant des restrictions sur les valeurs ou sur les cardinalités [Wach 01].

iii. l'instance : il indique l'exemple concret des concepts dans le domaine.

iv. les axiomes : ils représentent une déclaration qui est toujours vraie [Wach 01].

En prenant l'exemple d'ontologie du paludisme, nous pouvons citer comme plantes pouvant soigner cette maladie, l'écorce de nim, l'écorce de quinquina, la feuille de papaye, la feuille morte de cacao etc. Dans notre cas, les plantes représentent des concepts pour le paludisme et les relations sont ainsi données : l'écorce de nim soigne le paludisme, l'écorce de quinquina soigne le paludisme, la feuille de papaye soigne le paludisme. Les instances ici sont les couleurs de ces écorces qui peuvent être soit rouge, jaune ou marron. L'axiome ici, dans notre exemple est qu'une plante doit produire au moins une substance pour guérir le paludisme. Cette ontologie peut être spécifiée à différents niveaux de formalité en fonction de leur représentation :

1. La représentation très informelle exprimée dans un langage naturel ordinaire.

LE PALUDISME EST GUERI PAR L'ECORCE DE QUINQUINA

Figure 3 2 : Exemple de représentation très informelle.

2. La représentation semi informelle exprimée sous forme structurée du langage naturel.

L'ECORCE DE QUINQUINA GUERIT LE PALUDISME

Figure 3 3 : Exemple d'une représentation semi-informelle.

3. La représentation semi-formelle exprimée dans un langage formellement défini. Elle occasionne ainsi une économie de temps et surtout un gain de qualité dans la représentation des connaissances. De plus, la convivialité du langage semi-formel fait en sorte que des modèles de degré semi-formel peuvent être conçus par des informaticiens sans l'assistance d'un tradipraticien. Les informaticiens peuvent ensuite les formaliser avec la participation minimale du tradipraticien. Les opérations d'élicitations peuvent ainsi être intégrées de manière plus souple dans les activités des experts de contenu [Mich 09].



Figure 3 4 : Exemple de représentation semi-formelle.

4. La méthode formelle définit tous les termes de manière formelle, établit le théorème et les preuves. Il regroupe également un ensemble d'objets d'après les propriétés qu'ils partagent et de façon duale et regroupe également un ensemble de propriétés d'après les objets qui les possèdent [Bend 07] en s'appuyant sur la méthode de l'ACF (l'Analyse de Concepts Formels).

```

1  <owl:ObjectProperty
2  rdf:about="http://www.semanticweb.org/ngerard/ontologies/2014/0/untitled-ontology-18#SOIGNE">
3  <rdfs:domain
4  rdf:resource="http://www.semanticweb.org/ngerard/ontologies/2014/0/untitled-ontology-18#ECORCE_DE_QUINQUINA"/>
5  <rdfs:domain
6  rdf:resource="http://www.semanticweb.org/ngerard/ontologies/2014/0/untitled-ontology-18#PALUDISME"/>
7  <owl:propertyDisjointWith rdf:resource="&owl;topObjectProperty"/>
8  </owl:ObjectProperty>

```

Figure 3 5 : Exemple de représentation formelle d'une ontologie

L'usage d'un système de représentation plus conviviale permet ainsi d'élargir le bassin des personnes aptes à représenter leurs propres connaissances [Mich 09].

3.10. Construction d'ontologie

Les ontologies peuvent être construites en utilisant trois approches différentes : l'approche de l'ontologie unique, l'approche de l'ontologie multiple et l'approche hybride de l'ontologie.

L'approche de l'ontologie unique est la plus simple de toutes les approches. Elle utilise une seule ontologie globale pour toutes les sources d'information qui partagent le vocabulaire et la terminologie pour spécifier la sémantique. La limite de cette approche est qu'elle ne fournit pas une solution idéale pour l'intégration de l'information. Cette limite a ouvert la voie à

l'approche de l'ontologie multiple, où chaque source d'information est décrite par sa propre ontologie. Ainsi, chaque source aura sa propre ontologie locale. Le principal inconvénient de cette approche est la construction d'une ontologie individuelle.

L'approche hybride de l'ontologie est la combinaison d'approches uniques et multiples. Cette approche a été mise au point pour pallier les inconvénients de l'approche de l'ontologie unique ou multiple [Wach 01]. Elle s'appuie également sur l'analyse des corpus textuels construits. L'ensemble des n-grammes (séquence de n mots consécutifs) est extrait et analysé à l'aide d'un concordancier (Un concordancier est un logiciel qui permet de faire un tri rapide de tous les mots d'un texte ou d'un ensemble de textes), permettant le bon fonctionnement de certaines applications linguistiques [Ram 04]. C'est également un outil qui permet à l'utilisateur de formuler des requêtes sous forme d'expressions régulières dans le corpus et d'afficher toutes les parties du corpus contenant cette expression, pour enrichir l'ontologie au fur et à mesure de leur examen. L'ontologie utilisée est composée d'une hiérarchie de concepts potentiellement textuels. Elle est stockée sous forme d'un ensemble de fichiers textes extensibles afin de permettre à l'expert de rajouter ses propres informations. Chaque concept peut avoir des nœuds de niveau supérieur et inférieur (voir figure 3.7) [Der 08].

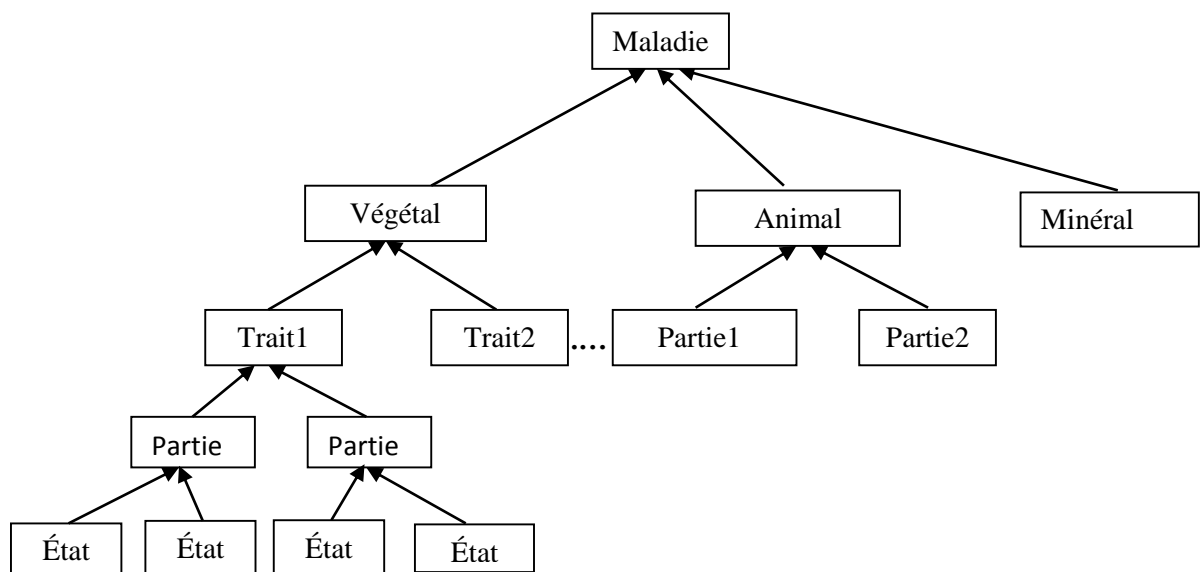


Figure 3 6 : Architecture de concepts d'une ontologie.

Un premier niveau de séparation est effectué entre les objets de type végétal, animal et minéral. Même si cette distinction n'a pas d'influence sur la description des concepts, elle permet de fournir une sémantique utile à l'expert du domaine.

Une fois la taxonomie établie, il convient de renseigner chaque concept par des contraintes, c'est-à-dire par des propriétés qui permettent d'identifier un élément du domaine comme étant une instance d'un concept. Comme c'est souvent le cas en gestion des connaissances, les

entretiens effectués avec des experts du domaine montrent que la formalisation de telles connaissances n'est pas aisée. Le fossé sémantique qui réside entre les descripteurs de haut niveau (abstrait) et de bas niveau (réalité) implique des difficultés de modélisation.

3.11. Algorithme de diagnostic d'une maladie

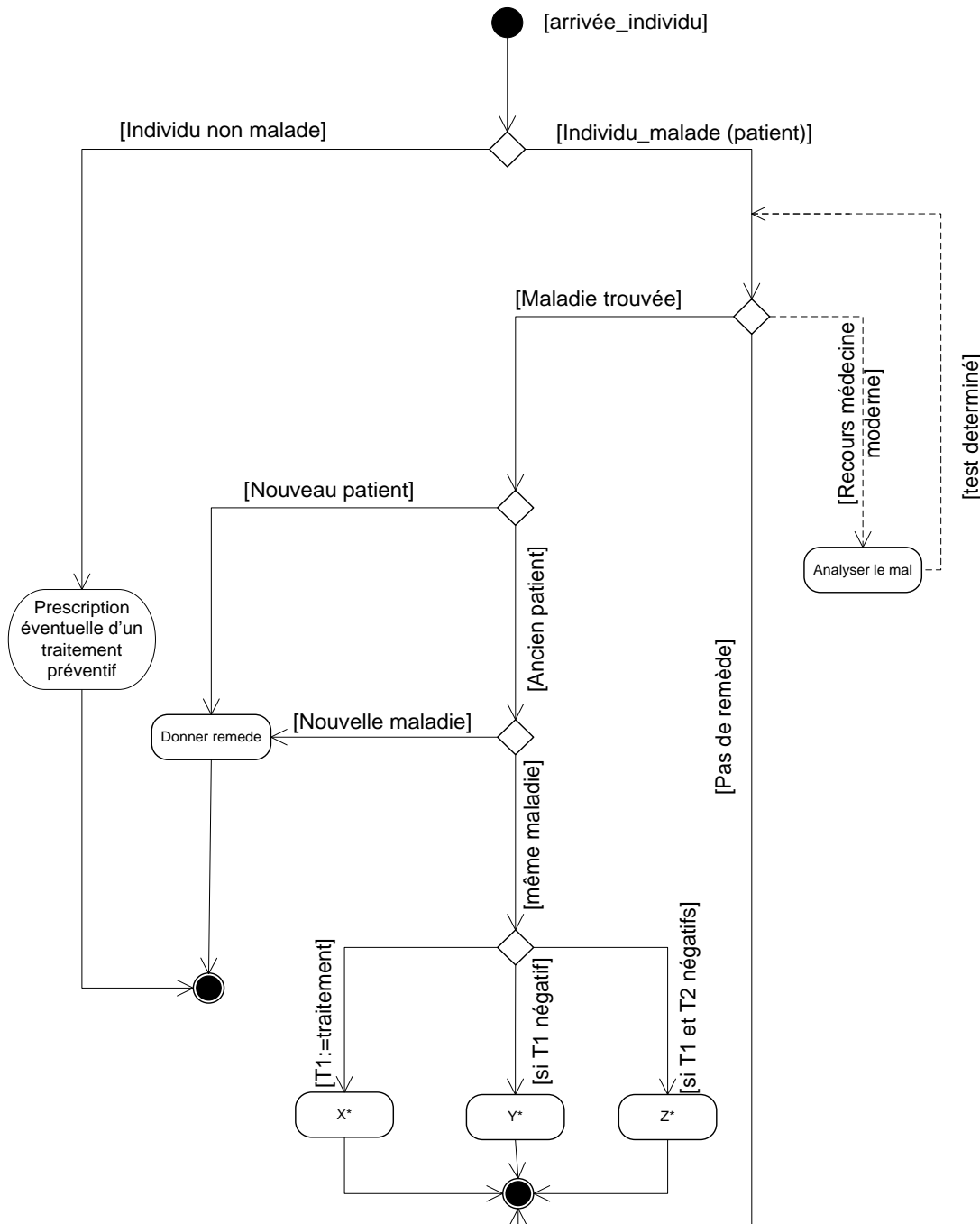


Figure 3 7 : Algorithme de diagnostic d'une maladie

3.11.1. Fonctionnement de l'algorithme proposé

L'algorithme de la figure 3.7 se présente comme suit :

Une fois que l'individu arrive chez le tradipraticien, celui-ci cherche à savoir le motif de la consultation. Après interrogatoire, si l'individu ne présente pas de signes avérés de maladie, le consultant peut éventuellement lui proposer un traitement préventif. Dans le cas contraire, le tradipraticien pose le diagnostic. Il se présente alors deux cas :

- Dans le premier cas la maladie est immédiatement identifiée. Dans ces conditions, le tradipraticien cherche à savoir s'il s'agit de la même maladie dont a souffert le patient au cours d'une consultation antérieure. Si c'est le cas, il lui propose un traitement $T2$ fondé sur l'ancien traitement $T1$ tiré de la base de connaissances ontologiques disponibles et représentées ici par (X qui exprime le végétal, Y l'animal et Z le minéral).

Mais, s'il s'agit d'une autre maladie, il lui prescrira un nouveau traitement $T1$.

- Dans le second cas, le tradipraticien a des doutes quant à la nature de la maladie dont souffre le patient ; dans ces conditions, il lui propose l'intervention de la médecine moderne au travers des examens complémentaires de laboratoire. C'est déjà un début de la collaboration entre les deux systèmes de médecine. En cas d'échecs persistants, le tradipraticien s'orientera vers d'autres formes d'intervention (moyens mystico-spirituelles, etc...).

3.11.2. Extensions Notationnelles

Les expressions notationnelles utilisées dans l'algorithme de la figure 3.7, constituent un outil puissant pour décrire des langages simples (rationnels). La nécessité de décrire de tels langages étant récurrente en informatique, ces formules sont donc utilisées avec de multiples extensions dans de nombreux outils d'usage courant [Fran 10]. Par exemple, « *grep* » est un utilitaire disponible sous « *UNIX* » pour rechercher les occurrences d'un mot (if) dans un fichier texte. En fait, « *grep* » va au-delà de la recherche d'occurrence d'un mot. En effet, à la place d'un mot unique, il est capable d'imprimer les occurrences de tous les mots d'un langage rationnel quelconque ; ce langage est défini sous la forme d'une expression rationnelle. La syntaxe des expressions rationnelles permises par « *grep* » fait appel aux caractères ' $*$ ', '+', etc. ($*$). Elle peut prendre entre $\{0, 1, \dots, N\}$ qui sont généralement des aspects formels de la théorie des automates. Ces extensions notationnelles permettent de simplifier grandement l'écriture des expressions rationnelles au prix de la définition de nouveaux caractères spéciaux.

3.12. Conclusion

Dans ce chapitre, il s'est agi d'élaborer un algorithme d'apprentissage de la science médicale traditionnelle. Pour ce faire, il a été question d'expliquer les grandes étapes de la méthodologie en présentant quelques scénarii portant sur la construction d'une ontologie basée sur certains aspects des éléments contenus dans la médecine traditionnelle. Ces scénarii ont permis de proposer un prototype d'architecture ontologique de cette médecine. C'est ainsi qu'a été élaboré cet algorithme d'apprentissage.

Ce chapitre permet de comprendre l'importance des ontologies dans la représentation des connaissances de même que dans le processus d'apprentissage et dans la pratique de la médecine traditionnelle.

Pour le prochain chapitre l'intérêt sera porté sur les mécanismes d'apprentissage à partir d'une ontologie.

Chapitre 4

APPRENTISSAGE A PARTIR D'UNE ONTOLOGIE

***Résumé.** Dans les systèmes éducatifs, l'analyse du mode d'apprentissage offre aux apprenants la possibilité d'évaluer la performance de ceux-ci en temps réel ; ainsi donc, cette interactivité entre l'apprenant et l'enseignant conduit à l'amélioration du système d'apprentissage et amène au développement des compétences. Dans ce chapitre, nous présentons d'un point de vue historique, les différentes étapes du processus d'apprentissage de manière générale et proposons un modèle d'architecture innovant pour l'apprentissage de la médecine traditionnelle, conforme à certains modèles d'apprentissage existant. Ce modèle participe du souci de promotion de la médecine traditionnelle par le biais des TICs.*

4.1. Introduction

L'informatique intervient dans tous les domaines et joue un rôle non négligeable dans le domaine de l'éducation [Ayc 13]. Ainsi, depuis les années 1970 [Bess 89], de nombreuses recherches portant sur la création de systèmes éducatifs liés aux Technologies de l'Information et de la Communication (TIC) ont vu le jour. Ces systèmes ont d'abord été appelés « Enseignement Programmé » puis « Enseignement Assisté par Ordinateur (EAO) » ; ils sont devenus par la suite des « Enseignements Intelligemment Assistés par Ordinateur (EIAO) » avec l'introduction des techniques d'Intelligence Artificielle pour aboutir à la fin aux « Environnements Informatiques pour l'Apprentissage Humain (EIAH) » pour lequel la notion de partenariat s'exprime désormais entre l'homme et la machine [Ami 12]. Dans les milieux nord-américains ou anglo-saxons, on parle de « Systèmes Tutoriels Intelligents (STI) ». Le but de ces opérations est de perfectionner les échanges entre l'utilisateur et l'ordinateur de manière à ce que ces échanges soient des plus naturels possibles. Ainsi, de nombreux chercheurs ont eu à proposer des typologies, des applications éducatives souvent appelées « Applications Pédagogiques de l'Ordinateur (APO) » [Moh 09]. Cette évolution sur l'apprentissage aux moyens des TICs a fait l'objet de plusieurs recherches. Nous pouvons citer entre autres Asma Ben Abacha et Pierre Zweigenbaum [Asm 12] qui se sont penchés sur l'analyse de questions dans le domaine médical. Dans le même ordre d'idée, Cimiano et *al.* [Cim 08] proposent un autre mécanisme de transformation des questions médicales en requête SPARQL à partir des ontologies. Ces évolutions technologiques prennent dès lors, une grande place dans la vie de tous les jours. L'homme utilise ces outils technologiques pratiquement dans tous les domaines (communication, éducation, santé etc.) [Ayc 13]. Actuellement, les bases de données relationnelles (c'est à dire un ensemble d'entrepôts de données dans lesquels les données sont inter-reliées) demeurent le moyen le plus populaire pour stocker, rechercher et manipuler des

données. Cependant, la structure et les contraintes d'intégrité du modèle relationnel sont définies par des schémas. Ceux-ci ne sont pas aussi expressifs que les ontologies, pour ce qui est de la représentation de la sémantique des données [Bens 06] du fait de l'apparition des TICs [Gar 06]. Malheureusement, l'apprentissage de la médecine traditionnelle reste méconnu dans le domaine de la recherche. Le manque de communication/apprentissage des vertus thérapeutiques des plantes reste l'un des défis à relever. Nos travaux ont pour but de proposer des techniques d'apprentissage basées sur l'importance quantitative des ontologies et leurs disponibilités.

4.2. Définition de l'Apprentissage

On peut définir l'apprentissage comme étant un ensemble de méthodes permettant d'acquérir des connaissances, du savoir-faire, du savoir-être en vue de répondre à des questions qui amènent à changer les comportements dans l'esprit de l'homme.

Dans le contexte de l'apprentissage l'on doit s'intéresser à des questions telles que :

1. Qui est l'apprenant ? qu'aura-t-il besoin de faire ?
2. Que lui faut-t-il apprendre pour être capable d'agir ?
3. Qui est l'enseignant ? Que doit-il disposer comme savoir-faire et comme expérience etc. ?
4. De quels outils doit-il disposer pour mener son action ?
5. Quelle est la nature des connaissances à transmettre ? Par quelle méthode ?
6. De combien de temps dispose-t-il pour transmettre la connaissance ?

4.3. Etapes historiques de l'apprentissage

La figure 4.2 donne un aperçu des principales étapes qui ont marqué l'histoire de l'informatique dans l'éducation. A cet effet, trois grandes périodes ont marqué cette histoire. Il s'agit des années 50, 70 et 80 [Cim 08].

Dans le courant des années 50, l'histoire de l'informatique a été marquée par l'apparition de l'enseignement programmé et par l'EAO ; la progression de cet enseignement programmé et de l'EAO a donné naissance aux techniques de l'intelligence artificielle (EAO et IA) au début des années 70. C'est au cours de cette même année 70 que les logos et des micros-mondes ont connu leur développement. Cependant, la continuité de cette progression a abouti, dans les années 80, à la création ou à la mise en place de tuteurs intelligents et à l'ouverture des environnements d'apprentissage. A cette même période, on assiste à l'émergence d'outils généraux tels que les micros-ordinateurs ; cette même années 80 a été l'année des tentatives de

synthèse pour ce qui est des environnements interactifs contrôlés avec l'apport des techniques d'hypertextes et la diffusion du multimédia. Cela a abouti à l'explosion récente des réseaux comme internet de même qu'arrivée des interfaces multi modèles [Bruil 03]. Dans ce schéma, la première colonne se réfère à des courants qui ont privilégié l'enseignement, c'est à dire pour lesquels la machine est principalement amenée à jouer le rôle du maître (de l'enseignement programmé aux tuteurs intelligents). La troisième colonne rend compte des recherches dans lesquelles l'ordinateur est un moyen permettant aux apprenants d'effectuer, de construire, d'explorer et de découvrir. La colonne centrale correspond aux tentatives de synthèse de ces deux approches, essayant de concilier un certain guidage par la machine dans des environnements largement contrôlés par les apprenants. Bien entendu, un tel découpage est schématique et réducteur et ne correspond que très imparfaitement à ce qui s'est réellement passé ; les idées ayant en effet traversé les différents courants de pensées [Bruil 03].

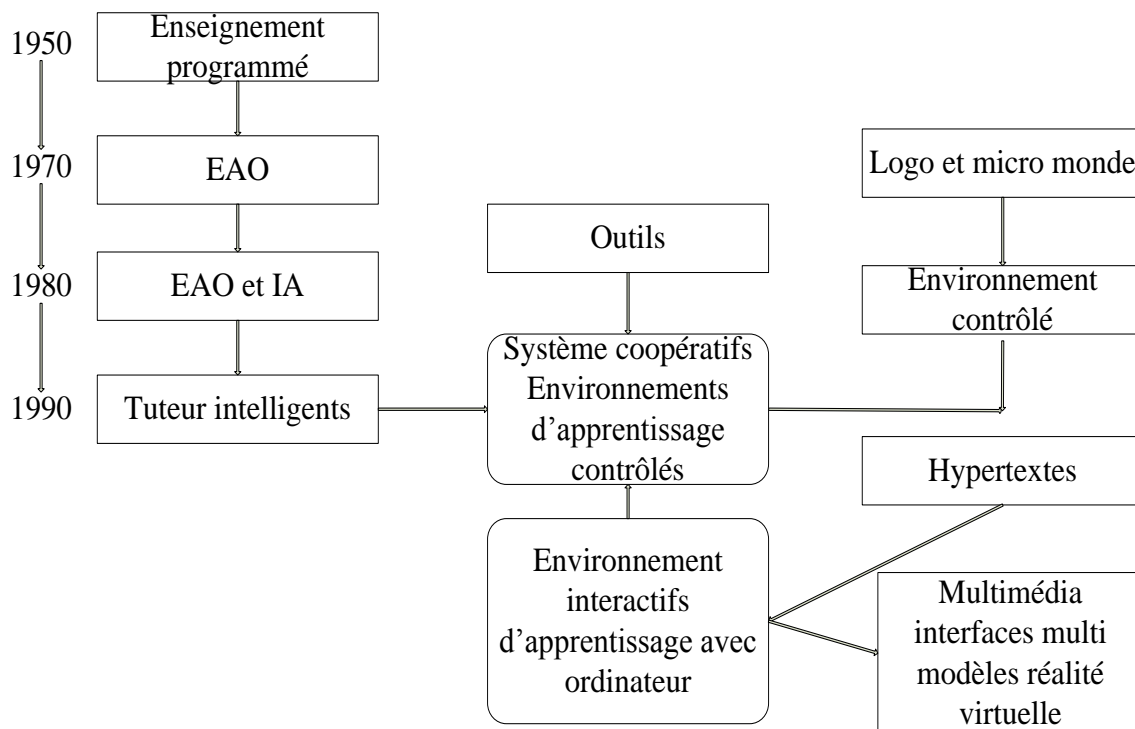


Figure 4.1 : Principaux courants en informatique et éducation

4.4. Conditions d'apprentissage

Avant la mise en place de tout processus d'apprentissage, il importe :

- de déterminer les besoins de l'apprenant,
- de déterminer les objectifs généraux et spécifiques de l'apprentissage,
- de définir la méthodologie appropriée et choisir les outils d'accompagnement qui s'imposent, et ainsi, développer efficacement les compétences à acquérir.

Par le respect scrupuleux de toutes ces conditions, on en arrive à assurer le progrès adéquat du système apprenant. Ainsi donc :

- *«Toute forme d'apprentissage nécessite la répétition des décisions dans le temps».* L'apprenant doit effectuer plusieurs essais, car tant qu'il y a plus de répétitions, il y aura plus d'expériences.
- *«Toute forme d'apprentissage nécessite un mécanisme de rétroaction environnementale».* Pour qu'un apprenant tire profit de tout système d'apprentissage, il faut qu'il reçoive un « feedback » (rétroaction) de son environnement à la suite de ses propres décisions et les décisions des autres.
- *«Toute forme d'apprentissage nécessite un mécanisme d'adaptation des décisions».* Selon cette même condition, non seulement l'acquisition de nouvelles informations par le système apprenant mène à un apprentissage, mais aussi, les renseignements sur lesquels ces informations peuvent être utilisées sont également nécessaires.
- *«Toute forme d'apprentissage nécessite l'existence d'un mécanisme de stockage de l'information ; la mémoire».* En effet, s'il n'y a pas une sauvegarde des connaissances obtenues dans le passé, le système ne bénéficie pas de ses décisions antérieures. Le stockage d'informations offre au système la possibilité de modifier sa décision sur la base de l'expérience passée et d'améliorer ses performances. En revanche, la mémorisation doit être guidée par des stratégies adéquates en tenant compte de l'espace mémoire comme un facteur critique.

4.5. Travaux existants

Plusieurs travaux de recherche ont évalué l'intérêt des systèmes multimédias dans le cadre des systèmes d'enseignement.

Nous pouvons noter entre autres, les applications Grid Learning Service (GLS) implémentées afin de permettre aux membres d'une communauté virtuelle de pouvoir créer et distribuer des objets d'apprentissage [Batc 07]. Elles reposent sur l'architecture «Open Grid Services Architecture» (OGSA) et facilitent ainsi le partage et les échanges des objets d'apprentissage [Aud 07]. Ces applications constituent des services importants dans le domaine de l'éducation tels que les services de qualification, de collaboration et de communication. Cependant, cette grille d'apprentissage a un accès limité puisque, pour y prendre part, il faut disposer de codes d'accès.

Alice Hermann et *al.* [Herm 12] [Tous 14] élaborent le dispositif «Technology Enhanced Learning Environment for Orthopedic Surgery» (TELEOS). Ce dispositif s'inscrit dans le

cadre des recherches sur le développement des EIAH ; il constitue aussi un environnement de transmission et d'acquisition de connaissances tels que : la vertébroplastie et le vissage ilio-sacré. C'est un Tutoriel Intelligent développé afin de soutenir l'enseignement et l'apprentissage de la chirurgie orthopédique percutanée. Cependant, ce dispositif est uniquement destiné aux étudiants chirurgiens ; il ne permet donc pas à d'autres disciplines de s'en servir. Pour Pierre Tchounikine [Tcho 02], les EIAH sont des environnements informatiques qui harmonisent l'interaction entre l'agent humain et l'ordinateur. Il affirme également que les EIAH donnent accès à des ressources formatives qui ont pour objectif de suivre et d'accompagner l'apprenant.

Le STI (Systèmes Tutoriels Intelligents) de Renaudie [Rena 05] porte un jugement sur les actions et les réponses de l'apprenant. Puis, il propose des explications afin que les erreurs détectées ne se reproduisent plus. En effet, c'est un système basé sur les connaissances d'un domaine quelconque. Pour ce faire, celui-ci dispose d'une représentation interne du problème à traiter et est donc capable d'évaluer le résultat de l'exercice, c'est à dire si l'objectif visé est atteint.

Alain DUREY et Daniel BEAUFILS [Dur 98] ont, quant à eux, développé un logiciel "didactiquement intelligent" qui permet de résoudre les problèmes de la mécanique secondaire. C'est un logiciel fondé sur les connaissances apportées par la didactique dans le domaine de la mécanique. En outre, Amina BAGHLI [Ami 12] propose une ontologie (vocabulaire RDF) nommée « Diag-K ». Cette ontologie regroupe les connaissances nécessaires pour intégrer les EIAH ; elle permet également d'apprécier le niveau de connaissances de l'apprenant dans le domaine enseigné.

Les systèmes « PRECISE » [Pop 03] et « ORAKEL » [Cim 08] pour leur part, analysent les questions et essaient de les transformer en requêtes SQL correspondant en F-logic (Frame logic) ou en requête SPARQL. En cas d'échec, ces systèmes reformulent une nouvelle requête. Cette requête reformulée attaque une nouvelle structure de graphes dans laquelle les mots de la question sont projetés sur les relations de la base de données puis sur les colonnes et les valeurs des graphes.

Alice Hermann et *al.* [Kaz 13] proposent le système « Updating Through Interaction in Logical Information Systems » (UTILIS). C'est une méthode d'aide à la création de nouveaux objets. UTILIS cherche des objets similaires au nouvel objet en appliquant des règles de relaxation et sa description.

4.6. Enseignement Assisté par Ordinateur (EAO)

Dans n'importe quel domaine, l'enseignement consiste à transférer différents types de connaissances d'un enseignant à un apprenant en utilisant des méthodes et/ou des techniques. Les chercheurs ont adapté cet enseignement par la mise sur pied d'un programme qui fonctionne avec un ordinateur ; ceci a donné naissance à un nouveau domaine de recherche en informatique nommé «Enseignement Assisté par Ordinateur» (EAO) [Kaz 13]. Cet enseignement se présente comme l'utilisation d'un matériel informatique et de logiciels, appelés logiciels éducatifs en vue d'assurer tout ou partie d'un processus de formation [Add 13]. Son principe fondamental s'appuie sur l'auto-éducation c'est-à-dire, l'auto-apprentissage associé à l'auto-évaluation. Cet ensemble est aussi considéré comme un processus d'acquisition de connaissances [Bess 89]. Au fil du temps et avec l'apparition de l'intelligence artificielle, les chercheurs ont tenté d'améliorer cette nouvelle approche (EAO) en introduisant les techniques de l'intelligence artificielle. Cela a conduit à l'apparition de nouvelles approches que sont les « Enseignements Intelligemment Assistés par Ordinateur (EIAO) » ou encore «Tuteurs Intelligents ». L'enseignement assisté par ordinateur (EAO) apparaît véritablement comme le mariage de l'enseignement programmé fondé sur des concepts fondamentaux de la cybernétique et du développement de la psychologie du comportement et de l'informatique. Il apparaît aussi comme une technologie éducative basée sur les trois composantes principales de l'action éducative : fixation des objectifs, sélection des moyens et des méthodes et évaluation des résultats [Kaz 13]. On aboutit au modèle classique d'une session d'EAO qui se déroule de manière cyclique :

- Présentation d'informations et d'une question ;
- Réponse de l'apprenant ;
- Analyse de la réponse ;
- Continuation ou branchement à une autre partie du cours.

L'EAO présente deux inconvénients majeurs à savoir :

- La préparation du cours requiert un temps considérable et les résultats obtenus ne sont pas très flexibles (difficilement modifiables) ;
- Le changement d'un ensemble de questions ou de réponses implique un gros travail de réécriture du programme.

4.7. Enseignement Intelligemment Assisté par Ordinateur (EIAO)

L'Enseignement Intelligemment Assisté par Ordinateur fait appel aux techniques et outils de l'Intelligence Artificielle et plus particulièrement au système expert c'est-à-dire à un outil capable de reproduire les mécanismes cognitifs d'un expert dans un domaine donné; Ceci afin d'étudier des systèmes "tuteurs intelligents" qui simulent deux modes alternatifs auteur-apprenant selon trois modalités : pédagogique, didactique, médiatique [Bess 89] comme le montre la figure 4.2. C'est une tentative développée à travers l'application des techniques d'intelligence artificielle.

J.F.Nicaud et M.Vivet [Nic 88] ont défini les EIAO comme étant une science interdisciplinaire faisant appel aux domaines suivants : pédagogie générale, didactique des disciplines, psychologie cognitive, ergonomie et informatique. Ceux-ci ont évolué à partir de la science informatique et non à partir de la psychologie ou de la technologie éducative. La mise en place d'un système EIAO consiste à réaliser un tuteur intelligent permettant de concevoir un programme autour d'un ordinateur dont la fonction est de se conduire comme un bon enseignant. Pour cela, il faut mettre en place un système doté de trois approches (domaine enseigné, enseignement et compétences, connaissances correctes et erronées de l'apprenant). L'expérimentation d'un tel système a pour but de généraliser les connaissances pédagogiques, didactiques et médiatiques. Généralement, le programme SCHOLAR est considéré comme le premier tuteur intelligent conçu pour l'enseignement de connaissances factuelles sur la géographie de l'Amérique du sud. Son originalité réside dans le type de dialogue instauré avec l'apprenant [Kaz 13].

Un système d'EIAO doit être capable de :

- résoudre les problèmes ;
- comprendre le raisonnement (correct ou erroné) de l'apprenant ;
- établir un modèle de l'apprenant ;
- proposer à l'apprenant des exercices en fonction de ce modèle [Dom 00].

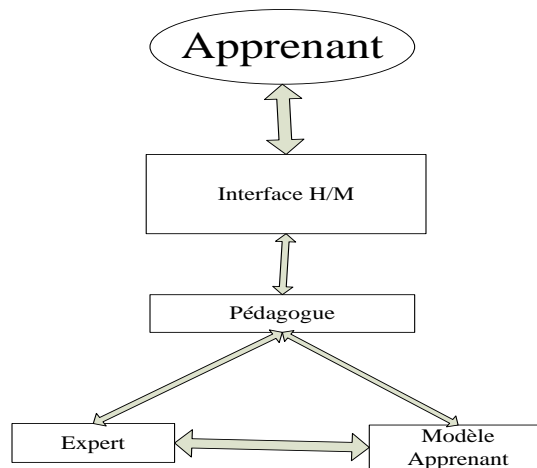


Figure 4.2 : Composantes d'un système d'EIAO

Le module expert : il contient les structures de connaissances du domaine enseigné ainsi que les méthodes permettant de rechercher les informations dans ces structures. Il est chargé de :

- la généralisation de problèmes ;
- la résolution de problèmes ;
- le test des réponses de l'apprenant ;
- et l'explication portant sur les erreurs, les énoncés d'exercices et la résolution de problèmes.

Le module pédagogique : il est constitué d'un ensemble de spécifications sur la manière dont le système doit construire ses interventions. Ce module interagit avec l'apprenant plus ou moins directement en sélectionnant les problèmes qu'il doit résoudre en le guidant vers la solution, en critiquant ses performances, en fournissant une aide appropriée lorsque l'élève le lui demande et en montrant des exemples.

Le modèle apprenant : Le modèle apprenant est constitué d'un profil de compétence de l'élève pour les divers éléments de connaissance. Le modèle apprenant regroupe les connaissances liées au domaine propre à chaque apprenant, les connaissances indépendantes du domaine enseigné telles que les mécanismes d'apprentissage préférés par l'apprenant. Il doit définir l'écart entre la propre connaissance (supposées) de l'apprenant et les connaissances cibles, enjeu de l'apprentissage [Kaz 13].

La façon de concevoir cet écart conduit à distinguer deux grandes classes de modèles :

- *Les modèles d'expertise partielle ou de superposition :* dans lesquels les connaissances de l'élève ne sont qu'un sous-ensemble de la connaissance cible.
- *Les modèles différentiels :* représentant les sous tâches mal apprises par l'étudiant.

4.8. Module tutorial

Un module tutorial est l'ensemble de spécifications concernant la matière à présenter par le système ainsi que la façon et le moment où il devrait la présenter. Dans les systèmes existants d'EIAO, les stratégies pédagogiques pour spécifier la présentation de matériel éducatif sont basées sur deux méthodes : La méthode SOCRATIQUE et la méthode de l'entraîneur [Kaz 13].

- La méthode SOCRATIQUE fournit aux apprenants une série de questions les guidant à travers le processus de «debugging», à partir de leur propre conception erronée. Durant ce processus, on suppose que les étudiants raisonnent à partir de ce qu'ils savent et de ce qu'ils ne savent pas et qu'ils modifient ainsi leurs conceptions.

- La méthode de l'entraîneur place l'apprenant dans un environnement où il s'engagera dans des activités comme les jeux informatisés pour maîtriser des habilités associées et les capacités générales de résolution de problèmes.

4.9. Apprentissage classique de la MT

La médecine traditionnelle est d'une diversité biologique et culturelle très riche. Sa pratique est très répandue et se fait sous des formes variées. C'est une médecine assez difficile à apprendre dans la mesure où la transmission des connaissances s'est toujours faite verbalement et de façon non structurée. Les tradipraticiens, avant de s'approprier ces connaissances et de les maîtriser, ont une longue phase d'apprentissage qui se caractérise essentiellement par un compagnonnage sous différentes formes : par imitation, par observation etc. dans le cadre :

- d'une approche mystico-spirituelle essentiellement dominées par l'acquisition des lois et règles de la cosmogonie traditionnelle,
- d'une approche thérapeutique (par apprentissage des moyens de traitement : végétale, animale, minérale...),
- d'une approche sémiologique (par apprentissage des signes et de la nature des maladies et de ses caractéristiques, du comportement du malade etc.).

En dehors de l'aspect informel du mode d'acquisition de connaissances dans la médecine traditionnelle, l'apprentissage dans ce domaine soulève bon nombre d'inconvénients ; en effet, du fait que le mode d'accès à ces connaissances n'est pas déterminé, que le contenu n'est pas formalisé, que l'évaluation est inexistante, les prestations que le tradipraticien est susceptible d'offrir peuvent être sujets à caution parce que :

- Les règles d'hygiène qui entourent la fabrication, le conditionnement et la conservation des produits peuvent être méconnues,
- Le diagnostic peut être approximatif,
- Les doses et les posologies peuvent être incertaines,
- Des plantes potentiellement toxiques peuvent être utilisées par ignorance etc.

Cependant, lorsque ces tradipraticiens ont plus ou moins maîtrisé ces connaissances, ils procèdent généralement dans leurs démarches par poser le diagnostic de la maladie en intégrant toujours l'aspect culturel et psychologique dans l'approche thérapeutique. Les tradipraticiens considèrent en effet le patient dans toutes ses dimensions y compris l'environnement dans lequel il vit. Ce principe est largement traditionnel à cause de l'utilisation d'une variété de méthodes, de pratiques, de croyances, de savoir et de savoir-faire. Ces méthodes sont basées sur l'expérience personnelle ou sur des connaissances transmises depuis plusieurs générations. Cette médecine regroupe plusieurs acteurs de différents domaines ; ce qui rend la transmission des connaissances encore plus complexes.

4.10. Apprentissage organisationnel

L'apprentissage organisationnel peut être appréhendé comme un processus social d'interactions individuelles qui a pour but l'acquisition ou la production de nouvelles connaissances organisationnelles. Il est communément admis de distinguer deux approches de l'apprentissage organisationnel : « **behavioriste** » et « **cognitif** » [Guy 07]. Parler d'apprentissage organisationnel suppose qu'il y ait capitalisation des savoirs dans la mémoire de l'organisation comme le précise GIRIN, « *Il y a apprentissage organisationnel lorsque les membres de l'organisation construisent des savoirs pertinents par rapport à la mission de l'organisation et que ces derniers sont codés ou mémorisés dans l'organisation* » [Dome 10]. En effet, tout être humain de tout âge peut apprendre. Avant de créer un milieu d'enseignement/apprentissage assisté par les nouvelles technologies, il faut préciser les tâches suivantes soit pour l'enseignant soit pour l'apprenant :

- Stimuler la motivation ;
- Gérer la méthode d'apprentissage ;
- Entraîner à l'analyse ;
- Renforcer l'animation ;
- Donner du Feedback ;
- Encourager l'autonomie de l'apprentissage.

L'acte d'apprendre consiste alors à retrouver ces connaissances et à faire des liens entre les informations consultées. Ainsi, pour un individu, les liens cognitifs qui lui permettent d'apprendre sont plus importants que la mémorisation d'informations [Dome 10].

4.11. Algorithme d'apprentissage

Un algorithme d'apprentissage A est un algorithme qui a pour fonction d'apprendre à effectuer une tâche à partir d'un ensemble S de données. Chaque donnée z_i est constituée d'un objet d'entrée x_i et une valeur de sortie y_i

$$S = \{Z\}_{i=1}^m = \{z_1, z_2 \dots z_m\}$$

L'objectif de l'algorithme d'apprentissage est de construire pour tout couple (x, y) une bonne fonction $h(x)$ appelée hypothèse qui représente la solution du problème en discussion, telle que $h(x) = y$. L'entrée de l'algorithme d'apprentissage A est un ensemble S et la sortie est une fonction appelée h .

On écrit : $A(S) = h$

4.12. Taxonomies des connaissances

La figure 4.3 de Tom Bekman [Teb 00] présente «comment la connaissance» fait partie d'une chaîne de transformation ou encore d'un processus comportant un certain nombre d'étapes. Les données (data) constituent la matière brute de la chaîne. L'information est une donnée dont la forme et le contenu sont utiles pour une tâche particulière. Les connaissances permettent l'exécution des tâches, la prise de décision et la résolution des problèmes. Les cas, les règles, les modèles sont trois types de connaissances. Elles s'acquièrent par accumulation d'informations qui s'organisent progressivement par rubriques dans l'ensemble des moyens de stockage : ouvrages, bases de données, supports magnétiques. Ainsi, les connaissances sont structurées, codifiées et subissent des concentrations successives ayant pour caractéristique d'accroître l'utilité de ces connaissances [Teb 00].

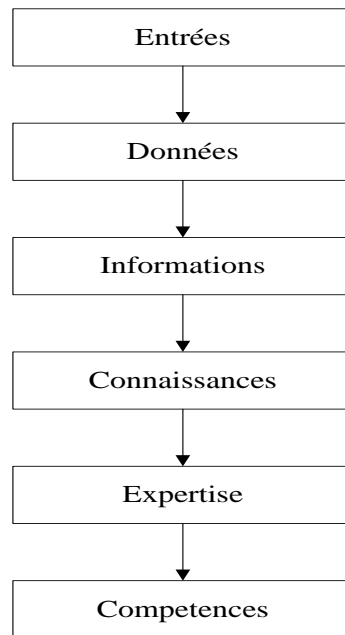


Figure 4.3 : La hiérarchie des connaissances

4.13. Dimension ontologique

La dimension ontologique fait passer du niveau individuel (la connaissance n'étant créée que par les individus) au niveau de l'organisation. La création de connaissances organisationnelles est alors le processus qui amplifie de façon organisationnelle les connaissances créées par les individus et les cristallise en tant que des parties d'un réseau de connaissances de l'organisation. Une distinction entre connaissances individuelles et collectives se précise [Dome 00].

- Les connaissances individuelles

C'est l'ensemble des croyances d'un individu, sur les relations de cause à effet entre phénomènes. Les philosophes se sont intéressés principalement à ce type de connaissances.

- Les connaissances collectives

Ces connaissances sont celles des organisations ou des groupes de travail. Elles sont stockées dans des règles, des procédures, des routines et des normes partagées. Elles sont aussi appelées connaissances «sociales ». Ainsi, dans l'entreprise, chaque individu ou groupe d'individus détient une connaissance particulière face à une problématique.

4.13.1. Modes de transformation des connaissances

Il existe plusieurs modèles de transmission des connaissances : Parmi les modèles les plus utilisés, nous pouvons citer le modèle d'Huber élaboré en 1991 [Teb 00] qui combine la gestion de l'information et celle de la connaissance à travers un processus d'apprentissage. Selon Huber, si une unité apprend, toute l'organisation pourra avoir accès à la connaissance nouvelle. Le

modèle décrit par Huber repose sur la relation entre l'apprentissage et le traitement de l'information par les individus. Ce processus passe par quatre étapes : l'acquisition de connaissances, la distribution de l'information, son interprétation et sa mémorisation.

4.13.2. Modes d'apprentissage

- L'acquisition de la connaissance ou de l'information

C'est une étape qui est constituée de 5 niveaux :

- *L'apprentissage congénital*

Il représente les connaissances héritées du passé de l'organisation. Une organisation fonctionne souvent sur des bases de connaissances élaborées depuis une grande période de temps.

- *L'apprentissage par l'expérience*

Il résulte le plus souvent d'efforts non-intentionnels. Si les apprentissages dus à des expérimentations sont fréquents, le hasard intervient couramment.

- *L'apprentissage par procuration*

Il constitue l'apprentissage par imitation. Non pas qu'il s'agisse d'un apprentissage comportementaliste, mais l'organisation prend exemple sur une autre (*le benchmarking*).

- *L'apprentissage par greffe* : il se déclenche par l'arrivée dans l'organisation de nouveaux individus. Que ce soit par embauche ou croissance externe, l'organisation apprend de ses nouveaux entrants.
- *La recherche* : elle est une source d'apprentissage constante pour l'organisation.

Elle se fait par la « veille stratégique », la recherche spécifique ou par un suivi des performances de l'entreprise.

- *La distribution de l'information*

C'est le processus par lequel les membres d'une unité regroupent les informations qu'ils ont acquises. L'information provenant de plusieurs sources est mise en commun avant de remonter ou d'être redistribuée à d'autres unités ou départements. Ici, la problématique est de déterminer le type d'information à distribuer. On remarque ainsi que les unités recueillent de nombreuses informations pas toujours pertinentes, mais qu'en plus, elles ne savent pas nécessairement tout ce qu'elles prétendent savoir. Ces lacunes peuvent être comblées par une distribution efficace de l'information. Pour Huber, l'apprentissage est d'autant plus efficace que l'ampleur de la distribution est importante. Cette étape est donc un élément moteur du processus.

- *L'interprétation de l'information*

C'est un processus qui donne du sens à l'information. Par ce processus, les individus cherchent à comprendre les événements, à les déchiffrer et à étudier les résultats de cette analyse. Cette

démarche entraîne un apprentissage si les membres de l'organisation assurent un partage des interprétations et des analyses. Une analyse commune permet à l'organisation de bénéficier d'une compréhension des théories en usage.

- *La mémorisation de l'information*

Une fois le processus arrivé à ce stade, il est nécessaire que l'organisation, par l'intermédiaire de ses membres, stocke la connaissance dans sa mémoire organisationnelle. Car les mémoires individuelles ne retiennent la connaissance que pendant le temps de présence de l'individu dans l'organisation. Les Technologies de l'Information et de la Communication (TIC), les bases de données et les systèmes de management des connaissances par codification jouent ici un rôle important même si les mémoires informatiques limitent le stockage d'informations tacites.

4.14. Ontologie et EIAO

Pour mieux comprendre la nature de la relation entre le concept d'ontologie et l'EIAO, nous proposons une large vision de celui-ci. Une ontologie est un méta base de connaissances déclaratives, rendues explicites par un exercice approfondi d'analyse conceptuelle. C'est également une méthode consistant à décrire les objets, les événements, les ressources, les contraintes afin de produire des consignes assignant les ressources et les temps aux objets et aux événements [Niv 13]. L'utilisation d'ontologies permet donc de spécifier la sémantique d'une activité et, le raisonnement bâti autour de toutes les activités permet de les identifier dans l'environnement. L'utilisation d'ontologies offre de nombreux avantages dans la reconnaissance d'activités dans un appartement intelligent. De façon générale, l'apport spécifique de l'ontologie pour les environnements d'aide à l'apprentissage serait la représentation formelle de connaissances déclaratives, couplée à un mécanisme d'inférence exploitable par ces environnements. Nous pensons que la richesse de l'apport de l'ontologie est cependant plus complexe et comprend plusieurs éléments tels que l'explication des connaissances déclaratives, la gestion de l'intégrité, la formation de consensus, l'orientation par l'usage, la communication dans les systèmes multi-agents, la sémantique riche, l'interopérabilité, les requêtes en direct [Psy 04].

Une ontologie offre un moyen de gérer la communication dans les systèmes multi-agents, en fournissant un référentiel commun aux agents qui peuvent ainsi l'exploiter de différentes façons sans altérer son intégrité. L'ontologie offre une meilleure performance puisqu'elle supporte une interaction en direct lors des requêtes d'information par les usagers. Ces caractéristiques mettent en relief la spécificité et l'originalité de l'importance des ontologies pour les EIOA.

4.15. Architecture du modèle EIAO

Dans cette section, nous présentons l'architecture du modèle EIAO proposée. Cette plateforme permet de favoriser la flexibilité, la souplesse de la modélisation et l'exécution du processus d'apprentissage. Elle permet d'analyser le déroulement du processus d'apprentissage. Elle permet également de vérifier sa correspondance avec le modèle ; autrement dit, elle permet d'améliorer le fonctionnement de l'apprentissage en garantissant une adaptation au changement et une facilité d'analyse du bon fonctionnement du processus.

L'architecture générale de cette plateforme est présentée dans la figure 4.4.

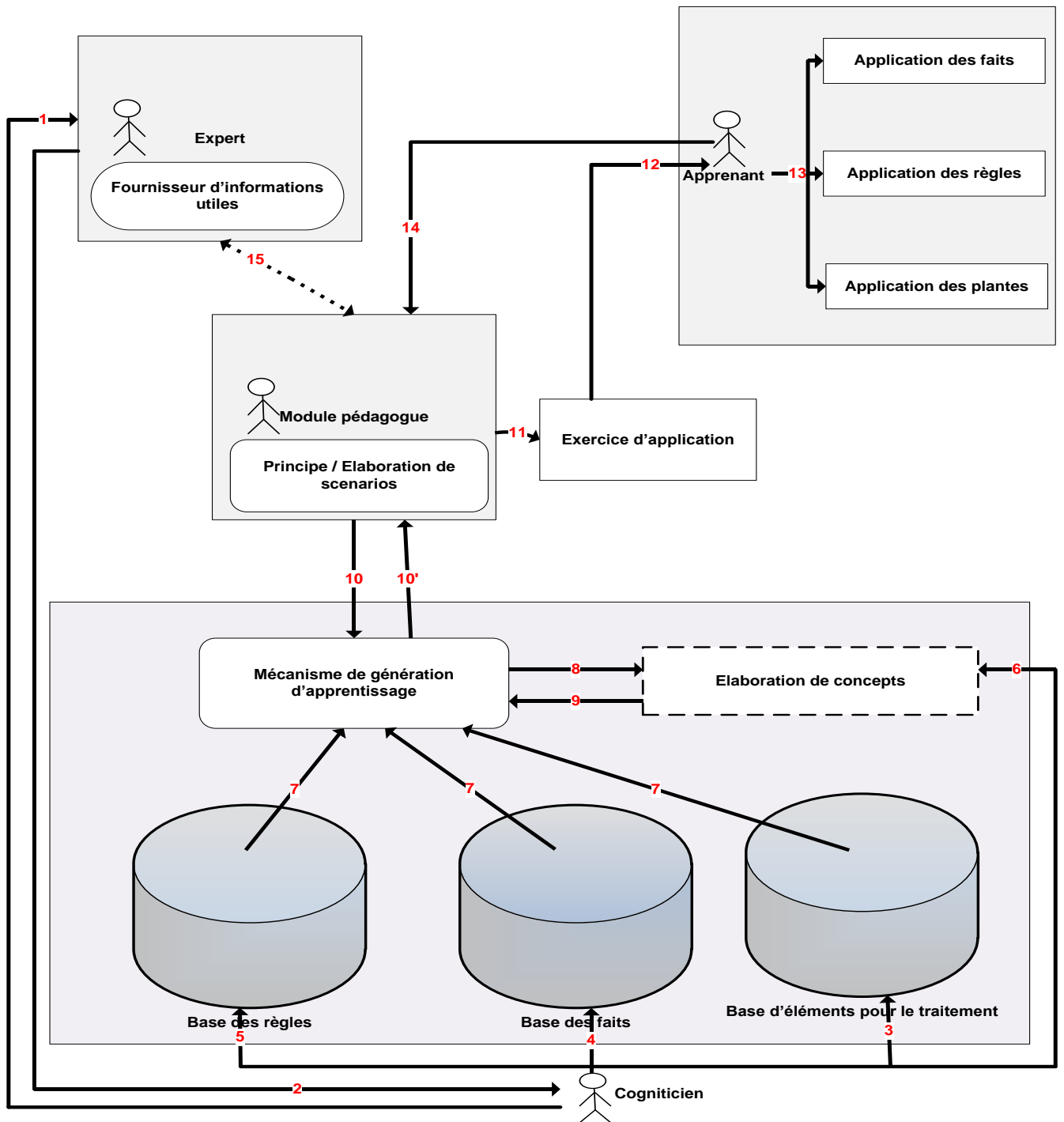
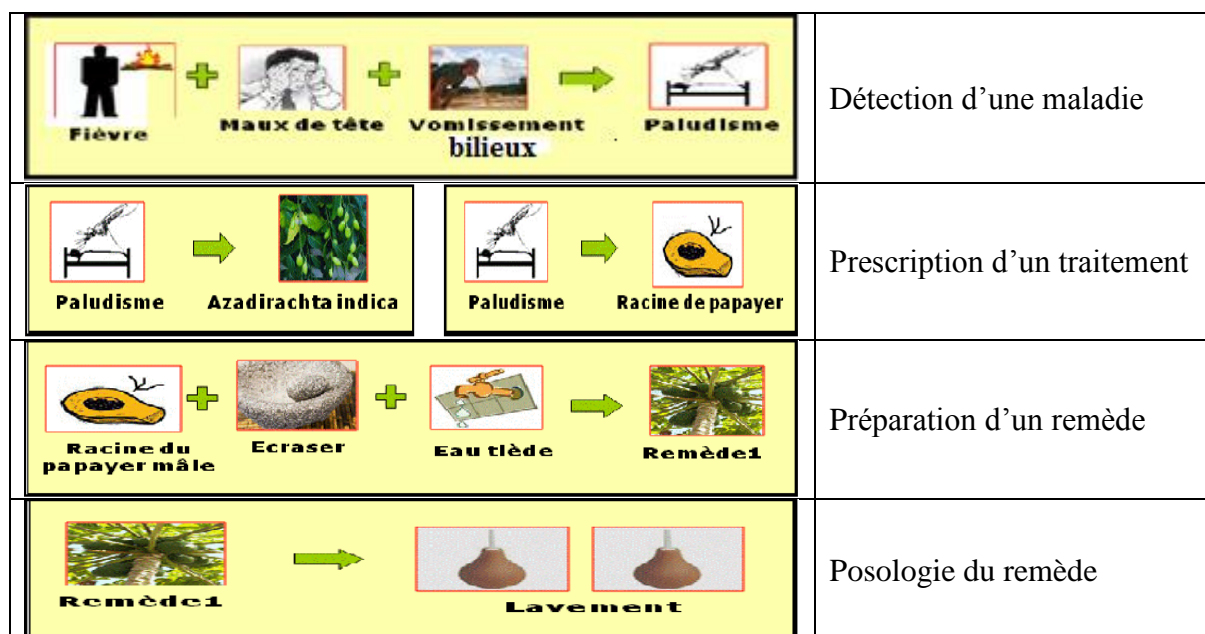


Figure 4.4 : Architecture d'apprentissage de la médecine traditionnelle par le mécanisme d'EIAO

L'exemple du tableau 4.1 tiré du modèle de l'article de [Brou 10] montre en image les éléments contenus dans les différentes bases.

Au travers de ces icones, le processus d'apprentissage et la pratique de la médecine traditionnelle peuvent apparaître relativement plus aisés.

Tableau 4.1 : Exemple d'éléments contenus dans : les bases de faits, les bases de règles et les bases de traitements.



4.15.1. Description de l'architecture

Notre système a été conçu en vue de répondre à la motivation précitée à savoir l'apprentissage de la médecine traditionnelle. Par ailleurs, ce système offre aux nouveaux apprenants des outils d'aide à la décision par rapport à leur vécu quotidien.

Le système s'articule autour des bases de connaissances suivantes :

- **Base des règles :** elle renferme les règles décrites par un expert face à une prise de décision, c'est-à-dire la connaissance « métier » des experts.
- **Base des faits :** cette base contient les faits propres à un problème à résoudre.
- **Base d'éléments pour le traitement:** dans cette base, tous les éléments susceptibles d'aider au traitement que le tradipraticien va fournir à l'expert sont stockées.

4.15.2. Description des flux

Le Tableau 4.2 présente les résumés des flux de l'architecture proposée dans la figure 4.4

Tableau 4.2 : Description des flux

Numéros de Flux	Désignation des Flux
(1)	Consultation d'un tradipraticien
(2)	Recueil d'informations utiles
(3)	Elaboration d'une base des plantes

(4)	Mise en place d'une base des faits
(5)	Elaboration d'une base des règles
(6)	Edition des concepts
(7)	Préparation d'un mécanisme d'apprentissage
(8)	Choix d'un concept
(9)	Mise en corrélation du concept par rapport au mécanisme d'apprentissage
(10)	Recherche d'un mécanisme d'apprentissage adéquat
(10')	Récupération d'un mécanisme d'apprentissage adéquat
(11)	Proposition d'exercices d'apprentissage
(12)	Récupération d'exercices par l'apprenant
(13)	Traitement d'exercices d'application
(14)	Demande d'explication ou restitution de l'exercice traité
(15)	Permutation de responsabilité

Telle que présentée, l'architecture semble bien répondre à des besoins de navigation, d'accès et de fourniture de matériels d'apprentissage. En effet, on voit clairement que les utilisateurs peuvent utiliser le contexte, la structure et le contenu (à travers des ontologies) pour formuler de façon adéquate leurs requêtes et ainsi rechercher efficacement un contenu d'apprentissage. Les fournisseurs de contenu de leur côté peuvent construire des objets d'apprentissage en se basant sur les mêmes règles (contexte, contenu, structure).

4.16. Conclusion

Le processus d'apprentissage évoqué dans ce chapitre met en exergue les différentes étapes qui ont servi à la mise en place de quelques systèmes d'apprentissage. Ces systèmes aident les apprenants dans la quête du savoir. Pour aboutir à la construction du système d'apprentissage en vue de la valorisation de la médecine traditionnelle, objet de notre travail, nous avons fait état de quelques grandes étapes de l'évolution de ces systèmes. Il s'est agi de faire l'historique de l'évolution des systèmes d'apprentissage en passant par l'enseignement programmé, par l'Enseignement Assisté par Ordinateur (EAO) pour aboutir enfin à la création d'Internet.

Eu égard aux difficultés que rencontrent l'apprenant et le pratiquant de la MT dans le mode classique, nous avons cherché à savoir comment celle-ci peut tirer profit des nouvelles méthodes d'apprentissage basées sur les Technologies de l'Information et de la Communication

(TIC). C'est ainsi que nous avons proposé l'utilisation d'un système d'apprentissage à base d'ontologie pour la formation et la pratique en médecine traditionnelle.

Le modèle d'architecture d'apprentissage proposé est basé sur le modèle EIAO.

Nous nous proposons dans le prochain chapitre de nous focaliser sur le processus de recherche d'informations dans une base de connaissances médicales traditionnelles.

Chapitre 5

PROCESSUS DE RECHERCHE D'INFORMATIONS DANS UNE BASE MEDICALE

Résumé. La recherche d'informations est un processus permettant de retrouver les résultats pertinents d'une requête posée par un utilisateur. Cependant la recherche dans la base médicale intervient lors de la phase d'indexation d'un document médical. La plupart des informations retournées lors de la phase de recherche sont retrouvées à l'aide des mesures de similarité entre le document et la requête émise. Dans ce chapitre, nous présentons l'impact des requêtes dans l'indexation d'une base médicale.

5.1. Introduction

Le développement des TICs (Technologies de l'information et de la communication) a été à la base de la naissance d'un grand nombre de nouvelles applications qui se développent autour des réseaux internet. La diversité de ces nouvelles applications ne favorise pas l'intégration des systèmes de recherche traditionnelle d'information (SRI) dans les TICs. En effet, les systèmes traditionnels de recherche d'information se font à travers des mots-clés choisis manuellement. Par contre, les nouvelles applications donnent la possibilité de parcourir et de retrouver tous les termes d'un document afin de les indexer automatiquement. La différence entre ces deux systèmes entraîne comme conséquence, au niveau de la recherche d'information traditionnelle, l'insatisfaction des besoins en informations spécifiques pour les utilisateurs [Daou 08]. A titre illustratif, nous pouvons citer parmi les nouvelles applications, le commerce électronique, la recherche d'information sur le web, les plates-formes pour calcul réparti, etc. Ces applications sont formées par des entités réparties et la bonne fonctionnalité nécessite la communication et les échanges entre ces entités [Seg 09, Daou 08].

Pour pallier les insuffisances du système de recherche traditionnelle, diverses méthodes et stratégies ont été proposées pour permettre la mise à jour des index lorsque la collection de documents est modifiée. Au nombre de ces stratégies et méthodes, nous avons l'indexation sémantique et l'indexation sémantique via les ontologies. Il s'agit, par exemple, de l'utilisation des délimiteurs, de la mise à jour incrémentale d'index et de la méthode différentielle. Ces méthodes considèrent une indexation de type « sac de mots », dans laquelle les termes issus des documents sont considérés comme indépendants. Cependant, il existe en réalité des relations (équivalence, subsomption, association, ...) entre les termes. De nouvelles approches tentent de les prendre en compte de façon automatique lors de l'indexation au travers, par exemple, de l'indexation sémantique et de l'indexation sémantique via les ontologies. A cet effet, plusieurs

ontologies de domaines spécifiques ont été conçues. Ces ontologies ont pour but de faire asseoir une recherche conceptuelle permettant de simplifier la navigation à travers les catégories sémantiques de la hiérarchie utilisée. Ceci exige l'utilisation d'un langage de représentation de connaissances permettant de spécifier le contenu des pages web selon une taxonomie de concepts [Daou 08]. Le processus de recherche d'information comprend différentes grandes fonctions qui distinguent les systèmes les uns des autres. La fonction d'indexation vise à construire une représentation réduite des contenus des documents et des requêtes [Mot 07] parfois nommées « cueillette de données » ou « collecte d'informations ». C'est une étape importante de l'élaboration d'un projet dont dépend la suite des événements. Les sources de l'information sont nombreuses et varient de classe en classe. Bien que les livres soient la source la plus exploitée actuellement, plusieurs autres sources présentent des possibilités intéressantes et sont parfois même plus appropriées à la nature de l'information recherchée. De même, dans la bibliothèque des archives, le souci majeur est d'établir des représentations de documents dans le but d'en récupérer des informations à travers la construction d'index. Ce procédé permet de retrouver rapidement les documents. Le même problème se pose en informatique médicale et plus précisément dans la médecine traditionnelle (MT) [Tra 12].

Nous nous intéressons, dans ce chapitre, à l'indexation sémantique c'est-à-dire à l'extraction des concepts et de leurs relations entre eux et ce, via les ontologies après nous être appuyés sur les technologies du web sémantique. Dans ce type d'approche, la connaissance du domaine (terminologique en particulier) est représentée sous forme d'ontologies, c'est-à-dire sous forme de concepts, d'instances de ces concepts et de leurs relations. C'est une nouvelle génération de moteurs de recherche basée sur la Recherche d'information (RI) ontologique dont l'objectif est de délivrer de l'information pertinente et appropriée au contexte de l'utilisateur qui a émis la requête. Nous exploitons les graphes conceptuels pour représenter le réseau sémantique et nous utilisons le moteur de recherche d'information SPARQL qui se base sur des fonctions de spécialisation, de généralisation et de projection pour offrir une recherche d'information en profondeur. La recherche d'information utilisée dans cette partie est basée sur la relation entre les graphes conceptuels et la représentation avec le formalisme RDF [Dja 05].

5.2. Recherche d'information (RI)

Dans les domaines de l'informatique et des sciences de l'information, la recherche d'information (« information retrieval » en anglais), désigne essentiellement la problématique de l'accès à un document ou à un fragment documentaire numérisé dont la localisation n'est pas connue mais dont l'existence ou la pertinence est acquise [Man 07]. Mais dans de nombreux

cas, la recherche d'information (RI) est une discipline de recherche qui intègre des modèles et des techniques dont le but est de faciliter l'accès à l'information pertinente pour un utilisateur ayant un besoin en information. De même, dans de nombreux cas, la recherche d'information s'inscrit dans la problématique de l'analyse d'une situation complexe dont on cherche à percevoir les contours sans savoir à l'avance s'il existe des ressources documentaires susceptibles de répondre au besoin. Le besoin en information de l'utilisateur est souvent formulé en langage naturel par une requête décrite par un ensemble de mots clés. Pour une requête utilisateur, un système de RI permet de retrouver un sous-ensemble de documents susceptibles d'être pertinents à partir d'une collection de documents, en réponse à cette requête [Our 11].

5.3. Indexation

L'objectif des systèmes de recherche d'information (SRI) est de fournir aux utilisateurs les documents pertinents par rapport aux besoins qu'ils expriment [Hub 09]. L'hypothèse sous entendue dans la mise en place d'un système à indexation est que ce dernier est plus apte à désigner des entités sémantiques ou des concepts que les mots uniques et constitue alors une meilleure représentation du contenu sémantique des documents [Jean 01]. La recherche d'informations est l'ensemble des opérations nécessaires pour répondre à la demande d'un utilisateur. De même, l'indexation est généralement associée à la recherche d'informations dans la mesure où la pose d'index permet d'améliorer la recherche de documents répondant à une requête donnée. L'indexation des documents peut être simple, par métadonnées et/ou par mots-clés (actuellement recommandés dans les documents HTML), ou plus complexe en utilisant des ontologies (soutenu par le consortium de web sémantique).

5.3.1. Indexation par une liste de descripteurs

Ce type d'indexation permet la représentation d'un texte par un ensemble de descripteurs simples et complexes. Ces descripteurs permettent de distinguer deux niveaux d'indexation :

5.3.1.1. Indexation libre

Dans ce cas, le lecteur choisit librement des mots ou des expressions qui lui semblent appropriés pour décrire le contenu d'un texte. C'est un mécanisme purement intuitif et personnel. En effet, aucun calcul précis n'est effectué sur la portée sémantique des mots du lexique et l'emplacement des mêmes mots dans le texte.

5.3.1.2. Indexation par un langage documentaire

L'indexation par un langage documentaire repose sur une liste nominative et limitée de descripteurs qui sont les « mots » du langage. Ces descripteurs peuvent être structurés sémantiquement et leur utilisation peut être régie par une syntaxe. Une indexation contrôlée ou indexation par un langage documentaire ne doit comporter que des éléments de la logique de descriptions. Ce qui implique que ce langage soit exhaustif quant à la représentation des divers sujets traités dans le domaine couvert par la collection de documents. C'est un langage formel et figé ; il comporte, non seulement une liste de descripteurs, mais aussi des règles d'utilisation de ces descripteurs. Son apprentissage se fait à la fois par l'indexeur et par l'interrogateur [Sah 02].

5.3.2. Indexation dynamique d'ontologie

L'indexation dynamique consiste à mettre à jour l'index après modification de la collection (ajout, modification et suppression de documents) [Hub 09] dans le domaine de la recherche et de la collecte incrémentale des pages web. Ces types d'indexation visent à permettre aux collections d'un moteur de recherche d'être plus synchronisées avec le web. L'exemple conçu par Google pour la recherche d'informations sur le W3C montre les atouts d'une indexation dynamique (voir figure 5.1). Il emploie plusieurs techniques (importance des pages ou Page Rank, structure des liens, texte des liens, polices de caractères, position des mots dans les documents, etc....) pour améliorer la qualité de recherche. L'indexation avec une ontologie permet d'exprimer les relations entre des expressions du document à l'aide de celles des concepts auxquels les expressions sont associées dans l'ontologie. De plus, l'utilisation des concepts d'ontologie comme vocabulaire de référence d'index sert à bien préciser les sens accordés aux termes d'un document.

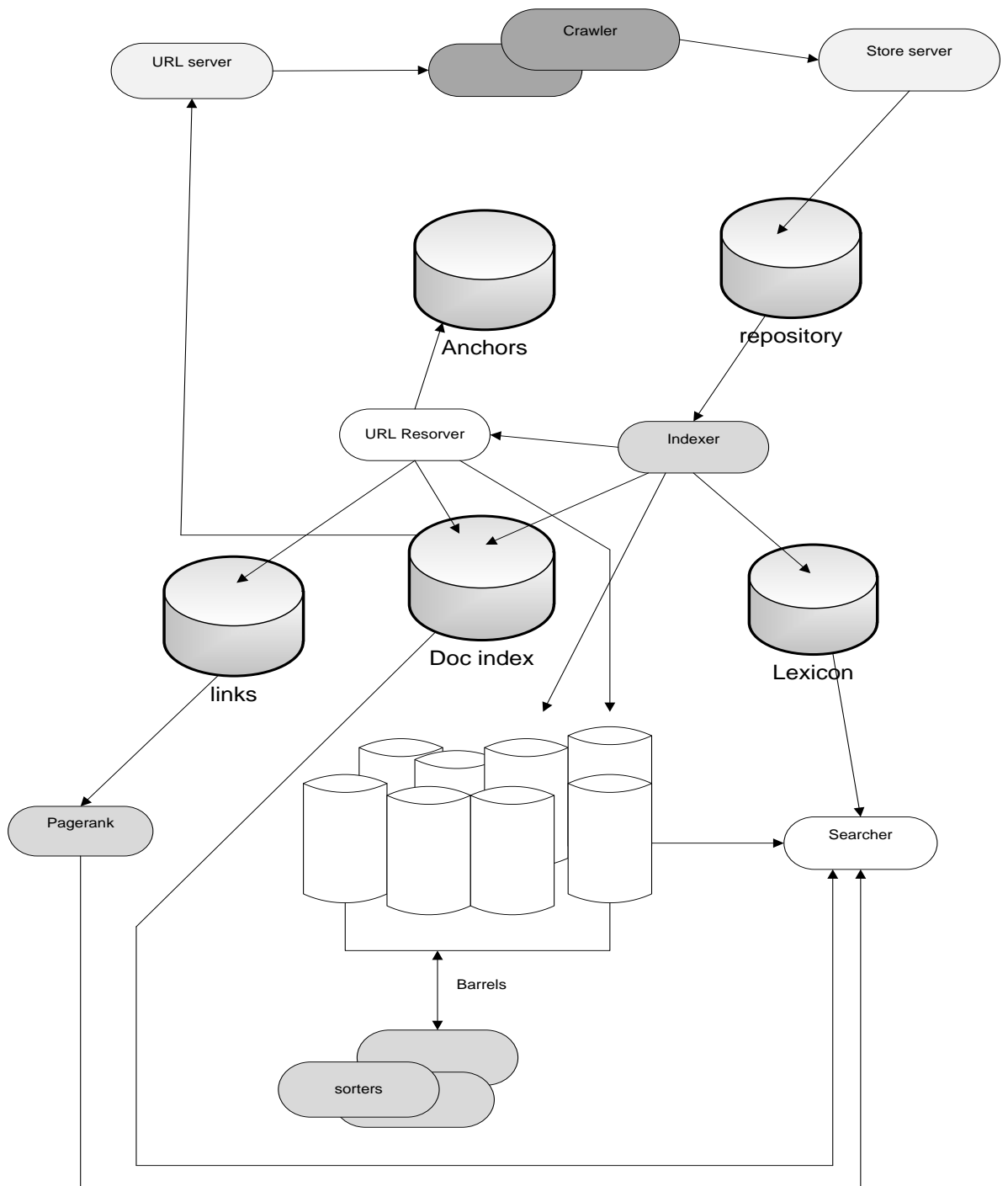


Figure 5.1 : Architecture du moteur de recherche Google

5.3.3. Indexation sémantique d'ontologie

L'utilisation d'une ontologie lors de la phase d'indexation permet de lever les ambiguïtés des sens des termes utilisés et de mieux représenter les connaissances inhérentes dans les documents. En terme d'indexation sémantique, des concepts de l'ontologie sont associés à chaque document selon les sémantiques qui y sont véhiculées. Différents travaux ont montré l'intérêt d'utiliser une indexation sémantique à base d'ontologie. Dans le domaine de

l'apprentissage en ligne, un auteur [Chang 07] propose une indexation basée à la fois sur une ontologie du domaine de l'apprentissage et sur une ontologie dérivée de LOM (Learning Object Metadata) qui représente les métadonnées décrivant les ressources pédagogiques. Dans le cadre des recherches d'objets pédagogiques relatifs aux mathématiques au cours secondaire, les résultats montrent une meilleure efficacité en matière de rappel et de précision par rapport aux mêmes recherches basées sur des mots-clés [Hub 09]. Dans ce type d'indexation, les structures de données utilisées permettent d'associer des concepts issus d'une ontologie aux documents de la collection. Cependant, le suivi de la dynamique (ajout, suppression, modification) des documents de la collection ainsi que l'impact de ces évolutions au niveau de l'index sémantique ne sont pas pris en compte.

5.4. Arbre conceptuel

Un arbre conceptuel est un formalisme qui offre une structure mettant en œuvre les relations sémantiques entre les concepts. Dans ce formalisme, il existe un certain nombre de fonctions d'inférence. Ce sont, pour la plupart des mécanismes d'aide à la recherche d'informations. Nous notons en particulier des fonctions de projection, de jointure, de généralisation et de spécialisation. Ce formalisme a été employé et validé dans un certain nombre d'applications. Ainsi, son efficacité est généralement approuvée. Le Moteur de recherche « SPARQL » est basé sur ce principe [Dja 05]. « SPARQL » est une recommandation du W3C dont la standardisation finale date du 15 Janvier 2008 [Nat 06].

5.5. Concepts de base de la recherche d'information

Chaque Système de Recherche d'Information (SRI) est jugé au travers de ses performances en terme de satisfaction de l'utilisateur sur les pertinences des documents retrouvés, du temps de réponse aux requêtes utilisateurs et la disponibilité du système. Ainsi, pour le fonctionnement de ce système, doivent être pris en compte certains paramètres afin de définir les structures de données à utiliser dans les listes inversées. Plusieurs autres paramètres entrent en jeu dans la performance des SRI. Ce sont :

- la taille du corpus (nombre de documents constituant la collection). Elle peut affecter la durée de ré-indexation ou même entraîner une lenteur du système, voire son indisponibilité pendant un certain temps.
- la fréquence de mise à jour de la collection et le format des documents ; ils peuvent avoir un impact sur la disponibilité de l'index.

- le format des documents affecte le temps d'indexation car les durées d'extraction des termes et expressions d'un document ne sont pas les mêmes pour tous les formats. Un document est un support physique destiné au stockage d'une information. Une information est un ensemble de connaissances ayant un sens bien déterminé dans un domaine spécifique.

5.6. Recherche sémantique

La recherche sémantique est différente de la recherche de mots-clés ordinaire. La recherche sémantique permet d'améliorer la précision de la recherche d'information à travers un langage de requête d'ontologie (SPARQL) [Hari 11] ; elle prend des données non structurées et les indexe dans un format de données structurées ; elle s'exécute sur la structure ontologique et rend les types de valeurs qui peuvent satisfaire les conditions dans la question. Cette méthode est une recherche Booléenne basée sur le modèle RDF.

5.7. Moteurs de recherche

Dans cette section, nous allons présenter quelques moteurs. Un moteur de recherche est une application permettant de retrouver des ressources associées à des mots quelconques. Les moteurs de recherche les plus connus sont : GOOGLE, YAHOO, ASK BRING et MSN ; ces moteurs sont basés sur une technologie complexe ayant tous en commun les mêmes principes et étapes de crawling, d'indexation et de ranking [Tra. 12]. Il faut noter que les moteurs de recherche ne sont pas destinés uniquement à la navigation internet. Il existe également des logiciels installés sur un PC (Personnal Computer) qui combinent la recherche parmi les fichiers stockés sur le PC et la recherche parmi les sites web tels que : Copernic, Desktop Search, etc.

5.8. Requêtes

Le processus de recherche d'information a pour but la mise en relation des informations disponibles d'une part et les besoins de l'utilisateur d'autre part. Ces besoins sont traduits de façon structurée par l'utilisateur sous forme de requêtes. Les requêtes retournent à l'utilisateur le maximum de documents pertinents attendus. La notion de pertinence est difficile à déterminer pour automatiser le besoin de l'utilisateur. Par contre, les requêtes soumises au moteur sont prédéfinies et les réponses pertinentes sont également connues à l'avance. En d'autres termes, la précision et le rappel restent des éléments essentiels dans la mesure où toutes les réponses ne sont pas fiables. Le processus de recherche d'information est communément présenté sur la figure 5.2 :

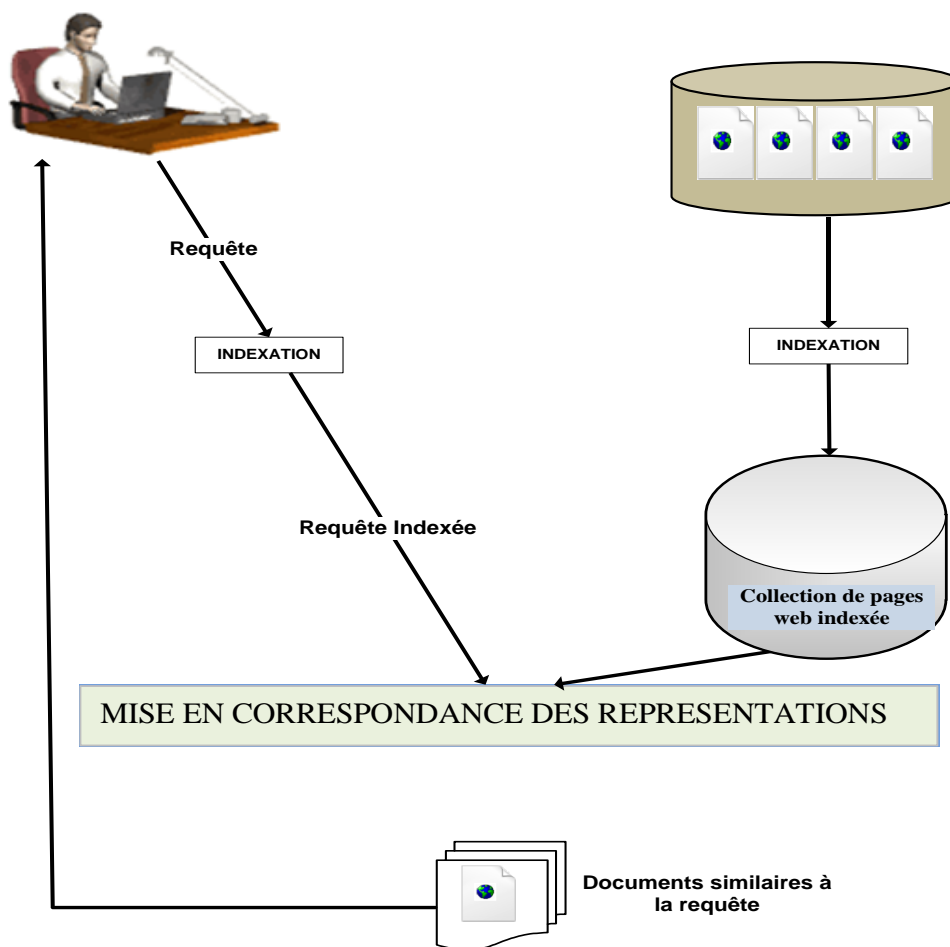


Figure 5.2 : Architecture d'un moteur de recherche

Une précision mesure la proportion de documents pertinents relativement à l'ensemble des documents restitués par le système. La précision est définie aussi comme un écart entre une valeur mesurée ou prédite et une valeur réelle. Elle est exprimée comme suit :

Précision = $\frac{N_{pr}}{N_r}$ avec N_r le nombre de documents renvoyés par un système pour une requête donnée et N_{pr} le nombre de documents pertinents renvoyés par le système.

Le rappel mesure la proportion de documents pertinents restitués par le système relativement à l'ensemble des documents pertinents contenus dans la base documentaire. Il est exprimé selon le rapport suivant :

Rappel = $\frac{N_{pr}}{N_p}$ avec N_p le nombre de documents pertinents dans la collection pour une requête.

En se basant sur cette formule, les taux de bruit et de silence sont donnés par les formules suivantes :

$$\text{Bruit} = 1 - \frac{N_{pr}}{N_r} \text{ ou } \text{Bruit} = 1 - \text{Precision.}$$

$$\text{Silence} = 1 - \frac{N_{pr}}{N_p} \text{ ou } \text{Silence} = 1 - \text{Rappel}$$

Dans le cas où tous les documents pertinents sont sélectionnés,

$$\text{Taux de Rappel} = \text{Taux de Précision.}$$

5.9. Outil de recherche sémantique

Le moteur de recherche opté dans notre cas est le SPARQL. C'est un langage de requêtes et un protocole pour l'accès RDF qui s'appuie sur les triplets le constituant. Il a été conçu par le groupe de travail du W3C RDF dont la dernière mise à jour date de janvier 2008. Il permet d'interroger un ensemble de descriptions RDF à partir d'opérations de mise en correspondance de patterns de graphes, de conjonctions, de disjonctions de patterns et de patterns optionnels. SPARQL intègre également des balises (tags) spécifiques, l'union et l'intersection, le filtrage, les opérateurs de comparaison des valeurs. Il permet d'effectuer des requêtes plus efficaces et flexibles. En d'autres termes, il est différent du modèle classique SQL qui est seulement adapté aux bases de données de type relationnel. La structure d'une requête SPARQL est très similaire à celle employée dans le langage SQL. Une requête SELECT, de type interrogatif, permet d'extraire du graphe RDF un sous-graphe correspondant à un ensemble de ressources vérifiant les conditions définies dans une clause WHERE. En effet, dans le langage SPARQL, il est possible de demander si une requête dispose d'au moins une solution. Pour ce faire, le SELECT est remplacé par un ASK. Il est aussi possible de construire un nouveau graphe RDF via le terme CONSTRUCT.

5.10. Comparaison SQL / SPARQL

Le tableau 5.1 fait une comparaison entre le langage SQL et le langage SPARQL

Tableau 5.1 : Comparaison des langages SQL et SPARQL

	SQL	SPARQL
Critère de comparaison	Base de données	Schémas et graphe
Élément de base	Enregistrement	Descripteur de Ressources

Critère d'accès	Identifiant unique	URI (Universal Resource Identifier) uniforme
Structure des éléments	Attributs des enregistrements	Liste des prédicats
Valeur du contenu	Valeurs des attributs ou des clés Etrangères	Valeurs sont des liens vers d'autres ressources

5.10.1. Modèle

Le modèle présenté à la figure 5.3 prend notamment en compte le double objectif d'actualisation dynamique des listes inversées et d'utilisation d'ontologies lors de l'indexation des documents.

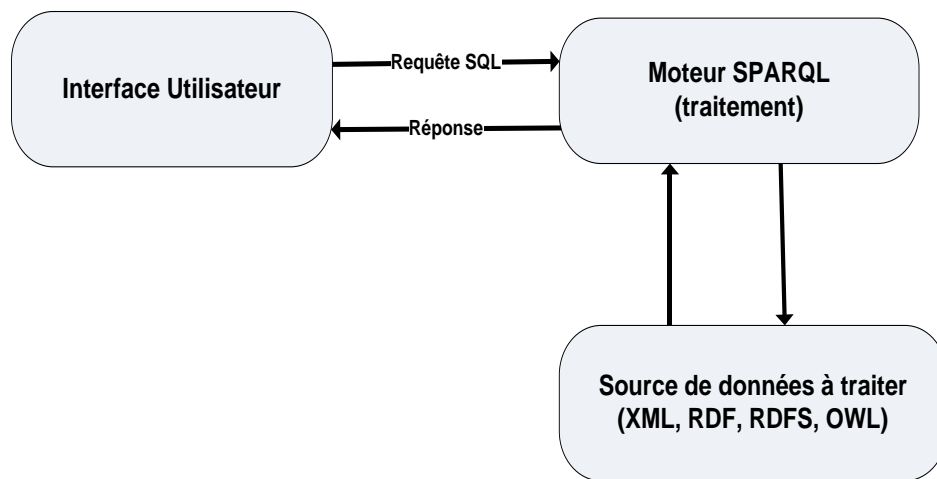


Figure 5.3 : Système d'interaction du SPARQL

5.10.2. Quelques aspects de l'outil

L'outil de recherche SPARQL minimise le délai de mise à jour dynamique d'index. Ceci limite donc les ajouts de nouveaux documents ou modification de documents dans le corpus et leur prise en compte dans les index. Le calcul des poids des expressions et concepts se fait au moment de l'évaluation de la requête c'est-à-dire de manière automatique. Il permet également de réduire le temps de réponse. Le temps qui s'écoule entre la saisie de la requête et l'affichage de résultat doit se situer dans la limite acceptable par l'utilisateur. Le fait de traiter dynamiquement l'indexation des documents ne doit pas pénaliser de trop le temps de réponse du système lors de la recherche de documents. La minimisation du temps d'indexation assure la disponibilité permanente de l'index. A la soumission d'une requête, le système permet la recherche sémantique sur les contenus. Le système de recherche dans l'ontologie prend en compte les concepts qui correspondent aux termes de la requête puis fournit les résultats des

concepts indexés. Il faut signifier que le format des documents chargé est pris en compte par le système de recherche. La figure 5.4 présente un exemple de requête SPARQL. Cette requête permet d'extraire le nom de toutes les personnes contenues dans une base de triplets.

```
1 PREFIX rdf : <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX owl : <http://www.w3.org/2002/07/owl#>
3 PREFIX xsd : <http://www.w3.org/2001/XMLSchema#>
4 PREFIX rdfs : <http://www.w3.org/2000/01/rdf-schema#>
5 PREFIX hum : <http://www.inria.fr/2007/09/11/humans.rdfs#>
6 select ?plantes ?nom
7 where {? plantes med: Name ?nom}
```

Figure 5.4 : Requête SPARQL retournant les noms de plantes présentes dans un fichier RDF.

5.11. Conclusion

Nous avons vu, dans ce chapitre, comment la recherche traditionnelle d'informations peut s'ouvrir sur les Technologies de l'Informations et de la Communication (TICs) dans le cadre d'un processus de recherche d'informations dans une base médicale ontologique. La recherche traditionnelle d'informations qui, autrefois, était basée sur l'usage de mots clés fait désormais place à une automatisation fondée sur les TICs. Ce processus a permis de prendre en compte les différentes méthodes proposées par ces nouvelles technologies. Nous avons, pour ce faire, présenté les différents modes de recherche moderne existants tels que l'indexation sémantique et l'indexation à base de ressources sémantiques comme les ontologies (thesaurus, bases de données sémantiques). Pour accéder à cette recherche d'information dans une base ontologique, il faut passer par un raisonnement ontologique. Ce raisonnement offre un niveau supérieur tel que la vérification de la connaissance et sa cohérence. De même, il offre un niveau supérieur à la recherche sémantique d'informations grâce à un ensemble de règles d'inférences.

Le chapitre suivant sera consacré à l'application des Réseaux Bayésiens au diagnostic et au traitement d'une maladie dans le cadre de la médecine traditionnelle.

Chapitre 6

RAISONNEMENT MEDICAL : APPLICATION DES RESEAUX BAYESIENS AU DIAGNOSTIC D'UNE MALADIE DANS LA MEDECINE TRADITIONNELLE.

Résumé. Le processus de raisonnement médical est très complexe. Il a été l'objet de nombreuses études notamment en psychologie cognitive. Les auteurs comme Nendaz et al. [Nen 05] ont mis en exergue le fait que les processus du raisonnement clinique sont hypothético-déductifs et/ou issus de la reconnaissance de similarité avec un cas déjà vécu. Dans le contexte du développement de la médecine traditionnelle, la capitalisation des expériences est fondamentale pour la pérennité de la science et l'harmonisation des diagnostics et des prescriptions. L'objet de ce chapitre est de proposer une approche de diagnostic d'une maladie en médecine traditionnelle à l'aide des Réseaux Bayésiens. Ce système s'appuie sur l'ontologie d'une maladie prenant en compte à la fois les signes cliniques et les plantes médicinales utilisées dans le traitement d'une maladie.

6.1. Introduction

L'action que mène le tradipraticien repose généralement sur sa capacité de raisonnement et son aptitude à prendre des décisions face à un problème ; alors que les informations utilisées sont potentiellement entachées d'incertitudes. Ces incertitudes sont d'origine multiple : possibilité d'erreur dans les données, ambiguïté de la représentation de l'information, incertitudes sur les relations entre les diverses informations. Une première approche à la représentation de la connaissance dans ce contexte d'incertitude a utilisé la théorie des probabilités. Ainsi, plusieurs études ont montré que les systèmes d'aide à la décision élaborés sous le modèle probabiliste pouvaient faire mieux que le médecin. Dans la pratique médicale, le diagnostic de la maladie dont souffre un patient est certainement l'une des activités les plus importantes du praticien. En effet, de ce diagnostic dépend le traitement à prescrire au patient. Si ce diagnostic est erroné, il va de soi que de nombreux problèmes peuvent survenir dont en particulier l'inadéquation du traitement au mal du patient. C'est dans ce cadre que le processus de raisonnement médical conduisant à un diagnostic a fait l'objet de nombreuses études. Dans son ouvrage intitulé « Raisonnement médical : la quête de l'explicite », Alain JUNOD explique très clairement les deux approches utilisées. La première est issue de la psychologie cognitive et la seconde approche, appelée généralement « analyse décisionnelle », résulte d'une formalisation du raisonnement par une approche probabiliste [Juno 07].

Dans ce chapitre nous rappelons quelques notions fondamentales des réseaux bayésiens ; puis, nous allons, par la suite présenter un modèle probabiliste proposé pour le diagnostic d'une maladie en pratique de médecine traditionnelle.

6.2. Raisonnement Bayésien

L'idée de base des théories sur le diagnostic médical est qu'il dépend en grande partie des connaissances de l'agent de santé effectuant le diagnostic. Si l'on peut formaliser cette connaissance du point de vue qualitatif, le raisonnement médical pourra être grandement facilité. Il s'agit d'une formalisation de connaissances qualitatives sur la base desquelles des raisonnements peuvent être effectués naturellement. Il est légitime de se demander, dans le cas de la médecine traditionnelle, quel comportement devrait adopter un praticien s'il ne maîtrise pas le traitement d'une maladie. Une telle préoccupation, si elle n'est pas prise en compte dans la modélisation, relèverait de la subjectivité. Or, l'un des intérêts de l'organisation de la médecine traditionnelle réside dans le fait que la part de subjectivité doit être réduite au strict minimum pour aboutir à des lois générales propres à l'usage des plantes médicinales. Il est donc nécessaire qu'un mécanisme soit introduit dans le raisonnement médical dans le but de rediriger le patient vers d'autres spécialistes dans le cas où le praticien ne maîtriserait pas les plantes médicinales entrant dans le traitement de la maladie dont souffre le patient.

Dans ce contexte, la technologie des Réseaux Bayésiens constitue l'outil privilégié pour la modélisation du raisonnement médical dans l'optique de diagnostiquer une maladie. Cette technologie présente l'avantage de modéliser la connaissance experte que l'on a du mal à appliquer et de s'en servir pour répondre à de nombreuses questions par un processus de simulation informatique correspondant à ce qu'il est convenu d'appeler un raisonnement automatique en Intelligence Artificielle.

6.3. Ontologie

Dans la littérature, les ontologies ont fait l'objet de nombreux travaux de recherches. Ces recherches portent sur l'apprentissage de plusieurs disciplines dont celui de la médecine qui nous intéresse principalement. En réalité, c'est le processus d'apprentissage et la pratique de la médecine traditionnelle qui font ici les objets de notre réflexion. Certains parlent d'apprentissage de manière générale sans l'adapter à la médecine traditionnelle ; d'autres par contre, donnent des explications sur la structure ontologique de la médecine traditionnelle sans toutefois approfondir le mécanisme d'apprentissage et la pratique de cette médecine traditionnelle. De même, la plupart des auteurs ne prennent pas en compte les problèmes liés à

la complexité du raisonnement en matière de diagnostic des maladies et suppose ces compétences comme acquises par les praticiens, ce qui est problématique. Pour prendre en compte le diagnostic des maladies dans la médecine traditionnelle, nous proposons une approche basée sur une ontologie. Pour les besoins de présentation de notre approche, nous utilisons une ontologie simplifiée du paludisme comme le montre la figure 6.1 suivante :

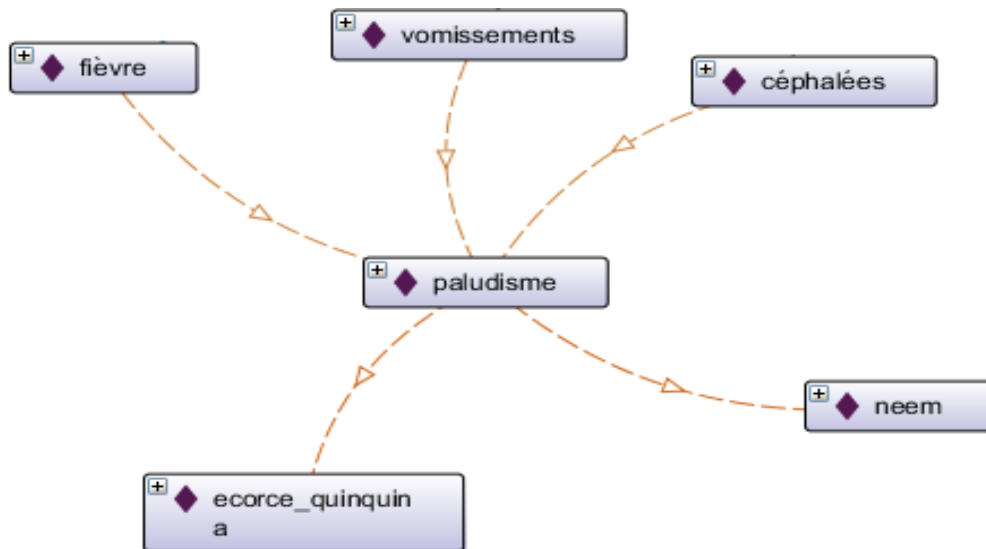


Figure 6.1 : Ontologie du paludisme

Cette ontologie présente le paludisme, les signes cliniques pris en compte pour son diagnostic (fièvre, céphalées et vomissements) et les plantes médicinales entrant dans le traitement de celui-ci.

6.4. Réseaux Bayésiens

Encore appelés Systèmes experts probabilistes, les Réseaux Bayésiens sont des outils de représentation des connaissances sur un système et de raisonnement automatique sur ces connaissances. Aussi, appelés diagrammes d'influence (belief network), les réseaux bayésiens constituent un des modèles de représentation des connaissances utilisables en analyse de décision. Un réseau bayésien, également appelé diagramme d'influence ou réseau causal est un graphe dirigé acyclique dans lequel les nœuds représentent les variables et les arcs précisent les dépendances probabilistes entre les variables. Il permet d'afficher graphiquement les variables d'un problème de décision et les relations ou influences entre ces variables, qui peuvent mener à des décisions complexes. Ils ont été introduits par Judea Pearl [Far 12] dans les années 1980 et se sont révélés être de puissants outils pratiques pour la représentation de connaissances incertaines et de raisonnement à partir d'informations incomplètes. Les Réseaux Bayésiens sont donc des outils de simulation qui permettent d'observer le comportement d'un système

complexe dans des contextes et des conditions qui ne sont pas nécessairement accessibles à l'expérimentation [Ler 06, Nai 07].

Techniquement, les réseaux Bayésiens sont des modèles graphiques combinant la théorie des graphes et la théorie des probabilités. L'apport de ce chapitre consiste à proposer une méthode permettant de modéliser le raisonnement médical par la technologie des Réseaux Bayésiens en se basant sur une ontologie de la maladie à diagnostiquer. La figure 6.2 suivante présente un exemple de Réseau Bayésien :

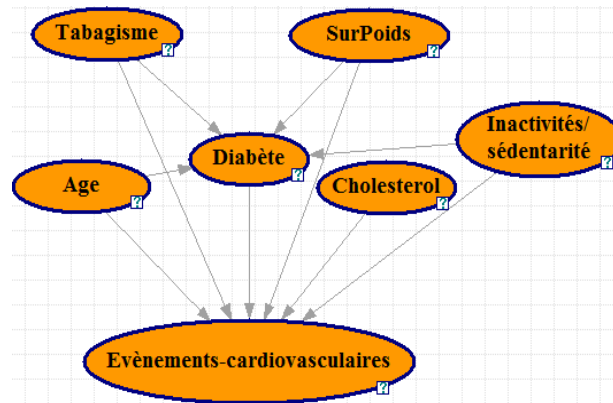


Figure 6.2 : Exemple de Réseaux Bayésiens événement

Cette figure 6.2 est un Réseau Bayésien permettant de modéliser le processus de déclenchement du diabète et d'un évènement cardiovasculaire. Ainsi donc, le tabagisme, le surpoids, l'inactivité, l'hyper cholestérimie sont des facteurs susceptibles de déclencher le diabète qui lui-même peut favoriser la survenue d'un évènement cardiovasculaire.

Comme le montre la figure 6.2, un Réseau Bayésien est un graphe orienté dans lequel les nœuds représentent les variables et les arcs symbolisent les relations de dépendance entre ces différentes variables. Chaque nœud est doté d'une table de probabilités conditionnelles qui est une modélisation des croyances en la survenue de tel ou tel cas lorsque nous sommes dans telle ou telle condition. Dans le cas du diagnostic médical, un tel graphe est fourni directement par l'ontologie composée des signes cliniques de la maladie et des connaissances en éléments médicaux entrant dans le traitement de cette maladie.

6.5. Modélisation

Considérons une maladie M donnée. Soit Z , la variable décrivant la présence ou l'absence de cette maladie. Soit $X = (X_i)_{1 \leq i \leq I}$ l'ensemble des signes cliniques de la maladie M et $Y = (Y_j)_{1 \leq j \leq J}$ l'ensemble des traitements les plus connus entrant dans la guérison de la maladie M .

6.5.1. Structure du Réseau

Intuitivement, il est évident que la qualité du diagnostic dépend d'un certain nombre de signes cliniques que le spécialiste recherche chez le patient. Il est donc clair que les signes cliniques $X = (X_i)$ déterminent directement le diagnostic du spécialiste en termes de présence ou d'absence de la maladie M , donc de la variable Z . Mais dans le cas de la médecine traditionnelle, on désire tenir compte du niveau de connaissances de l'ensemble des moyens de traitement dont dispose le tradipraticien pour affiner le diagnostic. En effet, nous estimons poser le diagnostic d'une maladie sur la base exclusive des signes cliniques et intégrer ensuite les modalités de traitement et ce, dans la mesure où les pratiques médicales ne sont pas harmonisées. La capacité d'une bonne prescription du traitement adéquat est fondamentale. Nous pensons donc que le modèle devrait pouvoir réagir si le niveau de connaissances et de la symptomatologie et des moyens de traitement du praticien est insuffisant.

L'hypothèse de base de la modélisation est que le praticien de la médecine traditionnelle utilisera toujours un traitement à base de plantes médicinales. Par conséquent, s'il ne connaît pas exactement l'ensemble des traitements entrant dans la guérison de la maladie M , il serait judicieux que la machine donne un résultat négatif. Cette précaution aura pour conséquence de conduire le patient vers une autre structure. L'introduction de cette précaution dans le modèle se fait en considérant simplement le diagnostic comme une conséquence de la connaissance que le praticien a de l'usage d'un certain nombre de traitements. Ainsi, si l'on s'en tient à l'ontologie basique proposée à titre d'exemple du paludisme, la structure du Réseau Bayésien correspondant est la suivante :

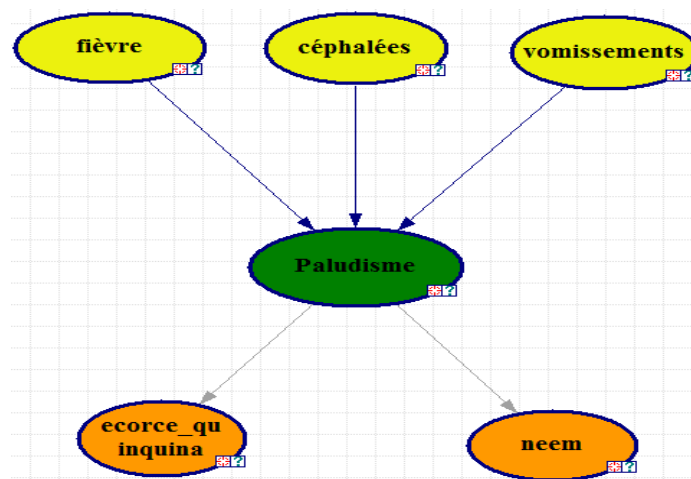


Figure 6.3 : Structure du Réseau Bayésien

La figure 6.3 est la structure du Réseau Bayésien correspondant à l'ontologie définie à la figure 6.1. Mais contrairement à l'ontologie où les relations sont constituées de sémantiques

diverses (les différents types de relations), les flèches de la figure 6.3 représentent un seul type de relation à savoir des relations de dépendances (causes à effets) entre les symptômes de la maladie, les plantes médicinales et le diagnostic de la maladie considérée.

6.5.2. Tables de probabilités

Les Réseaux Bayésiens permettent essentiellement de calculer des probabilités en se basant sur le théorème de Bayes. Si l'on considère les événements A et B , et si l'on note $\mathbb{P}(\cdot)$ l'opérateur de probabilité uniforme, on a les trois axiomes fondamentaux de la théorie des probabilités suivants $\forall X \in \{A, B\}$:

$$1) 0 \leq \mathbb{P}(X) \leq 1 \quad (1)$$

$$2) \mathbb{P}(X) = 1 \text{ si } X \text{ est un évènement certain} \quad (2)$$

$$3) \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \text{ si les évènements } A \text{ et } B \text{ sont disjoints} \quad (3)$$

En plus de ces axiomes, la probabilité de l'évènement A conditionnellement à la réalisation de l'évènement B est définie par :

$$\mathbb{P}(A \setminus B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad (4)$$

Muni de ces règles, la probabilité conjointe des événements A et B est donnée par la règle fondamentale suivante :

$$\mathbb{P}(A \cap B) = \mathbb{P}(B \setminus A) \mathbb{P}(A) \quad (5)$$

$$(1) \text{ et } (3) \Rightarrow \mathbb{P}(A \setminus B) = \frac{\mathbb{P}(A)}{\mathbb{P}(B)} \mathbb{P}(B \setminus A) \quad (6)$$

Ainsi, la connaissance, à priori, de la probabilité d'un évènement A et de la probabilité d'un évènement B conditionnellement à l'évènement A permet d'en déduire, à posteriori, celle de l'évènement A conditionnellement à l'évènement B . Toutefois, cette inférence nécessite la connaissance de la probabilité de réalisation de l'évènement B utilisé. En pratique, les évènements A et B peuvent revêtir des formes relativement complexes, comme c'est le cas dans les Réseaux Bayésiens. Pour utiliser le théorème de Bayes dans les Réseaux Bayésiens pour la modélisation du diagnostic médical des maladies dans la médecine traditionnelle, notons :

- K , le nombre d'états dans lesquels peut se trouver un patient relativement au symptôme X_i .
- L , le nombre de modalités de la variable test décrivant l'usage de la plante médicinale Y_j dans le traitement de la maladie M .
- x_{ik} , l'état du symptôme X_i dans lequel se trouve le patient (le k^{eme} état).

- y_{jl} , la modalité choisie par le praticien relativement au test de connaissance sur la plante médicinale Y_j pour le traitement de la maladie M (la l^{eme} modalité).

Si l'on considère le graphe du Réseau Bayésien tel que défini par la figure 6.3, l'usage du modèle pour l'estimation des probabilités sur lesquelles sera fondé le raisonnement automatique, nécessite la connaissance des trois sortes de probabilités suivantes :

- $\mathbb{P}(X_i = x_{ik}), \forall k \in [1, K]$ (7)

- $\mathbb{P}(Y_j = y_{jl}), \forall l \in [1, L]$ (8)

- $\mathbb{P}(Z = z | X_i = x_i)$ et $\mathbb{P}(Y_j = y_j | Z)$, $x_i = (x_{ik}, \dots, x_{iK_i})$ et $y_j = (y_{j1}, \dots, y_{jL_j})$ (9)

Les probabilités $\mathbb{P}(X_i = x_{ik})$ constituent des mesures de l'importance des signes cliniques dans le diagnostic de la maladie M . Il s'agit de la probabilité que le patient souffre de la maladie M s'il est dans l'état x_{ik} relativement au signe clinique X_i . Quant aux probabilités $\mathbb{P}(Y_j = y_{jl})$, elles fournissent une évaluation de l'importance du traitement Y_j de la maladie M . Il s'agit de la probabilité que le patient obtienne la guérison si le traitement Y_j est utilisé ($y_{jl} = 1$) ou pas ($y_{jl} = 0$) dans le traitement de la maladie M . L'ensemble de ces probabilités est évalué à l'aide de $I + J$ tables de probabilité, chaque table étant associée à un signe clinique ou à un traitement (les nœuds fils du nœud de la maladie).

Quant aux différentes probabilités fournies par $\mathbb{P}(Z = z | X_i = x_i)$ et $\mathbb{P}(Y_j = y_j | Z = z)$, elles constituent le fondement de l'apprentissage du Réseau Bayésien. Son calcul constitue un challenge car elle met en relation de nombreux paramètres. En effet, la table de probabilité associée comporte $I \times J \times 2$ valeurs de probabilités à estimer. En pratique, quelle que soit la complexité d'une telle loi de probabilité, les Réseaux Bayésiens disposent de nombreux algorithmes pour apprendre ces paramètres (estimer ces probabilités) en utilisant des données individuelles, c'est-à-dire collectées sur un ensemble de patients diagnostiqués dans notre cas. Et bien que les probabilités conditionnelles à priori peuvent être fournies par un ou quelques experts du domaine de la médecine traditionnelle maîtrisant la maladie dont on modélise le diagnostic, l'usage d'experts pour fournir ces probabilités à priori n'est pas souhaitable à cause du nombre important de probabilités à priori nécessaires.

6.5.3. Factorisation des probabilités

Pour effectuer un tel calcul très complexe, on fait l'hypothèse d'indépendance des différents signes cliniques les uns par rapport aux autres, des différentes plantes médicinales les unes par rapport aux autres ainsi que l'indépendance des plantes médicinales et des signes cliniques. Dans une modélisation plus complexe, on pourra envisager la levée de cette dernière

hypothèse. Cela modifiera la structure graphique du Réseau Bayésien sans changer les principes de modélisation sous-jacents. Techniquement, ces hypothèses permettent d'obtenir ces probabilités grâce à une règle fondamentale de factorisation des probabilités conditionnelles. En effet, Si l'on considère trois évènements A_1 , A_2 et A_3 , on peut déduire la règle de factorisation suivante :

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1 | A_2)\mathbb{P}(A_2) \quad (10)$$

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1 | \{A_2 \cap A_3\})\mathbb{P}(A_2 \cap A_3) \quad (11)$$

$$= \mathbb{P}(A_1 | \{A_2 \cap A_3\})\mathbb{P}(A_2 | A_3)\mathbb{P}(A_3) \quad (12)$$

D'une manière générale, ce principe de calcul obtenu avec l'équation (12) peut être aisément généralisé pour un nombre fini d'évènements quelconques. Ce principe de factorisation des probabilités jointes permet de simplifier considérablement le calcul du résultat du diagnostic.

En effet, dans le cas de notre modèle, tous les évènements $\{Z = z | X_i = x_i\}$ et $\{Y_j = y_j | Z = z\}$ sont tous deux des hypothèses indépendantes. Il s'en suit, en utilisant la règle de factorisation, que :

$$\begin{cases} \mathbb{P}(Z | X_{1:I}) = \prod_{i=1}^{i=I} \mathbb{P}(Z | X_i) & (13) \\ \mathbb{P}(Y_{1:J} | Z) = \prod_{j=1}^{j=J} \mathbb{P}(Y_j | Z) & (14) \end{cases}$$

Avec

$$\begin{cases} X_{1:I} = x_1, x_2, \dots, x_I \\ Y_{1:J} = y_1, y_2, \dots, y_J \end{cases}$$

Cette formule nous montre que, malgré les hypothèses d'indépendances, le nombre de paramètres (probabilités conditionnelles) nécessaires pour estimer correctement les paramètres de notre modèle est très élevé. Il n'est donc pas souhaitable de définir toutes ces probabilités par l'usage d'un expert du domaine de la médecine traditionnelle, même si celui-ci possède une très grande maîtrise des signes cliniques et des plantes médicinales utilisées dans le traitement d'une maladie. Il est donc nécessaire que les paramètres du Réseau Bayésien soient estimés à partir des données collectées auprès d'un certain nombre de spécialistes de la médecine traditionnelle.

6.6. Algorithme de la méthode

La méthode de modélisation proposée est générale et adaptée à toute maladie. Elle peut être synthétisée dans un algorithme. Pour ce faire, certaines opérations, à défaut de pouvoir les décrire dans les moindres détails, seront définies par des opérateurs génériques. Ainsi, considérons les opérateurs suivants :

- *Onto(.)* : C'est l'opérateur de construction d'une ontologie. Il correspond à toutes les étapes nécessaires à la construction d'une ontologie. Ici, on admettra que l'ontologie est construite à partir d'un graphe au sens mathématique du terme. Ce graphe sera défini en fournissant ses nœuds et ses arcs.
- *Graph(.)* : C'est l'opérateur de construction de la structure d'un Réseau Bayésien à partir de connaissances expertes.
- *Data(.)* : C'est l'opérateur désignant le processus de collecte de données nécessaires à l'apprentissage des paramètres du Réseau Bayésien. Les données sont collectées sur un ensemble de tradipraticiens spécialistes de la maladie et comporte des informations aussi bien sur les signes cliniques que sur les plantes médicinales à utiliser pour le test d'aptitude au traitement de la maladie.

• *BN(.)* : C'est l'opérateur d'apprentissage des paramètres des Réseaux Bayésiens à partir de la structure définie par expertise et des données collectées. L'algorithme suivant décrit les différentes étapes de la méthode de modélisation proposée pour chaque maladie M :

1) // Ensemble des traitements de la maladie M

$$\Omega_Y = \{Y_j \mid j \in [1, J]\}$$

2) for (Y_j in Ω_Y) {

i) // Ensembles des traitements Y_j

$$\mathcal{G}_j = \{Y_l \in \Omega_Y \mid Y_j \Rightarrow Y_l\};$$

$$\Theta_j = \emptyset \quad // \text{Ensemble des relations significatives de } \mathcal{G}_j$$

For (Y_j in \mathcal{G}_j) {

$$\Phi(Y_j) \in \{0, 1\} // \text{Evaluation de la force de la relation}$$

$$\text{If } (\Phi(Y_j) = 1) \Theta_i \leftarrow \Theta_i \cup \{Y_j\}$$

}

}

3) // Ensemble des signes cliniques de la maladie M

$$\Omega_X = \{X_i \mid i \in [1, I]\}$$

4) for (X_i in Ω_X) {

i) // Ensemble des relations significatives avec X_i

$$\Delta_i = \phi$$

ii) // Ensembles des signes cliniques qu'entraîne X_i

- $$S_i = \{X_k \in \Omega_X | X_i \Rightarrow X_k\};$$
- For (X_k in S_i) {
- $$\gamma_{ik} = \Gamma(X_i \Rightarrow X_k) \in \{0, 1\} // \text{Evaluation de la force de la relation}$$
- relation
- $$\text{if } (\gamma_{ik} == 1) \Delta_i \leftarrow \Delta_i \cup \{X_k\}$$
- }
- }
- 5) // Ensemble des nœuds et arcs du graphe des liens significatifs
- $$\mathcal{O} = \left\{ \bigcup_{i=1}^{i=I} (X_i, \Delta_i) \right\} \cup \left\{ \bigcup_{j=1}^{j=J} (Y_j, \Delta_j) \right\}$$
- 6) $\mathfrak{S} = \text{Onto}(\mathcal{O}) // \text{Ontologie associée au graphe.}$
- 7) $\mathfrak{S}_{BN} = \text{Graph}(\mathfrak{S}) // \text{Structure du Réseau Bayésien correspondant à l'ontologie.}$
- 8) $\mathfrak{D}_{BN} = \text{Data}(X, Y) // \text{Données nécessaires à l'estimation des probabilités.}$
- 9) $\mathfrak{RB}_M = \text{BN}(\mathfrak{S}_{BN}, \mathfrak{D}_{BN}) // \text{Réseaux Bayésiens à paramètres estimés.}$

6.7. Exploitation du modèle

Pour des raisons d'illustration, des données ont été générées de façon aléatoire. Ces données permettent d'illustrer l'usage des Réseaux Bayésiens à des fins de raisonnement médical en termes de diagnostic d'une maladie. En pratique, ces données sont recueillies chez un praticien expert du domaine. Un exemple de résultat de modélisation par Réseaux Bayésiens est donné par la figure 6.4 suivante :

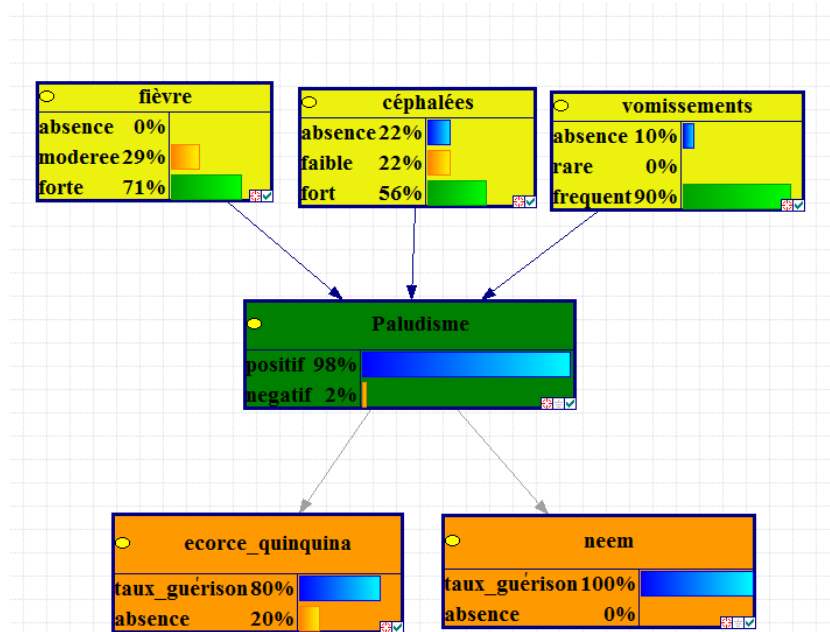


Figure 6.4 : Diagnostic et traitement du paludisme

Cette figure présente un exemple d'utilisation du modèle développé. Il traduit le diagnostic d'un patient qui dit avoir très souvent de fortes fièvres alternées de fièvres modérées. Le patient affirme avoir très souvent de violentes céphalées qui peuvent s'atténuer et parfois même, dans de rares cas, disparaître. Il affirme aussi avoir très fréquemment des vomissements.

Le praticien de la médecine traditionnelle qui effectue le diagnostic est certain que le *neem* est utilisé dans le traitement du paludisme. Mais en ce qui concerne le traitement à base d'*ecorce de quinquina*, il a un petit doute sur l'usage de cette plante pour obtenir la guérison de la dite maladie. Puis, une fois les signes cliniques intégrés au modèle, le tradipraticien arrive à la conclusion selon laquelle le patient souffre de la maladie c'est à dire du paludisme (dans notre cas) avec une certitude à 98%. Sur cette base, le traitement adéquat serait l'usage soit du neem, soit de l'ecorce de quinquina soit des deux et soit d'autres substances médicamenteuses.

Le modèle a été développé en utilisant le logiciel GeNIes³ car il est très complet et surtout gratuit aussi bien pour les travaux de recherches que pour les travaux à caractère commercial. En effet, GeNIes est un logiciel open source dédié aux Réseaux Bayésiens et à ses extensions (Réseaux Bayésiens Dynamiques et Diagrammes d'influence). Il comporte de nombreux algorithmes d'apprentissage des paramètres comme de la structure à partir des données. Il dispose d'une interface très conviviale et peut facilement être utilisé par des non spécialistes en modélisation, comme les praticiens de la médecine traditionnelle.

Du fait de l'importance des Réseaux Bayésiens, de nombreux logiciels ont été développés pour leur implémentation et leur utilisation dont un nombre important est « open source ». L'essentiel de l'usage de la modélisation consiste, une fois les paramètres estimés, à calculer automatiquement le diagnostic à partir d'un nouveau patient sur lequel l'on dispose des modalités des signes cliniques. Les paramètres liés aux plantes médicinales dépendent du praticien qui effectue le diagnostic. Et comme nous l'avons déjà souligné, il s'agit d'un mécanisme de protection du patient contre la méconnaissance des plantes médicinales nécessaires au traitement de la maladie.

6.8. Comparaison au niveau des Réseaux bayésiens

La méthode des réseaux bayésiens a été utilisée très récemment en médecine moderne pour le traitement de la probabilité de survenue d'une certaine maladie. Les travaux de Michel Ducher et *al.* [Duch 15] indiquent ce qui suit : « en utilisant le réseau bayésien ou la régression logistique, la sensibilité de la prédiction d'un AldosteRone sur Renine (ARR) élevé atteint

³ Site Web de GeNIes : <https://dslpitt.org/genie/>

respectivement (100 % vs 63,3 %) et la spécificité (89,6 % vs 67,2 %). L'aire sous la courbe ROC (outil d'évaluation et de comparaison des modèles) obtenue avec le réseau bayésien est significativement plus élevée que celle obtenue par la régression logistique (c'est à dire par comparaison des sujets malades à des sujet sains) $0,93 \pm 0,015$ vs $0,70 \pm 0,03$, $p < 0,001...$ ».

C'est cette même méthode que nous avons utilisée dans le cadre de notre travail pour estimer la probabilité de survenue d'une certaine maladie à partir des pourcentages d'apparition des signes cliniques et la probabilité de l'usage d'un traitement donné en fonction des pourcentages de résultats satisfaisants obtenus.

6.9. Conclusion

La technologie des Réseaux Bayésiens est une approche adaptée au raisonnement médical pour le diagnostic des maladies. Sa souplesse et sa capacité à intégrer aussi bien les connaissances expertes que les connaissances issues des données en font une technique de modélisation efficace. Dans le cadre de la médecine traditionnelle, les Réseaux Bayésiens sont d'autant plus adaptés qu'ils peuvent être utilisables lorsque l'on se trouve confronté à des données incomplètes, imprécises et incertaines. La modélisation du processus de diagnostic des différentes maladies dans la médecine traditionnelle doit être une priorité ; de nombreux modèles doivent donc être construits. En effet, il s'agit d'un gage de bonnes pratiques et d'uniformisation des pratiques en matière de diagnostic, opérations jusqu'ici laissées à la subjectivité des praticiens sans aucune garantie de fiabilité.

Cependant, les modèles pratiqués doivent faire l'objet d'un minimum de consensus sur les signes cliniques et les plantes médicinales entrant dans le traitement de chaque maladie ; les interactions entre les différents signes cliniques d'une part, les différentes plantes médicinales d'autre part, de même que les interactions entre les plantes médicinales et les signes cliniques doivent également se soumettre à un minimum d'accord. Mais la prise en compte de ces différentes interactions se limitera aux éléments les plus significatifs qui s'y trouvent pour ensuite faire l'objet de modélisation sous la forme d'une ontologie visuelle. L'estimation des paramètres doit être effectuée en privilégiant les données les plus significatives.

De nos jours, la médecine traditionnelle a besoin d'intégrer les outils informatiques dans sa pratique ; il est plus que nécessaire que les modèles que nous proposons aient la possibilité de réagir, si le niveau de connaissances du praticien dans des cas donnés, est limité. Cette mesure de précaution est très importante pour protéger les patients. Certes, cette méthode participe d'une nouvelle culture ; cependant, elle reste un moyen plus que rapide

d'uniformisation des pratiques, de transmission des connaissances et de professionnalisation scientifique du secteur.

Dans le processus d'apprentissage de la médecine traditionnelle et dans sa pratique, lorsque la symptomatologie conduit à la définition de plusieurs maladies c'est à dire au diagnostic de celles-ci on se trouve dans une situation de « conflit » ; dans ces conditions, il est question de gestion de conflit, objet du prochain chapitre.

Chapitre 7

GESTION DE CONFLITS DANS L'EXERCICE DE LA MEDECINE TRADITIONNELLE A BASE D'ONTOLOGIE MEDICALE

***Résumé :** Dans cette partie, nous nous intéressons à la gestion des conflits lors de la pratique de la médecine traditionnelle. L'objectif est de poser un diagnostic solide afin de mettre en place un traitement efficace et minimiser le risque de déséquilibre physique et social des patients. Nous analysons en particulier les différentes techniques en se basant sur la sémantique d'ontologie afin d'éviter l'indexation similaire ou conflit d'un concept (les posologies, les vertus thérapeutiques). Nous analysons également l'influence de la prise en compte des différentes parties d'une requête.*

7.1. Introduction

Le principe du fonctionnement de l'outil informatique est de recevoir en entrée une masse d'informations, de l'analyser à l'aide de programmes informatiques spécifiques, puis de fournir un résultat qui tient compte de la requête formulée.

Dans notre contexte, il s'agit d'informations provenant de la MT qui sont formalisées sous forme d'ontologies ; celles-ci sont fournies à l'ordinateur dans le but d'en faire l'analyse pour aboutir à un résultat recherché. Il peut s'agir, soit du diagnostic, soit du traitement d'une maladie donnée. Cette approche constitue l'élément essentiel dans l'exercice au quotidien de la médecine traditionnelle. Les données fournies à l'ordinateur sont multiformes, variées et toujours très complexes dans leur contenu. Aussi, peut-il arriver que la machine fournisse des résultats qui soient parfois contradictoires ou incohérents créant ainsi ce que nous avons appelé « conflit ». On entend formellement par conflit, le fait que l'outil informatique, à partir d'un ensemble fini d'informations relatives à une maladie, identifie non pas cette seule maladie, mais fait apparaître plusieurs autres. Devant une telle situation, le praticien se trouve en face d'un dilemme ; il y a donc conflit.

Pour résoudre un tel cas que nous avons appelé « gestion de conflit », nous proposons l'utilisation d'un modèle mathématique basé sur un algorithme de détection de conflit et un autre algorithme de résolution de conflit.

Nous commencerons par la présentation d'un exemple de cas de conflit suivi d'une modélisation mathématique pour la résolution de ce conflit puis nous terminerons par une conclusion.

7.2. Ontologie médicale : exemple de cas de conflit

Nous utilisons l'ontologie que nous avons développée sur « protégé 2000 » comme le montre la figure 7.1. Cette ontologie s'est faite à partir de connaissances issues de l'ensemble des symptômes propres à deux différentes maladies (P et Z).

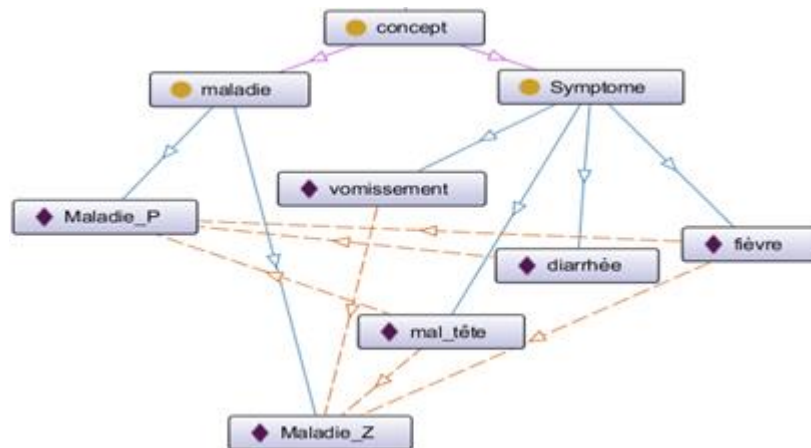


Figure 7.1 : Ontologie de la maladie_Z et de la maladie_P

Le graphe utilisé dans l'exemple de la figure 7.1 contient des maladies et des symptômes sémantiquement liés aux formats « XML » et « RDF ». En effet, nous avons considéré, à titre d'exemple 4 symptômes et 2 cas de maladies (P et Z). Dans ce graphe, une flèche partant d'un symptôme tel que la fièvre vers une maladie_P, suggère que celui-ci est l'élément déclencheur de cette maladie.

Un fichier « RDF » permet de décrire les ressources web de manière à faciliter le traitement automatique et l'interopérabilité des informations.

A travers cette ontologie, quelques données ont été relevées (voir tableau 7.1) ; celles-ci sont fournies à titre indicatif ; elles ne reflètent pas les maladies. Ce sont des cas de maladies mises en évidence à titre d'illustration. Ici, un apprenant ou un pratiquant par exemple, fait une requête composée de quelques signes cliniques présentés avec leurs pourcentages d'apparition (voir tableau 7.1). Le système retrouve deux(2) maladies à savoir les maladies Z et P. Dans le cas d'espèce, le système rencontre une situation de conflit ; du fait d'un déficit de symptômes symbolisés par les points d'interrogation (?) il y a lieu de poursuivre les investigations à la recherche de symptômes complémentaires pour définir exactement la maladie en question.

Tableau 7.1 : Présentation de cas de conflit

Maladie_Z	%	Maladie_P	%
Mal de tête	5%	Mal de tête	5%
Fièvre	5%	Fièvre	5%
Vomissement	80%	Diarrhée	70%
?	%	?	0%

La section suivante présente la modélisation de cette nouvelle approche.

7.3. Modélisation

7.3.1. Filtrage collaboratif

Comme nous l'avons mentionné au niveau de la première partie, le diagnostic dépend de la connaissance du praticien. En effet, si nous arrivons à formaliser cette connaissance, le raisonnement médical sera aisé.

Dans les points qui suivent, nous avons utilisé le filtrage collaboratif en vue de filtrer les informations basées sur les préférences des apprenants. Les systèmes basés sur le filtrage collaboratif (FC) aident un utilisateur à trouver l'information qui l'intéresse à partir des jugements d'autres utilisateurs. En général, les systèmes de FC demandent aux utilisateurs de fournir des jugements sur des articles de leur choix, exprimés sous la forme d'une note, un réel positif borné. L'objectif visé, en utilisant ce système, est de comparer les symptômes de chaque maladie afin de prédire les symptômes manquants non fournis au système.

7.3.1.1. Hypothèses du problème

Soit $S = \{S_j, 1 \leq j \leq n\}$ tel que $S_j = (s_{j1}, s_{j2}),_{j1, j2 \in N}$, l'ensemble des symptômes connus de notre base d'apprentissage. En effet, une base d'apprentissage permet la construction de processus décisionnel avec des temps de réponse courts pour la réalisation d'applications ayant des contraintes de temps réel élevées. Chaque symptôme S_j est caractérisé par s_{j1} une chaîne de caractère qui le décrit et un paramètre s_{j2} qui décrit l'importance du symptôme par rapport à une maladie donnée. La description d'un symptôme se résume au nom de ce symptôme. Les valeurs s_{j1} et s_{j2} sont définies par l'utilisateur du système selon une échelle de classification définie au moment de la mise en place du système. Nous supposons que l'ensemble S est fini. Soit $|S| = \text{card}(S) = n$. Soit $M = (M_r)_{r \in I}$ avec I un ensemble fini de cardinal t . Chaque maladie M_r est définie par un ensemble de n symptômes et caractérisée par la probabilité d'occurrence du symptôme S_j . La quantification de la maladie doit être proposée par des spécialistes. Nous sommes partis de l'hypothèse qu'une telle base existe, est déjà utilisée et est issue de l'expérience des consultations médicales. Nous définissons notre base

d'apprentissage à partir de m échantillons représentant les paramètres caractéristiques (les probabilités d'apparition des symptômes) des différentes maladies. Donc nous avons à considérer les matrices suivantes :

$$\text{La matrice d'apparition des symptômes } X = \begin{pmatrix} (x_1)^T \\ (x_2)^T \\ \vdots \\ (x_t)^T \end{pmatrix}$$

Dans laquelle les $x_i = (x_{ij}/1 \leq i \leq t, 1 \leq j \leq n)$ sont des vecteurs $\in \mathcal{M}_{n,1}(R_+)$. Chaque x_{ij} est la probabilité d'observation du symptôme s_j de la maladie M_i .

Les probabilités d'apparition d'un symptôme pour une maladie sont calculées à partir de l'estimation suivante.

$$P(s_j|m_i) = \frac{\text{nombre de patients ayant le symptôme } j \text{ et la maladie } i}{\text{nombre de patients ayant la maladie } i}$$

Le nombre de patients est le cardinal de notre base d'apprentissage. Il est donc naturel que nous devrions disposer d'une base assez consistante afin d'avoir des estimations de qualité.

Nous définissons la matrice $Y = (y_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$ dans laquelle chaque y_{ij} représente le degré d'importance (donné selon une échelle de classification variant de 0 à 5 pour absent, très faible, faible, moyen, élevé, très élevé) du symptôme s_j pour la maladie M_i . $y_{ij} = 0$ si un symptôme s_j ne fait pas partie des symptômes de la maladie M_i .

$$\text{Nous définissons également la matrice } Z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{pmatrix},$$

dans laquelle chaque z_i représente un nombre entier affecté à une maladie de notre base d'apprentissage.

Le système reçoit en entrée un sous-ensemble Λ de l'ensemble des parties $P(S)$ composé d'un certain nombre fini de symptômes. Soit $q = \text{card}(\Lambda) \leq \text{card}(S)$. Pour nos travaux nous allons poser $q \geq 3$ pour mieux discriminer le résultat.

Il y a conflit lorsque le résultat de la sortie de notre algorithme appliqué à l'ensemble Λ identifie plusieurs maladies. Afin de permettre un diagnostic efficace, nous ne pouvons pas permettre de tels conflits. Nous proposons, dans la suite, notre algorithme en deux étapes afin de lever les éventuels conflits :

- Détection de la présence de conflit ;
- Résolution du conflit.

7.3.1.2. Détection du conflit

A partir de notre base d'apprentissage, définissons la matrice binaire $R = (R_{ij})_{1 \leq i \leq t, 1 \leq j \leq n}$ telle que $R_{ij} = 1$ si le symptôme s_j est un symptôme de la maladie M_i . 0 sinon. La matrice R représente la matrice d'adjacence entre l'ensemble des symptômes de notre base d'apprentissage et l'ensemble des maladies connues de notre base d'apprentissage. Le nombre d'échantillons est différent du nombre de maladies car deux patients différents peuvent ne pas présenter le même nombre de symptômes pour la même maladie.

Pour détecter le conflit, nous cherchons à déterminer le cardinal de l'ensemble

$$\xi = \{i \in \{1, 2, 3, \dots, t\} \mid \sum (R_{ij}, 1 \leq j \leq n) = q\}.$$

Ces cas peuvent se produire :

- $card(\xi) = 0$. Dans ce cas, l'ensemble des symptômes donnés ne correspond à aucune maladie. Cela correspond à un cas d'aberration. Soit la base ne contient pas cette maladie et elle est à compléter, soit il y a une incohérence dans les symptômes et il est impossible que de tels symptômes décrivent une maladie. Ce cas est hors de notre domaine d'étude.
- $card(\xi) = 1$. Dans ce cas les symptômes entrés décrivent une unique maladie. Il n'y a pas de conflit. L'indice de la maladie nous donne la maladie correspondante.
- $card(\xi) > 1$. Nous avons détecté un conflit. Nous lançons alors la procédure de gestion des conflits.

7.3.1.3. Gestion des conflits

Cette étape contient deux sous-phases. Une phase de « traitement des données manquantes ». Le nom de cette phase est emprunté aux statistiques qui, dans cette phase, estiment à partir des données recueillies et des données de la base d'apprentissage, un ensemble de valeurs les plus vraisemblables possibles afin de classifier la donnée. De plus, dans la phase de traitement des données manquantes, nous faisons l'hypothèse que les symptômes sélectionnés ainsi que les symptômes manquants sont mutuellement indépendants.

Cette phase est suivie de la phase de classification de nos symptômes à partir de l'algorithme des K-moyennes⁴.

Soit $Z = i \in \{1, 2, 3, \dots, t\}$ tel que $z_i \in \xi$

Modifions les matrices Y en ajoutant en dernière ligne les valeurs des symptômes s_j de S .

7.3.1.4. Traitement des données manquantes

Le but est de passer d'un espace de dimension q à un espace de plus grande dimension afin de pouvoir renseigner un classifieur linéaire généralement basé sur les méthodes de régression combinées à un algorithme de gradient descendant. Un algorithme à gradient impose la non-négativité des résultats des observations dans l'expression. La méthode de régression combinée permet la superposition d'un graphique et expose l'influence de chaque variable. En effet, elle permet de résoudre les problèmes de multi-colinéarité.

Nous pouvons alors indiquer la valeur de ces données avec le minimum d'erreur et évaluer la précision d'une série de symptômes par la méthode des moindres carrés. Le modèle théorique consiste en une famille de fonctions $f(x_i, \theta)$ d'une ou plusieurs variables muettes x_i , indexées par une famille de paramètres θ inconnus. La méthode des moindres carrés permet de sélectionner parmi ces fonctions celle qui reproduit le mieux les données expérimentales.

La méthode consiste en la construction de la fonction $f(x_i, \theta)$ qui minimise la somme quadratique des déviations des mesures aux prédictions de $f(x_i, \theta)$. Si, par exemple, nous disposons de N variables (y_i) avec $i = 1, 2, 3, \dots, N$, les paramètres θ « optimaux » au sens de la méthode des moindres carrés sont ceux qui minimisent la quantité :

$$S(\theta) = \sum_{i=1}^N (y_i - f(x_i; \theta))^2 = \sum_{i=1}^N r_i^2(\theta)$$

Pour cela déterminons les ensembles suivants :

Soit $Z = i \in \{1, 2, 3, \dots, t\}$ tel que $z_i \in \xi$

Υ l'ensemble des symptômes des maladies de la base d'apprentissage, dont l'indice i est dans Z et pour lesquels $x_{ij} \neq 0$ et qui n'appartiennent pas à S . Posons $\Omega = \Lambda \cup \Upsilon$ dans lequel les symptômes sont rangés dans l'ordre croissant de leurs indices j .

⁴ Richard O. Duda, Peter E. Hart, David G. Stork, *Pattern classification*, Wiley-interscience, 2001 ([ISBN 0-471-22361-1](https://doi.org/10.1002/9781118130401))

Soit $\Upsilon = \emptyset$. Nous calculons le vecteur d'occurrence x_{m+1} correspondant aux symptômes d'entrée, puis nous passons à l'étape de classification.

Soit $\Upsilon \neq \emptyset$. Alors nous allons chercher à compléter les données manquantes puis indiquer la classe correspondante à notre vecteur de symptômes d'entrée.

$$\text{Définissons la matrice } M = (m_{ij}) = \begin{cases} 1 & \text{si } i \in Z, 1 \leq j \leq n \text{ et } x_{ij} \neq 0 \\ 1 & \text{si } i = m + 1 \text{ et } s_i \ s_{j2} \neq 0 \\ 0 & \text{sinon} \end{cases}$$

Le tableau 7.2 suivant illustre la situation dans laquelle nous sommes. Les variables s_3, s_6 et s_{11} représentent les paramètres d'entrée de notre système. Ces variables sont les seules communes aux maladies z_1, z_2, z_{10}, z_{25} et z_{50} . Nous sommes donc en présence d'un conflit car $\{s_3, s_6, s_{11}\}$ identifie plusieurs maladies dans notre base d'apprentissage (cf. paragraphe sur la détection des conflits). Par exemple dans l'ensemble $\Upsilon = \{s_1, s_5, s_8, s_{13}\}$, la maladie z_2 possède le plus grand nombre de symptômes donc $\Phi = 13$. La dernière ligne de notre tableau représente les valeurs des symptômes entrées par l'utilisateur et nous nous proposons de prédire à partir des valeurs connues de la base d'apprentissage les valeurs manquantes illustrées par les points d'interrogation.

Tableau 7.2 : Exemple de matrice Y représentant le degré d'importance des symptômes dans les maladies de notre base d'apprentissage

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}
z_1	2	2	1	0	1	0	2	0	3	4	0	2	1.5	5	0	0
z_2	1.5	0	2	0	2.5	5	0	2	0	0	1	0	2.5	0	0	0
⋮																
z_{10}	0	0	3	0	3	4	0	1	0	0	5	0	3.5	0	0	0
⋮																
z_{25}	2.5	0	3.5	0	3.5	4	0	1.5	0	0	2.5	0	0	0	0	0
⋮																
z_{50}	2	0	2	0	5	5	0	2	0	0	3	0	3	0	0	0
⋮																
?	?	0	3	0	?	4	0	?	0	0	2.5	0	?	0	0	0

Nous allons utiliser le principe de la régression linéaire avec régularisation pour approximer le mieux possible les valeurs manquantes. Cela va nous permettre de minimiser l'erreur quadratique sur l'approximation faite sur les données observées. Les données manquantes seront exprimées par $Y_{m+1j} = (\theta_j)^T X_{m+1}$ pour j tel que $1 < j < \Phi$. Pour cela,

nous allons déterminer la famille optimale de paramètres θ et le vecteur X_{m+1} qui minimisent la solution de l'équation :

$$J(x_{m+1}, \theta_1, \theta_2, \dots, \theta_\phi) = \frac{1}{2} \sum_{(i,j):m_{ij}=1} ((\theta_j)^T x_i - y_{ij})^2 + \frac{\lambda}{2} \sum_{k=1}^n (x_{m+1k})^2 + \frac{\lambda}{2} \sum_{j=1}^\phi \sum_{k=1}^n (\theta_{jk})^2$$

Pour tout i dans M_1 l'algorithme est le suivant :

- 1) Initialisation de $x_{m+1}, \theta_1, \theta_2, \dots, \theta_m$ par de petites valeurs aléatoires.
- 2) Minimisation de la fonction $J(x_{m+1}, \theta_1, \theta_2, \dots, \theta_\phi)$.

Pour $1 < j \leq \phi$ et $1 < i \leq m_z$,

$$x_{m+1k} = x_{m+1k} - \alpha (\sum_{j:m_{ij}=1} ((\theta_j)^T x_i - y_{ij}) \theta_{kj} + \lambda x_{m+1k})$$

$$\theta_{kj} = \theta_{kj} - \alpha (\sum_{i:m_{ij}=1} ((\theta_j)^T x_i - y_{ij}) x_{ik} + \lambda \theta_{kj})$$

- 3) Pour les $j = 1, 2, 3, \dots, n$ tels que y_{ij} n'est pas défini, nous posons $y_{ij} = (\theta_j)^T x_i$ qui est le produit scalaire des vecteurs θ_j et x_i

Remarque : Les paramètres sont choisis très petits typiquement des valeurs de 0.01 ou 0.03. Quant au paramètre lambda, il est choisi typiquement entre 1 et 5. Pour nos tests $\alpha = 1$ et $\lambda = 1.5$

7.3.1.5. Classification de la nouvelle donnée

Partant de notre base d'apprentissage, nous devons déterminer à quelle classe appartient notre nouveau vecteur, c'est-à-dire déterminer z_{m+1} . Pour ce faire, à l'aide des algorithmes des k-moyennes, nous allons déterminer les centroïdes des classes puis comparer la distance séparant les centroïdes C_k de notre vecteur Y_{m+1} . Les centroïdes sont des critères d'approximation pour optimiser les fonctions sous-jacente dans les K-moyens. Cela revient à déterminer l'indice k de Z minimisant la distance $d(Y_{m+1}, C_k)$ avec la distance d prise comme distance euclidienne. Plus précisément nous cherchons à résoudre $argmin_{k \in M_1} d(x_{m+1}, C_{g(k)})$ ou $g(k)$ est la fonction qui, à k élément de Z , lui affecte un indice compris entre 1 et K selon l'ordre croissant. Ainsi $g(i_s) = K$.

L'algorithme est le suivant :

1. Réorganiser par ordre croissant les éléments de M_1 et leur assigner des index compris entre 1 et $\text{card}(Z) = K$.
2. Déterminer par l'algorithme des k-moyennes les C_k centroïdes.
3. Résoudre $\text{argmin}_{1 \leq k \leq K} d(y_{m+1}, C_k)$.

7.4. Illustration du modèle.

Pour mettre en exergue le modèle proposé, nous avons opté pour l'utilisation de MATLAB. En effet, Matlab est un logiciel de calcul numérique convivial et de visualisation graphique destiné aux ingénieurs et aux scientifiques. Il possède un langage de programmation puissant et simple à utiliser avec lequel l'utilisateur peut effectuer des calculs en ligne ou par l'intermédiaire d'un fichier de commandes. Les résultats de modélisation à travers la méthode collaborative sont donnés en Annexe A. Nous avons considéré 20 symptômes en rapport avec 10 maladies et 1682 échantillons d'individus.

7.5. Comparaison au niveau de la gestion de conflit

Les modèles de gestion de conflit que nous proposons pour résoudre les cas de multiplicité de résultats suite à une convergence d'informations introduites dans le système est une démarche personnelle. En effet, la gestion des conflits, rencontrée dans les autres sciences est à l'évidence de toute autre nature et ne correspond pas à la vision que nous avons. Il est difficile, dans ces conditions, d'établir des tableaux de comparaison.

7.6. Conclusion

Dans ce chapitre, nous avons pu nous rendre compte que des difficultés pouvaient surgir avec l'utilisation de l'outil informatique ; avec la survenue du conflit, l'objectif recherché qui est de poser un diagnostic fiable suivi d'un traitement approprié peut être compromis. Mais, avec la mise en place d'un modèle mathématique fondé sur l'utilisation des algorithmes nous pouvons parvenir à gérer efficacement ces conflits et ainsi parvenir à un résultat qui soit satisfaisant et qui réponde à nos attentes.

Suite à ces opérations, il est toujours possible que l'ordinateur nous fournisse une nouvelle information dont les caractéristiques sont susceptibles de nous aider à affiner les éléments de notre requête. Il s'agit, à partir d'un cas déjà connu et existant dans notre base de connaissances, d'apprécier le comportement de cette nouvelle information par une approche de similarité basée sur le raisonnement. C'est l'objet de notre prochain chapitre.

Chapitre 8

RAISONNEMENT MEDICAL : APPLICATION DU RAISONNEMENT A PARTIR DE CAS EN MEDECINE TRADITIONNELLE

Résumé. *Le raisonnement à partir de cas (RàPC) est un paradigme de résolution de problèmes fondé sur la réutilisation d'expériences passées stockées dans une base de cas existant. Ce type d'outils de l'Intelligence Artificielle provient des travaux effectués par Schank (Schank, 1982) sur la mémoire dynamique. Le RàPC a été utilisé dans de nombreux systèmes pour résoudre des problèmes dans des domaines divers dont celui du diagnostic médical. Dans les systèmes de RàPC, la notion de similarité est très importante car elle sert de fondement à tout le système. Cependant, l'efficacité des mesures de similarité utilisées dépend du problème abordé. Dans ce chapitre, nous proposons une mesure de similarités pour le développement de systèmes de RàPC en vue du diagnostic automatique de maladies dans le domaine de la médecine traditionnelle.*

8.1. Introduction

Le système de gestion des fichiers contenus dans la base de connaissances médicales est depuis un certain temps, l'objet d'un intérêt croissant pour le monde des TICs. L'enseignement et la pratique de la médecine traditionnelle ne pouvaient rester en dehors de cette évolution technologique.

Dans le cas particulier de la pose du diagnostic des maladies, les pratiques actuelles en médecine traditionnelle, ne sont pas uniformes et sont, pour la plupart du temps, laissées à la subjectivité du praticien. Aussi, a-t-il donc recours à ses expériences antérieures tant pour effectuer ses diagnostics que pour prescrire ses traitements. Il est évident que de tels diagnostics ainsi que les traitements proposés sont intrinsèquement liés aux capacités cognitives de l'individu, notamment en ses possibilités de remémoration. Dans ce contexte, une aide à l'élargissement des capacités de remémoration du praticien peut être un gage d'efficacité plus accrue. Pour ce faire, la mise en place de systèmes de raisonnement à partir de cas (RàPC) peut être d'une aide précieuse aussi bien dans l'apprentissage que dans la pratique de cette science médicale en plein essor en vue d'améliorer la qualité des prestations, de diminuer les risques d'erreurs médicales en offrant des soins mieux appropriés.

La technique du RàPC n'ayant presque jamais été expérimentée dans la médecine traditionnelle, son usage exige la prise en compte des éléments spécifiques à ce domaine puisque les systèmes de RàPC sont adaptés aux différentes réalités à travers les mesures de similarités. On définit la mesure de similarité comme étant un concept permettant d'identifier un sous-groupe dans

un groupe d'objet ; du reste, c'est par la différence de distances entre deux objets qu'on mesure leurs degrés de similarité. Notre travail consiste à développer de telles mesures pour la médecine traditionnelle même si celles-ci sont généralement très complexes. On y trouve, en effet, des variables numériques, des variables qualitatives et des variables textuelles. Le contenu de la médecine traditionnelle étant en réalité moins structuré, ces trois types de variables doivent coexister dans les systèmes de RàPC qui y seront développés. Cependant, la forme de collaboration entre ces différentes variables doit être bien spécifiée pour une appropriation plus rapide de la technologie.

8.2. Etat des lieux

8.2.1. Approche du raisonnement à partir de cas

Comme nous l'avons déjà dit, la méthode du RàPC n'a pratiquement jamais été appliquée en médecine traditionnelle si ce n'est dans certaines disciplines telles que l'aviation, la mécanique etc ; c'est une méthode introduite par Minsky et Schank [Iva 08] dès la fin des années 70 ; le RàPC a fait l'objet de plusieurs travaux de recherche dont celui de Bitar [Ibra 14] qui propose une approche dénommée « CBR420SD », basée elle aussi sur le raisonnement à partir de cas. Son approche permet donc d'aller dans le même sens que la découverte des services du web sémantique. Ceci permet à de nouveaux clients de pouvoir déterminer leurs besoins fonctionnels et non fonctionnels. Quant à Paula et dhouib [Paul 12, Dial 11], ils proposent un mécanisme décisionnel basé également sur le raisonnement à partir de cas (RàPC). Ce mécanisme permet, quant à lui, de résoudre les problèmes complexes en matière de maintenance et plus particulièrement lorsque les experts travaillant en collaboration se retrouvent confrontés à un problème difficile. Mohamed [Moh 10] de son côté développe une méthode de diagnostic qualitative s'apparentant à une classification orientée vers des expériences. Au regard de tous ces travaux qui attestent de l'évolution de la méthode du RàPC, on estime qu'il serait temps que la médecine traditionnelle s'engage dans les nouvelles technologies particulièrement par le biais de la technique du raisonnement à base de cas.

En définitive, le Raisonnement à Partir De Cas (RàPC) est un modèle de raisonnement exploitant les expériences de résolution de problèmes antérieurs pour en résoudre de nouveaux (figure 8.1) [Coj 11]. Ces expériences relèvent donc de connaissances de situations particulières dans lesquelles il n'y a plus d'efforts à faire, a priori, pour les généraliser à d'autres situations.

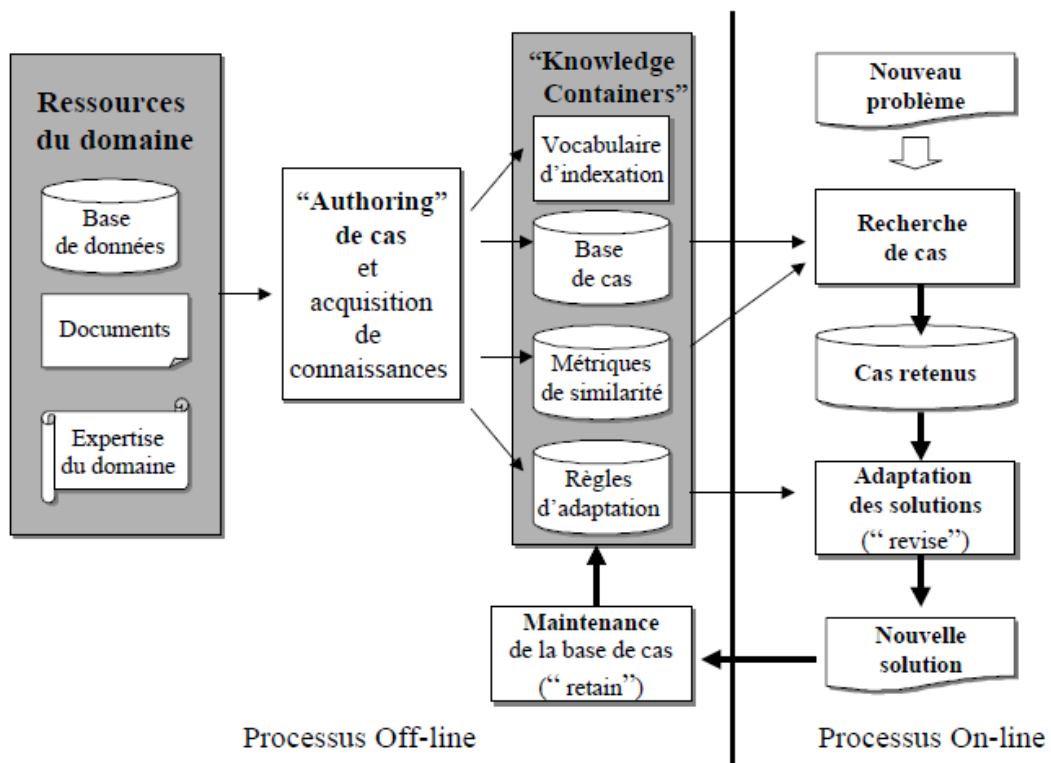


Figure 8 1 : Architecture d'un système de RàPC

Dans l'architecture d'un système de RàPC telle que décrit dans la figure ci-dessus, la partie « on-line » comporte les phases du cycle de RàPC ; la partie « off-line » implique les ressources utilisées, la phase d'acquisition et de représentation des connaissances ainsi que des containers de connaissances. La partie construction (authoring) du cas et celle d'acquisition de connaissances guident la structuration initiale de la base de cas et des autres connaissances du système à partir de différentes ressources telles que les documents, les bases de données ou les experts du domaine. Nous nous intéressons à ce modèle dans lequel nous exploitons les containers de connaissances ainsi que les phases du cycle de RàPC. Sur la base de l'architecture du RàPC portant sur l'enseignement, nous avons essayé d'adapter notre situation à celle-ci en présentant des cas de maladie (figure 8.2).

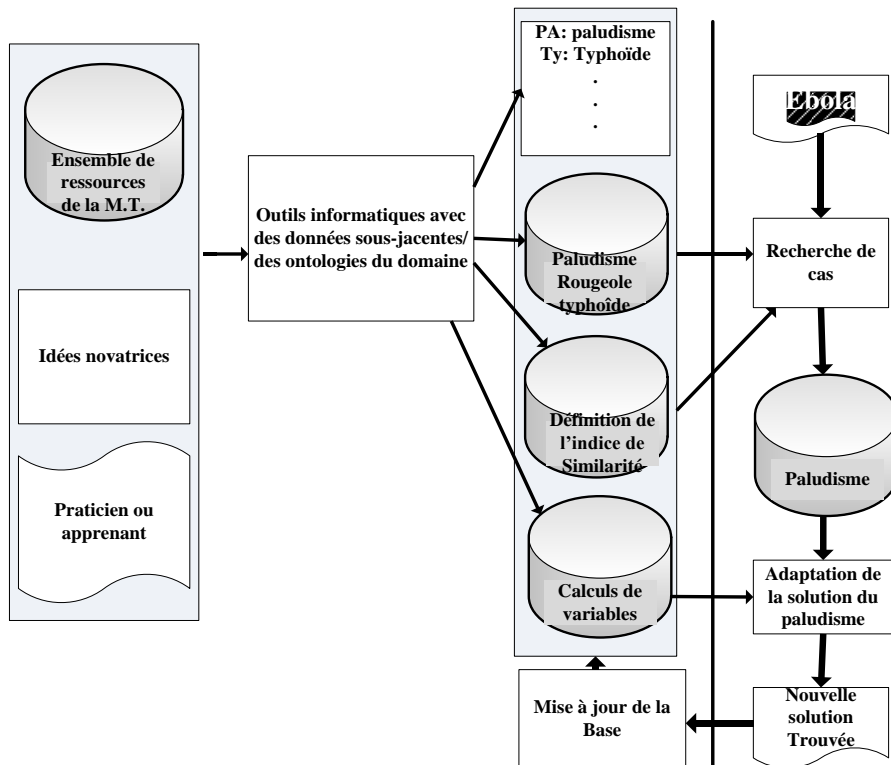


Figure 8 2 : Exemple d'adaptation de cas

Selon la figure 8.2 nous nous trouvons devant une situation où il existe une base de données dans laquelle sont stockées des informations sur des maladies ayant été traitées et guéries par la science médicale traditionnelle. Si, par la suite, il advenait que d'autres signes apparaissent qui ne se trouvent pas dans la base de données, on poursuivrait les recherches pour aboutir à un diagnostic différentiel permettant de déterminer une autre maladie (*ebola*) nécessitant un nouveau mode de traitement qui viendrait enrichir la base de connaissances existante.

Lorsqu'il s'agira d'utiliser le RàPC, des descripteurs appelés mesures de similarité vont être définis puis appliqués selon un certain nombre de principes.

8.2.2. Principes de la mesure de similarité

Un système de RàPC doit résoudre, dans un domaine particulier, de nouveaux problèmes en adaptant des solutions préexistantes qui ont été déjà utilisées pour la résolution d'anciens problèmes. La remémoration de cas sources (problèmes déjà résolus) pour chercher la solution à un nouveau cas cible (nouveau problème) est le point focal du cycle RàPC. En effet, les systèmes de RàPC utilisent différentes techniques pour comparer la description d'un cas cible avec celle d'un des cas sources déjà connus. L'utilisateur donne une description du nouveau problème et le système cherche dans sa base, le cas source dont la description est la plus similaire possible à la description du nouveau problème cible. Ensuite, à travers une session de consultation des questions et réponses des

cas sources, le système propose les cas susceptibles d'être les plus proches ; l'utilisateur sélectionne le cas qui lui semble être le plus approprié.

Une mesure de similarité dans un système de Raisonnement à Partir de Cas doit posséder les propriétés suivantes :

- **Réflexivité** : Un cas étant, par principe, similaire à lui-même, la mesure de similarité doit être nécessairement réflexive.
- **Symétrie** : Si un cas A est similaire à un autre cas B alors il est nécessaire que le cas B soit aussi similaire au cas A en utilisant la mesure de similarité.
- **Non-transitivité** : Une mesure de similarité ne doit pas nécessairement être transitive. Par conséquent, Si un cas A est similaire à un cas B ; et si le cas B est similaire à un cas C, il n'est pas impératif que l'application de la mesure de similarité permette de déduire que le cas A est obligatoirement similaire à C. La non exigence de la transitivité provient du fait que les caractéristiques qui définissent les similarités entre A et B et entre B et C ne sont pas nécessairement les mêmes.

8.3. Modélisation de la similarité du diagnostic

Soit $\mathcal{C} = (C_i)_{i \in I}$ l'ensemble des I patients diagnostiqués, traités et guéris par un praticien de la médecine traditionnelle. Ces individus constituent les cas sources et forment la base de cas. Chaque cas peut être décomposé en un couple $C_i = (S_i, \Gamma_i)$ où S_i est l'ensemble des symptômes présentés par le patient et Γ_i représente la suite médicale donnée aux symptômes observés sur le patient par le tradipraticien. Cette suite est un couple $\Gamma_i = (M_i, T_i)$ dans lequel M_i désigne la principale maladie diagnostiquée et T_i le traitement prescrit au patient et ayant conduit à son rétablissement.

Notons :

- $\mathcal{S} = (X_1, \dots, X_S) = (X_j)_{1 \leq j \leq S}$, l'ensemble des S signes cliniques considérés chez les patients, toutes les maladies confondues.
- \mathcal{M} l'ensemble des M individus malades enregistrés ayant eu recours à la médecine traditionnelle (les cas).
- \mathcal{N} l'ensemble des N signes cliniques numériques (température, pression artérielle, tension, etc.),
- \mathcal{Q} l'ensemble des Q signes cliniques catégorielles (fièvre, migraine, etc.) et
- \mathcal{T} l'ensemble des T variables de description textuelle de signes divers non standards.

$$\text{On peut écrire que } \mathcal{S} = \mathcal{N} \cup \mathcal{Q} \cup \mathcal{T} \quad (1)$$

$$\text{et on a } S = N + Q + T \quad (2)$$

Soit $I_s \in \mathcal{C}$ un cas source et I_c un nouveau patient de la médecine traditionnelle appelé cas cible dans le contexte du RàPC. Si l'on note $Sim(\cdot)$ la mesure de la similarité entre ces deux patients et Φ une réécriture de $Sim(\cdot)$, on a :

$$\begin{aligned} Sim(I_s, I_c) &= Sim(C_s, C_c) \\ &= Sim(\{S_s, \Gamma_s\}, \{S_c, \Gamma_c\}) \\ &= Sim(\{S_s, S_c\}, \{\Gamma_s, \Gamma_c\}) \\ &= Sim(S_s, S_c) \\ &= Sim(\{\mathcal{N}_s \cup \mathcal{Q}_s \cup \mathcal{T}_s\}, \{\mathcal{N}_c \cup \mathcal{Q}_c \cup \mathcal{T}_c\}) \\ &= \Phi(\{(\mathcal{N}_s, \mathcal{N}_c), (\mathcal{Q}_s, \mathcal{Q}_c), (\mathcal{T}_s, \mathcal{T}_c)\}) \\ &= \Psi(\alpha(\mathcal{N}_s, \mathcal{N}_c), \beta(\mathcal{Q}_s, \mathcal{Q}_c), \gamma(\mathcal{T}_s, \mathcal{T}_c)) \\ &= \Psi(\alpha(I_s, I_c), \beta(I_s, I_c), \gamma(I_s, I_c)) \end{aligned} \quad (3)$$

Les fonctions α, β, γ désignent des distances entre deux vecteurs de variables respectivement numériques, qualitatives et textuelles. Celles-ci s'appliquent à chacune des entités uniformes (numérique, qualitative, textuelle) du vecteur d'un cas. Quant à la fonction Ψ , elle désigne l'expression mathématique finale de la mesure de similarité.

8.4. Différentes fonctions de la mesure de similarité

8.4.1. Fonction α

α étant une distance entre deux vecteurs numériques, il semble naturel de la définir en utilisant une distance de Minkowski. La distance de Minkowski d'ordre p notée L_p permet de quantifier une distance entre deux vecteurs. Elle est définie comme :

$L_p = (\sum_{i=1}^N |Ix_i - Iy_i|^p)^{1/p}$ Où N représente la dimension du vecteur \vec{I}_{x_i} et \vec{I}_{y_i} . Trois cas de la norme L_p ont un intérêt particulier à savoir : La distance de City-Block ($p = 1$), distance euclidienne ($p = 2$) et distance de Chessboard ($p = \infty$).

$$\text{Par conséquent, on a : } \alpha(I_s, I_c) = \sqrt[p]{\sum_{X_k \in \mathcal{N}} (X_{ks} - X_{kc})^p}. \quad (4)$$

On pourra se limiter à $p = 2$ qui correspond à la distance euclidienne. Ainsi, on prendra comme distance entre les variables numériques, la fonction définie par :

$$\alpha(I_s, I_c) = \sqrt{\sum_{X_k \in \mathcal{N}} (X_{ks} - X_{kc})^2} \quad (5)$$

La métrique α telle que définie possède les propriétés de réflexivité et de symétrie escomptées. En effet, on a :

$$\alpha(I_s, I_s) = \sqrt{\sum_{X_k \in \mathcal{N}} (X_{ks} - X_{ks})^2} = 0 \quad (6)$$

$$\text{Symétrie : } \alpha(I_s, I_c) = \sqrt{\sum_{X_k \in \mathcal{N}} (X_{ks} - X_{kc})^2} = \sqrt{\sum_{X_k \in \mathcal{N}} (X_{kc} - X_{ks})^2} = \alpha(I_c, I_s) \quad (7)$$

8.4.2. Fonction β

β étant une distance entre deux vecteurs de variables qualitatives, il semble naturel de considérer la distance du Khi-deux. La distance de Khi-deux permet de déterminer les espèces les plus importantes. Cette distance attribue autant d'importance à l'absence qu'à la présence d'une espèce et donne généralement des résultats satisfaisants. Elle est définie comme suit :

$$\text{Khi} - 2 = 2 * \frac{\sum(\text{effectif observés} - \text{effectifs théoriques})}{\text{effectifs théoriques}}$$

Il y a ici davantage de variables qualitatives. Nous proposons donc de construire une distance de Khi-Deux par variable qualitative, puis de faire leur somme pondérée par l'inverse du nombre de degrés de liberté.

$$\text{Soit } \mathcal{C}^c = \mathcal{C}_{s \leftarrow c} = \{\mathcal{C} \setminus \{I_s\}\} \cup \{I_c\} = \mathcal{C}_{I_s}^I \cup \{I_c\} \quad (8)$$

la base de cas modifiée en remplaçant le cas source I_s par le cas cible I_c est

$$\mathcal{C}^s = \mathcal{C}_{c \leftarrow s} = \{\mathcal{C}^c \setminus \{I_c\}\} \cup \{I_s\} = \mathcal{C} \quad (9)$$

la base de cas obtenue par l'opération inverse.

$$\text{Par construction, on a : } \text{Card}(\mathcal{C}^c) = \text{Card}(\mathcal{C}^s) = \text{Card}(\mathcal{C}) = S. \quad (10)$$

$\forall X_k \in \mathcal{Q}$, notons N_k le nombre de modalités de la variable qualitative X_k et x_{kj} la j^{eme} modalité de X_k . Notons : $N_{kj}^E = \text{card}\{I_i \in E | X_k = x_{kj}\}$ (11)

$$\chi_k^2(I_s, I_c) = \sum_{j=1}^{N_k} \frac{(N_{kj}^{\mathcal{C}^s} - N_{kj}^{\mathcal{C}^c})^2}{N_{kj}^{\mathcal{C}^s} N_{kj}^{\mathcal{C}^c}} \quad (12)$$

$$\omega_k = \frac{1}{\text{ddl}(X_k)} = \frac{1}{N_k + 1} \quad (13)$$

En utilisant l'Eq. (11), Eq. (12) et Eq. (13), la métrique β est Définie par:

$$\beta(I_s, I_c) = \sum_{k=1}^{k=Q} \omega_k \chi_k^2(I_s, I_c) = \sum_{k=1}^{k=Q} \sum_{j=1}^{N_k} \frac{(N_{kj}^{\mathcal{C}^s} - N_{kj}^{\mathcal{C}^c})^2}{(N_k + 1) N_{kj}^{\mathcal{C}^s} N_{kj}^{\mathcal{C}^c}} \quad (14)$$

La métrique β telle que définie possède les propriétés de réflexivité et de symétrie escomptées. En effet, on a :

$$\beta(I_s, I_s) = \sum_{k=1}^{k=Q} \omega_k \chi_k^2(I_s, I_s) = \sum_{k=1}^{k=Q} \sum_{j=1}^{N_k} \frac{(N_{kj}^{\mathcal{C}^s} - N_{kj}^{\mathcal{C}^s})^2}{(N_k + 1) N_{kj}^{\mathcal{C}^s} N_{kj}^{\mathcal{C}^s}} = 0 \quad (15)$$

$$\beta(I_s, I_c) = \sum_{k=1}^{k=Q} \omega_k \chi_k^2(I_s, I_c) = \sum_{k=1}^{k=Q} \sum_{j=1}^{N_k} \frac{(N_{kj}^{\mathcal{C}^s} - N_{kj}^{\mathcal{C}^c})^2}{(N_k + 1) N_{kj}^{\mathcal{C}^s} N_{kj}^{\mathcal{C}^c}} = \sum_{k=1}^{k=Q} \sum_{j=1}^{N_k} \frac{(N_{kj}^{\mathcal{C}^c} - N_{kj}^{\mathcal{C}^s})^2}{(N_k + 1) N_{kj}^{\mathcal{C}^c} N_{kj}^{\mathcal{C}^s}} = \beta(I_c, I_s) \quad (16)$$

8.4.3. Fonction γ

La fonction γ est une distance entre deux vecteurs de variables textuelles. Pour la définir, nous considérerons la distance appelée « kernel string » que l'on peut traduire imparfaitement en français par distance à « noyaux textuels ». Comme plusieurs variables textuelles sont en jeu, nous proposons de construire une distance à noyaux textuels par variable, puis de faire leur somme pondérée par le poids relatif de chaque variable textuelle relativement aux séquences de mots communs. Les méthodes à noyaux permettent de trouver des fonctions de décision non linéaires, tout en s'appuyant fondamentalement sur des méthodes linéaires.

$\forall X_k \in \mathcal{T}$, on note Δ_{ks} et Δ_{kc} les corpus de la variable textuelle X_k correspondant respectivement au cas source I_s et au cas cible I_c à partir d'une même suite d'opérations linguistiques, en particulier la même opération de segmentation de textes (tokenisation). Dans le cas de la médecine traditionnelle, nous conseillons les opérations de mise en minuscule des textes, de suppression des espaces vides, de suppression des valeurs numériques, de suppression des ponctuations, de suppression des mots vides et de segmentation par l'usage de l'algorithme des tri-grams. L'utilisation de l'algorithme du tri-gram pour la tokenisation est vivement conseillée. En effet, nous supposons un mot composé « document » subdivisé comme suit : *Do*, *cum* et *ent*.

Ainsi donc, on veut trouver la séquence *cu* en appliquant l'algorithme de tri-grams on aura : $\mathbb{P}(cu|Doc)$ et avec la probabilité conditionnelle sur ces éléments indépendants on aura :

$$\mathbb{P}(cu|Doc) = \mathbb{P}(D) * \mathbb{P}(o|D) * \mathbb{P}(c|Do)$$

Remarque : un n-gramme permet de déterminer les corpus d'une séquence de n mots. Par exemple, l'ensemble de tri-gram (N = 3) du terme (document) = {_do, doc, ocu, cum, ume, men, ent, nt_}, ou le caractère "_" représente le caractère blanc.

Ainsi, une fois la segmentation de texte effectuée, l'ensemble des tri-grams de X_k communs aux cas source I_s et au cas cible I_c est donnée par $\Delta_k = \Delta_{ks} \cap \Delta_{kc}$. (17)

Notons $\tau_{kj} \in \Delta_k$ un tri-gram rencontré δ_j^{ks} fois dans Δ_{ks} et δ_j^{kc} fois dans Δ_{kc} . Le nombre de fois que le tri-gram τ_{kj} est rencontré dans Δ_k est $\delta_{kj} = \text{Min}(\delta_j^{ks}, \delta_j^{kc})$. (18)

La distance cosinus correspondant entre le cas source et le cas cible peut être définie par:

$$\text{cosine}_k(I_s, I_c) = \frac{\langle X_{ks}, X_{kc} \rangle}{\|X_{ks}\| \cdot \|X_{kc}\|} = \frac{\sum_{j=1}^{\text{card}(\Delta_k)} \delta_j^{ks} \cdot \delta_j^{kc}}{\sqrt{\sum_{j=1}^{\text{card}(\Delta_k)} \delta_j^{ks}} \sqrt{\sum_{j=1}^{\text{card}(\Delta_k)} \delta_j^{kc}}} \quad (19)$$

$$\text{En considérant les poids } \lambda_k = \frac{\text{Card}(\Delta_k)}{\sum_{k=1}^T \text{Card}(\Delta_k)}, \quad (20)$$

on obtient :

$$\begin{aligned}
\gamma(I_s, I_c) &= \sum_{k=1}^{k=T} \lambda_k | \ln(\cosine_k(I_s, I_c)) | \\
&= \frac{1}{\sum_{k=1}^{k=T} Card(\Delta_k)} \sum_{k=1}^{k=T} Card(\Delta_k) | \ln\left(\frac{\sum_{j=1}^{card(\Delta_k)} \delta_j^{ks} \cdot \delta_j^{kc}}{\sqrt{\sum_{j=1}^{card(\Delta_k)} \delta_j^{ks}} \sqrt{\sum_{j=1}^{card(\Delta_k)} \delta_j^{kc}}} \right) | \quad (21)
\end{aligned}$$

Par construction, la métrique γ telle que définie possède les propriétés de réflexivité et de symétrie escomptées. En effet, on a :

$$\gamma(I_s, I_s) = \sum_{k=1}^{k=T} \lambda_k | \ln(\cosine_k(I_s, I_s)) | = \sum_{k=1}^{k=T} \lambda_k | \ln\left(\frac{\langle X_{ks}, X_{ks} \rangle}{\|X_{ks}\| \cdot \|X_{ks}\|} \right) | = 0 \quad (22)$$

$$\gamma(I_s, I_c) = \sum_{k=1}^{k=T} \lambda_k | \ln\left(\frac{\langle X_{ks}, X_{kc} \rangle}{\|X_{ks}\| \cdot \|X_{kc}\|} \right) | = \sum_{k=1}^{k=T} \lambda_k | \ln\left(\frac{\langle X_{kc}, X_{ks} \rangle}{\|X_{kc}\| \cdot \|X_{ks}\|} \right) | = \gamma(I_c, I_s) \quad (23)$$

8.5. Mesure de similarité globale

8.5.1. Fonction Ψ

Les mesures α , β et γ sont des distances. Pour en déduire des mesures de similarité, on utilisera une interpolation polynomiale de Lagrange sur l'intervalle $[0, 1]$. En effet, notons d_α , d_β et d_γ les valeurs normalisées de α , β et γ . On a donc $d_v = 1 - \frac{1}{1+v}$ avec $v \in \{\alpha, \beta, \gamma\}$. (24)

$$\text{et posons } \mu = \frac{d_\alpha + d_\beta + d_\gamma}{3} = 1 - \frac{\frac{1}{1+\alpha} + \frac{1}{1+\beta} + \frac{1}{1+\gamma}}{3}. \quad (25)$$

Par construction, $\mu \in [0, 1]$. (26)

On veut construire la fonction Ψ de telle sorte que $\Psi(0) = 100$, $\Psi(\frac{1}{4}) = 75$, $\Psi(\frac{3}{4}) = 25$ et $\Psi(1) = 0$

On dispose donc d'un ensemble de 5 points $(x_i, y_i)_{1 \leq i \leq 5}$ qui sont $\{(0, 100), (\frac{1}{4}, 75), (\frac{1}{2}, 50), (\frac{3}{4}, 25), (1, 0)\}$. Le polynôme de Lagrange associé à chaque (x_i, y_i) est

$$\text{donné par } l_i(X) = \prod_{j=0, j \neq i}^{j=5} \frac{x - x_j}{x_i - x_j}, \quad (27)$$

L'application de l'équation Eq. (27) permet d'obtenir la fonction polynomiale qui respecte la condition qui est:

$$\Psi(x) = \sum_{i=1}^{i=5} y_i l_i(x) = 100(1 - x) \quad (28)$$

8.5.2. Application générale de la mesure

Si on se conforme aux notations définies en 8.3, 8.4 et 8.5.1, on peut prendre comme mesure de similarité globale, la quantité définie par :

$$\begin{aligned}
Sim(T_s, T_c) &= \Psi(\mu(T_s, T_c)) \\
&= 100\left(1 - \frac{\frac{1}{1+\alpha(T_s, T_c)} + \frac{1}{1+\beta(T_s, T_c)} + \frac{1}{1+\gamma(T_s, T_c)}}{3}\right) \quad (29)
\end{aligned}$$

Avec :

$$\alpha(I_s, I_c) = \sqrt{\sum_{X_k \in \mathcal{N}} (X_{ks}^2 + X_{kc}^2)}$$

$$\beta(I_s, I_c) = \sum_{k=1}^{k=Q} \sum_{j=1}^{N_k} \frac{(N_{kj}^{\mathcal{C}} - N_{kj}^{\mathcal{C}_{s \leftarrow c}})^2}{(N_k - 1) N_{kj}^{\mathcal{C}_{s \leftarrow c}}}$$

$$\gamma(I_s, I_c) = \frac{1}{\sum_{k=1}^{k=T} \text{Card}(\Delta_k)} \sum_{k=1}^{k=T} \frac{\sum_{j=1}^{\text{card}(\Delta_k)} \delta_j^{ks} \cdot \delta_j^{kc}}{\sqrt{\sum_{j=1}^{\text{card}(\Delta_k)} \delta_j^{ks}} \sqrt{\sum_{j=1}^{\text{card}(\Delta_k)} \delta_j^{kc}}} \cdot \text{Card}(\Delta_k)$$

Où $\alpha(I_s, I_c)$, $\beta(I_s, I_c)$ et $\gamma(I_s, I_c)$ sont ceux définir respectivement dans Eq. (5), Eq. (14) et Eq. (21).

L'indice de similarité tel que définit est décroissant avec la divergence du cas cible par rapport aux cas sources, réflexifs, symétriques et bornés entre 0 et 100. En effet :

$$\text{Sim}(I_s, I_s) = 100 \left(1 - \frac{\frac{1}{1+\alpha(I_s, I_s)} + \frac{1}{1+\beta(I_s, I_s)} + \frac{1}{1+\gamma(I_s, I_s)}}{3} \right) = 0 \quad (30)$$

$$\begin{aligned} \text{Sim}(I_s, I_c) &= 100 \left(1 - \frac{\frac{1}{1+\alpha(I_s, I_c)} + \frac{1}{1+\beta(I_s, I_c)} + \frac{1}{1+\gamma(I_s, I_c)}}{3} \right) \\ &= 100 \left(1 - \frac{\frac{1}{1+\alpha(I_c, I_s)} + \frac{1}{1+\beta(I_c, I_s)} + \frac{1}{1+\gamma(I_c, I_s)}}{3} \right) \\ &= \text{Sim}(I_c, I_s) \end{aligned} \quad (31)$$

$$\text{Le maximum de } \text{Sim}(I_c, I_s) \text{ est obtenu si } \alpha(I_c, I_s) = \beta(I_c, I_s) = \gamma(I_c, I_s) = \infty. \quad (32)$$

$$\text{Dans ce cas, on a } \text{Sim}(I_c, I_s) = 100(1 - 0) = 100. \quad (33)$$

De façon similaire, le minimum de $\text{Sim}(I_c, I_s)$ est obtenu si

$$\alpha(I_c, I_s) = \beta(I_c, I_s) = \gamma(I_c, I_s) = 0. \quad (34)$$

$$\text{Dans ce cas, on a } \text{Sim}(I_c, I_s) = 100(1 - 1) = 0. \quad (35)$$

Par conséquent, $\text{Sim}(I_c, I_s)$ est bornée entre 0 et 100.

8.6. Simulation

Pour tester notre approche, nous avons considéré un échantillon de maladies. Celles-ci sont définies par leurs symptômes, leurs signes et les éléments textuels qui les caractérisent. Ces maladies ont été traitées et guéries par la médecine traditionnelle et stockées dans la base de données (voir tableau 8.1).

Tableau 8.1 : Ensemble de cas de maladies traitées et guéries par la médecine traditionnelle

	Température...C.	Tension	Sexe	Qualité	Âges	Commentaire1	Commentaire2
1	34	13	Masculin	Bonne	12	Traitement à base de racine d'annona senegalensis	Traitement à base de feuille d'azadirachta indica
2	33	20	Féminin	Mauvaise	23	Traitement à base de feuille d'athonotha crassifo>	Traitement à base d'églantier
3	39	18	Masculin	Moyenne	45	Traitement à base de feuilles de papaye	Traitement à base d'églantine
4	39	20	Masculin	Mauvaise	12	Traitement à base de piments	Traitement à base d'écorce d'érable
5	33	17	Masculin	Moyenne	34	Traitement à base de neem	Traitement à base de feuilles d'eucalyptus
6	37	21	Masculin	Mauvaise	56	Traitement à base de poivre	Traitement à base de l'écorce de figuier
7	34	22	Masculin	Mauvaise	76	Traitement à base d'oignons	Traitement à base de feuilles de flamboyant
8	36	18	Féminin	Moyenne	23	Traitement à base de fleurs femelles de maïs	Traitement à base de forsythia
9	40	17	Féminin	Moyenne	12	Traitement à base d'abricotier	Traitement à base de la fougère
10	41	17	Masculin	Moyenne	40	Traitement à base d'acacia	Traitement à base de rameaux nauclea latifolia
11	33	12	Féminin	Bonne	50	Traitement à base de feuille de cassia alata	Traitement à base de graine de carapa procera
12	40	17	Féminin	Moyenne	45	Traitement à base d'aloès	Traitement à base de fuchsia
13	38	21	Féminin	Mauvaise	65	Traitement à base de quinquina	Traitement à base de gazania
14	40	21	Masculin	Mauvaise	75	Traitement à base d'althéa	Traitement à base de genêt
15	37	13	Masculin	Bonne	45	Traitement à base d'amandier	Traitement à base de gentiane
16	41	15	Masculin	Bonne	45	Traitement à base d'amaryllis	Traitement à base de gerbera
17	42	16	Féminin	Bonne	63	Traitement à base d'ancolie	Traitement à base de giroflée
18	38	16	Féminin	Bonne	9	Traitement à base d'anémone	Traitement à base de gueule de loup
19	37	14	Féminin	Bonne	16	Traitement à base d'anthémis	Traitement à base d'hellébore
20	37	19	Féminin	Mauvaise	14	Traitement à base d'anthurium	Traitement à base de feuille d'hêtre
21	32	21	Féminin	Mauvaise	17	Traitement à base d'arum	Traitement à base d'hibiscus
22	37	17	Féminin	Moyenne	18	Traitement à base d'aster	Traitement à base d'hortensia
23	41	21	Masculin	Mauvaise	19	Traitement à base d'aubépine	Traitement à base de graines d'houx
24	41	17	Masculin	Moyenne	20	Traitement à base d'azalée	Traitement à base d'ipomée
25	33	20	Masculin	Mauvaise	20	Traitement à base de Bambou	Traitement à base d'iris
26	35	21	Masculin	Mauvaise	12	Traitement à base de Bananier	Traitement à base de jacinthe
27	36	19	Masculin	Mauvaise	10	Traitement à base de feuille de baobab	Traitement à base de jasmin
28	36	16	Masculin	Moyenne	11	Traitement à base de bégonia	Traitement à base de jonquille
29	34	13	Masculin	Bonne	13	Traitement à base de bleuet	Traitement à base de laurier
30	39	17	Masculin	Moyenne	13	Traitement à base de bougainvillier	Traitement à base de laurier-rose
31	41	16	Féminin	Moyenne	17	Traitement à base de bouleau	Traitement à base de lavande

Nous avons calculé les mesures de similarité des différentes distances entre les individus des cas source (voir tableau) et d'un cas cible. Un exemple d'algorithme du calcul de ces métriques est présenté en annexe C.

La figure 8.3 nous montre le diagramme de la classification hiérarchique des individus de notre métrique selon leur ressemblance.

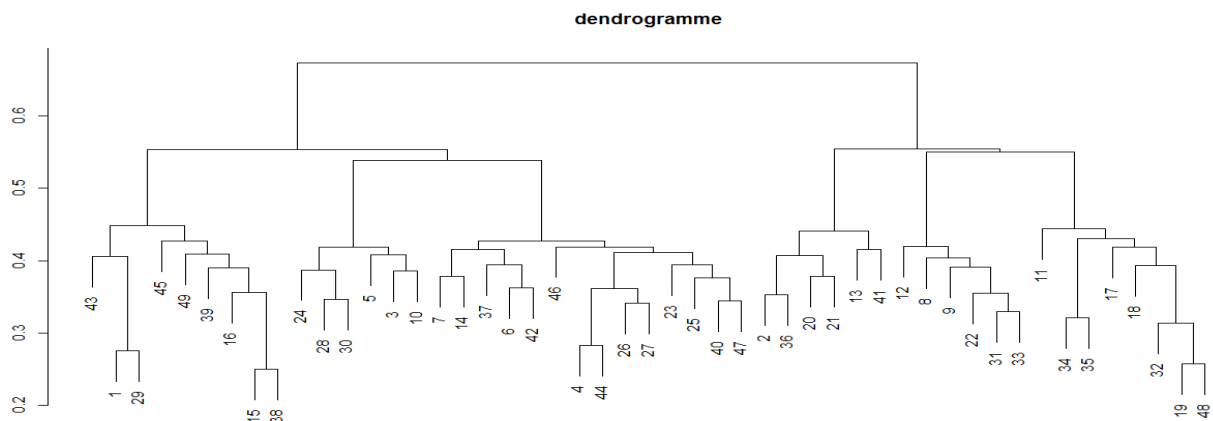


Figure 8 3 : Classification hiérarchique avec notre métrique

La figure 8.4 permet de regrouper les différents cas similaires. Ce regroupement nous amène à retrouver la position du cas cible. Nous nous basons alors sur le traitement reçu par les individus de ce groupe pour simuler le traitement de ce nouveau cas.

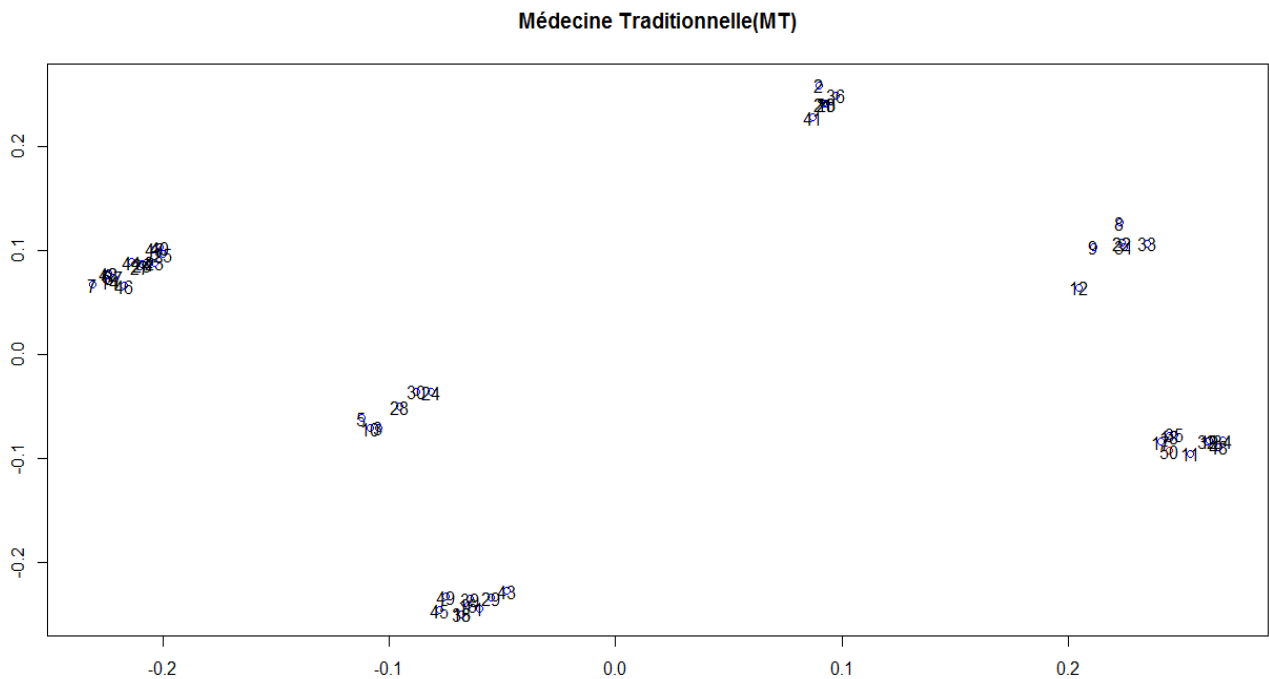


Figure 8 4 : Positionnement multidimensionnel avec notre métrique

8.7. Comparaison au niveau de la mesure de similarité

La technique du RàPC a été fréquemment utilisée en médecine moderne et plus spécifiquement en chirurgie corrective pour aider à la prise de décision préopératoire. Cette technique se fonde sur l’histoire d’un patient opéré avec des résultats positifs, est comparé à un nouveau cas similaire. Le nœud du problème repose sur la représentation de ce cas et la mémorisation de cas similaires antérieurs. Les données sur le cas en question font l’objet de comparaison avec de nouveau cas au moyen d’indices de similarité ; concrètement le malade à opérer est comparé à des malades ayant déjà été opérés avec succès selon un certain nombre d’indices de similarité. Les travaux de Chusseau et *al.* [Chus 00] illustrent parfaitement cette approche et donnent des résultats plus que satisfaisants.

Nous avons utilisé la même approche dans le cadre de la MT pour aboutir à des résultats qui pourraient être aussi satisfaisants que ceux obtenues en médecine moderne.

8.8. Conclusion

L’application du raisonnement à partir de cas (RàPC), dans le cadre de la médecine traditionnelle se révèle être une opération très complexe du fait que les informations textuelles, les données qualitatives et le nombre de modalités de chaque variable considérée sont difficilement appréhendables. Il est donc nécessaire de proposer un indice de similarité prenant en compte la diversité des types d’informations que l’on peut prendre sur chaque cas.

L'intérêt de la mesure de similarité proposée dans ce chapitre est double. Cette mesure prend en compte d'une part, la diversité des types de données généralement recueillies dans le cadre de la médecine traditionnelle (numérique, qualitative et textuelle) et, présente d'autre part, l'avantage de préserver les deux propriétés théoriques de base d'un tel type d'indicateur à savoir la réflexivité et la symétrie. En outre, son caractère borné entre les valeurs 0 et 100 peut lui conférer une interprétation pratique en terme de pourcentage. Le sens exact de cette interprétation pourrait faire l'objet de recherches ultérieures.

Eu égard à la complexité des modalités d'approche que soulève l'exercice de la médecine traditionnelle et, animé du souci de rationaliser au maximum la pratique de cette science, nous nous sommes proposés d'introduire dans notre démarche un système en rapport avec l'utilisation d'un langage de requête particulier appelé SPARQL. L'usage de ce langage SPARQL dans notre démarche a l'avantage d'éviter aux utilisateurs la manipulation d'un trop grand nombre de fichiers médicaux.

Le chapitre qui suit va s'employer à expliquer le mode de fonctionnement de cette technique.

Chapitre 9

METHODOLOGIE D'APPROCHE VIA LE LANGAGE SPARQL.

***Résumé.** Nous proposons dans ce chapitre, une méthode complémentaire de pratique de la médecine traditionnelle, essentiellement fondée sur l'utilisation du langage de requête SPARQL (SPARql Query Language). Il y est également question de montrer comment les ressources sémantiques appartenant en propre à cette discipline, peuvent, lorsqu'elles font l'objet d'une recherche, être obtenues par l'utilisation d'une interface de requêtes textuelles.*

9.1. Introduction

De nos jours, il y a possibilité pour tout un chacun d'accéder à une masse importante de données disponibles sur le web ; ainsi, avec le développement quantitatif des entrepôts de triplets RDF, l'utilisation de l'interfaçage de moteur SPARQL s'avère nécessaire car il est pratiquement impossible pour un utilisateur final de maîtriser la complexité des schémas de données d'un domaine tel que celui de la médecine traditionnelle. En effet, pour pouvoir exprimer une requête valide sur les ressources du web sémantique, l'utilisateur doit connaître le langage SPARQL, les ontologies et la forme des graphes RDF utilisées. Celui-ci se charge alors d'introduire une ontologie et saisit les requêtes en rapport avec ses besoins. Les requêtes des utilisateurs sont exprimées dans un langage naturel et subissent une transformation en langage formel grâce à des systèmes existants tel que le « SemanticWeb Interface with Patterns » (SWIP) [Pra 13]. Le système alors se charge de déterminer le type de la requête saisie ainsi que sa réponse en exécutant un ensemble de règles prédéfinies. Si les règles ne permettent pas de trouver une réponse à la requête, l'assistance d'un expert peut être sollicitée ; celui-ci, soit reformule la requête, soit recharge un référentiel sémantique qui sera exploité pour la recherche de la réponse pertinente. De manière générale, le but de l'utilisation des ontologies est de permettre un dialogue entre la machine et l'utilisateur afin de ne pas fournir à l'ordinateur des informations disparates. Cette pratique apparaît comme un excellent moyen d'apprentissage et de pratique de la médecine traditionnelle du fait que les ressources disponibles sont stockées et regroupées par concepts sémantiques.

Dans ce chapitre, nous décrivons quelques scénarii où la solution proposée résulte d'une approche architecturale basée sur la construction de requêtes par le biais d'un langage de requête naturelle SPARQL.

9.2. Architecture Globale du Système Proposé

La phase la plus importante dans n'importe quel processus de développement d'un logiciel est la phase de conception. Celle-ci définit les processus, les principes et les modèles d'un système. Elle renferme suffisamment de détails permettant d'aboutir à la conception et à la réalisation d'un système.

Nous présentons notre architecture agent c'est-à-dire une architecture basée sur des programmes informatiques baptisés système multi-agents (S.M.A.). C'est une technique de communications et de partage d'informations basée sur le principe de l'Enseignement Intelligemment Assisté par Ordinateur (E.I.A.O.) (figure 9.1). Cette technique présente notre système en utilisant le paradigme multi-agent et fait distribuer l'expertise sur un ensemble d'agents qui modélisent les différents modules de l'E.I.A.O. Ce système est subdivisé en deux modules :

- Le module de gestion de la base de connaissances renferme toutes les composantes d'une base de données ;
- Le module d'interrogation est, quant à lui, destiné aux utilisateurs pour exprimer leurs requêtes et recueillir les réponses.

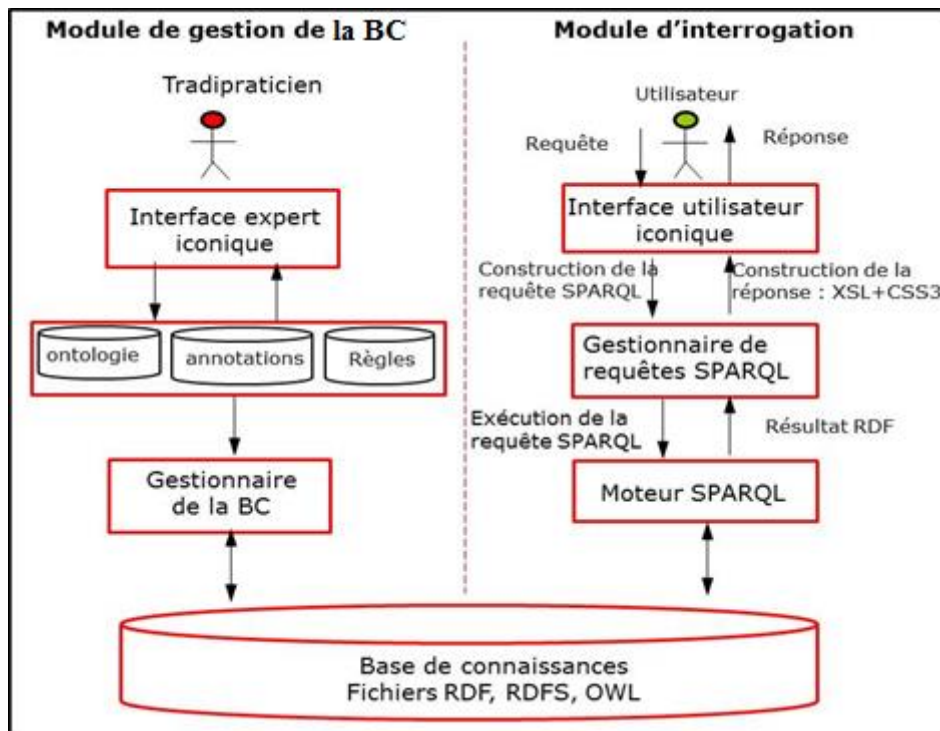


Figure 9.1 : Architecture globale du système

9.2.1. Module de gestion

9.2.1.1. Gestionnaire de la BC

La gestion informatisée de l'ensemble des connaissances susceptibles d'être maîtrisées par le tradipraticien doit faire l'objet de numérisation à partir des documents et des données issues de la médecine traditionnelle. Cette gestion prend également en compte l'analyse et le partage des connaissances de très haut niveau de même qu'elle prend en compte les données et les informations factuelles c'est-à-dire celles qui s'en tiennent aux faits et/ou aux concepts [Paq 03].

Une base de connaissances est avant tout une banque de ressources pédagogiques. Elle permet à l'utilisateur de référencer les ressources d'un domaine donné (en l'occurrence la médecine traditionnelle) à l'aide de spécifications standards fondées sur une ontologie du domaine ; puis de les intégrer à la banque de ressources en se basant sur la description d'un ensemble d'activités à des fins de recherche et d'enseignement. Les ressources sont référencées par des métadonnées utilisant une ontologie du domaine et leur recherche est guidée par des agents informatiques décrivant les activités dans lesquelles ces ressources sont utilisées ou produites.

Le gestionnaire de la Base de Connaissances fournit des services d'analyse, d'extraction et de stockage de connaissances et permet de générer automatiquement les fichiers de la B.C. à partir de l'ontologie du domaine, des annotations et des règles.

9.2.1.2. Module d'interrogation

Le module d'interrogation permet de retrouver une ressource à partir de ses caractéristiques tout en gardant ses performances les plus acceptables. Il est chargé de la construction des requêtes, de leur exécution ainsi que de la présentation du résultat à l'utilisateur. L'interface utilisateur permet à celui qui en a la charge de créer une requête SPARQL sans que celui-ci connaisse nécessairement l'ontologie du domaine.

9.2.1.3. Gestionnaire de requêtes SPARQL

Un gestionnaire de requêtes est basé sur les expressions de correspondance pour réécrire une requête exprimée sur le schéma d'un pair spécifique ou sur les schémas des pairs liés. Cependant, les tables de correspondance, bien que plus faciles à établir qu'un processus complet de découverte de correspondances entre schémas, sont jusqu'ici créées manuellement par des spécialistes du domaine ; ce qui peut être coûteux en temps. Le gestionnaire de requête

SPARQL traduit une requête utilisateur dans sa forme correspondante en SPARQL avant de la passer au Moteur SPARQL. Ensuite, il reçoit la réponse de la requête du Moteur SPARQL et transforme les résultats en données RDF avant de retourner le résultat en RDF à l'interface utilisateur.

9.2.1.4. Moteur SPARQL

Parmi les nombreux langages de requêtes sur des données RDF (Resource Description Framework) (wikipédia anglophone en recense une dizaine) existant sur le web, ceux qui semblent le plus en vue sont RDF Data Query Language (RDQL) et Sparql Protocol And RDF Query Language (SPARQL) [Bern 01].

SPARQL est le standard recommandé par le W3C depuis janvier 2008. Il permet d'exécuter la requête SPARQL et de délivrer le résultat RDF au Gestionnaire de requête SPARQL.

9.3. Accès aux données contenues dans l'entrepôt sémantique

Nous avons opté pour la requête SPARQL afin de pouvoir produire de manière efficace et précise les résultats d'un diagnostic. L'objectif de cette approche est non seulement de réduire le temps et le coût de la recherche vu l'importance des données que traite la médecine traditionnelle, mais également de disposer d'un protocole et d'un langage de requête similaire à SQL adapté à RDF. Cette approche est généralement basée sur l'interprétation des ontologies. Les processus SPARQL pour twinkle, pour le java ou RDFAPI for MySQL existent déjà et sont très complets et utilisables. La portion de code de la figure 9.2 est celle du langage de requête SPARQL. Elle permet de visualiser notre ressource RDF chargée dans le serveur.

```
1.  select DISTINCT ?subject ?property ?object
2.  where { ?subject ?property ?object }
```

Figure 9.2 : Requête du langage SPARQL

9.4. Interface graphique de recherche d'informations

La figure 9.3 suivante représente l'interface graphique de la plateforme de recherche d'informations médicales. Cette plateforme est mise en œuvre en utilisant le langage PHP (nous avons présenté quelques codes sources en Annexe B). Le « PHP » est un langage de script (coté serveur) qui a été conçu spécifiquement pour le web. Il est inclus dans une page « HTML » et est interprété au niveau du serveur web. Notre choix s'est porté sur ce langage parce qu'il fait partie du projet « open-source » dont le code source est non seulement d'accès et d'utilisation

faciles mais aussi modifiable et gratuitement disponible. Cette interface présentée ci-après est implémentée suivant l'architecture de la figure 9.3. Elle est subdivisée en deux grandes parties :

- La partie supérieure est l'interface destinée à l'utilisateur ; celui-ci l'utilise pour exécuter ses différentes requêtes sur les ressources ontologiques de la médecine traditionnelle.
- La partie inférieure est réservée à l'affichage de la ou des réponse(s) attendue(s) par l'utilisateur.

En dehors de ces deux parties, il existe une autre partie réservée au chargement de la ressource ontologique comme le montre la figure 9.4

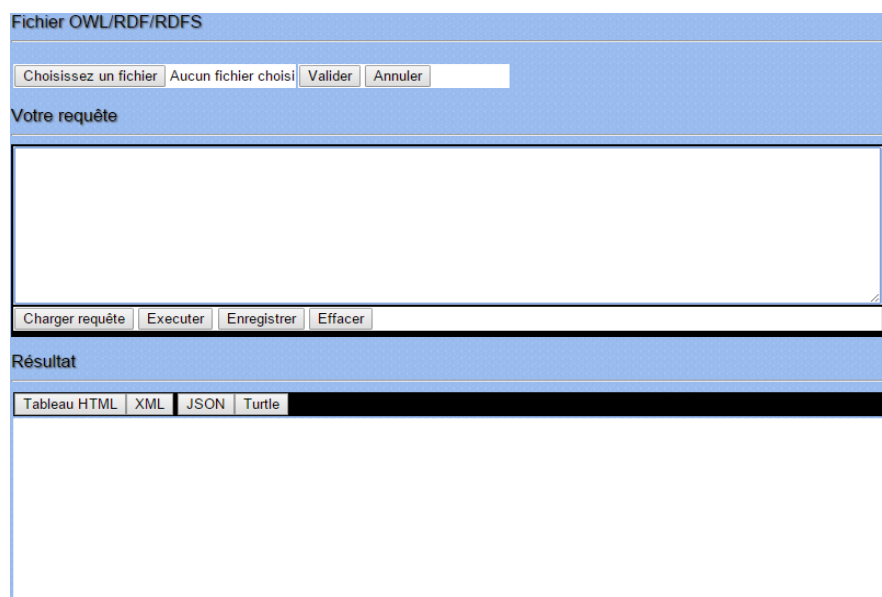


Figure 9.3 : Interface d'utilisation

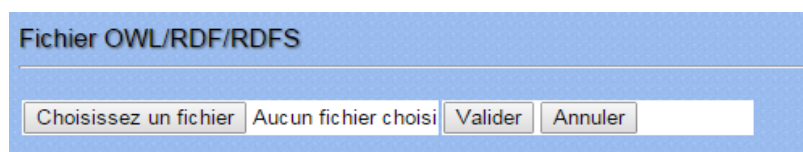


Figure 9.4 : Chargement d'une base ontologique

A titre de démonstration, nous utilisons au point suivant une ressource de connaissance tirée de la médecine traditionnelle sous la forme d'une ontologie du paludisme.

9.5. Ontologie du paludisme

Pour faciliter le processus de réflexion sur une maladie comme le paludisme dans le cadre de la médecine traditionnelle, nous avons utilisé l'ontologie du paludisme (figure 9.5). L'acquisition de cette connaissance sera facilitée par le langage SPARQL. Ce langage ne

s'accommode qu'aux fichiers de type « RDF, OWL, XML, RDFS ». Ainsi donc, cette technique offre aux utilisateurs la possibilité de réutiliser ou d'enrichir leurs bases de connaissances. Elle permet également à ces mêmes utilisateurs ainsi qu'à de nouveaux apprenants de partager entre eux leurs connaissances sur le paludisme ; ce qui devrait, en principe, favoriser la pérennisation de l'ensemble des connaissances capitalisées sur le paludisme dans la pratique de la médecine traditionnelle.

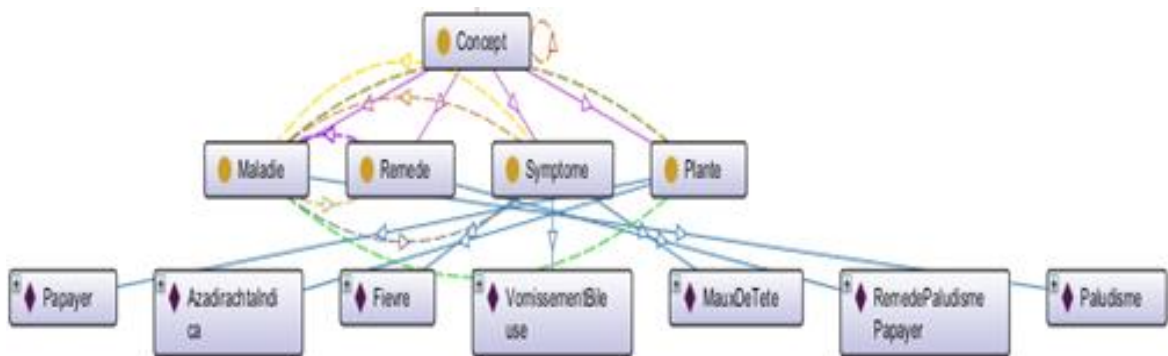


Figure 9.5 : Ontologie du paludisme

Le point suivant, présente une démonstration de requête au travers d'une interface (figure 9.6) basée sur l'ontologie du paludisme (figure 9.5).



Figure 9.6 : Exécution de la requête

On charge la ressource ontologique du paludisme ; ensuite on reformule une requête en langage SPARQL puis on l'exécute afin d'obtenir les informations recherchées.

9.6. Résultats d'exécution

L'interface de la figure 9.7 nous donne un aperçu des résultats. L'ensemble des informations (ressources) sur le paludisme contenues dans le fichier d'ontologie qui a été chargé dans le serveur, s'affiche sur cette figure. Pour parvenir à ce résultat, nous nous sommes basés sur la requête de la figure 9.2.

R�sultat			
Tableau HTML	XML	JSON	Turtle
http://www.semanticweb.org/gerard/ontologies/2012/8/medtrad#Paludisme	http://www.semanticweb.org/gerard/ontologies/2012/8/medtrad#langue		Fran�ais
http://www.semanticweb.org/gerard/ontologies/2012/8/medtrad#Paludisme	http://www.semanticweb.org/gerard/ontologies/2012/8/medtrad#terme		Paludisme
http://www.semanticweb.org/gerard/ontologies/2012/8/medtrad#Paludisme	http://www.semanticweb.org/gerard/ontologies/2012/8/medtrad#terme		Le paludisme est une maladie provoquee par...

Figure 9.7 : Le r sultat de l'ex cution de la requ te SPARQL

9.7. Evaluation des diff rents mod les

Dans la litt rature, de nombreuses interfaces (telles que « Prot g  », « corese ») ont  t  mises au point comme approche m thodologique pour aider   l'acquisition de connaissances dans des disciplines de la m decine moderne. Cette m thodologie a facilit  la mobilisation de connaissances et rehauss  le niveau de cr dibilit  de la science m dicale moderne. Certaines de ces interfaces sont structur es pour la navigation, l'interpr tation, l'enregistrement et la visualisation des donn es m dicales ; citons par exemple le « BioAssay Ontology (BAO) » [Vis 11] qui est une interface qui couvre un large champ d'exploitation allant de la biologie animale   la m decine traditionnelle africaine. Cependant, le fait qu'elles soient stock es sur le web, est un s rieux handicap pour un bon nombre d'utilisateurs qui ne peuvent en faire usage, faute d' quipements appropri s et disponibles. En d'autres termes, lorsque l'utilisateur n'a pas de connexion internet, par exemple, il ne peut faire usage de cette application.

Certaines autres interfaces, con ues en langage de programmation « java, AJAX et javascript ... » ont une structuration tellement difficile   comprendre qu'il faut, pour obtenir des r sultats probants, faire appel   d'autres types de langage de programmation. La principale question soulev e est donc la suivante :

« Quel type de langage de programmation serait appropri  pour la mise en place d'une interface d'interrogation susceptible de conduire   des r sultats pertinents dans le cadre de la m decine traditionnelle? »

« Comment cr er une interface adapt e au tradipraticien pour interroger une base de connaissances ? »

9.8. Evaluation de notre approche

Comme expliqu  au point 9.4 on s'accorde   dire que l'utilisation du langage PHP est celle qui est susceptible de r pondre   la premi re question pr c demment pos e du fait qu'il s'agit d'un langage universel libre et disponible pour tout le monde. Le mod le d'interface

proposée est assurément efficace et répond également à la deuxième question posée si on utilise les ressources ontologiques issues de la médecine traditionnelle sans nécessairement avoir recours à d'autres éditeurs d'interface. Il nous semble que, dans notre contexte, l'apport des icônes (c'est à dire les images de taille petite ayant pour but de représenter ou d'illustrer quelque chose) pourraient être un moyen encore plus efficace d'atteindre les objectifs poursuivis à savoir l'obtention de données fiables sur une connaissance quelconque.

9.9. Conclusion

L'interface utilisateur est une composante très importante dans les systèmes d'informations. C'est l'un des éléments qui affecte considérablement l'efficacité de la recherche d'informations et par voie de conséquence, nuit à toute prise de décision. Ainsi, notre logiciel présente une interface adaptative pour l'exploration et l'analyse de données multi sources par navigation et par requête (l'implémentation de la procédure utilisée et de la plateforme d'utilisation). Des éléments du code sources sont présentés en Annexe B.

Notre modèle a donné des résultats encourageants. Il permet de retrouver les besoins médicaux du praticien ainsi que ceux du patient. D'après l'expérience de la figure 9.7, il est possible de retrouver des informations pertinentes suite à une requête. Les informations obtenues en réponse, doivent être exactes afin d'éviter tout risque d'erreurs médicales graves. Des améliorations devraient cependant être faites en terme d'indexation multidimensionnelle et d'optimisation des requêtes afin d'atteindre des performances élevées. Il semble que le pouvoir d'expression des requêtes « SPARQL », contrairement à sa syntaxe, soit largement suffisant pour répondre à notre attente

CONCLUSIONS ET PERSPECTIVES

Les travaux de recherche présentés dans ce mémoire ont porté sur la revalorisation de la MT au travers d'une démarche de rationalisation et de capitalisation de connaissances articulées autour d'un ensemble de méthodologies axées sur l'utilisation d'outils informatiques. Nous avons eu pour objectif de proposer des algorithmes et des modélisations pour faciliter l'apprentissage et la pratique de cette science. Par conséquent, nos travaux ont porté sur l'étude de la mise en place d'une modélisation d'aide au diagnostic basée sur les approches du raisonnement à partir de cas (RàPC), la méthode collaborative, le raisonnement à travers les réseaux bayésiens et le langage de requête SPARQL.

Pour cela, nous avons commencé par présenter les notions de base de la médecine traditionnelle, les difficultés qui sont les siennes en termes d'apprentissage et de pratique, de même que les principes fondamentaux des approches qui entrent en ligne de compte. Il a été question ensuite de faire l'état de l'art des travaux se rapportant aux différents systèmes technologiques orientés vers la médecine traditionnelle, vers les ontologies ainsi que vers les différentes étapes du processus d'apprentissage et d'exercices de cette science.

Par la suite, nous avons présenté nos propositions concernant les méthodes d'acquisition des connaissances de la médecine traditionnelle et de ses pratiques.

Selon nos recherches sur l'apprentissage, plusieurs méthodes ont été développées. Malgré le succès de certaines d'entre elles, celles relatives à l'apprentissage à partir d'outils informatiques montrent leurs limites pratiques en raison de la faible capacité de ceux-ci. C'est ainsi que sont apparus les machines à états finis, les automates cellulaires et les machines de Turing pour contribuer à la résolution des difficultés rencontrées dans l'apprentissage.

La principale contribution de ce travail consiste à proposer des méthodes d'apprentissage et d'exercices de la MT à partir des ontologies dans le souci d'une valorisation de cette science. Notre contribution porte spécifiquement sur les points suivants :

1. Possibilité de poser le diagnostic d'une maladie et d'en proposer la thérapeutique par l'utilisation du modèle défini ;
2. Utilisation d'un mécanisme de gestion des conflits apparaissant au moment de la recherche d'une information médicale capitale et spécifique ;
3. Elaboration du traitement d'une nouvelle maladie à partir de cas déjà traités et guéris par le praticien dans le cadre de la médecine traditionnelle ;
4. Facilitation de la recherche d'une information médicale dans une base de connaissances appartenant à la médecine traditionnelle.

Ce mémoire de thèse est une base de travail à partir de laquelle de nouvelles activités de recherche peuvent être conduites afin d'en améliorer le contenu.

Les perspectives de recherche pourraient s'orienter dans les directions suivantes :

- S'agissant des réseaux bayésiens, la recherche pourrait porter sur la conduite de tests effectués sur des données réelles avec des maladies bien ciblées. Si l'approche proposée était validée sur le terrain, on pourrait alors cerner les contours d'un ensemble de problèmes pratiques dont notamment le consensus sur les signes cliniques, l'utilisation des plantes médicinales pour déterminer l'aptitude du praticien et la définition de l'échantillonnage à utiliser pour la collecte des données. Ces différentes applications pratiques pourront donner lieu à d'autres travaux de recherches.
- S'agissant de la gestion des conflits, les éléments sur lesquels nous nous sommes basés pour bâtir notre modèle sont relativement réduits ; la recherche pourrait se poursuivre par l'extension du nombre d'éléments informatifs afin de mesurer la fiabilité réelle du modèle.
- Il faut aussi noter que notre modèle sur le Raisonnement médical à Partir des Cas (RàPC), s'est essentiellement axé sur la problématique du développement d'un indice de similarité. Il pourrait donc être souhaitable, dans une perspective de recherche, que cette métrique puisse être confrontée à des données réelles afin d'en mesurer toute la portée. Ces données pourront permettre une analyse plus fine d'autres propriétés mathématiques éventuelles de cet indice de similarité. Il ne faut pas perdre de vue l'objectif final qui est de développer des systèmes de raisonnement à partir de cas dans le domaine de la médecine traditionnelle. Aussi, l'intégration de cette métrique aux outils permettant de construire des systèmes de RàPC est une problématique qui doit être envisagée.
- De par sa nature globale, l'utilisation de la mesure de similarité développée en dehors du cadre du raisonnement à partir de cas peut être aussi envisagée ; il s'agit notamment des domaines de la statistique multidimensionnelle, de l'apprentissage artificiel, de la fouille de données et plus particulièrement de la science des données dans laquelle la variété des types de données est une caractéristique fondamentale.
- Une recherche pourrait s'orienter vers l'utilisation d'un index sémantique sur les données de l'interface du système. On pourrait envisager la possibilité de mettre en place un système iconique permettant la génération automatique des expressions formelles du langage de requête SPARQL.

Nous n'avons pas la prétention, par les méthodes que nous proposons, de résoudre toutes les difficultés inhérentes à l'exercice et à la pratique de la MT. Notre contribution se limite modestement à une tentative de rationalisation, de structuration et de stockage de toutes les connaissances que renferme cette science aux fins d'en rendre la pratique plus aisée en la soustrayant de toutes notions subjectives.

Bibliographie

- [Abd 06] Abdel K. K., Catherine R., Robert L. (2006) “Un outil d’aide à la construction d’ontologies pré-consensuelles: le projet Towntology” *In INFORSID (pp. 911-926).*
- [Abde 06] Abderrazak M. (2006) “XML et travail collaboratif: vers un Web sémantique”. *Revue maghrébine de documentation et d’information, Vol. 16,- ISSN : 0330-9274, pp. 231-252, 2006, ISD, Tunis, 21 pages.*
- [Abdu 11] Abdul K., Edi L. N., Adhi S., Paulus I. S. (2011) “Leaf Classification Using Shape, Color, and Texture Features”, *International Journal of Computer Trends and Technology- July to Aug Issue 2011*
- [Adam 07] Adams A., Pineau F., Rivier A., Samyn C. (2009) “ Technologies du Web Sémantique pour l’Entreprise 2.0 ” *PhD thesis, Université Paris IV-Sorbonne.*
- [Add 13] Addour D. (2013) “Développement d’une plateforme pour la découverte de services web” (*Doctoral dissertation, Université M’hamed Bougara de Boumerdès.*)
- [Ahs 05] Ahsan-ul M., Ramanjit S. (March 2005) “Evaluation and Ranking of Ontology Construction Tools” *Technical Report DIT-05-013, University of Trento (March 2005)*
- [Akil 11] J. Akilandeswari, Subhashini, R. (2011, Jan.) “A Survey On Ontology Construction Methodologies” *International Journal of Enterprise Computing and Business Systems, Vol. 1(1), p.60-72.*
- [Alain 12] M. Sanou, A. Jean, Michel M., Dominique P., Yunsan Meas, Chantal Enguehard, Leila Moret, Augustin Emame (2012) “ Conventional medical attitudes to using a traditional medicine vodou-based model of pain management: survey of French dentists and the proposal of a pain model to facilitate integration.” *Journal of Chiropractic Humanities, Volume 19, Issue 1, December 2012, Pages 24-35*
- [ALJC 04] Alain Léger et Jean Charlet (2004) “Applications du web sémantique” *Revue I3s, http://www.irit.fr/journal-i3/hors_serie/annee2004/revue_i3_hs2004_01_08.pdf.*
- [Am 08] Ammar M., Xavier M., Christine G., and Bernard G. (2008) “A Hybrid System for the Semantic Annotation of Sulco-Gyral Anatomy in MRI Images” *MICCAI 2008, 11th International Conference, United States*
- [Ami 12] Amina Baghli (avr. 2012) “Diag-K: Ontologie pour l’élaboration du modèle de connaissances de l’apprenant dans les EIAH” <http://dSPACE.univ-tlemcen.dz/handle/112/336>
- [Anh 07] P. Viet Anh, HO Tuong Vinh (Juil. 2007) “Classification de documents à l’aide de cartes auto-organisatrices (SOM) basée sur une ontologie” *laboratoire TIMC-IMAG institut d’ingénierie de l’information de santé faculté de Médecine, 38700la Tronche-France.*
- [Ant 05] Antoine I. (2005) “Conception et utilisation d’ontologies pour l’indexation de documents audiovisuels” *Doctoral dissertation, Université de la Sorbonne Paris 4.*
- [Aria 10] Ariane A. K., Mels G., Choquet R., Charlet J., & Jaulent, M. C. (2010). “Une approche ontologique pour l’exploitation de données cliniques.” *In IC 2010 (pp. 183-194). Ecole des Mines d’Alès.*
- [Arian 10] Ariane A. K., Giovanni M., Rémy C., Jean C. et M. C. Jaulent: (2010) “Une approche ontologique pour l’exploitation de données cliniques”, *In IC 2010 (pp. 183-194). Ecole des Mines d’Alès*

- [Arm 12] Armel A., Souleymane K., Laure P. F., Balira O. Konfé (2012) “Using METHONTOLOGY to Build a Deep Ontology for African Traditional Medicine: First Steps” *IC, 2012*.
- [Asm 12] Asma B. A., Pierre Z. (Mars 2012) “Analyse et transformation des questions médicales en requêtes SPARQL”, *CORIA 2012, pp. 111–122, Bordeaux, 21-23 mars 2012*
- [Aud 07] M. Audet et R. Nkambou (2007) “Les services de collaboration à travers la grille d’apprentissage” *Doctoral dissertation, Université du Québec à Montréal département d’informatique (LGDAC)*.
- [Ayc 13] A. Aycan (2013) “l’enseignement du fle aux enfants assiste par le multimedia” *International Periodical for the Languages, Literature and History of Turkish or Turkic Vol. 8/10 p. 77-96, Ankara-Turkey*.
- [Bag 04] Baget J.F., Étienne C., Jérôme E. et Mohand S.H. (2004) “Les langages du Web sémantique”. *Revue I3, Hors Série 2004*
- [Bak 02] B. R. Baker. (1982, Sept.) “Minspeak A semantic compaction system that makes self-expression easier for communicatively disabled individuals” *Byte, vol.7, no. 9, p. 186-2002*
- [Batc 07] B. Batchakui, C. Tangha, R. Nkambou, G. Kouamou (2007 Nov.) “Environnement de collaboration basé sur le Grid Learning Services(GLS) pour les communautés de formation à centre d’intérêt commun.” *In 4th Annual Scientific Conference of the Research Network LORNET (pp. 4-7)*.
- [Bati 06] Batista F., Pardal J., P. Mamede, P. V. N., & Ribeiro, R. (2006, Feb.) “Ontology construction: cooking domain” *Artificial Intelligence: Methodology, Systems, and Applications Vol. 41, p.1-30*.
- [Ben 12] Ben Hamadi Olfa (Juil.2012) :“management des connaissances et mémoire organisationnelle: entre perdre connaissance et retrouver la mémoire. le cas d’hydro-quebec”, (*Doctoral dissertation, Université du Québec à Trois-Rivières*).
- [Ben 13] Benkherourou Chafika (11 sept 2013) “Le langage Human MI comme support de communication entre agents”, <http://dspace.univ-ouargla.dz/jspui/handle/123456789/592>
- [Bend 07] Bendaoud R., Hacene M. R., Toussaint Y., Delecroix B., & Napoli A. (Jul. 2007). “Construction d’une ontologie à partir d’un corpus de textes avec l’ACF”, *Dans Frankie Trichet, éditeur : Actes des 18èmes Journées francophones d’Ingénierie des Connaissances (IC 07) Cépaduès*,
- [Benj 09] Benjamin H.K.B. (29 Oct. 2009) “Indexation de documents pédagogiques: fusionner les approches du Web Sémantique et du Web Participatif” *Doctoral dissertation, Université Henri Poincaré-Nancy I*.
- [Bens 06] S. M. Benslimane, D. Benslimane, M. Malki, Y. Amghar, F. Gargouri (2006). “Construction d’une ontologie à partir d’une base de données relationnelle: approche dirigée par l’analyse des formulaires HTML” *Laboratoire EEDIS, Université Djillali Liabes de Sidi Bel-Abbes In INFORSID (Vol. 6, pp. 991-1010)*.
- [Berg 13] J. Bergeron (Déc. 2013) “Intégration de services de raisonnement automatique basés sur les logiques de description dans les applications d’entreprise”. <http://hdl.handle.net/1866/9101>
- [Berm 11] BERMES, Emmanuelle Centre G. Pompidou Paris, France (Mai 2011) “Publier les référentiels sur le Web sémantique Tutoriel”; *Journée AFNOR / BnF Mai 2011*.

- [Bern 01] Berners-Lee, J. Hendler, and O. Lassila (2001, May) “The Semantic Web: A New Form of Web Content That Is Meaningful to Computers Will Unleash a Revolution of New Possibilities” *Scientific Am.*, vol. 284, no. 5, pp 28-37.
- [Bern 01] Berners-L. T., Hendler J. and Lasilla O., (May, 2001) “the Semantic Web”, *Scientific American*, 284(5) P 34-43.
- [Bess 89] M. N. Bessagnet, M.F. Canut (1989) “sur la conception de tuteurs intelligents” *Bulletin de l'EPI (Enseignement Public et Informatique)*, (56), p.95-110.
- [Blon 04] Blondel X., Rambuteau D. L, & letourmy. (15 oct. 2004) "Les systèmes de gestion de contenu: description, classification et évaluation" *Mémoire d'Ingénieur C.N.A.M. Paris*,
- [Bon 83] Bonnin, A., Pallo J. M. (1983). “A-transformation dans les arbres n-aires.” *Discrete Mathematics*, Vol. 45(2), p.153-163. *Dijon, France*
- [Bour 02] D. Bourigault, Nathalie A.G., Jean C. (2002) “Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas”, *Revue d'Intelligence Artificielle. Volume X – n° X*, pages 1 à X.
- [Bout 04] Sabri Boutemedjet (10 févr. 2004). “Web Sémantique et e-Learning”, *Université de Montréal*, <http://www.iro.umontreal.ca/~aimeur/cours/ift6261/Presentations-etudiants/WebSemantiqueE-learning.pdf>. Hiver2014.
- [Brou 10] M. K. Brou, Ibrahim LOKPO, Tra GOORE BI (2010) “Medtrad+ : un système expert pour la MEDecine TRADitionnelle”, *CARI, yamoussoukro*, P.381-388.
- [Brou 97] M. K. Brou (20 Nov. 1997) “base de concepts: contribution à la représentation et à l’exploitation hypertexte de concept -le système CodBD-Web” *PhD Thesis of University of Pau, France*,
- [Bruil 03] Bruillard E. (2003) “Enseignement sur mesure, Aspects: EAO, Tuteurs, micromondes et hypertextes ” <http://www.edusud.org/ressources/ntic/esm>
- [Caun 00] Caune J., (2000) “La médiation culturelle: une construction du lien social”, *Les enjeux de l'information et de la communication, en ligne*, <http://w3.ugrenoble3.fr/les_enjeux/2000/Caune/index.php>.
- [Ceus 05] Ceusters W., Smith B., Goldberg L. (2005) “A terminological and ontological analysis of the NCI thesaurus” *Methods of information in medicine* Vol. 44(4), p.498–507
- [Chan 04] Chan Le Duc. (2004)“Transformation d'ontologies basées sur la logique de description : Application dans le commerce électronique” *Human-Computer Interaction. Université Nice Sophia Antipolis. French*. <tel-00214145>
- [Chang 07] Chang B., Ham D.H., Moon D. S., Choi Y. S., Cha J. (2007) “Using Ontologies to Search Learning Resources.” *Computational Science and Its Applications – ICCSA 2007, LNCS 4705, Springer Berlin / Heidelberg*, p. 1146-1159
- [Chus 00] S. Chusseau, A.P. G. Maquinghen, F.X. Lepoutre, D. Chopin (2000) “Sadépo : remémoration de cas pour l’aide à la décision en chirurgie correctrice de la scoliose ” *ITBM-RBM; Editions scientifiques et médicales Elsevier SAS. Tous droits rCservCs* Vol. 21 : p.78-83
- [Cim 08] Cimiano P., Haase P., Heizmann J., Mantel M., Studer R (2008) “Towards portable natural language interfaces to knowledge bases: The Case of the ORAKEL system”, *Data Knowl. Eng.*, 65(2):325–354, p. 325-354.

- [Cim 08] Cimiano P., Haase P., Heizmann J., Mantel M., Studer R. (2008) “Towards portable natural language interfaces to knowledge bases: The Case of the ORAKEL system”, *Data Knowledge Engineering (DKE)*, 65(2), p. 325-354.
- [Coj 11] Julien Cojan. (2011) “Application de la théorie de la révision des connaissances au raisonnement à partir de cas”. *Artificial Intelligence. Université Henri Poincaré - Nancy I, French*.
- [Coul 08] Coulet A. (2008). “Construction et utilisation d'une base de connaissances pharmacogénomique pour l'intégration de données et la découverte de connaissances” (*Doctoral dissertation, Université Henri Poincaré-Nancy I*).
- [Daou 08] M. Daoud, Tamine L., L., Boughanem M., & Bilal C. (2008, March). “Construction des profils utilisateurs à base d'une ontologie pour une recherche d'information personnalisée”. *In francophone en Recherche d'Information et Applications (CORIA 2008)*.
- [Den 98] Denis M., Brigitte LE P. et Jonathan Y. (1998) “ La prédiction lexicale et syntaxique à partir du contexte gauche: Application au système HandiAS” *Revue Informatique et Statistique dans les Sciences Humaines, vol. 33*.
- [Der 08] Derivaux S., Forestier G., Wemmert C., & Lefèvre S. (2008). “Extraction de détecteurs d'objets urbains à partir d'une ontologie. Atelier Extraction de Connaissance à partir d'Images (ECOI)”, *Journées Francophones Extraction et Gestion des Connaissances (EGC 2008), Sophia Antipolis, France, p.71-81*.
- [Des 05] Dessus P., Peraya D. (2005). “Le rôle des connaissances dans l'identification automatique d'icônes: une comparaison avec des humains [The effect of knowledge in the automatic recognition of icons: a comparison with humans]” *Revue d'Intelligence Artificielle, vol 19, p 195-214*.
- [Dial 11] Diala Dhoub (2011) “Aide multicritère au pilotage d'un processus basée sur le raisonnement à partir de cas”, *Engineering Sciences. Université Paris VIII Vincennes-Saint Denis, French*
- [Dja 05] C. Djaiz, N. Matta, O. Corby, W. Léone, W. Lesguillier (May 2005) “Flexibilité de la recherche d'information dans les projets de conception” ; *In Proc. of the Workshop Reasonner le web sémantique avec des graphes, AFIA Platform, Nice*.
- [Dom 00] Dominique Pastre (2000) “l'intelligence artificielle définition - généralités - historique - domaines” *Université Paris 5 vol. 1, p. 973-979*.
- [Dome 10] J.C. Domenget, I. Szilagy, F. Greffier (2010) “Web sémantique et système actif d'apprentissage: apports d'un wiki sémantique”; *Université de Franche-Comté*.
- [Duch 15] M. Duchera, C. M. Véhierd, P. Lantelmea, B. Vaissef, J. P. Baguetg, J. P. Fauvel (May 2015) “ Fiabilité d'un réseau bayésien pour prédire un rapport aldostérone sur rénine élevé” *Archives of Cardiovascular Diseases, Volume 108, Issue 5, , Pages 293–299*
- [Dur 98] A. Durey, D. Beaufls (1998). “L'ordinateur dans l'enseignement des sciences physiques : questions de didactique” *Actes des 8ème Journées Nationales Informatique et pédagogie des Sciences Physiques, Udp et INRP, p.63-74*.
- [Fab 06] Fabien Gandon (2006) “Le Web sémantique n'est pas antisocial” *In IC-17èmes Journées francophones d'Ingénierie des Connaissances (pp. 131-140)*.
- [Fab 08] Fabien L. Gandon (5 Nov. 2008) “Graphes RDF et leur Manipulation pour la Gestion de Connaissances” *Habilitation A dirigé des recherches, University of Nice - Sophia Antipolis*

- [Fay 01] Faye S. N. (2001). "Médecine traditionnelle et dynamiques interculturelles: les implications socio-anthropologiques de la formalisation de la tradithérapie à l'Hôpital Traditionnel" de Keur Massar". *Mémoire de Maitrise, Université Gaston Berger, Saint Louis, SN. 124p.*
- [Far 12] Mohammed Farouk Bouaziz. (2012) "Contribution à la modélisation Bayésienne de l'état de santé d'un système complexe : application à l'industrie du semi-conducteur." *Other. Université de Grenoble, French. <NNT: 2012GRENT109>. <tel-00993732>*
- [Flor 07] Florence Amardeilh (Mai 2007) "Web Sémantique et Informatique Linguistique : propositions méthodologiques et réalisation d'une plateforme logicielle" *Thèse de doctorat. Discipline : Informatique. Université Paris X – Nanterre, Mai 2007*
- [Fon 09] Fontaine R., Patrice L. et Alexandre A. (2009) "Analyse des Performances et Modélisation d'un Serveur Web". *Proc. 5th International IEEE of Science of Electronics, Technologies of Information and Telecommunication, Hammamet, Tunisia Mars 2009, ISBN 978-9973-0-0122-1.*
- [Fran 03] Franz Baader, Werner Nutt (2003, January) "Basic Description Logics". *In Description logic handbook (pp. 43-95).*
- [Fran 10] François Y. et Akim Demaille (Janv. 2010) "Théorie des langages Notes de cours" *EPITA Research and Development Laboratory, EPITA.*
- [Fras 86] C. Frasson M. Erradi (1986, June) "An icon-based language for applications in Medicine." *Université of Montréal, département d'informatique et de recherche opérationnelle Vol. 15 Issue 2, p. 144-152.*
- [Gab 11] Gabriel M. Torres, Alexandre L., Vitor R., Rafael P. da Rocha, Mara A. (2011) "Collaborative Construction of Visual Domain Ontologies Using Metadata Based on Foundational Ontologies" *In ONTOBRAS-MOST (p. 201-206).*
- [Gand 10] Gandon, F. L. (2010). "Software Components". *Ontologies in Computer Science: These New Ontology Theory, Management and Design: Advanced Tools and Models: Advanced Tools and Models, P.1.*
- [Gar 06] Él Garrot (2006, May) "Un système pour conseiller le tuteur de situations d'apprentissage collaboratives" *In Rencontres Jeunes Chercheurs en EIAH (RJC-EIAH 2006) (pp. p-99).*
- [Ghis 07] A. Ghislain., Pavón, J. (2007). "Intelligent Environment for Medical Practices" *in African Traditional Medicine. In 6th International Workshop on Practical Applications of Agents and Multiagent Systems (IWPAAMS 2007), Salamanca, Spain (pp. 101-108).*
- [Ghis 08] Ghislain A., Juan P. (Oct. 2008) "An Ontology for African Traditional Medicine" *International Symposium on Distributed Computing and Artificial Intelligence (DCAI'08): Series: Advances in Intelligent and Soft Computing, Vol. 50 P. 329-337. Salamanca, Spain,*
- [Gir 95] Girin, J (1995) "Les agencements organisationnels, les savoirs en action." *(Sous la dir. de F. Charue-Duboc) (pp.234-294).*
- [Golb 08] C. Golbreich, O. Dameron, B. Gibaud, A. Burgun (2003) "Comment représenter les ontologies pour un Web Sémantique Médical ?" *In Journées Francophones de la Toile (JFT'2003), Tours,*
- [Grig 08] Grigoris A., Frank V. H. (2008) "La vision du web sémantique", *Mit Press,*

- [Guy 07] Guy Politzer (2007) “La logique, le langage et les variétés du raisonnement” *S. Rossi et J.-B. Van der Henst (éds), Psychologies du raisonnement, Bruxelles, De Boeck, p. 9-37.*
- [Hab 04] Habert, B., Illouz, G., & Folch, H. (2004). “Dégrouper les sens: pourquoi, comment.” *Actes de JADT, 1, p. 565-576.*
- [Hair 00] S. L'haire (Fév. 2000). “L’Enseignement Assisté par Ordinateur et le Traitement Automatique du Langage Naturel” *mémoire de diplôme d’études supérieures linguistique informatique, Université de Genève.*
- [Haj 12] A. Hajjam A. A. Benyahia, E. Andres, S. Erve I. (2012) “e-Care: Vers une interopérabilité des ontologies pour le monitoring des maladies chroniques” *Université d’été de la e-Santé, Castres.*
- [Harb 07] K. Harbouche, M. Djoudi (2007, Nov.). “Assistant virtuel à base d’agents pour un enseignement à distance”. *In 4th International Conference on Computer Integrated Manufacturing CIP (pp. 3-4).*
- [Hari 11] Harish, D.V.N, Y. Srinivas, K.N.V.S.S.K Rajesh and P. Anu radha (2011) “Image annotations using machine learning and features of ID3 algorithm.” *Int. J. Comput. Appli, Vol 25: n°5 p. 0975-8887.*
- [Herm 10] Hermes, D. Andler (2004) “Logique, Raisonnement Et Psychologie” *Introduction aux sciences cognitives, p.315-405.*
- [Herm 12] A. Hermann, Sébastien F. et Mireille D. (2012) “Aide à la création d’objets dans une base RDFS avec des règles de relaxation” *In Journées francophones d’ingénierie des connaissances p 301-316.*
- [Hub 09] G. Hubert, J. Mothe, B. Ralalason, B. Ramamonjisoa (2009, May) “Modèle d’indexation dynamique à base d’ontologies” *In CORIA, P.169-184.*
- [Ibra 14] Ibrahim EL Bitar (2004) “CBR4WSD : Une approche de découverte de services Web par Raisonnement à Partir de Cas ”, *Artificial Intelligence. Ecole Mohammadia d’Ingénieurs - Université Mohammed V de Rabat - Maroc. French*
- [Iva 08] Ivana Rasovska (20 Féb. 2008) “Contribution à une méthodologie de capitalisation des connaissances basée sur le raisonnement à partir de cas: Application au diagnostic dans une plateforme d’e-maintenance.”, *HAL <https://tel.archives-ouvertes.fr/tel-00257893>*
- [Jean 01] Jean-P. C., H. Haddad (2001) “Proposition d’un modèle relationnel d’indexation syntagmatique: mise en œuvre dans le système iota” *Laboratoire CLIPS-IMAG Grenoble Cedex 9, France In INFORSID (p. 465).*
- [Jean 03] Jean Charlet, Philippe Laublet & Chantal Reynaud (Déc. 2003) “le Web sémantique” *Rapport final. Action spécifique 32 CNRS / STIC. J. Charlet, P. La ublet et C. Reynaud. p. 143. 2003*
- [Jean 06] Lamy Jean. B. (2006). “Conception et évaluation de méthodes de visualisation des connaissances médicales: mise au point d’un langage graphique et application aux connaissances sur le médicament” *Doctoral dissertation, Paris 6.*
- [Jer 10] Jérôme Champavère (10 Mars 2010) “De la représentation des connaissances au Web sémantique” *http://www.grappa.univ_lille3.fr/~champavere/Enseignement/0809/l3miashs/ia/rc-ws.pdf.*

- [Jon 12] Jonathan M. M., Matthew H., Mark A. M., and Natalya F. N. (2012) “Modest Use of Ontology Design Patterns in a Repository of Biomedical Ontologies”, *Stanford Center for Biomedical Informatics Research Stanford University, Stanford CA 94305, USA In WOP*
- [Jon 98] D. Jones, Trevor Bench-Capon and Pepijn V. (1998). “methodologies for ontology development” In *Proceedings of IT&KNOWS Conference of the 15th IFIP World Computer Congress* (Chapman-Hall), pp, 20-35.
- [Jos 00] José Rouillard (11 Jan. 2000) “hyperdialogue sur internet le système HALPIN”. *thèse de doctorat de l'Université Joseph Fourier, CLIPS-IMAG*
- [Jos 92] Joseph William Yoder, (1992) “the role of human-computer interaction in medical information systems: principles and implementation of MEDIGATE” Unpublished Master Thesis, University of Illinois at Urbana-Champaign.
- [Juno 07] Junod, A.F., (2007) *Décision médicale ou la quête de l'explicité. Seconde ed.*, Genève: *Médecine et Hygiène. 280.*
- [Kaz 13] O. Kazar (2013) “Un système multi-agents pour L'E.I.A.O”. <http://dspace.univ-biskra.dz:8080/jspui/handle/123456789/386>
- [Kob 10] Kobangue L., Mozouloua D. (31 Août 2010) “La décennie de la médecine traditionnelle, progrès et perspectives”. *conference-debat de la 8eme journee africaine de medecine traditionnelle*
- [Kof 09] N.L.Koffi, E. B. Assoumou (2009) “conception et réalisation d’une interface iconique pour Medtrad”, *rapport de projet 2e année ingénieur INP-HB, 2009*
- [Laub 04] Laublet, P., Charlet, J., & Reynaud, C. (2004). “ Introduction au web sémantique. ” *Revue Information, Interaction, Intelligence I3.*
- [Leg 10] Legendre V., Petitjean G. & Lepatre T. (2010). “Gestion des règles « métier » ”. *Génie logiciel, 92, 43–52. vii, 51, 54*
- [Ler 06] Philippe Leray, (2006) “Réseaux Bayésiens : apprentissage et modélisation de systèmes complexes, HDR” *Dans Soutenance Habilitation à Diriger les Recherches.*
- [Ler 08] Philippe Leray. (2008) “Réseaux Bayésiens, Introduction et apprentissage, Modélisation et découverte de Connaissances. Tutoriel” *8èmes Journées Francophones Extraction et Gestion des Connaissances. Sophia Antipolis.*
- [Mag 13] Magesh and Thangaraj (2013) “comparing the performance of semantic image retrieval using sparql query, decision tree algorithm and lire”, *Journal of Computer Science Vol 9 (8): p.1041-1050*
- [Man 07] Manuel Zacklad (2007, May). “Classification, thésaurus, ontologies, folksonomies: comparaisons du point de vue de la recherche ouverte d’information (ROI)”, In: *CAIS/ACSI 2007, 35e Congrès annuel de l'Association Canadienne des Sciences de l'Information. Partage de l'information dans un monde fragmenté: Franchir les frontières, sous la dir. de C. Arsenault et K. Dalkir. Montréal: CAIS/ACSI, 2007 Université de Technologie de Troyes.*
- [Mart 12] S. Martin., Jean A., Marjolet M., Pécaud D., Meas Y., Enguehard C., & Emame, A. (2012). “Conventional medical attitudes to using a traditional medicine vodou-based model of pain management: survey of French dentists and the proposal of a pain model to facilitate integration”. *Journal of chiropractic humanities, 19(1), p.24-35.*
- [Mest 07] Mestiri, Mohamed Amine (2007-07) “Vers une approche web sémantique dans les applications de gestion de conférences” *Doctoral dissertation, Université Laval.*

- [Mich 09] Michel H., Gilbert P., Josianne B. (2009, April) “Méthodologie assistée de conception d’une ontologie à partir d’une conceptualisation consensuelle semi-formelle” *In IC 2009: 20es Journées Francophones d’Ingénierie des Connaissances* (p. 61).
- [Mich 98] Michel Léchère (1998) “Raisonnement avec des définitions de types dans le modèle des graphes conceptuels” *Revue d’Intelligence Artificielle*, 12(2):P. 243-278. Hermès
- [Moh 02] Mohamed M. M. (2002) “le rôle de la logique floue dans le web sémantique” *Secrétariat du DEA d’Informatique Université de la Méditerranée Faculté des Sciences de Luminy, MARSEILLE*.
- [Moh 05] Mahmoud, Y., Mianpeurem, T., Dolmazon, R., Bouchu D., & Fenet B. (2005). “Amphiphile triterpénoïdes from *Hymenocardia acida* Tul. Phytoantimalarial and anti-inflammatory activities?”. *Curr. Top. Phytochem*, 7, p.61-66.
- [Moh 09] Mohamed A. Touré, M. Mbangwana, Rocare, P. A. Sène, Ineade R. S. (2009) “Que sont les TIC: Typologies des outils et systèmes” *Intégration pédagogique des TIC: Stratégies d’action et pistes de réflexion*, p.33-56.
- [Moh 10] Mohamed Karim Haouchine (24 Mar. 2010) “Remémoration guidée par l’adaptation et maintenance des systèmes de diagnostic industriel par l’approche du raisonnement à partir de cas.”, <https://tel.archives-ouvertes.fr/tel-00466560>
- [Mot 07] Mothe J., & Tanguy L. (2007). “Unités d’indexation et taille des requêtes pour la recherche d’information en français” *In Congrès Informatique des Organisations et Systèmes d’Information et de Décision (INFORSID 2007)* (pp. 225-241).
- [Mou 09] Mouna K., Nathalie A.-G. (2009, May) “Construction automatique d’ontologies à partir de spécifications de bases de données” *In IC 2009: 20es Journées Francophones d’Ingénierie des Connaissances «Connaissance et communautés en ligne»* (pp. 85-96).
- [Nad 03] C. Nadine, P. Christine, S. Spaccapietra, C. Vangenot (2003) “Des SIG aux ontologies géographiques”, *Revue internationale de géomatique*, Vol 0-n°0 13(3), pp. 285-306
- [Nai 07] P. Naim, P. Leray, et al, (2007) “Réseaux Bayésiens” 3ème édition, Eyrolle.
- [Nat 06] Nathalie N. C. Kuicheu. (2006, June) “Système de Communication Iconique: Application MEDITRA”. *Master's Thesis, The University of Yaounde I, Yaounde, Cameroon*.
- [Nath 05] Nathalie H. (6 Déc. 2005) “Ontologies de domaine pour la modélisation du contexte en recherche d’information” *Thèse de doctorat, Institut de Recherche en Informatique de Toulouse*.
- [Nen 05] Nendaz, M., et al., (2005) “Le raisonnement clinique: données issues de la recherche et implications pour l’enseignement”. *Pédagogie Médicale*, Vol 6: p. 235-254.
- [Nge 15] Betti J. L., & Lejoly J. (1999). “Importance en médecine traditionnelle de *Combretum mucronatum* Shum. & Thon (Combretaceae) dans le Dja (Cameroun).” *La Gestion des Forêts Denses Africaines Aujourd’hui. Original submitted in on 13th February 2015. Published online at www.m.elewa.org on 30th April 2015* <http://dx.doi.org/10.4314/jab.v8i1.6>
- [Nic 88] Nicaud J.F., Vivet M. (1988) “les Tuteurs Intelligents: Réalisation Et Tendances De Recherches” *TSI, numéro spécial: Application De L’informatique A La Formation, Dunod-Afcet, Vol.7, n°1, p.21-45*.

- [Niv 13] Pierre-Yves Nivollet (29 Nov. 2013) “Un Habitat Capable De Reconnaître Les Activités Planifiées Dans Un Calendrier Electronique” <http://savoirs.usherbrooke.ca/handle/11143/46>
- [Noel 10] M. Noëlle, Gaio, Eric Kergosien (2010) “Extraction de termes, reconnaissance et labellisation de relations dans un thésaurus Vers une ontologie”, *Faculté des Sciences, Département Informatique, 64000 PAU arXiv preprint arXiv:1002.0215*.
- [Oliv 13] Olivier Steichen (Juin, 2013). “Utilisation d’une observation médicale informatisée à d’autres fins que les soins Ingénierie des connaissances, évaluation des pratiques, recherche clinique” *Doctoral dissertation, Université Pierre et Marie Curie-Paris VI*.
- [Om 13] Omotosho, L. O., Odejebi, O. A., & C. O Akanbi. (March 2013). “A Framework For Knowledge-Based Ontology Model In African Traditional Medicine”. *African Journal of Computing & ICT, Vol 6 No. 1, P.43-48*.
- [Our 11] R.B. Ourdia (12 Déc. 2011) “Accès contextuel à l’information dans un environnement mobile: approche basée sur l’utilisation d’un profil situationnel de l’utilisateur et d’un profil de localisation des requêtes” (*Doctoral dissertation, Université de Toulouse, Université Toulouse III-Paul Sabatier*).
- [Paq 03] G. Paquette, J. Bourdeau, F. Henri, J. Basque, M. Leonard, M. Maina (2003): “Construction d’une base de connaissances et d’une banque de ressources pour le domaine du téléapprentissage” *Revue des Sciences et Technologies de l’Information et de la Communication pour l’Education et la Formation, Vol 10*.
- [Paul 12] Paula P. R., D. Noyes, K. F. Bernard (Jun 2012) “Raisonnement collaboratif à partir de cas dans la résolution de problèmes en maintenance”, *9th International Conference on Modeling, Optimization et SIMulation, french*.
- [Pha 05] Phan Quang Trung Tien Promotion X, (Juil. 2005) “Ontologies et Web Services” *Hanoi*,
- [Phil 02] Philippe L., Chantal R., Jean C. (2002) “ Sur quelques aspects du Web sémantique”, *Actes des deuxièmes assises du GdR 13, pp 59-78, Cépaduès Editions*
- [Pier 11] Pierre-Yves Vandebussche (2011) “Définition d’un cadre formel de représentation des systèmes d’organisation de la connaissance” (*Doctoral dissertation, Université Pierre et Marie Curie-Paris VI*).
- [Pop 03] Popescu A, Etzioni O, Kautz H (2003) “towards a theory of natural language interfaces to databases”, *Proceedings of the International Conference on Intelligent User Interfaces (IUI’03), p. 149-157*.
- [Pra 13] C. Pradel, Ollivier H., Nathalie H. (Juil. 2013) “Passage de la langue naturelle à une requête SPARQL dans le système SWIP” *IC – 24 émes Journées francophones d’Ingénierie des Connaissances, Lille, France*.
- [Priv 10] Priya Shetty (juin 2010) “Place de la médecine traditionnelle dans le système de santé: Faits et chiffres” <http://www.scidev.net/afrique-sub-saharienne/maladie/article-de-fond/place-de-la-m-decine-traditionnelle-dans-le-syst-me-de-sant-faits-et-chiffres.html>
- [Psy 03] Psyché V., Mendes O., Bourdeau J.: (2003) “Apport de l’Ingénierie Ontologique aux Environnements de Formation à Distance.” *Revue des Sciences et Technologies de l’Information et de la Communication pour l’Education et la Formation STICEF Vol. 10, p.89-126*

- [Psy 04] Psyché, V., Nkambou, R., & Bourdeau, J. (2004) “Proposition d'une méthode d'ingénierie ontologique pour les EIAH: application aux systèmes auteurs”, {vpsyche@liceftelug.quebec.ca}
- [Rad 09] Radja Messai (09 Nov. 2009) “Ontologies et services aux patients: Application à la reformulation des requêtes ” *Doctoral dissertation, Université Joseph Fourier – Grenoble I, French.*
- [Raim 13] Raimbault T., Maillot P. (2013). “ Vues d'ensembles de documents RDF ”. In *INFORSID* (pp. 387-402).
- [Ram 04] Ramzi Abbes (13 déc. 2004) “la conception et la réalisation d'un concordancier électronique pour l'arabe” *Doctoral dissertation, Thèse de Doctorat, Institut national des sciences appliquées de Lyon, Lyon, France.*
- [Rena 05] D Renaudie. (2005) “ Méthodes d'apprentissage automatique pour la modélisation de l'élève en algèbre ” *PhD thesis, Institut National Polytechnique de Grenoble.*
- [Rok 07] Rokia B., Mohamed R. H., Yannick T., Bertrand D., Amedeo N. (2007) “Construction d'une ontologie à partir d'un corpus de textes avec l'ACF” *Actes IC 07.*
- [Rol 03] D.-G. et Rolland D. (2003) “La nouvelle économie. Où? Quoi? Comment?” *Québec: Presses de l'Université du Québec Sainte-Foy, Laval.*
- [Rwa 93] Rwangabo, P. C. (1993). “La médecine traditionnelle au Rwanda”. *KARTHALA Editions.*
- [Sab 04] Sabri Boutemedjet (Hiver 2004) “ Web Sémantique et e-Learning” *Université de Montréal, (Canada).*
- [Sah 02] Sahbi Sidhom (2002) “Plate-forme d'analyse morpho-syntaxique pour l'indexation automatique et la recherche d'information : de l'écrit vers la gestion des connaissances ”. *Human-Computer Interaction. Université Claude Bernard - Lyon I, French.*
- [Sal 09] Salem Yassine, K Louisa, S Mohamed (2009) “Conception d'un Tuteur Intelligent en langage naturel français pour l'enseignement de la modélisation conceptuelle des données d'un système d'information” *Contemporary Engineering Sciences, Vol. 2, no. 4, p.191 - 208.*
- [Sbi 09] Sbihi, B. (2009). “ Web 2+: vers une nouvelle version du web 2.0”. *Journal of Information and Communication Technologies, [Online], 35(2), p.12-24.*
- [Seb 08] Sébastien D., Germain F., Cédric W. et Sébastien L. (2008) “Extraction de détecteurs d'objets urbains à partir d'une ontologie” *Atelier Extraction de Connaissance à partir d'Images (ECOI), Journées Francophones Extraction et Gestion des Connaissances (EGC 2008), Sophia Antipolis, France, p.71-81*
- [Seg 09] N. B Seghir, O. Kazar (2009) “ Une Architecture Basée Agents Mobiles Pour la Recherche D'information dans des Sources Hétérogènes et Réparties” in *Proc. 2nd Conférence Internationale sur l'Informatique et ses Applications, vol. 547.*
- [Sin 04] Sinuhé A., Rubén L., Ying D., Michael S., Dieter F.: (2004) “semantic web languages – strengths and weakness” In: *International Conference in Applied computing .*
- [Syl 13] Sylvie R. (11 juil. 2013) “Les ontologies comme support à l'interaction et à la personnalisation dans un processus décisionnel. Exploitation de la sémantique pour favoriser l'automatisation cognitive” *Doctoral dissertation, Université Montpellier II-Sciences et Techniques du Languedoc.*

- [Tan 04] Tanasa D., Trousse B., Masségli F. et AxIS P. (2004). “ Application des techniques de fouille de données aux logs web: Etat de l’art sur le Web Usage Mining ”. Dans Eric Guichard editor, *Mesures de l'Internet*, PP 126-143, Les Canadiens en Europe.
- [Tcho 02] Pierre Tchounikine (Déc. 2002) “Quelques éléments sur la conception et l'ingénierie des EIAH” In *Actes des 2ème assises nationales du GdR I3-Groupe de Recherche Information Interaction Intelligence*, pp. 13.
- [Teb 00] Tebourbi N. (2000). “L'apprentissage organisationnel: penser l'organisation comme processus de gestion des connaissances et de développement des théories d'usage” *Note de recherche de la Chaire Bell en Technologies et organisation du travail, sous la direction de D.G. Tremblay, Télé-Université, Université du Québec*
- [Thag 97] P. Thagard (1997) “The Concept of Disease: Structure and Change” *Philosophy Department University of Waterloo Waterloo, Ontario, N2L 3G1 communication and cognition, Vol. 29, P 445-478*
- [Thom 12] Thomas Cerqueus (15 Nov. 2012) “contribution au problème d’hétérogénéité sémantique dans les systèmes pair-à-pair: application à la recherche d’information” *Doctoral dissertation, Université de Nantes LINA*
- [Ton 06] Tonio Wandmacher, Jean-Yves Antoine (2006) “Training language models without appropriate language resources: experiments with an AAC system for disabled people”, *Proc. 5th European Conference on Language Resources and Evaluation (LREC')*
- [Tous 14] B.M. Toussaint, V. Luengo, L. Vadcard et J. Tonetti (27 déc. 2014) “Apprentissage de la chirurgie orthopédique assisté par ordinateur. Le cas du Système Tutoriel Intelligent TELEOS” *Field Actions Science Reports. The journal of field actions, Special Issue vol.9.*
- [Tra 12] Traore Issa (19 Jul 2012) “recherche d’information multimédia dans des documents semi-structures: cas des archives médicales”, *Doctoral dissertation université cheikh ANTA DIOP de Dakar.*
- [Tric 00] F. Trichet, M. Leclère, C. Choquet (IC’2000) “Construire un Système à Base de Connaissances de type Tâche/Méthode à l’aide des Graphes Conceptuels” *Actes des Journées Francophones d’Ingénierie des Connaissances P. 13-21, Toulouse*
- [Trus 12] Jean-Luc Trussart, conseiller technopédagogique Cégep régional de Lanaudière à L’Assomption (15 Oct 2012) “Wolfram|Alpha, le Web sémantique à son meilleur” *bulletin collegial des technologies de l’information et de la communication, <http://clic.ntic.org/cgi-bin/aff.pl?page=article&id=2263>*
- [Vis 11] Ubbo Visser¹, Saminda Abeyruwan¹, Uma Vempati², Robin P Smith², Vance Lemmon²³ and Stephan C Schürer²⁴: (2011) "BioAssay Ontology (BAO) a semantic description of bioassays and high-throughput screening results" *BMC Bioinformatics Vol. 12, p.257*
- [Wach 01] H. Wache, T. Vogele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann and S. Hubner (2001, August) “Integration of ontology-based information-A Survey of Existing Approaches” In *IJCAI-01 workshop: ontologies and information sharing (Vol. 2001, pp. 108-117).*
- [Xav 05] Xavier Lacot (Juin 2005) “Introduction à OWL, un langage XML d'ontologies Web” *DOI: http://lacot.org/public/introduction_a_owl.pdf*.
- [Xuez 04] Xuezhong Z., Zhaohui W., Aining Y., Lancheng W., Weiyu Fa., Ruen Z. (2004) “Ontology Development for Unified Traditional Chinese Medical Language System” *Artificial Intelligence in Medicine Vol. 32 p.15-27.*

Annexes

Annexe A : Exemple d'Application de Gestion de Conflit

```
16 %n=nombre de symptomes
17 %t=nombre de maladies
18 %m=nombre d'echantillons
19
20 n =20;
21 t=10;
22 m=1682;
23
24 R=zeros(n,t);
25 Z=zeros(m,1);
26 X=zeros(m,n);
27 Yinit=zeros(1682,943);
28
29 %%%%%%. création de la matrice y
30 Dstore = load('Donnees.txt');
31 idSymp = Dstore(:,1);
32 idmal = Dstore(:,2);
33 rating = Dstore(:,3);
34 N=size(idSymp,1);
35 for k=1:N
36     Yinit(idmal(k),idSymp(k))= rating(k);
37 end;
38 Y=Yinit(:,1:20);
39 disp(Y(1:10,1:10));
40 %%%%%%. determination des 50 cantroides correspondants aux 50 maladies
41 [IDX, Centroids]=kmeans(Y,10,'replicates',10);
42 %disp('=====Matrice des maladies====');
43 Maladie=round(Centroids);
44 %disp(Maladie)
```

Annexe A.1 : Présentation du code source de l'implémentation

Command Window

-0.2573	3.0503	0.7695	1.0233	1.8134	-1.8455	0.9404	0.8272	0.8227
0.4910	1.8336	-0.0347	0.3570	0.3585	-1.2568	0.5990	0.6965	3.2644
-0.0560	-0.0997	-0.3582	1.0211	2.1956	0.5321	-0.0381	0.9101	0.2423
-1.2864	1.5872	0.0489	1.7678	1.4829	1.1664	0.5537	0.8073	0.0909
0.8214	2.1440	-1.2413	0.2476	-0.8579	1.1114	0.8870	0.4588	1.1437
0.5931	2.9366	0.6092	0.3345	-0.9754	2.1797	0.0412	0.7609	-0.8532
0.5915	4.2501	0.9138	0.8335	1.2131	-1.1593	0.5059	-0.8433	1.6657

Annexe A.2 : Le code de la vérification des symptômes

```

Valeurs attendues pour les données manquantes:
Valeur prevue 3.8449 pour le symptome 1 valeur en entrée 4
Valeur prevue 1.9291 pour le symptome 2 valeur en entrée 0
Valeur prevue 0.33855 pour le symptome 3 valeur en entrée 0
Valeur prevue 2.377 pour le symptome 4 valeur en entrée 0
Valeur prevue 0.82819 pour le symptome 5 valeur en entrée 0
Valeur prevue 2.4922 pour le symptome 6 valeur en entrée 2
Valeur prevue 4.9133 pour le symptome 7 valeur en entrée 5
Valeur prevue 2.2562 pour le symptome 8 valeur en entrée 3
Valeur prevue 1.1745 pour le symptome 9 valeur en entrée 0
Valeur prevue 3.6278 pour le symptome 10 valeur en entrée 0
Valeur prevue 3.2778 pour le symptome 11 valeur en entrée 0
Valeur prevue 1.8544 pour le symptome 12 valeur en entrée 0
Valeur prevue 4.1276 pour le symptome 13 valeur en entrée 0
Valeur prevue 3.5883 pour le symptome 14 valeur en entrée 0
Valeur prevue 2.5007 pour le symptome 15 valeur en entrée 0
Valeur prevue 2.481 pour le symptome 16 valeur en entrée 0
Valeur prevue 1.3149 pour le symptome 17 valeur en entrée 0
Valeur prevue 2.885 pour le symptome 18 valeur en entrée 0
Valeur prevue 0.70086 pour le symptome 19 valeur en entrée 0
Valeur prevue 1.1462 pour le symptome 20 valeur en entrée 0

```

Annexe A.3 : Les valeurs attendues pour les données manquantes

```

Columns 12 through 20

    1.8544    4.1276    3.5883    2.5007    2.4810    1.3149    2.8850    0.7009    1.1462

===Maladies détectées===
    1    2    10

====Maladie trouvée====
    2

=Maladie originale===
    1

```

Annexe A.4 : Indication de la maladie

Annexe B : Les quelques codes sources de la plate forme du moteur de recherche SPARQL

Les codes sources que nous présentons ici ont été programmés en PHP. Nous les avons utilisés pour la génération de l'interface SPARQL.

1. Génération du code source du chargement de l'ontologie

```
<?php
include_once("../arc/ARC2.php");
$config = array(
    /* db */
    'db_name' => 'arc2test',
    'db_user' => 'arc2test',
    'db_pwd' => 'arc2test',
    /* store */
    'store_name' => 'sandbox',
    /* stop after 100 errors */
    'max_errors' => 100,
);
$store = ARC2::getStore($config);
if (!$store->isSetUp()) {
    $store->setUp();
}
/* $dossier = '/Tamp/'; */
$file_name = $_FILES["fichier"]["name"];
$file_tmp = $_FILES["fichier"]["tmp_name"];
$file_type = $_FILES["fichier"]["type"];
$file_size = $_FILES["fichier"]["size"];
$file_error_msg = $_FILES["fichier"]["error"];
if (!$file_tmp) { // if file not chosen
    echo $file_type.'message derreur';
    exit();
}
if (move_uploaded_file($file_tmp, '../Tamp/'.$file_name)) {
    echo $file_name.' chargé avec succès';
    $store->query('LOAD <../Tamp/'.$file_name.'>');
} else {
    echo 'le chargement du fichier a échoué';
}

?>
```

2. Génération du code source d'exécution de la requête

```
<?php
include_once("../arc/ARC2.php");
include_once("../config.php");
$store = ARC2::getStore($arc_config);
if (!$store->isSetUp()) {
    $store->setUp(); /* create MySQL tables */
}
if(isset($_POST['requete'])){
    $q = $_POST['requete'];
    $rows = $store->query($q, 'rows');
    $r = "";
    if ($rows = $store->query($q, 'rows')) {
        $r = '<table border=1>
<th>lien_concept</th><th>relation</th><th>indication</th>'. "\n";
        foreach ($rows as $row) {
            $r .= '<tr><td>'. $row['subject'] . '</td><td>'. $row['property'] .
'</td><td>'. $row['object'] . '</td></tr>'. "\n";
        }
        $r .= '</table>'. "\n";
    }
    else{
        $r = '<em>No data returned</em>';
    }
    echo $r;
}
?>
```

3. Code permettant la connexion au serveur

```
<?php
include_once('../config.php');
/* instantiation */
$sep = ARC2::getStoreEndpoint($arc_config);
if (!$sep->isSetUp()) {
    $sep->setUp(); /* create MySQL tables */
}
/* request handling */
$sep->go();
$config = array(
    /* db */
    'db_host' => 'localhost',
    'db_name' => 'arc2test',
    'db_user' => 'arc2test',
    'db_pwd' => 'arc2test',
    /* store name */
    'store_name' => 'sandbox',
);
?>
```

Annexe C : Extrait d'algorithme de la métrique du RàPC

Les différents algorithmes ci-dessous ont été développés avec le logiciel de programmation R. En effet, c'est un logiciel de statistique créé dans les années 1988 par Ross Ihaka Robert Gentleman. Il est à la fois un langage informatique et un environnement de travail : les commandes sont exécutées grâce à des instructions codées dans un langage relativement simple, les résultats sont affichés sous forme de texte et les graphiques sont visualisés directement dans une fenêtre qui leur est propre.

Algorithme de la fonction alpha

```
1 alphaf=function(i=1,j=2,num=NULL,data){
2   if(is.null(num)) {
3     resultat = 0
4   } else {
5     resultat=as.vector(dist(rbind(data[i,num],data[j,num]),method="euclidian"))
6   }
7   return(resultat)
9 }
```

Annexe C.1 : Algorithme de la fonction alpha

Algorithme de la fonction beta

```
1 betaf=function(i=1,j=2,cat=NULL,data){
2   vars=names(data[,cat])
3   if(is.null(cat)) {
4     resultat = 0
5   } else {
6     calc =
7     function(k){
8
9     tab = rbind(table(data[i,k]),table(data[j,k]))
10      res = ifelse(2 %in% apply(tab,2,sum),1,0)
11      return(res)
12    }
13    calcul = Vectorize(calc)
14    calcValq = calcul(cat)
15    resultat = (length(calcValq)-sum(calcValq))/(1+sum(calcValq))
    }
    return (resultat)
  }
```

Annexe C.2 : Algorithme de la fonction beta

Algorithme de la fonction gamma

```
1 gamma <-function(i,j,text,data){
2   require(tm)
3   base <-data[c(i,j), text]
4   for(k in text){
5     cosine(k)<-function(k){
6       src <- VectorSource(data[,k])
7       crp <- Corpus(src)
8       crp <-tm_map(crp,removeNumbers)
9       crp <- tm_map(crp, removePunctuation)
10      crp <- tm_map(crp, removeTolower)
11      crp <- tm_map(crp, Plaintext)
12    }
13  }
14 }
15 }
```

Annexe C.3 : Algorithme de la fonction gamma

Algorithme de la fonction phi

```
1 disFin = function(i=1,j=2,v,data, func=alphaf){
2   vect = 0
3   if((length(i)!=1)|(length(j)!=1)){
4     for(r in i){
5       for(c in j) {
6         vect = c(vect,func(r,c,v,data))
7       }
8     }
9     vect = vect[-1]
10    resultat = matrix(vect,nrow=length(i),byrow=TRUE)
11  } else {
12    resultat = func(i,j,v,data)
13  }
14  return(resultat)
15 }
16 alphah = function(i=1,j=2,num=NULL,data){
17   resultat = disFin(i,j,v=num,data, func=alphaf)
18   return(resultat)
19 }
20 betah = function(i=1,j=2,cat=NULL,data){
21   resultat = disFin(i,j,v=cat,data, func=betaf)
22   return(resultat)
23 }
24 gammah = function(i=1,j=2,text=NULL,data){
25   resultat = disFin(i,j,v=text,data, func=gammaf)
26   return (resultat)
27 }
28 phiSim = function(i=1,j=2,num=NULL,cat=NULL,text=NULL,data){
```

```

29  alphaVal = alphah(i,j,num,data)
30  betaVal = betah(i,j,cat,data)
31  gammaVal = gammah(i,j,text,data)
32  phi = 100*(1 - (((1-1/(1+alphaVal)))+(1-1/(1+betaVal)))+(1-1/(1+gammaVal))))/3))
33  return(phi)
34  }
35  phiDist = function(i=1,j=2,num=NULL,cat=NULL,text=NULL,data,
36  lower=FALSE,upper=FALSE){
37  resultat = (100 - phiSim(i,j,num,cat,text,data))/100
38  if(lower == TRUE) {
39  resultat[lower.tri(resultat, diag = TRUE)]=0
40  }
41  if(upper == TRUE) {
42  resultat[upper.tri(resultat, diag = TRUE)]=0
43  }
44  return (resultat)
45  }

```

Annexe C.4 : Algorithme de la fonction phi