

الجمهورية الجزائرية الديمقراطية الشعبية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جامعة أبي بكر بلقايد - تلمسان

Université Aboubakr Belkaïd - Tlemcen -

Faculté de TECHNOLOGIE



MEMOIRE

Présenté pour l'obtention du **diplôme** de **MASTER**

En : Génie biomédicale

Spécialité : Informatique Biomédicale

Par : Belloula Nor el houda et Benkhantech Hayet

Sujet

**La création d'une ontologie Pharmaceutique pour la
recherche d'information.**

Soutenu le 19 juin 2016, devant le jury composé de :

M. Gaouar Adil Université de Tlemcen Président

Mme Hamza Cherif Souad Université de Tlemcen Encadreur

M. Habib dahou Université de Tlemcen Examineur

Remerciement :

Avant tout, Nous tenons à remercier le bon DIEU le tout puissant et clément de nous avoir illuminé le chemin du savoir et de nous avoir donné le courage, la puissance et la volonté pour accomplir ce modeste projet.

Nous tenons particulièrement à exprimer notre profonde gratitude à notre encadreur Madame Hamza Cherif Souad pour sa patience, sa collaboration et ses précieux conseils qui étaient très bénéfiques pour la réalisation de ce travail.

Nous désirons témoigner notre reconnaissance et nos remerciements les plus sincères au Chef de département du Génie Biomédical et le Doyen de la faculté de la Technologie.

Nous tenons aussi à remercier vivement les examinateurs pour avoir accepté d'examiner ce travail et leurs participations au jury.

Mes remerciements vont également à l'endroit de Missoum Sid Ahmed, médecin assistant au CHU de Sidi Bel Abbes pour sa disponibilité à mon égard.

Enfin, nous ne saurons terminer ces remerciements sans y associer toute personne qui, de près ou de loin, nous a apporté son aide ou sa sympathie.

Dédicaces

Tous les mots du monde ne sauraient exprimer l'immense amour que nous vous portons, ni la profonde gratitude que nous vous témoignons pour tous les efforts et les sacrifices que vous n'avez jamais cessé de consentir pour notre instruction et notre bien-être.

C'est à travers vos encouragements que nous avons opté pour cette noble profession, et c'est à travers vos critiques que nous nous sommes réalisés.

Nous espérons avoir répondu aux espoirs que vous avez fondés en nous.

Nous vous rendons hommage par ce modeste travail en guise de notre reconnaissance éternelle et de notre infini amour.

Vous résumez si bien le mot parents qu'il serait superflu d'y ajouter quelque chose.

Que Dieu tout puissant vous garde et vous procure santé, bonheur et longue vie pour que vous demeuriez le flambeau illuminant le chemin de vos enfants.

À Tous ceux qui nous ont aidés de près ou de loin.

Finalement à tous ceux qui nous portent dans leurs cœurs.

Liste des Figures

Figure I.1: Exemple d'ontologie

Figure I.2: cycle de vie d'une ontologie

Figure I.3 : Pyramide du Web sémantique

Figure I. 4 : Processus de développement d'ontologie

Figure I.5 : Pyramide du Web sémantique

Figure II.1 : Les deux premiers processus transparents à l'utilisateur

Figure II.2 : L'architecture générale d'un SRI « processus en U»

Figure II.3 : Architecture d'un SRI utilisant le modèle vectoriel

Figure III.1 : TERMONTO : interface d'accès aux données de SYNTAX-UPERY

Figure III.2 : Modélisation de deux expressions du thésaurus de spécialité

Figure III.3 : Interface de l'outil de codage

Figure III.4 : Correspondance entre les termes WHO-ART et les termes SNOMED-CT

Figure III.5 : Hiérarchie des concepts primitifs et définis

Figure III.6 : Combinaison de la recherche par mots clés et de la recherche sémantique

Figure III.7 : Sur lignage des termes dénotant les concepts de la requête

Figure IV.1. Le processus de développement de l'application proposée

Figure IV.2. Architecture générale de notre approche

Figure V.1. Figure V.1. Editeur d'Ontologie Protégé-2000

Figure V.2. Page d'accueil Jena.

Figure V.3. Vue globale de nos ontologies

Figure V.4. Vue globale de notre base de données

Figure V.5. Fenêtre principale d'application

Figure V.6 fenêtre d'accueil d'application

Figure V.7. Fenêtre gestion médicament

Figure V.8. Fenêtre gestion patient

Figure V.9. Fenêtre gestion patient à l'étape de recherche du médicament

Liste des Tableaux

Tableau III.1. Statistique des corpus utilisés

Tableau III.2. Les 10 syntagmes nominaux les plus fréquents dans les corpus anglais et français

Tableau III.3. Résultats de leur différents runs

Tableau III.4. Précision de Mgrep et MetaMap avec deux dictionnaires

Tableau IV.1. Spécification des besoins de l'ontologie

Tableau IV.2. Extrait du glossaire des termes

Tableau IV.3. Glossaire des sous-classes.

Tableau IV.4. Glossaire de relation

Tableau IV.5. Glossaire d'attributs

TABLE DU MATIERE

Résumé.....	1
--------------------	----------

Introduction générale.....	3
-----------------------------------	----------

Chapitre I : Les ontologies

1. Introduction.....	6
----------------------	---

2. Définitions de l'ontologie.....	6
------------------------------------	---

3. Rôles des ontologies.....	8
------------------------------	---

3.1. Modularité et réutilisation des connaissances.....	8
---	---

3.2. Communication.....	8
-------------------------	---

4. Composantes d'une ontologie.....	9
-------------------------------------	---

4.1. Les concepts.....	9
------------------------	---

4.2. Les relations et les fonctions.....	9
--	---

4.3. Les axiomes.....	10
-----------------------	----

4.4. Les instances.....	10
-------------------------	----

5. Le cycle de vie des ontologies.....	10
--	----

5.1. Besoins et évaluations.....	11
----------------------------------	----

5.2. Conception et évolution.....	11
-----------------------------------	----

5.3. Diffusion.....	11
---------------------	----

5.4. Utilisation.....	12
-----------------------	----

5.5. Gestion.....	12
-------------------	----

6. Niveau de détail de l'ontologie.....	13
6.1. Typologie selon le niveau de complétude.....	13
6.2. Typologie selon le niveau du formalisme.....	13
6.3. Typologie selon l'objet de conceptualisation.....	14
6.3.1. Ontologie supérieure ou de Haut niveau.....	14
6.3.2. Ontologie Générique.....	14
6.3.3. Ontologie du Domaine.....	14
6.3.4. Ontologie de Tâches.....	14
6.3.5. Ontologie d'Application.....	15
7. Les méthodes de Construction d'ontologie.....	15
7.1. La méthodologie METHONTOLOGY.....	16
7.2. La methodologies On-To-Knowledge.....	17
7.3. La méthode OntologyDevelopment 101.....	17
8. Les domaines d'applications des ontologies.....	18
8.1. Système d'information.....	18
8.2. Web sémantique.....	19
9. Les avantages d'une ontologie.....	19
10. Les difficultés rencontrées dans l'élaboration d'une ontologie.....	20
11. Conclusion.....	21

Chapitre II : La recherche d'information

1. Introduction.....	23
2. Les Systèmes de Recherche d'Information.....	23

2.1. Définition de la recherche d'information.....	23
2.2. Définition de système de recherche d'information.....	23
2.2.1. Architecture générale des systèmes de recherche d'information.....	24
3. Concepts clés de la recherche d'information (RI).....	25
4. Les quatre principaux modes de recherche d'information.....	26
5. Principaux modèles de recherche d'information.....	27
6. Utilisation des ontologies dans le domaine de la recherche d'information.....	28
6.1. Lien entre les ontologies et la recherche d'information.....	28
6.2. Apports de l'ontologie dans le domaine de la recherche d'information.....	29
7. Conclusion.....	30

Chapitre III : Les travaux connexes

1. Introduction.....	32
2. Ontologie de la pneumologie pour l'aide au codage de l'information médical: Onto-Pneumo.....	33
2.1 . Objectifs et hypothèse.....	33
2.2. Matériel : outils et corpus.....	34
2.3. Méthode.....	35
2.4. Formalisation de l'ontologie avec les logiques de description.....	35
2.5. Résultats.....	36
3. Ontologie construite spécifiquement pour augmenter les signaux de détections de	

causalité possibles entre un médicament et un effet indésirable : Pharmacovigilance.....	37
3.1. Objectif.....	38
3.2. Hypothèse.....	39
3.3. Matériel.....	39
3.4. Méthode.....	39
3.5. Validation.....	41
3.6. Résultats.....	41
4. Contribution à la construction d'ontologies et à la recherche d'information : application au domaine médical.....	42
4.1. La constitution des corpus.....	43
4.1.1. L'extraction des candidats termes.....	43
4.1.2. Prétraitement du texte.....	43
4.1.3. La construction du noyau ontologique.....	43
4.1.4. L'enrichissement de l'ontologie.....	43
4.2. Résultats.....	44
4.3. L'enrichissement de l'ontologie.....	45
4.4. Conclusion.....	45
4.5. Application pour la mise en œuvre du portail SemBiP.....	45
4.5.1. Le portail SemBiP.....	46
4.5.2. La phase de recherche d'information.....	46

5. Indexation et intégration de ressources textuelles à l'aide d'ontologie : application au domaine biomédical.....	47
5.1. L'index de ressource biomédicale du NCBO.....	48
5.2. Cas d'utilisation et scénarios supportés.....	49
6. Comparaison entre les travaux.....	50
6.1. Discussion des résultats.....	52
7. Conclusion.....	52

Chapitre IV : Proposition d'une application médicale

1. Introduction.....	53
2. Problématique.....	54
3. Approche proposée.....	55
3.1. Conception de l'ontologie du domaine.....	55
3.2. Un processus de recherche d'information.....	60
3.3. Création de l'interface de l'application.....	60
3.4. Connexion entre l'ontologie et l'interface.....	61
4. Fonctionnement de l'application.....	61
5. Conclusion.....	63

Chapitre V : Implémentation et résultat

1. Introduction.....	65
2. Implémentation de l'ontologie.....	65

2.1. OWL (Ontology Web Language).....	65
2.2. Protégé.....	66
3. implémentation de l'interface.....	67
3.1. Netbeans.....	67
3.2. Construction de la base de données.....	67
3.2.1. MySql.....	67
4. Connexion ontologie-interface.....	68
4.1. Jena.....	68
5 .Architecture logicielle.....	68
5.1. Ontologie.....	68
5.2. La base de données.....	69
5.3. L'interface.....	69
6. Exemple.....	72
6.1. Gestion patient.....	72
6.2. Gestion médicament.....	72
7. Analyse et résultats.....	74
8. Conclusion.....	74
Conclusion générale perspectives.....	75
Bibliographie.....	76

Résumé :

Les ontologies associent les concepts de base d'un domaine précis et les relations entre ces concepts, tout cela d'une manière compréhensible par les machines. Elles encodent la connaissance d'un domaine particulier ce qui permet de rendre les connaissances réutilisables et elles retracent le champ de recherche de l'information et elles permettent le partage et la compréhension commune de la structure de l'information entre les personnes impliquées dans le domaine. Dans ce contexte, nous avons proposé dans ce mémoire de créer une application utilisant un moteur de recherche basé sur une ontologie dans le domaine pharmaceutique afin d'aider les pharmaciens et médecins à faire des prescriptions de médicament a des patients avec un minimum d'effet indésirables pour ces derniers. Surtout qu'il s'agit d'un véritable enjeu de santé publique car les effets indésirables liés aux médicaments EIM causent aujourd'hui plus de 3 % des hospitalisations et sont une des dix principales causes de décès dans la plupart des pays du monde entier.

Mots-clés : ontologie, recherche d'information, médicament, effet indésirables.

Abstract:

Ontologies combine the basic concepts of a specific domain and the relationships between these concepts, all this in a manner understandable by machines. They encode knowledge in a particular field allowing making that knowledge reusable and they trace the information search fields and allow sharing and common understanding of the structure of information between those involved in the field. In this context, we have proposed in this paper to create an application using a search engine based on ontology in the pharmaceutical field to help pharmacists and doctors to drug prescriptions to patients with minimal effect undesirable for the latter. Especially that it is a true public health issue because adverse events related to EIM drugs cause more than 3% of hospitalizations and today are one of the ten leading causes of death in most countries around the world .

Keywords: ontology, information retrieval, medication side effect

ملخص:

الأنطولوجيا هي الجمع بين المفاهيم الأساسية لمجال معين والعلاقات بين هذه المفاهيم، كل هذا بطريقة مفهومة من قبل الأجهزة. أنها ترميز المعرفة في مجال معين مما يسمح ويمكن إعادة استخدامها المعرفة وأنها تتبع حقول البحث عن المعلومات والسماح تقاسم وفهم مشترك لبنية المعلومات بين المشاركين في المجال. في هذا السياق، اقترحنا إنشاء تطبيق باستخدام محرك البحث على أساس الأنطولوجيا في مجال المستحضرات الصيدلانية لمساعدة الصيادلة والأطباء إلى وصفات الدواء للمرضى مع الحد الأدنى من التأثير غير مرغوب فيه لهذا الأخير. خاصة أنه يمثل قضية صحية عامة حقيقية بسبب الأحداث السلبية المتعلقة بالآثار الجانبية EIM تسبب أكثر من 3% من المستشفيات، واليوم هي واحدة من الأسباب الرئيسية العشرة للوفاة في معظم البلدان في جميع أنحاء العالم.

كلمات البحث: الأنطولوجيا، استرجاع المعلومات، الآثار الجانبية، الدواء.

INTRODUCTION GENERALE

La Recherche d'Information (RI) peut être définie comme une activité dont la finalité est de localiser et de délivrer des granules documentaires à un utilisateur en fonction de son besoin en informations. Afin de faciliter la mise en place de systèmes pouvant gérer de grandes quantités d'information et de s'adresser à un maximum d'utilisateurs, de nombreuses suppositions pragmatiques et simplifications ont été avancées dans la littérature. Actuellement, une énorme quantité d'informations est disponible dans toute sorte de domaines. Ce volume d'informations croissant avec une production abondante de données numériques. De plus, ces informations sont généralement exprimées en langage naturel (sous forme de textes) et donc dans un format non structuré (près de 80% de données dans le domaine médical selon les estimations) ; ce qui rend leur traitement automatique difficile.

Les systèmes informatiques disponibles sur le marché utilisent généralement des thésaurus et de portée limitée. Pour pallier à cette insuffisance il est nécessaire de décrire la sémantique des objets et de l'organisation du domaine médical, et ceci dans le but de développer des modèles conceptuels qui ne sont ni contextuel ni ambiguë, et dont le sens est inscrit dans la structure même du modèle. Cette modélisation est appelé «ontologie» Les ontologies sont un moyen de représenter la connaissance et elles contiennent des concepts et des relations entre concepts ainsi qu'un lexique permettant de référencer les concepts et les relations mais n'intègrent pas d'axiomes dans leur formalisation

Une ontologie fournit une référence pour la communication entre les machines mais aussi entre humains et machines en définissant le sens des objets et l'utilisation d'ontologies dans un modèle de Recherches 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.

Notre travail se situe dans le domaine médical, plus précisément dans la représentation des connaissances pharmaceutique, ou nous tentons de créer une application afin d'aider le médecin ou pharmacien à donner le bon médicament au patient avec un minimum d'effets secondaire et indésirables. Surtout qu'il s'agit d'un véritable enjeu de santé publique car les effets indésirables liés aux médicaments EIM causent aujourd'hui plus de 3 % des hospitalisations et sont une des dix principales causes de décès dans la plupart des pays du monde entier.

Pour ce faire nous allons tenter de créer une ontologie des médicaments qui va servir de base de connaissance pour un moteur de recherche qui va l'exploiter au mieux afin de répondre aux requêtes des utilisateurs.

Ce mémoire est rédigé sur deux parties : la première partie est plutôt théorique et regroupe les 3 premiers chapitres qui abordent les concepts de base traités à savoir : le premier chapitre est consacré à introduire les ontologies, leurs définitions, et caractéristiques. Quant au deuxième chapitre il s'intéresse à la recherche d'information et enfin dans le troisième chapitre nous présentons quelques travaux connexes, qui se basent sur l'utilisation des ontologies dans un moteur de recherche.

La 2ème partie de ce mémoire regroupe les 2 derniers chapitres. En outre donc dans le quatrième chapitre nous avons fait une description globale de l'approche suivie pour la conception de l'application. Et enfin le 5^{ème} chapitre est consacré à l'implémentation de l'application, et la description du fonctionnement.

CHAPITRE I :

Les ontologies

1. Introduction

La recherche sur les ontologies qui a débuté vers les années 90 du siècle dernier dans les domaines de l'intelligence artificielle, de l'ingénierie des connaissances, et de la représentation des connaissances, a vu son champ d'application s'élargir de façon considérable, et fait partie des axes de la recherche actuelle, en particulier dans le domaine de la modélisation des systèmes d'information où les recherches utilisant les ontologies sont de plus en plus nombreuses. Les ontologies sont des systèmes formels dont l'objectif est de représenter les connaissances d'un domaine au moyen de concepts, définis et structurés les uns par rapport aux autres. La représentation ontologique des connaissances assure le maintien de la cohérence des axiomes et de l'intégrité du système, ainsi que l'évolution de la représentation sans changer la structure. La connaissance visée par ces ontologies est un thème de recherche dans divers axes tels que l'ingénierie des connaissances, la recherche d'information, les systèmes d'information coopératifs, l'intégration intelligente d'information, la gestion des connaissances.

Dans ce chapitre nous allons voir une présentation de la notion des ontologies, leurs différents types, les composants, et quelques domaines d'application. Nous avons rappelé aussi quelques méthodes pour construire une ontologie.

2. Définitions de l'ontologie

Les ontologies sont développés dans un contexte informatique – que ce soit celui de l'Ingénierie des connaissances, de l'Intelligence artificielle ou de la gestion et des systèmes d'information ou plus spécifiquement le contexte du Web sémantique.

Dans le cadre de l'intelligence artificielle [Neeches et ses collègues 93] furent les premiers à proposer une définition à savoir « une ontologie définit les termes et les relations de base de vocabulaire d'un domaine, ainsi que les règles qui indiquent comment combiner les termes et les relations de façon à pouvoir étendre le vocabulaire».

[GRUBER 93] propose la définition suivante: «une ontologie est une spécification explicite d'une conceptualisation». [GUARINO 95] a modifié légèrement la définition de GRUBER, et la définit par: «une ontologie est une spécification partielle et formelle d'une conceptualisation». En 1997, ces deux dernières définitions sont regroupées dans celle de [BORST 97] comme « une ontologie est définit comme étant une spécification formelle d'une conceptualisation partagée».

La notion d'ontologie a été définie aussi comme Un « vocabulaire et des définitions des concepts d'un domaine » [USH 96]. La même notion est également développée dans [GUARINO

97] : une ontologie est une théorie logique dont les modèles contraignent une certaine conceptualisation, sans la spécifier exactement. La définition fondatrice de [GRU 93], a été actualisée dans [STU 98] sous la forme suivante « spécification formelle et explicite d'une conceptualisation partagée».

- Formelle : l'ontologie doit être lisible par une machine, ce qui exclut le langage naturel.
- Explicite : la définition explicite des concepts utilisés et des contraintes de leurs utilisations.
- Conceptualisation : le modèle abstrait d'un phénomène du monde réel par identification des concepts clefs de ce phénomène.
- Partagée : l'ontologie n'est pas la propriété d'un individu, mais elle représente un consensus accepté par une communauté d'utilisateurs.

Charlet en propose une définition complémentaire [CHA 02] : *Une ontologie est une spécification normalisée représentant les classes des objets reconnus comme existant dans un domaine. Construire une ontologie, c'est aussi décider d'une manière d'être et d'exister des objets de ce domaine.* La même notion est également développée par Gomez comme : « *une ontologie fournit les moyens de décrire de façon explicite la conceptualisation des connaissances représentées dans une base de connaissances* ». Ainsi, une ontologie répond à des exigences complémentaires et symétriques : (i) en tant que spécification, elle définit une représentation formelle des connaissances permettant son exploitation par un ordinateur ; (ii) en tant que reflet d'un point de vue – partiel – sur un domaine, que l'on cherche le plus consensuel possible, elle fournit une sémantique qui doit permettre de relier la forme exploitable par la machine à sa signification pour les humains.

En analysant toutes ces définitions, on constate que des divergences règnent entre les auteurs et les chercheurs du domaine, mais une unanimité est établie autour de deux principes : une ontologie est relative à un domaine, et est constituée de concepts et de relations les reliant les uns aux autres.

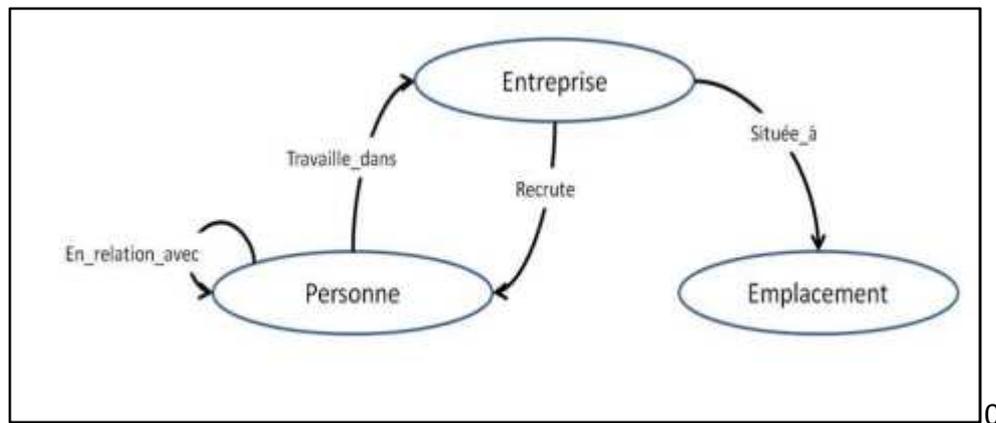


Figure 0.1.Exemple d'ontologie

3. Rôles des ontologies

3.1. Modularité et réutilisation des connaissances

Les ontologies sont surtout utilisées pour la représentation de connaissances et l'application de raisonnements sur ces connaissances. Cependant une ontologie possède des caractéristiques qui, au-delà de cette représentation, favorisent la réutilisation et le partage de données. Gruber [TRGU 95] insistait sur le rôle que pouvaient tenir les ontologies pour favoriser la modularité et la réutilisation dans les systèmes informatiques. En effet, ces ontologies permettent l'étude de conceptualisations, indépendamment du formalisme choisi pour les représenter et doivent être définies indépendamment du langage utilisé pour la programmation des applications, de la plate-forme utilisée et des protocoles de communication (protocoles réseaux).

3.2. Communication

La communication entre humain pose surtout des problèmes quand les acteurs de cette communication ne sont pas du même domaine et ne parlent donc pas forcément le même langage. La réutilisation, le partage de connaissances et d'ontologies, suppose que plusieurs utilisateurs soient d'accord sur les ontologies partagées. C'est ce que FENSEL [FDES 98] nomme un onto group. Philippe Martin [PHMA 95] propose d'aider les spécialistes de l'ingénierie de la connaissance, en utilisant la terminologie définie dans Word Net comme base de la communication, car c'est un standard. Dans le domaine pédagogique c'est la communication entre auteurs et informaticiens qui est parfois difficile, d'où l'intérêt d'utiliser des ontologies

dans les environnements auteur pour la définition d'un vocabulaire convivial et précis dans la définition des tâches pédagogiques [IHLB 99]. L'ontologie joue alors le rôle d'un méta-modèle.

Les ontologies peuvent également être utilisées pour harmoniser la communication entre différentes applications où entre différents agents [WCR 99]. Cette idée, également sous-jacente dans les publications de Gruber [TRGU 95], repose souvent sur une ontologie du domaine. Pourtant Mizoguchi [MIZO 98] veut aller plus loin en dotant les agents d'une connaissance sur une ontologie de tâche indépendante du domaine.

4. Composantes d'une ontologie

Les ontologies fournissent un vocabulaire commun d'un domaine et définissent la signification des termes et des relations entre elles. La connaissance dans les ontologies est principalement formalisée en utilisant les cinq types de composants [GPBV 99] : concepts (ou classes), relations (ou propriétés), fonctions, axiomes (ou règles) et instances (ou individus).

4.1. Les concepts

Aussi appelés termes ou classe de l'ontologie, correspondent aux abstractions pertinentes d'un segment de la réalité (le domaine du problème) retenus en fonction des objectifs, Un concept peut représenter un objet matériel, une notion, une idée. Un concept peut être divisé en deux parties [UMKM 95] :

- Un terme(ou plusieurs).
- Une notion et un ensemble d'objets. La notion, également appelée intension du concept, contient la sémantique du concept, exprimée en termes de propriétés et d'attributs, de règles et des contraintes, également appelé extension du concept.

4.2. Les relations et les fonctions

L'interaction entre les notions d'un domaine définit les relations, On trouve des relations binaires, tertiaire. Certaines relations binaires entre des objets sont considérées comme des rôles comme le définit [BORG 96].Les fonctions sont des cas particuliers de relations dans lesquelles un élément de la relation est défini à partir des autres éléments.

4.3. Les axiomes

Les axiomes désignent des vérités indémonstrables qui doivent être admises. Ce sont des affirmations considérées comme évidentes sans preuve. Ils permettent de contraindre les valeurs de classes ou d'instances.

4.4. Les instances

Les instances représentent les éléments des concepts et des relations dans un domaine donné [BLAS et AL. 98].

5. Le cycle de vie des ontologies

Un cycle de vie d'une ontologie débute du besoin qui se transforme en idée, la concrétisation de l'idée qui se traduit par la conception qui est diffusée pour son utilisation. Vient ensuite l'étape de l'évaluation qui donne naissance, le plus souvent à une étape d'évolution. et de maintenance du modèle. Une réévaluation de l'ontologie et des besoins devra se faire après chaque utilisation significative [Dien et al. 01].

L'ontologie peut être étendue et, si nécessaire, en partie reconstruite. La validation du modèle de connaissances est au centre du processus et se fait de manière itérative [Pere 99].

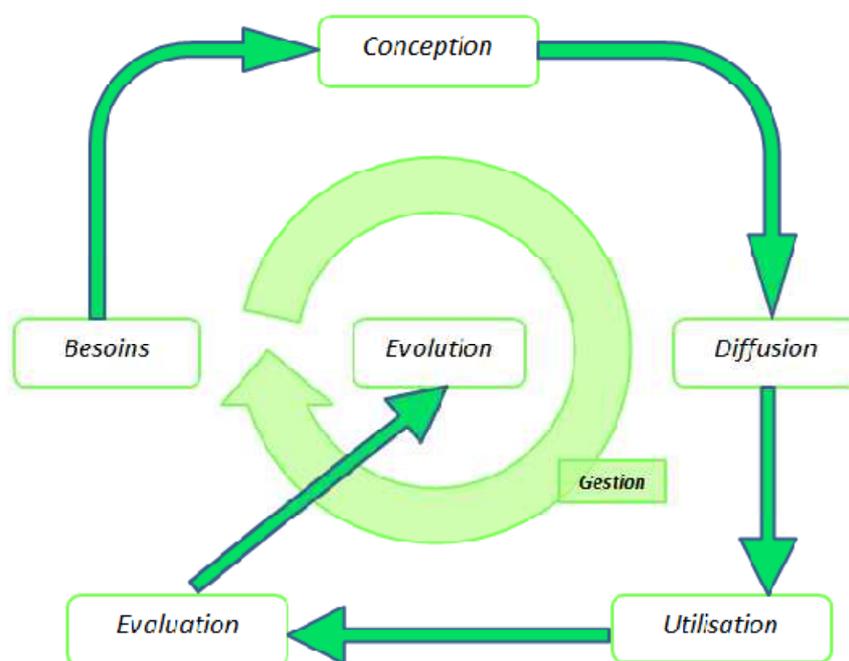


Figure 0.2. Cycle de vie d'une ontologie

5.1. Besoins et évaluation

Dans l'utilisation d'une ontologie, des problèmes méthodologiques de recueil d'information, d'analyse et d'identification peuvent surgir. Afin d'éviter ces problèmes dans l'utilisation d'une ontologie, un état des lieux initial approfondi doit être élaboré dans l'étape de détection des besoins, car cette étape ne peut reposer sur des études précédentes ou des retours d'utilisation comme dans le cas de l'activité d'évaluation.

5.2. Conception et évolution

Les phases de conception et d'évolution ont en commun un certain nombre de points :

- Spécification des solutions
- Conceptualisation et modélisation
- Formalisation (logiques de description, graphes conceptuels, formalismes du web sémantique RDF, RDF(S) et OWL)
- Intégration de ressources existantes
- Implantation (graphes conceptuels, logiques de description)

Le choix de représentation et de conceptualisation faits dans l'ontologie représente un problème de conception et d'évolution. Notons aussi que l'évolution pose le problème de la maintenance de ce qui repose déjà sur l'ontologie. L'ontologie est à la fois un ensemble évolutif et un ensemble de primitives pour décrire des faits et des algorithmes sur ces faits. Ses changements donc, ont un impact direct sur tout ce qui a été construit sur la base de cette ontologie. Le maintien de la cohérence dans une ontologie est un des points clés dans son utilisation. La maintenance de l'ontologie soulève donc des problèmes d'intégration technique et des problèmes d'intégration aux usages. Les domaines de la maintenance comme l'historique et la gestion des versions, la propagation des changements après modification, sont des questions importantes de la recherche actuelle.

5.3. Diffusion

Le déploiement et à la mise en place de l'ontologie interviennent dans la phase de diffusion, où les problèmes sont fortement contraints par l'architecture des solutions. Pour l'application Web on utilisera des technologies adéquates et pour le partage de fichiers, nous pouvons utiliser des architectures distribuées ou l'architectures Peer to Peer. Dans toutes les architectures dans

les serveurs web, services web, Peer to Peer, la distribution des ressources et leur hétérogénéité du point de vue syntaxiques, sémantiques, protocolaires, contextuelles, ou autres posent des problèmes de recherche sur l'interopérabilité et le passage à l'échelle (larges bases, optimisation d'inférences, propagation de requêtes, etc.).

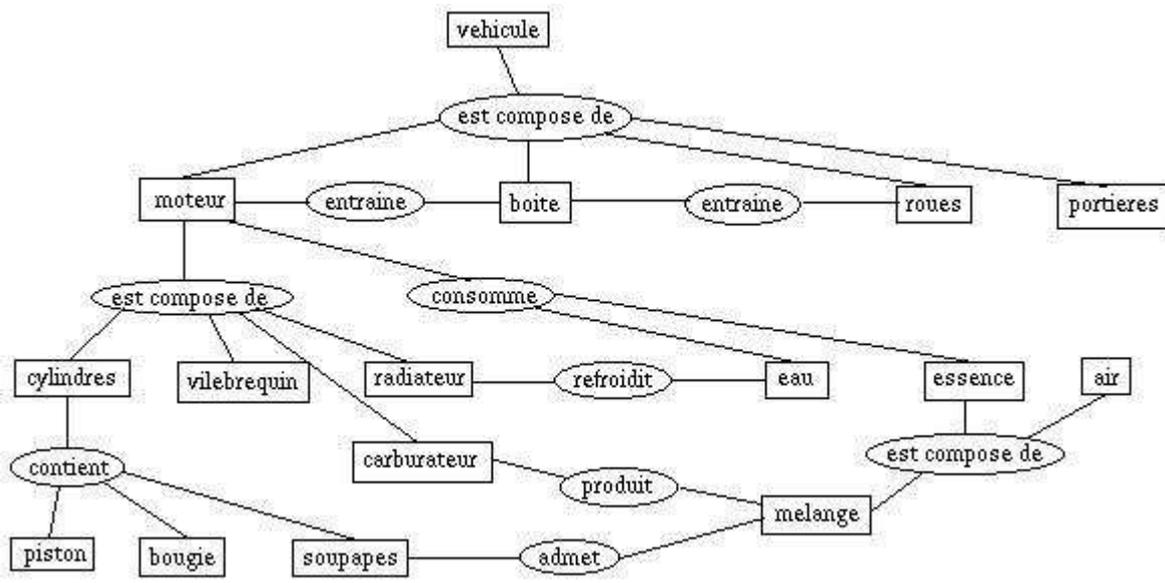
5.4. Utilisation

Les activités reposant sur la disponibilité de l'ontologie, comme l'annotation des ressources, la résolution de requête, la déduction de connaissances, etc. constituent la phase d'utilisation.

Les problèmes de la conception des interactions avec l'utilisateur et de leur ergonomie concernant les interfaces dynamiques, les profils et contextes d'utilisation, sont posés dans toutes ces activités. Sur ce point, l'ontologie apporte à la fois de nouvelles solutions et de nouveaux problèmes.

5.5. Gestion

L'existence d'une activité permanente de gestion et planification est importante pour assurer une pérennité dans le travail de suivi et la politique globale pour la détection, la préparation l'évaluation des itérations d'un cycle et s'assurer que l'ensemble de ces causes améliorent les systèmes d'information



Figurel.3.Pyramide du Web sémantique

6. Niveau de détail de l'ontologie

Par rapport au niveau de détail utilisé lors de la conceptualisation de l'ontologie, en fonction de l'objectif opérationnel envisagé pour l'ontologie, deux catégories au moins peuvent être identifiées [GUARI 97] :

- Granularité fine : correspondant à des ontologies très détaillées, possédant ainsi un vocabulaire plus riche capable d'assurer une description détaillée des concepts pertinents.
- Granularité large : correspondant à un vocabulaire moins détaillé comme par exemple dans les scénarios d'utilisation spécifiques où les utilisateurs sont déjà préalablement d'accord à propos d'une conceptualisation sous-jacente.

6.1. Typologie selon le niveau de complétude :

Le niveau de complétude a été abordé par [MIZO 98] et [BACHI 00]. À titre d'exemple, nous décrivons la typologie de [BACHI 00]. Ce dernier propose la classification sur trois niveaux suivante :

- Niveau 1-Sémantique : Tous les concepts (caractérisés par un terme/libellé) doivent respecter les quatre principes différentiels :
Communauté avec l'ancêtre.
Différence (spécification) par rapport à l'ancêtre.
Communauté avec les concepts frères (situés au même niveau).
Différence par rapport aux concepts frères (sinon il n'aurait pas lieu de le définir).
- Niveau 2 -Référentiel : Outre les caractéristiques énoncées au niveau précédent, les concepts référentiels (ou formels) se caractérisent par un terme/libellé dont la sémantique est définie par une extension d'objets
- Niveau 3- Opérationnel : Outre les caractéristiques énoncées au niveau précédent, les concepts du niveau opérationnel ou computationnel sont caractérisés par les opérations qu'il est possible de leur appliquer pour générer des inférences (engagement computationnel).

6.2. Typologie selon le niveau du formalisme

Par rapport au niveau du formalisme de représentation, du langage utilisé pour rendre l'ontologie opérationnelle, [GRFO 95] propose une classification comprenant quatre catégories :

- Informelles : ontologies opérationnelles dans un langage naturel (sémantique ouverte).
- Semi informelles : utilisation d'un langage naturel structuré et limité.
- Semi formelles : langage artificiel défini formellement.
- Formelles : utilisation d'un langage artificiel contenant une sémantique formelle, ainsi que des théorèmes et des preuves des propriétés telles la robustesse et l'exhaustivité [GPBV 99].

6.3. Typologie selon l'objet de conceptualisation

Les ontologies classifiées selon leur objet de conceptualisation, sont de la façon suivante :

6.3.1. Ontologie supérieure ou de Haut niveau

Cette ontologie est une ontologie générale. Son sujet est l'étude des catégories des choses qui existent dans le monde, soit les concepts de haute abstraction tels que: les entités, les événements, les états, les processus, les actions, le temps, l'espace, les relations, les propriétés. L'ontologie de haut de niveau est fondée sur : la théorie de l'identité, la méréologie (theory of whole and parts role) et la théorie de la dépendance.

6.3.2. Ontologie Générique

Cette ontologie aussi appelée, méta-ontologies, véhicule des connaissances génériques moins abstraites que celles véhiculées par l'ontologie de haut niveau, mais assez générales néanmoins pour être réutilisées à travers différents domaines.

6.3.3. Ontologie du Domaine

Contient de la connaissance se rapportant à un domaine. Elle décrit les concepts et leurs relations, les théories et principes valides dans le domaine, elle se doit d'être réutilisable. Ainsi, il est nécessaire de construire ce type d'ontologie par agrément

6.3.4. Ontologie de Tâches

Ce type d'ontologies est utilisé pour conceptualiser des tâches spécifiques dans les systèmes, telles que les tâches de diagnostic, de planification, de conception, de configuration, de tutorat, soit tout ce qui concerne la résolution de problèmes.

6.3.5. Ontologie d'Application

Cette ontologie est la plus spécifique. Les concepts dans l'ontologie d'application correspondent souvent aux rôles joués par les entités du domaine tout en exécutant une certaine activité.

7. Les méthodes de Construction d'ontologie

Les activités exécutées lors de la construction des ontologies forment le processus de développement d'ontologie. Les ontologies sont construites par des équipes coopératives géographiquement distantes, et pour cette raison il est indispensable d'identifier ces activités. Le processus de développement d'ontologie est une proposition basée sur le standard IEEE pour le développement logiciel [FER, GOM 97]. Il est recommandé d'effectuer les trois catégories des activités indiquées ci-dessous et d'orienter clairement le processus des constructions :

- Les activités de gestion incluant la programmation, le contrôle et la garantie de la qualité.
- Les activités orientées développement regroupant les activités de pré-développement, développement et post-développement.
- Les activités supports incluent une série d'activités exécutées en même temps que les activités orientées développement, sans lesquelles l'ontologie ne pourrait être construite. Ces activités incluent l'acquisition de connaissance, l'évaluation, l'intégration, la fusion, l'alignement, la documentation et la gestion de version.

Pendant le pré-développement, une étude de l'environnement est effectuée pour connaître l'environnement dans lequel l'ontologie sera utilisée, les applications dans lesquelles l'ontologie sera intégrée, etc. Aussi pendant le pré-développement, l'étude de faisabilité répond aux questions telles que : est-il possible ou est-il approprié de construire l'ontologie.

Dans le développement, les activités à réaliser sont :

- La spécification.
- La conceptualisation.
- La formalisation.
- L'implémentation.

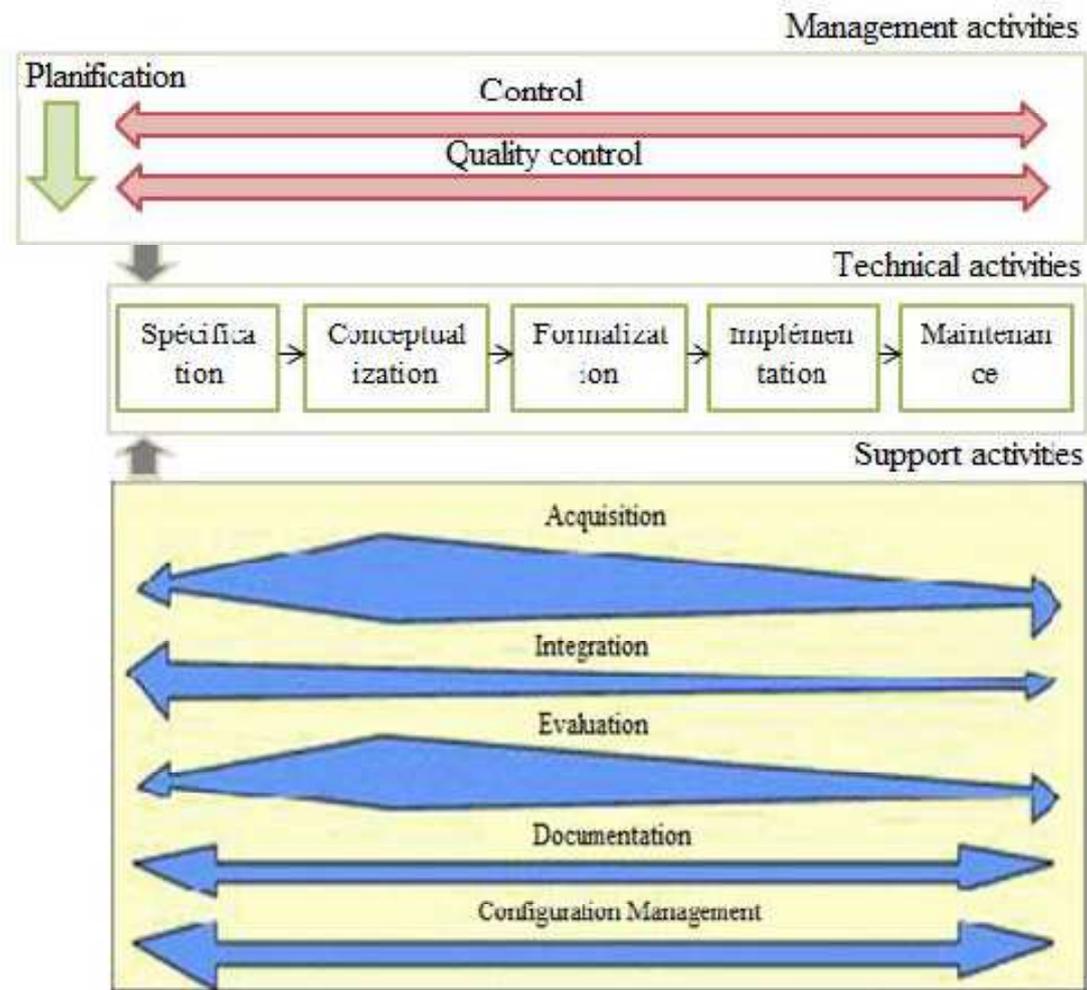


Figure 0.3. Processus de développement d'ontologie

Pendant le post-développement, les activités de maintenance et de (ré) utilisation sont réalisées. L'analyse, nous montre que le processus de développement de l'ontologie identifie les différentes activités à exécuter.

Cependant, il n'identifie pas l'ordre dans lequel les activités doivent être exécutées. Le cycle de vie de l'ontologie identifie à quel moment les activités doivent être effectuées, c'est à dire, il identifie l'ensemble des étapes à travers lesquelles l'ontologie évolue, décrit quelles activités doivent être exécutées dans chaque étape et comment les étapes sont reliées.

7.1. La méthodologie METHONTOLOGY

Cette méthodologie a été développée par le groupe d'Ontologie à l'Université Polytechnique de Madrid. METHONTOLOGY [FER, GOM 97], elle prend ses racines dans les activités principales identifiées par le processus de développement logiciel et dans les

méthodologies d'ingénierie de connaissance. Cette méthodologie inclue: l'identification du processus de développement, un cycle de vie basé sur des prototypes évolutifs, et les techniques pour effectuer chaque tâche dans les activités de gestion, de développement et de support.

METHONTOLOGY a été proposée pour la construction d'ontologie par la FIPA (Foundation for Intelligent Physical Agents), qui favorise l'interopérabilité à travers les applications.

METHONTOLOGY inclut une méthode de réingénierie pour résoudre certains des problèmes liés à la construction d'une ontologie par la réutilisation d'une autre ontologie [FER, GOM 04].

7.2. La méthodologies On-To-Knowledge

On-to-Knowledge recommande un procédé itératif de développement, et comporte quatre phases principales: une phase de spécification de condition, une phase d'amélioration, une phase d'évaluation et une phase d'application et d'évolution [FER, GOM 04].

On-To-Knowledge propose l'acquisition des connaissances en spécialisant une ontologie générique. Elle propose de construire l'ontologie en tenant compte de la manière dont elle sera utilisée dans d'autres applications. Par conséquent, les ontologies développées avec cette méthodologie sont fortement dépendantes de l'application.

On-To-Knowledge recommande la suite d'outils Onto Studio comme support de développement des ontologies. Le principal auteur est GRÜNINGER [M_{elle} BOUARROUDJ.S 09]. Il indique clairement les relations entre un ensemble de techniques, méthodes et principes pour chacun de ses processus (par exemple, ordre recommandé, entrée/sorties) [SUR , ERDM 02].

7.3. La méthode OntologyDevelopment 101

OntologyDevelopment 101 [NOYN, GUIN 01] a été développée à l'Université de Stanford, elle cherche à construire des ontologies formelles par la reprise et l'adaptation des ontologies déjà existantes, et propose de suivre les démarches ci-après:

- Déterminer le domaine et la portée de l'ontologie.
- Considérer la réutilisation des ontologies existantes.
- Énumérer les termes les plus importants dans l'ontologie.
- Définir les classes et hiérarchie des classes.
- Définir les propriétés des classes.

- Définir les facettes des attributs.
- Construire les instances.

8. Les domaines d'applications des ontologies

Il existe plusieurs domaines qu'on applique les ontologies selon nos besoins comme :

8.1. Système d'information

L'intégration d'une ontologie dans un système d'information vise à réduire, voire éliminer, la confusion conceptuelle et terminologique à des points clefs du système, et à tendre vers une compréhension partagée pour améliorer la communication, le partage, l'interopérabilité et le degré de réutilisation possible, ce qui permet de déclarer formellement un certain nombre de connaissances utilisées pour caractériser les informations gérées par le système, et de se baser sur ces caractérisations et la formalisation de leur signification pour automatiser des tâches de traitement de l'information.

L'ontologie retrouve maintenant dans une large famille de systèmes d'information. Elle est utilisée pour :

- Décrire et traiter des ressources multimédia.
- Assurer l'interopérabilité d'applications en réseaux.
- Piloter des traitements automatiques de la langue naturelle.
- Construire des solutions multilingues et interculturelles.
- Permettre l'intégration des ressources hétérogènes d'information.
- Vérifier la cohérence de modèles.
- Permettre les raisonnements temporel et spatial.
- Faire des approximations logiques.

Ces utilisations des ontologies se retrouvent dans de nombreux domaines d'applications tel que :

- Intégration d'information géographique.
- Gestion de ressource humaine.
- Aide à l'analyse en biologie, suivi médicale informatisé.
- Commerce électronique.
- Enseignement assisté par ordinateur.
- Bibliothèque numériques.

- Recherche d'informations.

8.2. Web sémantique

Un courant particulièrement prometteur pour l'expansion des systèmes à base d'ontologies est celui du Web sémantique. Il s'agit d'une extension du Web actuel, dans laquelle l'information se voit associée à un sens bien défini, améliorant la capacité des logiciels à traiter l'information disponible sur le Web. L'annotation des ressources d'information du Web repose sur des ontologies, elles sont aussi disponibles et échangées sur le Web. Grâce au Web sémantique, l'ontologie a trouvé un jeu de formalismes standards à l'échelle mondiale, et s'intègre dans de plus en plus d'applications Web. Cela se fait au profit des logiciels qui à travers les ontologies et les descriptions qu'elles permettent, peuvent proposer de nouvelles fonctionnalités exploitant les effets d'échelles du Web pour en améliorer les effets.

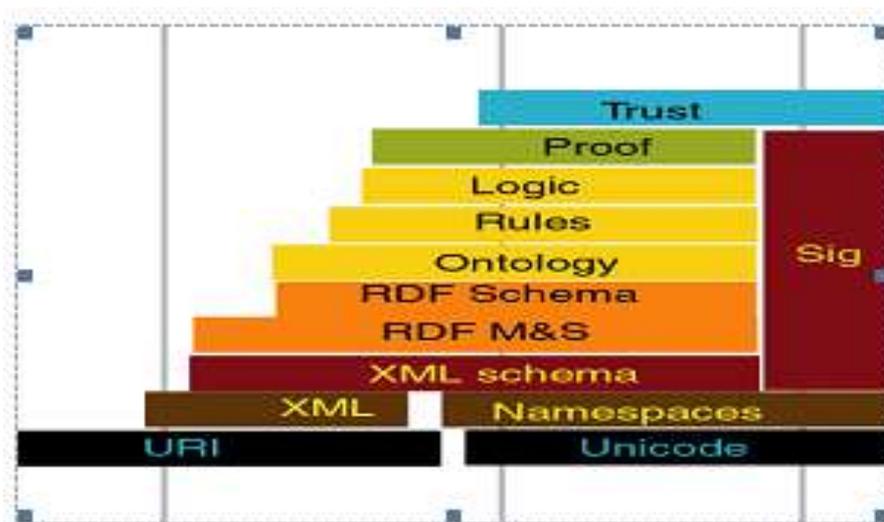


Figure 0.4. Pyramide du Web sémantique

9. Les avantages d'une ontologie

Parmi les avantages des ontologies la compréhension commune de la structure de l'information entre les personnes ou les fabricants de logiciels, assurer l'interopérabilité entre systèmes, permettre l'échange de connaissances entre Systèmes, Permettre la réutilisation du savoir sur un domaine: créer et conserver des bases de connaissances réutilisables, Expliciter ce qui est considéré comme implicite sur un domaine, Distinguer le savoir sur un domaine du savoir opérationnel et Analyser le savoir sur un domaine.

10. Les difficultés rencontrées dans l'élaboration d'une ontologie

Tout d'abord, il est toujours difficile d'avoir accès aux connaissances, à la constitution du corpus.

La deuxième difficulté est l'organisation conceptuelle d'une ontologie. Les principes d'organisation sont souvent abstraits et font appel à des notions philosophiques. Il faut donc avoir une théorie soit sur la connaissance, soit sur le monde. On se retrouve souvent à faire ce qu'on appelle l'upperontology, c'est à dire les concepts les plus abstraits, ceux qui vont organiser le reste de l'ontologie.

La troisième difficulté est de pouvoir modéliser des concepts, non pas fondateurs de l'ontologie et qui sont de niveau philosophique, mais des concepts du domaine. Là, il faut avoir l'intelligence du domaine. Il va falloir travailler avec des spécialistes qui soient capables de confirmer, infirmer, suggérer les concepts qui sont importants pour structurer la connaissance de leur domaine.

Une fois qu'on a les concepts qui organisent le domaine, les concepts philosophiques qui organisent l'ontologie, il suffit de remplir celle-ci. En médecine, par exemple, on va énumérer tous les os humains une fois qu'on a défini la notion de squelette et la notion d'anatomie. L'anatomie et le squelette sont les concepts structurants du domaine. Par contre on n'a pas besoin d'être un grand spécialiste pour lister les médicaments ou les os.

11. Conclusion

Au long de ce chapitre, nous avons essayé d'éclaircir la notion d'ontologie en présentant certaines définitions. Nous avons montré aussi les étapes de construction des ontologies, leurs avantages, et leurs domaines d'application.

Dans ce qui a été présenté précédemment, il ressort que la notion d'ontologie constitue l'une des approches les plus efficaces pour représenter et analyser et traiter des connaissances. A cet effet, nous allons présenter dans le chapitre suivant la notion de la recherche d'information (RI) et l'apport des ontologies dans ce domaine.

CHAPITRE II :

La recherche d'information

1. Introduction

Le domaine de la recherche d'information (RI) remonte au début des années 1950, peu après l'invention des ordinateurs. A l'origine, les chercheurs s'intéressaient à la problématique de l'indexation des documents. Classiquement, la recherche d'information est définie comme étant l'ensemble des actions, méthodes et procédures ayant pour objet d'extraire d'un ensemble de documents les informations voulues. Le système qui permet de mettre en œuvre ces différentes opérations est appelé Système de Recherche d'Information (SRI). Son but est de retrouver les documents pertinents dans une grande collection de documents.

Dans ce chapitre nous allons présenter quelle que notions sur la recherche d'information (RI) et les principaux modèles utilisés par les systèmes de recherche d'information (SRI). Puis on va cerner sur le domaine des systèmes de recherche d'information, et ces notions de base de ce domaine, l'architecture générale des SRI, les concepts clé sur lesquels ces systèmes sont basés, les principaux modèles de recherche d'informations existants et l'apport de l'ontologie sur la RI .

- 2. Les systèmes de recherche d'Information :

2.1 . Définition de la recherche d'information

La recherche d'information (RI) est une branche de l'informatique qui s'intéresse l'acquisition, l'organisation le stockage, la recherche et la sélection d'information.

Généralement, l'information recherché est contenue dans un document qui peut avoir différents formes : textes, hypertexte ou media, et il peut être aussi : non structuré, semi structuré ou quasi structuré.

2.2. Définition de système de recherche d'information

Les systèmes de recherche d'information (SRI), servent d'interface entre une source (collection) contenant des quantités considérables de documents et des utilisateurs cherchant via des requêtes, des informations susceptibles de se trouver dans cette collection. Les SRI intègrent un ensemble de techniques permettant de sélectionner ces informations.

Elles peuvent être résumées en quatre fonctions, qui sont le stockage de l'information, l'organisation de ces informations, la recherche d'informations en réponse à des requêtes utilisateurs et la restitution des informations pertinentes pour ces requêtes.

Un Système de Recherche d'Information (SRI) est sensé gérer un volume d'informations relativement considérable. Il intègre un ensemble de modèles et de processus permettant de représenter, d'organiser, de questionner ce volume d'informations et de restituer l'information qui correspond le mieux au besoin de l'utilisateur, exprimé via une requête.

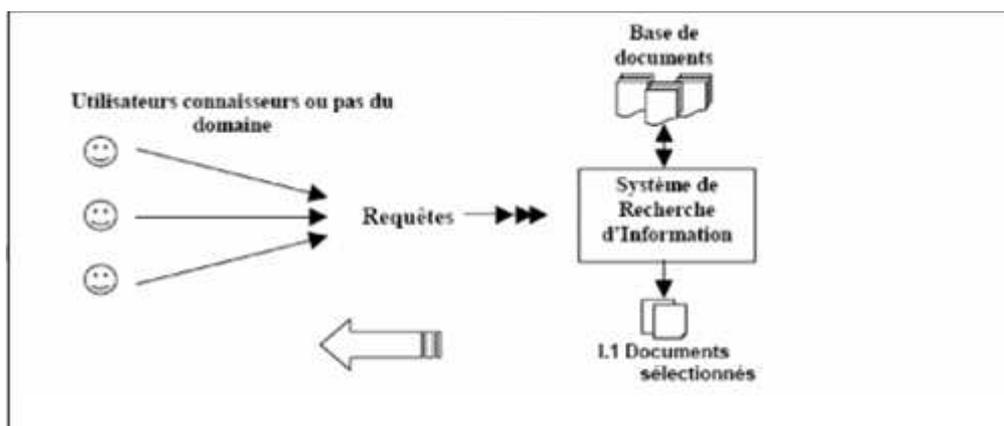


Figure 0.1. Les deux premiers processus transparents à l'utilisateur

2.2.1. Architecture générale des systèmes de recherche d'information

Le processus de recherche d'informations pertinentes que le SRI est sensé restituer à un utilisateur, consiste en la mise en correspondance des représentations des informations contenues dans un fond documentaire et des besoins de cet utilisateur exprimés par une requête. Cette notion de pertinence peut être appréhendée à deux niveaux :

- Niveau utilisateur : la pertinence correspond à la satisfaction de l'utilisateur de l'ensemble des documents restitués par le SRI.
- Niveau système : le système mesure un degré de pertinence, une valeur de similitude entre un document et une requête.

Le but de tout SRI est de rapprocher la pertinence système de la pertinence utilisateur. Pour effectuer de façon efficace cette fonction, le SRI doit réaliser l'indexation des documents, la formulation de la question, la comparaison question - documents et enfin la reformulation de la requête (processus non toujours présent mais important).

Nous pouvons représenter schématiquement un SRI, comme illustré par la Figure II.2, par ce qui est appelé communément le processus en U de recherche d'information.

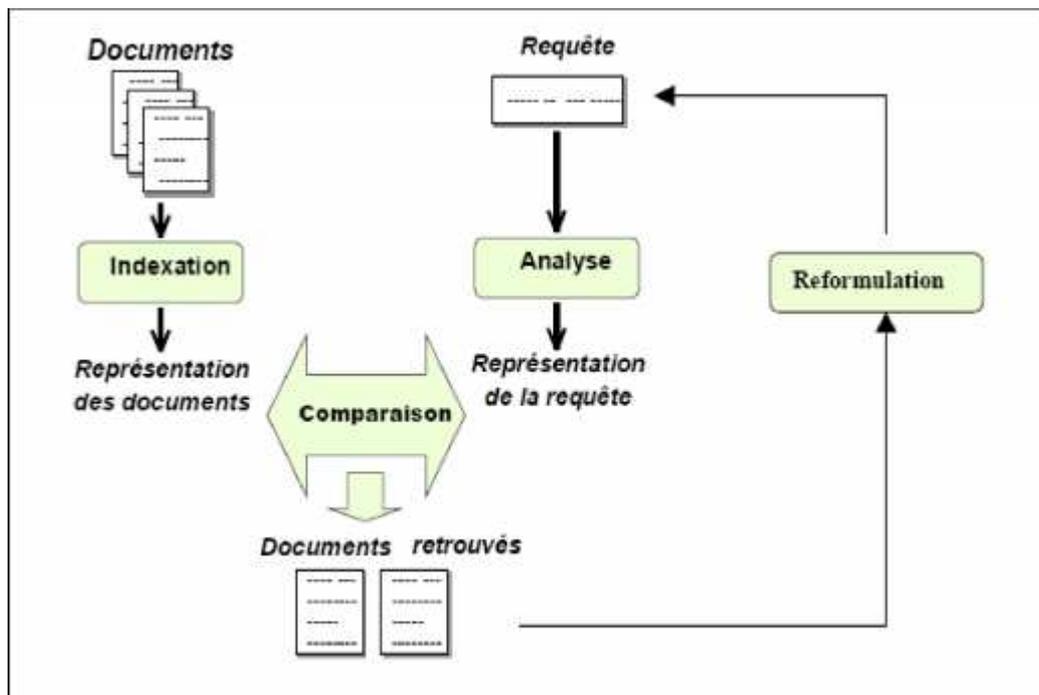


Figure 0.2.L'architecture générale d'un SRI « processus en U »

3. Concepts clés de la recherche d'information (RI) :

Du schéma précédent, on peut dégager les concepts suivants :

Document : le document représente l'information élémentaire d'une base documentaire. Cette information élémentaire, appelée aussi granule de document, peut représenter tout ou une partie d'un document.

Requête : représente l'interface entre le SRI et l'utilisateur. Divers types de langages d'interrogation sont proposés dans la littérature. Une requête est un ensemble de mots clés, mais elle peut être exprimée en langage naturel, booléen ou graphique.

Modèle de représentation : un modèle de représentation est un processus permettant d'extraire d'un document ou d'une requête, une représentation paramétrée qui couvre au mieux son contenu sémantique. Ce processus de conversion est appelé indexation.

Modèle de recherche : Représente le modèle du noyau d'un SRI. Il comprend la fonction de décision fondamentale qui permet d'associer à une requête, l'ensemble des documents

pertinents à restituer. Il est utilisé pour la recherche d'informations proprement dite et est étroitement lié au modèle de représentation des documents et requêtes.

Ces modèles de recherche représentent ce qui diffère le plus entre les SRI. Ils sont inspirés de concepts mathématiques afin de pouvoir évaluer certaines relations, notamment la relation d'appariement entre les termes et les documents, qui permettent au système d'obtenir une valeur de pertinence pour chaque document de la base à partir de laquelle il sélectionne ou non ce dernier. On peut définir différents besoins d'information correspondant à différentes stratégies de recherche :

La demande précise naît quand l'utilisateur sait parfaitement ce qu'il cherche, voire connaît le document dont il a besoin. Le documentaliste doit alors chercher précisément.

La demande exploratoire naît quand l'utilisateur veut se faire une idée d'une collection donnée sans à priori. Il s'agit alors de lui proposer des extraits jugés représentatifs de la base.

La demande thématique est destinée à illustrer un thème. Le type de raisonnement alors suivi par l'utilisateur est un raisonnement pas association d'idées stimulé par la visualisation des documents.

4. Les quatre principaux modes de recherche d'information

Selon la façon dont s'effectue la recherche d'information, on distingue :

- La recherche par navigation arborescente : ce type de recherche s'effectue par menus successifs. L'information est structurée en partant du général au particulier (l'annuaire de Google ou de Yahoo !, la table des matières d'un livre).
- La recherche par navigation hypertextuelle : elle s'effectue par navigation dans un réseau de nœuds et de liens (navigation à travers un ensemble de sites web).
- La recherche par requête sur les métadonnées du document : dans ce mode, l'information est au préalable indexée, en général manuellement. La recherche s'effectue en spécifiant différents champs, souvent liés par des opérateurs booléens (catalogue de bibliothèque).
- La recherche par requête sur le texte intégral : dans ce mode, une démarche d'analyse linguistique est effectuée au préalable avant de procéder à une représentation automatique des documents sur lesquels la recherche doit s'effectuer.

5. Principaux modèles de recherche d'information

Le rôle d'un modèle est tout d'abord de donner une signification au résultat de l'indexation.

Un document est représenté par un ensemble de termes significatifs et leurs poids. S'il est communément admis qu'un terme d'un document présent dans l'index est censé être important pour ce document, la façon d'interpréter son poids (ou son importance) pour le document, peut varier.

Un modèle théorique est censé d'une part donner un cadre d'interprétation précis à ce poids, mais aussi les relations possibles qui peuvent exister entre les termes d'indexation. Ceci nous ramène à la notion de représentation des documents (ou des requêtes, considérés aussi comme des documents). Le modèle doit déterminer la relation entre le document et la requête à partir de leurs représentations, en procédant le plus souvent à un calcul de similarité.

➤ Les modèles booléen et booléen étendu

Le modèle booléen, propose la représentation d'une requête sous forme d'une expression logique. Les termes d'indexation sont reliés par les connecteurs logiques ET (\wedge), OU (\vee) et NON (\neg).

➤ Modèle basé sur les ensembles flous

Une autre extension du modèle booléen est basée sur la théorie des ensembles flous. Dans ce modèle, un document est représenté comme un ensemble de termes pondérés.

➤ Le modèle vectoriel (Vector Space Model)

Dans ce modèle, les requêtes et les documents sont représentés dans l'espace vectoriel engendré par les termes d'indexation L'espace est de dimension N (N étant le nombre de termes d'indexation de la collection de documents). Chaque document est représenté par un vecteur.

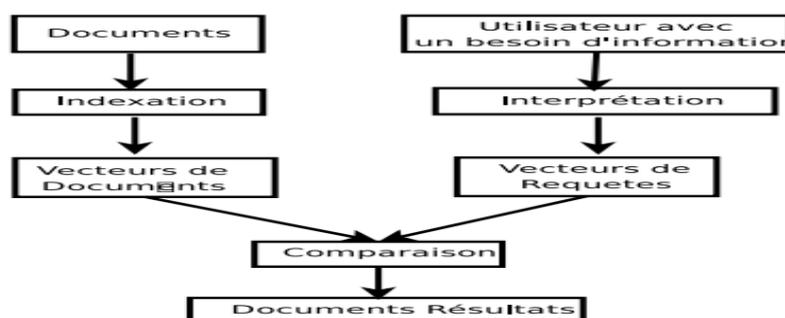


Figure 0.3. Architecture d'un SRI utilisant le modèle vectoriel

➤ **Le modèle probabiliste :**

Le modèle de recherche probabiliste utilise un modèle mathématique fondé sur la théorie de la probabilité. Le processus de recherche se traduit par calcul de proche en proche, du degré ou probabilité de pertinence d'un document relativement à une requête.

➤ **Modèle connexionniste :**

Ce type de modèle se base sur le formalisme des réseaux de neurones. Ce réseau est construit à partir des représentations initiales des documents et des informations descriptives associées (termes, auteurs, mots clés).

L'approche connexionniste permet d'apporter plusieurs fonctionnalités souhaitées dans la recherche d'informations. Elle va d'une simple comparaison des requêtes et des documents à des techniques associatives basées sur des associations de documents pour l'expansion de la réponse (sélection de nouveaux documents).

- 6. Utilisation des ontologies dans le domaine de la recherche d'information

6.1. Lien entre les ontologies et la recherche d'information

Dans le contexte de la recherche d'information, un besoin existe de partager la signification de termes dans un domaine donné (parler le même langage), entre l'utilisateur et le contenu de la collection de documents. Or de nos jours, toute activité humaine spécialisée développe son propre jargon (langue de spécialité) sous la forme d'une terminologie et d'une conceptualisation associée spécifiques.

L'existence de tels jargons entraîne des problèmes de compréhension et des difficultés à partager des connaissances entre les acteurs de l'entreprise, les services d'une entreprise et les entreprises d'une industrie, qui font des métiers différents.

Fondamentalement, le rôle des ontologies est d'améliorer la communication entre humains, mais aussi entre humains et ordinateurs et finalement entre ordinateurs. Nous nous intéresserons dans ce qui suit, à étudier le lien entre les ontologies et la recherche d'information. On peut d'abord se demander si l'utilisation des ontologies dans la recherche d'information est un phénomène récent ou pas. Nous utilisons le principe d'ontologie pour la recherche d'information dans notre vie quotidienne sans se rendre compte. Comme exemple élémentaire, ce que font les utilisateurs d'ordinateurs par exemple lorsqu'ils cherchent un fichier sur leur

disque. Pour accéder au fichier Excel de facture correspondant au mois de janvier de l'année 2001, l'utilisateur de l'ordinateur utilise des connaissances préalables pour distinguer le bon répertoire à chaque niveau de l'arborescence.

6.2. Apports de l'ontologie dans le domaine de la recherche d'information :

De manière générale, ce qui est attendu d'une ontologie, est qu'elle assure la réutilisation de connaissances. Nous donnerons dans ce qui suit quelques gains apportés par l'utilisation des ontologies :

- Les ontologies doivent réduire le silence dans les réponses aux requêtes

Le but est de trouver autant de documents pertinents que possible dans une collection donnée.

À cet effet, les relations et les axiomes d'une ontologie, devraient fournir les moyens de rechercher quelques concepts qui ne sont pas explicitement écrits dans la requête. Pour une requête donnée (éventuellement floue), tous les concepts possibles sont inférés.

- Les ontologies doivent aider à réduire le nombre de réponses bruitées

L'idée est d'ignorer les documents contenant les mots de la requête, mais avec un sens différent. Parce qu'elle contient des définitions non ambiguës, une ontologie devrait être assez précise pour fournir une définition unique à des termes, ou pour traiter la synonymie et l'ambiguïté d'une manière satisfaisante que ça soit dans la représentation d'ontologie ou dans l'annotation de documents. À cette fin, la conception de beaucoup d'ontologies pour la RI est contrôlée par une seule personne.

- Avec l'aide de l'ontologie, l'utilisateur peut exprimer son besoin plus facilement.

Afin de guider l'utilisateur, des étapes peuvent lui être suggérées pour préparer sa requête ou une nouvelle formulation avec des termes plus appropriés. L'ontologie permet d'établir une interface qui le guidera. Le parcourt de l'ontologie, mène à choisir des concepts et à définir une requête composée de concepts choisis et de leur description. Le fait d'exprimer une requête, revient alors à une instanciation des concepts de l'ontologie.

- Les ontologies doivent faciliter la recherche d'information

Dans des sources de données variées et hétérogènes et dans des domaines ouverts pratiquement des travaux dans des domaines ouverts en général, conduisent à de petites avancées dans leurs résultats :

Classer les documents prend du temps et peut mener à quelques erreurs, sauf si un système de désambiguïsation est disponible ; la réponse aux requêtes est beaucoup moins rapide qu'avec les systèmes traditionnels surtout si plusieurs ontologies sont consultées.

7. Conclusion

Dans les systèmes de recherche d'information actuels, il existe des problèmes dans les requêtes et l'utilisation des ontologies dans le domaine de la recherche d'information peut remédier à la résolution de ses problèmes. L'ontologie dans le processus de recherche d'information joue le rôle entre la requête de l'utilisateur et le module chargé de la recherche des documents dans la base.

Dans ce chapitre nous avons parlé de domaine des systèmes de recherche d'information, et ces notions de base, l'architecture générale des SRI, les concepts clé sur lesquels ces systèmes sont basés et la relation entre les ontologies et la recherche d'information, où l'ontologie peut s'avérer comme un interlocuteur entre l'utilisateur et le SR et doivent aider à réduire le nombre de réponses bruitées à cet effet nous allons présenter dans le chapitre suivant les travaux connexes du domaine d'utilisation d'ontologie dans les systèmes de recherche.

Dans le contexte de la recherche d'information, un besoin existe de partager la signification de termes dans un domaine donné (parler le même langage), entre l'utilisateur est le contenu de la collection de documents. Or de nos jours, toute activité humaine spécialisée développe son propre jargon (langue de spécialité) sous la forme d'une terminologie et d'une conceptualisation associée spécifiques.

L'existence de tels jargons entraîne des problèmes de compréhension et des difficultés à partager des connaissances entre les acteurs de l'entreprise, les services d'une entreprise et les entreprises d'une industrie, qui font des métiers différents.

CHAPITRE III:

Les travaux connexes

1. Introduction

L'intelligence artificielle (IA) via l'ingénierie des connaissances et l'informatique médicale s'est spécifiquement intéressée aux ontologies qui semblaient pouvoir résoudre un certain nombre de problèmes amont de la représentation des connaissances. Cet intérêt a été d'autant plus fort en médecine que celle-ci a une forte tradition de nommage des objets médicaux. En effet, les connaissances médicales se structurent depuis le XVIIe siècle autour de l'anatomie. Ainsi, les médecins nomment les structures anatomiques et tous les concepts qui font la médecine. Ils construisent pour cela de nombreux thésaurus. Depuis une dizaine d'années, ces réflexions se sont insérées dans le contexte du renouveau sur la représentation des connaissances et, plus largement, sur la sémantique des objets du web, à savoir le web sémantique. La médecine est alors un terrain d'expérimentation des technologies, langages et standardisations du web sémantique. Terrain d'autant plus intéressant qu'elle conserve ses spécificités en termes de rapports entre l'homme et la machine et de création de thésaurus et classifications en tous genres.

Pour cela, nous allons présenter trois applications caractéristiques, La première application décrite, le projet Onto Pneumo vise à construire une ontologie de la pneumologie pour l'aide au codage médical. La deuxième application est une celle de la pharmacovigilance fondée sur une ontologie construite spécifiquement pour augmenter les signaux de détections de causalité possibles entre un médicament et un effet indésirable. La troisième Application est la Contribution à la construction d'ontologies et à la recherche d'information : application au domaine médical qui permis de créer une ontologie bilingue de la maladie d'Alzheimer, et à la fin c'est l'Indexation et intégration de ressources textuelles à l'aide d'ontologies : application au domaine biomédical.

2. Ontologie de la pneumologie pour l'aide au codage de l'information médical: Onto-Pneumo

Le projet Onto-Pneumo montre comment le web sémantique médical peut (et devrait) prendre en compte le caractère majoritairement documentaire et linguistique de la médecine. La construction de l'ontologie Onto-Pneumo utilise fortement les ressources terminologiques du domaine à modéliser pour rendre compte, le plus précisément possible, non seulement des pratiques médicales actuelles en pneumologie mais également des vocabulaires utilisés par les médecins.

2.1. Objectifs et hypothèse :

L'objectif de ses recherches est de développer un environnement de codage de l'information médicale dépassant le seul cadre médico-économique du PMSI (Le Programme de médicalisation du système d'information), vise à introduire des concepts de comptabilité analytique dans la gestion administrative des hôpitaux.

L'hypothèse de recherche est qu'une telle exigence est atteignable en développant un environnement de codage à base d'ontologie s'appuyant sur l'analyse de travaux existants. En effet, une fois construite et acceptée par une communauté particulière (ici celle des pneumologues), une ontologie doit traduire un consensus explicite et un certain niveau de partage au sein de la communauté, deux aspects essentiels pour permettre l'exploitation des ressources construites par différentes applications ou agents logiciels. Ils ne souhaitent pas complètement automatiser la procédure de codage.

L'idée est plutôt de proposer un système de représentation des connaissances avec lequel le pneumologue puisse interagir pour construire la représentation du patient qu'il désire, en tenant compte de ses propres capacités de choix et d'interprétation mais en l'aidant dans sa tâche. Ils appellent cette représentation des connaissances médicales « codage médical ». A partir d'un tel codage, ils peuvent par exemple proposer au pneumologue de rechercher tous les cas de patients atteints de « sténose serrée de la trachée à la fois par compression extrinsèque et par envahissement de la muqueuse » diagnostiquée par « une endoscopie bronchique ». Pour l'instant aucun outil ne permet de faire de telles recherches. Ils proposent également le codage PMSI, plus classique en médecine mais obligatoire, prenant en compte d'autres critères que les seuls aspects médicaux.

Pour représenter les données propres au patient ils modélisent le thésaurus de spécialité de

la pneumologie sous la forme de concepts primitifs et/ou définis dans l'ontologie. Ainsi, ils ont à même d'identifier et de coder les pathologies listées dans les comptes rendus d'hospitalisation. Le problème du codage se situe au niveau du passage au formalisme, toujours difficile. L'expression des connaissances, qui se présente le plus souvent et le plus naturellement sous forme textuelle, doit être transformée en un codage qui, lui, est toujours réducteur en termes de représentation du sens. Mais c'est parce qu'il est réducteur que ce codage permet au médecin une représentation résumée de ce qu'il sait sur le patient et, par là, un rappel rapide. Il s'agit donc proposer des outils qui permettent le codage en même temps qu'ils permettent aux médecins d'assumer le caractère réducteur de ce processus. De plus, il faut aussi qu'ils puissent interagir avec le système formel dans les termes de leur domaine de spécialité, répertoriés dans les thésaurus de spécialité.

2.2. Matériel : outils et corpus

Ils utilisent le logiciel SYNTEX-UPERY comme outils d'analyse et de traitement (automatique) du langage (TAL). Un outil comme SYNTEX se nourrit d'un corpus de texte qui doit couvrir avec le plus d'exhaustivité possible l'activité étudiée, ici la pneumologie. Dans ce but, ils ont collecté des comptes rendus d'hospitalisation (corpus intitulé [CRH]) dans six hôpitaux de l'Assistance Publique-Hôpitaux de Paris. Au total, ils disposent de 1 038 CRH. Ce premier corpus [CRH] compte environ 417 000 mots. Sachant qu'il a été établi dans [Le Moigno et al. 02b] que 350 000 mots est un minimum pour obtenir de bons résultats avec nos outils, le corpus [CRH] semble être une bonne base d'expérimentation. Le second corpus, intitulé [LIVRE], est construit d'après un ouvrage pédagogique et correspond environ à 823 000 mots.

SYNTEX est un module d'analyse syntaxique fondée sur l'hypothèse que les mots qui ont un sens proche se caractérisent par des dépendances similaires [Bourigault et al, 02]. A la fin du traitement nous obtenons un réseau de dépendances syntaxiques – ou réseau terminologique – dont les éléments sont les candidats termes qui vont-ils servir pour construire l'ontologie. Le module UPERY met ensuite en œuvre le principe de l'analyse distributionnelle [Harris, 68] : il calcule des proximités distributionnelles entre les candidats termes du réseau sur la base des contextes syntaxiques partagés. Ils atteindront un réseau de candidats termes, leurs proximités contextuelles et leurs liens avec le corpus source.

Les résultats de l'analyse sont visualisables dans TERMONTO, l'interface d'accès et de traitement des données du logiciel (figure III.1)

productivité T	E	nb voisins T	E	nb var	nbdoc	freq	fsc1	fec2	cat	terme	validité
22	30	9	27		538	746	741	5	SNom	voies aériennes	⊗⊗⊗⊗⊗
83	57	11	38		585	661	168	493	SNom	Gaz du sang	⊗⊗⊗⊗⊗
44	54	7	5		538	597	40	557	SNom	air ambiant	⊗⊗⊗⊗⊗
41	61	25	65		460	561	397	164	SNom	embolie pulmonaire	⊗⊗⊗⊗⊗
128	31	30	9		267	554	0	554	SNom	1 ep	⊗⊗⊗⊗⊗
44	68	7	37		476	515	184	331	SNom	examen clinique	⊗⊗⊗⊗⊗
53	66	39	88		467	503	345	158	SNom	insuffisance respiratoire	⊗⊗⊗⊗⊗
98	59	67	107		426	482	249	233	SNom	panchement pleural	⊗⊗⊗⊗⊗
30	38	10	1		465	475	87	388	SNom	état général	⊗⊗⊗⊗⊗
79	161	17	71		426	473	92	361	SNom	scanner thoracique	⊗⊗⊗⊗⊗
39	97	8	53	1	400	462	200	262	SNom	radiographie de thorax	⊗⊗⊗⊗⊗
37	88	8	43		439	461	131	330	SNom	radiographie thoracique	⊗⊗⊗⊗⊗
260	37	52	12		375	454	0	454	SNom	cure de chimiothérapie	⊗⊗⊗⊗⊗
66	42	20	10		373	403	83	320	SNom	murmure vésiculaire	⊗⊗⊗⊗⊗
29	73	7	9		339	403	84	319	SNom	membres inférieurs	⊗⊗⊗⊗⊗
46	8	18	25		332	369	269	100	SNom	hypertension artérielle	⊗⊗⊗⊗⊗
21	11	3	8		350	361	140	221	SNom	insuffisance cardiaque	⊗⊗⊗⊗⊗
10	4				353	353	5	348	SNom	ionogramme sanguin	⊗⊗⊗⊗⊗
82	37	13	55		307	337	141	196	SNom	douleur thoracique	⊗⊗⊗⊗⊗
19	19				320	331	23	308	SNom	Bilan hépatique	⊗⊗⊗⊗⊗
23	29	2	36		255	315	298	17	SNom	cancer broncho-pulmonaire	⊗⊗⊗⊗⊗

Figure III.1. TERMONTO : interface d'accès aux données de SYNTAX-UPERY

L'éditeur DOE (The DifferentialOntology Editor), permet de construire notre ontologie selon la sémantique différentielle. L'ontologie est exportée en OWL (Web Ontology Language), un langage de représentation des connaissances préconisé par le consortium W3C. Ce format garantit l'interopérabilité de notre modèle et rend l'ontologie disponible pour l'éditeur PROTÉGÉ.

2.3. Méthode

La méthodologie mise en œuvre permet de décrire les variations des sens des termes considérés en contexte. Ils distinguent quatre étapes : 1) la constitution du corpus des connaissances et son analyse par des outils de TAL, 2) la normalisation sémantique des termes du domaine grâce à la mise en œuvre des principes différentiels, 3) l'engagement ontologique qui permet de formaliser les concepts, 4) l'opérationnalisation de l'ontologie dans un langage de représentation des connaissances interprétable par l'ordinateur [Bachimont et al, 02]. Ils s'intéressent aux étapes trois et quatre.

2.4. Formalisation de l'ontologie avec les logiques de description

Les étapes de formalisation et d'opérationnalisation sont faites à l'aide de l'éditeur d'ontologies PROTÉGÉ 3.16 qui permet de formaliser l'ontologie en utilisant les logiques de description. L'étape de formalisation permet d'introduire des axiomes logiques qui définissent le comportement des individus qui constituent les extensions des concepts formels. Ainsi, nous formalisons précisément chaque concept de la hiérarchie de l'ontologie et définissons les rapports qu'entretiennent les concepts les uns avec les autres par le biais des relations. Par

exemple, l'ingénieur des connaissances décide des relations qui peuvent lier le concept de Médecin et celui d'Asthme : Diagnostiquer, observer. Cette étape de formalisation permet, par exemple, de créer un nouveau concept formel Personne Pneumologue défini qui hérite de son concept père Etre Humain et se spécialise avec la relation A-Une-Spécialité dont le domaine de valeur est Spécialité Médicales qui prend, ici, la valeur Pneumologie.

La liste des Personne Pneumologue recoupe celle des Etre Humain. L'ajout de nouveaux concepts à ce stade de développement modifie la structure hiérarchique car on passe d'une arborescence fondée sur des relations de similarités et de différences à une arborescence fondée sur une logique d'inclusion ensembliste. La structure de l'ontologie n'est plus un arbre mais un treillis.

2.5.Résultats

L'enrichissement manuel de l'ontologie effectué à l'aide des 337 expressions contenues dans le thésaurus de spécialité garantit que presque toutes les expressions du thésaurus sont modélisées (figure III.2). En effet, chacune de ces expressions a un lien (de synonymie, de variances lexicales) avec les concepts de l'ontologie. Par exemple, la partie du thésaurus consacrée aux broncho-pneumopathies obstructives chroniques (groupe des maladies qui implique une limitation du passage de l'air dans les poumons) contient 17 expressions, telle que la bronchite chronique simple. Le niveau de couverture de ce chapitre est de 92 % et implique la modélisation de 38 concepts et 16 relations. Les 8 % restant sont dus à des problèmes d'imprécision. Par exemple, ils peuvent représenter la bronchite chronique simple mais pas la surinfection de la bronchite chronique parce qu'Onto-Pneumo contient le concept infection mais pas le concept surinfection.

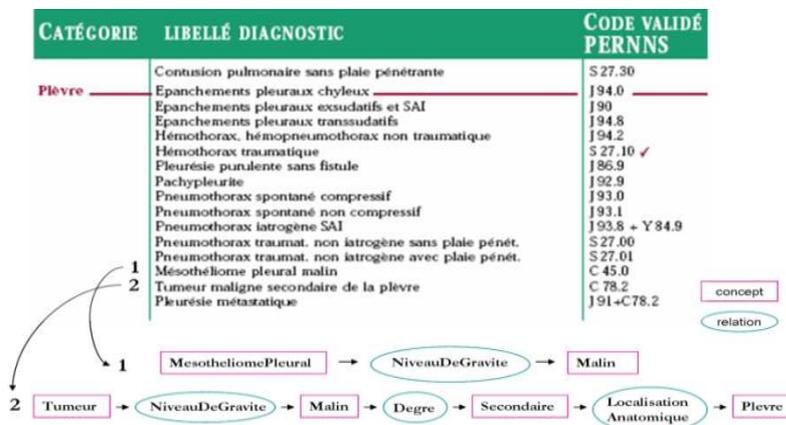


Figure III.2. Modélisation de deux expressions du thésaurus de spécialité

Onto-Pneumo est utilisée au cœur d'un outil de codage médical semi-automatique proposant deux types de codages : a) le codage médical représente graphiquement les informations relatives aux pathologies du patient et, à terme, ces informations serviront de descripteur pour indexer intelligemment les comptes rendus d'hospitalisation. b) le codage PMSI repose au médecin pneumologue une liste de codes médico-économiques générée en fonction des pathologies pertinentes identifiées. Ils obtiennent un rappel de 80 % et une précision de 87 %⁹. La figure 3.3 est une copie d'écran de l'interface de l'outil et en montre les principales fonctionnalités : 1) le chargement du compte rendu et l'identification des informations pertinentes, 2) la représentation graphique du codage médical, 3) les codes médico-économiques proposés et 4) les fonctionnalités d'interaction entre l'utilisateur et l'interface de codage médical.

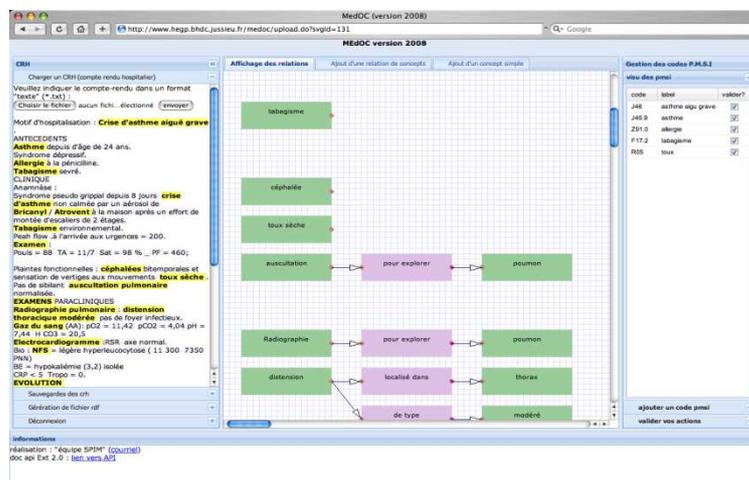


Figure III.3. Interface de l'outil de codage. A gauche, le compte rendu analysé avec Sur lignage des mots repérés comme renvoyant à des concepts. Au milieu, affichage des concepts et de leurs relations. A droite, proposition de codes PMSI

3. Ontologie construite spécifiquement pour augmenter les signaux de détections de causalité possibles entre un médicament et un effet indésirable : Pharmacovigilance

La pharmacovigilance a pour objet la détection, l'évaluation et la prévention des effets indésirables liés aux médicaments (EIM). Il s'agit d'un véritable enjeu de santé publique car les EIM causent aujourd'hui plus de 3 % des hospitalisations en France et sont une des dix principales causes de décès aux Etats-Unis. Ces effets indésirables sont signalés par les professionnels de santé qui envoient les déclarations spontanées aux centres régionaux de

pharmacovigilance ou à l'unité de pharmacovigilance du laboratoire pharmaceutique qui commercialise le produit suspect. L'OMS (Organisation Mondiale de la Santé) définit un signal en pharmacovigilance comme toute information sur une relation causale possible entre un médicament et un effet indésirable, la relation étant préalablement inconnue ou mal documentée [Edwards et al, 94].

La détection d'un signal nécessite le codage reproductible et fiable des effets indésirables; codage sur lequel peuvent ensuite se reposer des regroupements permettant de signaler des EIM de même type. Les terminologies WHO-ART (World Health Organization – Adverse Reaction Terminology) et MedDRA (Medical Dictionary for Drug Regulatory Activities) sont utilisées en pharmacovigilance pour le codage et pour l'analyse statistique des données [Brown, 02].

3.1. Objectif

Son objectif est de construire une ontologie des effets indésirables dans laquelle chaque terme est défini formellement. La conséquence de ces définitions formelles est de structurer les terminologies pour faciliter le regroupement sémantique des concepts qui sont en relation selon certains critères médicaux. L'idée sous-jacente est de pouvoir construire des groupes de termes en fonction de la demande de l'utilisateur en s'appuyant sur la classification automatique des concepts à partir de leur définition formelle. L'étude qui est présentée ici est restreinte à la terminologie WHO-ART.

Cette section consacrée à la pharmacovigilance décrit l'application de plusieurs méthodes du web sémantique :

- l'alignement d'une terminologie (WHO-ART) avec une ontologie,
- la modélisation ontologique d'une liste de termes au moyen d'une méthode d'extraction des connaissances à partir d'un réseau sémantique,
- la structuration de l'ontologie résultante au moyen de concepts primitifs et de concepts définis avec le langage OWL,
- le raisonnement sur l'ontologie afin de construire des groupes de termes qui désignent une condition clinique similaire au moyen du moteur d'inférence RACER [Haarslev et al, 01].

L'étude qui est présentée ici est restreinte à la terminologie WHO-ART (WHO-ART compte environ 1 900 termes préférés et est entièrement incluse dans MedDRA qui compte environ 15 000 termes préférés).

3.2.Hypothèse

Ils ont formulé l'hypothèse que les définitions formelles des concepts de WHO-ART peuvent être obtenues en ajoutant de nouvelles propriétés extraites d'un système terminologique plus sophistiqué tel que la SNOMED-CT. Il s'agit d'une ontologie généraliste destinée à la description des informations dans le dossier patient informatisé qui fait l'objet d'une commercialisation au niveau international et dispose par conséquent de moyens financiers et humains conséquents. Par ailleurs, ils considèrent que L'UMLS (Unified Medical Language System), qui propose une fusion de nombreux produits terminologiques médicaux (dont SNOMED-CT, MedDRA et WHO-ART) est le cadre approprié pour extraire ces définitions formelles.

3.3.Matériel

UMLS est un projet de la « National Library of Medicine (NLM) ».La raison d'être d'UMLS est de regrouper les différents noms/termes et les différentes vues d'un même concept qui proviennent de terminologies médicales diverses ainsi que d'identifier les relations utiles entre ces différents concepts.

WHO-ART est développé par l'OMS et s'organise sur trois niveaux hiérarchiques. Seul le niveau des termes préférés nous intéresse puisque c'est celui qui est utilisé pour coder les effets indésirables. Il existe 1857 termes préférés (par exemple : «Thrombophlebitisleg ») qui sont répartis dans 32 classes organes.

SNOMED-CT, est une ontologie dont l'objectif est le codage de l'information médicale. Elle a le mérite de fournir une description formelle des concepts qu'il inclut [Spackman et al, 98].SNOMED-CT descend de la SNOMED internationale qui est un système terminologique poly-hiérarchique. Son utilisation augmente ainsi la possibilité de regroupements.

Les versions des ressources que nous utilisons sont : UMLS 2005AA ; WHO-ART 2004- third quarter, RACER PRO 1.8, PROTÉGÉ 3.1 et PROTÉGÉ OWL Java API.

3.4.Méthode

Ils proposent une méthode générique pour structurer WHO-ART en s'appuyant sur la hiérarchie de SNOMED-CT. Notre méthode est organisée en plusieurs étapes. Tout d'abord, les termes WHO-ART sont mis en correspondance avec les termes SNOMED-CT en utilisant le méta thésaurus d'UMLS. Deuxièmement, ils extraient la sous-hiérarchie de SNOMED-CT qui correspondra à la hiérarchie des concepts primitifs de son ontologie. Troisièmement, les

définitions des termes WHO-ART (qui désignent les concepts définis de notre ontologie) sont écrites en OWL et automatiquement classées pour produire l'ontologie des EIM. La quatrième étape est la validation.

- Alignement de WHO-ART avec SNOMED-CT

Au cours de la 1^{re} étape, ils créent une table de correspondance entre les termes préférés WHO-ART et leurs synonymes dans la SNOMED-CT. Cette étape s'appuie sur le fait qu'un concept UMLS regroupe des termes synonymes de plusieurs terminologies médicales dont WHO-ART et la SNOMED-CT.

- Extraction de la hiérarchie de concepts primitifs :

A partir des termes SNOMED-CT obtenus dans l'étape précédente, ils reconstruisent la partie de la structure de SNOMED-CT qui inclut tous les termes de WHOART et uniquement ceux-là Figure III.4

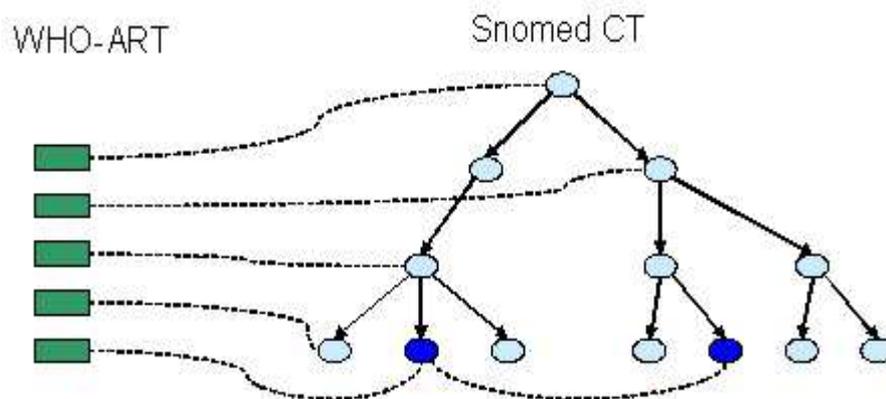


Figure III.4.Correspondance entre les termes WHO-ART et les termes SNOMED-CT

Les concepts de SNOMED-CT qui sont liés aux termes WHO-ART deviennent les concepts primitifs de notre ontologie. Ils considèrent une seule propriété, la synonymie, pour extraire les définitions formelles. La définition d'un concept WHOART, notée Concept W est l'union des restrictions existentielles de tous les concepts SNOMED-CT « synonymes » du concept WHO-ART. Le choix de l'opérateur union provient du fait qu'il permet d'obtenir plus de regroupements que l'intersection (mais il ramène aussi plus de bruit). La condition nécessaire « partial » exprime que le concept défini appartient à la hiérarchie WHO-ART. Ainsi, la hiérarchie des concepts primitifs (SNOMED-CT) est clairement séparée des concepts définis (WHO-ART) (Figure III.5).

Les définitions établies de cette manière sont traitées par un moteur d'inférence afin d'être classées. Le résultat de cette procédure est une nouvelle structure WHOART liée à la SNOMED-CT qui indique, par exemple, que le concept WHO-ART « Thrombosisarterial », est synonyme du concept SNOMED-CT « Arterialthrombosis ».

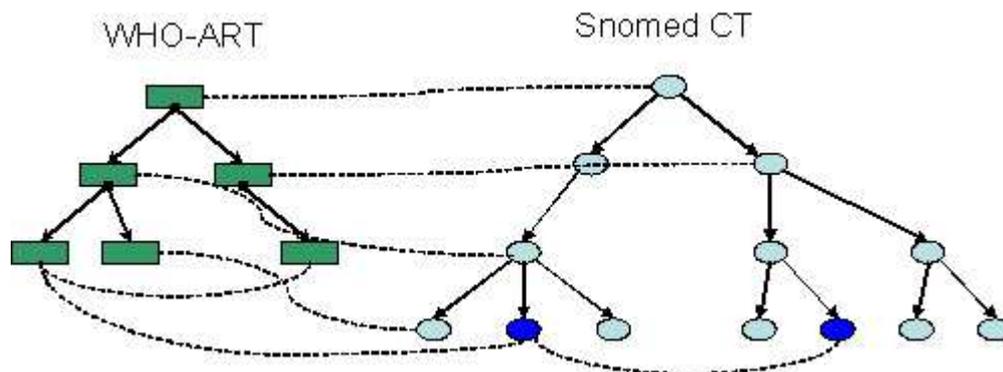


Figure III.5. Hiérarchie des concepts primitifs et définis

3.5. Validation

L'idée principale pour valider la nouvelle structure de WHO-ART est de comparer des groupes de termes élaborés manuellement par les experts et disponibles dans MedDRA avec les groupes de termes obtenus par raisonnement à partir de notre ontologie. Cette stratégie est possible en conservant uniquement les termes MedDRA qui sont également des termes WHO-ART (rappelons que WHO-ART est inclus dans MedDRA) et qui ont un synonyme dans la SNOMED-CT.

Deux situations sont ensuite comparées : la correspondance et la composition. Dans le cas le plus fréquent, la correspondance, le groupe de termes peut être directement créé en sélectionnant un concept SNOMED-CT et en récupérant tous les concepts qu'il subsume. C'est le cas lorsque le terme qui désigne le groupe de termes a directement un synonyme dans la SNOMED-CT. Dans le cas où un seul concept SNOMED-CT n'est pas suffisant, la composition, nous considérons l'union de plusieurs concepts SNOMED-CT.

3.6. Résultats

Au cours de la première étape de la méthode, 85.9%(1 597) des termes WHO-ART ont été mis en correspondance avec les termes SNOMED-CT. L'ontologie contient 7 357 concepts dont 1

596 concepts définis. La hiérarchie de concepts primitifs extraite pour les effets indésirables représente 1.6 % de la SNOMED-CT.

Ils illustrent leurs propos avec l'exemple d'un médicament X qui semble associé à des épisodes d'hémorragie ou de perforation de la partie supérieure du tube digestif. Afin d'extraire de sa base de pharmacovigilance l'ensemble des déclarations spontanées décrivant cette condition clinique, ils ont besoin de constituer le groupe de termes correspondant. Ils recherchent alors les concepts SNOMEDCT qui couvrent le sens de ce groupe de termes et réalisent donc l'union des concepts SNOMED-CT décrivant d'une part les hémorragies digestives hautes et d'autre part les perforations digestives hautes.

Ainsi, la définition correspondant au groupe de termes « saignement ou perforation haute gastro-intestinale » est :

WSC1 = (is_synonym_with some "upper gastrointestinal hemorrhage" or (is_synonym_with some "gastrointestinal perforation"))

La constitution du groupement par les deux concepts SNOMED-CT, upper gastro-intestinal hemorrhage et gastro-intestinal perforation et les concepts subsumés par ceux-là, permet de retrouver 13 concepts. Ce résultat est à comparer aux 18 retrouvés manuellement par les experts. Il correspond à un taux de 76 %. Sur 8 groupes de termes évalués, le taux est en moyenne de 71 % avec des écarts entre 20 et 94 %.

4. Contribution à la construction d'ontologies et à la recherche d'information : application au domaine médical

Ce travail vise à permettre un accès efficace à des informations pertinentes malgré le volume croissant des données disponibles au format électronique. Pour cela, ils ont étudié l'apport d'une ontologie au sein d'un système de recherche d'information (RI).

Ils ont tout d'abord décrit une méthodologie de construction d'ontologies. Ainsi, ils ont proposé une méthode mixte combinant des techniques de traitement automatique des langues pour extraire des connaissances à partir de textes et la réutilisation de ressources sémantiques existantes pour l'étape de conceptualisation. Ils ont par ailleurs développé une méthode d'alignement de termes français anglais pour l'enrichissement terminologique de l'ontologie. L'application de leur méthodologie a permis de créer une ontologie bilingue de la maladie d'Alzheimer.

Ensuite, ils ont élaboré des algorithmes pour supporter la RI sémantique guidée par une

ontologie. Les concepts issus d'une ontologie ont été utilisés pour décrire automatiquement les documents mais aussi pour reformuler les requêtes.

Ils ont intéressé à :

- 1) l'identification de concepts représentatifs dans des corpus.
- 2) leur désambiguïsation.
- 3) leur pondération selon le modèle vectoriel, adapté aux concepts.
- 4) l'expansion de requêtes. Ces propositions ont permis de mettre en œuvre un portail de RI sémantique dédié à la maladie d'Alzheimer.

4.1. La constitution des corpus

Afin de gérer l'aspect bilingue de la ressource à construire, ils ont constitué deux corpus textuels à partir de la base bibliographique Biblio Démences : un corpus anglais composé des titres et des résumés des articles et un corpus français regroupant les titres traduits, les synthèses et les commentaires de ces mêmes articles.

4.1.1. L'extraction des candidats termes

Avant de pouvoir utiliser l'outil Syntaxe pour extraire les candidats termes à partir de ces corpus, le texte doit être préalablement traité par un étiqueteur grammatical

4.1.2. Prétraitement du texte

L'étiquetage grammatical consiste à associer aux mots du texte des informations grammaticales, telles que leur nature (nom, adjectif, verbe, article, etc.) et éventuellement leur forme canonique. Pour étiqueter ses corpus, ils ont utilisé TreeTagger[Schmid, 1994] qui est l'un des étiqueteurs les plus couramment utilisés pour réaliser cette tâche. Il repose sur une méthode d'étiquetage qui utilise les arbres de décisions pour déterminer ces informations.

4.1.3. La construction du noyau ontologique

Cette étape consiste à regrouper les termes synonymes en concepts et à les organiser via des relations sémantiques.

4.1.4. L'enrichissement de l'ontologie

La phase d'enrichissement comprend deux étapes : l'alignement des candidats termes et l'intégration de nouveaux concepts

4.2. Résultats

Dans cette partie, ils présentent les résultats de l'application de leur méthodologie sur le domaine spécifique de la maladie d'Alzheimer.

Un ensemble de 49 390 syntagmes nominaux et 8 844 noms ont été extraits à partir du corpus anglais. De manière similaire, dans le corpus français, 69 505 syntagmes nominaux et 11 688 noms ont été extraits. Le tableau 1 présente les dix syntagmes nominaux les plus fréquents dans les corpus anglais et français. Après filtrage, 2 916 candidats termes anglais ont été retenus, correspondant à 1083 candidats termes simples (c'est-à-dire composés d'un seul mot) et 1833 candidats termes complexes (c'est-à-dire composés de plusieurs mots). Dans le corpus français, 3 152 candidats termes ont été trouvés : 1 196 candidats termes simples et 1 956 candidats termes complexes. A titre d'exemple, les candidats termes *priorresearchet optimal mechanism* sont été élagués car leur fréquence dans le corpus était inférieure à sept.

Syntagme nominal en anglais	Fréquence dans le corpus anglais	Syntagme nominal en français	Fréquence dans le corpus français
Alzheimer's disease	1736	maladie d'Alzheimer	1452
cognitive impairment	764	déclin cognitif	586
mild impairment	437	sujets âgés	428
mild cognitive impairment	419	facteurs de risque	311
cognitive decline	375	performances cognitives	298
cognitive function	273	fonctions cognitives	298
risks factors	241	troubles cognitifs	293
confidence interval	221	risque de démence	275
patients with disease	186	niveau d'études	269
risk of dementia	183	personnes âgées	257

Tableau III.1. Les 10 syntagmes nominaux les plus fréquents dans les corpus anglais et français

- 4.3. L'enrichissement de l'ontologie

Une table de traduction de 492 556 lignes a été générée avec Moses à partir du corpus parallèle. En filtrant ces alignements sur les candidats termes extraits et retenus des corpus, augmentés de leurs synonymes extraits du Meta thesaurus, Il obtient un ensemble de 1959 paires de candidats termes alignés. Les résultats de l'alignement varient en 90 fonctions du seuil de probabilité fixé pour le filtrage. Plus la probabilité est élevée, plus les candidats termes sont susceptibles d'être en relation de traduction. Ainsi, plus le seuil est élevé, plus la précision augmente tandis que le rappel décroît et vice versa.

4.4. Conclusion

L'approche de construction d'ontologies basée sur l'exploitation de corpus textuels et la réutilisation de ressources sémantiques existantes qu'ils ont fait c'est une approche simple qui a permis de générer une ontologie partiellement bilingue de la maladie d'Alzheimer de manière semi-automatique.

Après avoir constitué les corpus de texte, ceux-ci ont été analysés avec l'outil de TAL Syntaxe pour extraire les candidats termes du domaine. L'UMLS a ensuite été utilisé pour regrouper les candidats termes extraits en concepts et les structurer Grâce à des relations taxonomiques et transversales. Toutefois, malgré sa richesse, l'UMLS reste une ressource imparfaite contenant notamment des redondances et des incohérences. Ainsi, des actions correctives ont été proposées pour traiter ces différents problèmes. Par ailleurs, leur approche intègre une phase d'enrichissement comprenant l'alignement de termes de langues différentes ou ils ont utilisé des techniques de traduction automatique et l'intégration automatique de nouveaux concepts au sein de l'ontologie en exploitant les dépendances syntaxiques entre les termes associés à ces concepts. L'ensemble des connaissances ont été validées par des spécialistes du domaine.

L'ontologie développée est actuellement utilisée pour supporter un portail pour la RI et la navigation sémantiques sur une collection de documents scientifiques dédiée à la maladie d'Alzheimer.

4.5. Application pour la mise en œuvre du portail SemBiP

L'objectif du portail SemBiP66 est de permettre à différents utilisateurs, tels que des étudiants, médecins généralistes ou spécialistes, chercheurs et des usagers du grand public, d'accéder facilement à une synthèse critique de la littérature mondiale de référence sur la

maladie d'Alzheimer et les syndromes apparentés

- 4.5.1. Le portail SemBiP

Dans le portail SemBiP, interviennent différents acteurs :

- 1) un administrateur qui assure la gestion des différentes ressources
- 2) les relecteurs qui se chargent de réaliser les analyses critiques des articles sélectionnés
- 3) les utilisateurs de différents types (médecins, étudiants, grand public, etc.) qui peuvent réaliser des recherches sur le portail.

Différentes tâches doivent ainsi pouvoir se réaliser via SemBiP : l'attribution des documents aux relecteurs pour leur analyse, la validation et l'indexation des documents analysés et la fonction principale qui est l'accès à la bonne information. Dans ce qui suit, ils présentent le module de RI du portail.

Pour la mise en œuvre de SemBiP, ils ont exploité les algorithmes décrits précédemment implémentés en Java et disponibles sous forme d'une bibliothèque réutilisable. Au sein de SemBiP sont mises en œuvre :

Une indexation conceptuelle automatique des ressources du portail basée sur

L'ontologie Onto AD, complétée par une indexation en texte libre, en utilisant la bibliothèque open source Apache Lucene67.

Une aide à la saisie de requête pour aider l'utilisateur à exprimer ses besoins grâce à l'implémentation d'une technique d'auto-complétion qui utilise les termes associés aux concepts de l'ontologie Onto AD; Une fonctionnalité de surlignage lors de la présentation des résultats, de manière à ce que l'utilisateur comprenne directement la raison pour laquelle tel ou tel document lui sont retournés.

- 4.5.2. La phase de recherche d'information

Dans la phase de recherche, l'utilisateur exprime son besoin en information sous la forme d'une requête. Cette dernière peut être exprimée en texte libre ou via un formulaire où l'utilisateur peut sélectionner facilement les concepts correspondant à son besoin. Dans le premier cas, sa requête est analysée et les concepts correspondants sont identifiés. Dans le deuxième cas, l'utilisateur sélectionne lui-même les concepts qui vont constituer sa requête.

Après le traitement de la requête de l'utilisateur, cette dernière est soumise au SRI, qui retourne l'ensemble des documents jugés pertinents pour la requête. Ainsi, chaque document retrouvé est associé à un score qui représente sa pertinence par rapport à la requête. Pour cela, à l'instar des

documents, la requête est d'abord représentée par un vecteur de concepts.

Ensuite, la mesure du cosinus est utilisée pour calculer sa pertinence pour chaque document de la collection. Les documents sont retournés par ordre décroissant de leur pertinence.

Pour améliorer les performances des SRI, différentes stratégies permettant d'optimiser la correspondance entre les documents et les requêtes peuvent être envisagées. Dans ce travail, ils proposent d'implémenter deux techniques : une technique d'expansion de requêtes exploitant la similarité sémantique et une technique de combinaison de la recherche sémantique et la recherche par mots clés.

5. Indexation et intégration de ressources textuelles à l'aide d'ontologie : application au domaine biomédical

L'utilisation d'ontologies pour indexer et intégrer les ressources de données est un moyen de valoriser la connaissance d'un domaine en facilitant la recherche et la fouille de données. Dans cet article ils nous présentent un mécanisme d'indexation de ressources de données textuelles dirigé par les ontologies. Ils nous détaillent la création et l'utilisation d'un index de ressources de données biomédicales qui fournit un accès uniforme à plus d'une vingtaine de ressources indexées avec plus de 200 ontologies.

Cet index est accessible via la plateforme Web Bio Portal du Centre National pour les Ontologies Biomédicales (NCBO) : <http://bioportal.bioontology.org/>.

Au sein du projet NCBO, ils ont développé un système d'indexation de données dont l'objectif est d'indexer les éléments de diverses ressources biomédicales publiques non pas par mot clé, mais par concept d'ontologie. Pour chaque élément d'une ressource de données, le mécanisme d'indexation repose sur la génération d'annotations sémantiques et sur l'agrégation de ces annotations.

Le processus d'indexation est composé de quatre étapes :

1. Récupération 'contextualisée' des données :
2. Génération des annotations directes
3. Expansion sémantique des annotations :
4. Agrégation et score des annotations :

- 5.1. L'index de ressource biomédicale du NCBO :

Le NCBO développe et maintient une application Web appelée *Bio Portal* (<http://bioportal.bioontology.org>) qui permet d'accéder, visualiser, rechercher et commenter des ontologies biomédicales. Le portail contient une grande collection d'ontologies, et permet de gérer leur évolution dans le temps. Les utilisateurs ont accès à la plateforme soit via l'interface graphique, soit via une interface de service Web.

Bio Portal fournit un accès uniforme à un grand nombre d'ontologies. Ainsi, ils ont mis en œuvre le système d'indexation présenté précédemment avec une des plus grandes collections d'ontologies biomédicales publiques disponibles (215 ontologies au 04/02/2010). Ces ontologies permettent de générer un dictionnaire contenant 3 653 128 concepts et 7 061 411 termes. Comme outil de reconnaissance de concept, le système utilise Mgrep [Dai et al., 2008], développé à l'Université du Michigan qui présente un degré de précision élevé pour la reconnaissance par exemple de noms de maladie [Xuan et al., 2007]. Ils ont réalisé une évaluation comparative de Mgrep avec MetaMap [Shah et al., 2009a], l'outil de reconnaissance de concept qui sert de référence dans le domaine biomédical. Dans cette évaluation ils ont utilisé quatre ressources de données et quatre dictionnaires différents pour évaluer les deux outils en termes de précision, rapidité d'exécution, et de passage à l'échelle. Le tableau 1 reporte certains de ces résultats pour deux des dictionnaires. Mgrep s'est montré très rapide avec une précision plus élevée pour la plupart des ressources de données traitées et pour les quatre dictionnaires. En outre, Mgrep est ouvert à tout type de dictionnaire et n'est pas limité, comme MetaMap, au méta thesaurus UMLS (Unified Medical Language System). Ils ont donc pu utiliser Mgrep pour traiter les ontologies NCBO qui sont généralement disponibles dans les formats OBO et OWL.

Ressource de données biomédicale	Maladies (termes d'UMLS)		Biological processes (GO)	
	Mgrep	MetaMap	Mgrep	MetaMap
ClinicalTrials.gov	0,87	0,71	0,6	0,63
Gene Expression Omnibus	0,88	0,755	0,93	0,73
ARRS GoldMiner	0,73	0,548	0,58	0,33

Tableau III.2:Précision de Mgrep et MetaMap avec deux dictionnaires

Ils ont indexé 22 ressources biomédicales publiques (au 04/02/2010 cf. http://rest.bioontology.org/resource_index/resources/list/) de différentes tailles (ex : <800k éléments et <180Mo). Les domaines couverts par ces ressources sont divers et comprennent des données d'expression génétique (e.g, ArrayExpress, GEO), des descriptions d'essais cliniques

(Clinicaltrials.gov), ou des légendes et descriptions d'images radiologiques (e.g. ARRS Goldminer). L'index contient plus de 2 milliards d'annotations agrégées et 10 milliards d'annotations. L'index fournit des annotations pour tous les éléments des ressources traitées. Le nombre moyen d'annotations est situé entre 359 et 824 par élément, avec une moyenne de 27% d'annotations directes. L'index est interfacé par une API[§] services Web REST (REpresentational State Transfer) et les annotations peuvent être retournées aux utilisateurs dans différents formats: texte, tab-delimited, XML ou RDF/OWL. Le contenu de l'index est également accessible via l'interface graphique du Bio Portal.

- 5.2. Cas d'utilisation et scénarios supportés

Les annotations de l'index peuvent être accédées *par concept(s)* ou *par élément*. Ces deux modes permettent différents cas d'utilisation :

1. l'obtention, pour un concept donné, de l'ensemble des éléments d'une ou plusieurs ressources annotés avec ce concept ex : les éléments de GEO et ArrayExpress annotés avec DOID:1909.
2. l'obtention, pour un ensemble de concepts donnés, de l'intersection ou de l'union des ensembles d'éléments annotés avec ces concepts ex : les éléments de GEO et ArrayExpress annotés à la fois par DOID:1909 et CL:0000148.
3. l'obtention, pour un élément d'une ressource, de l'ensemble des concepts d'une ou plusieurs ontologies qui annotent cet élément ex : les concepts du NCI thesaurus qui annotent l'élément GDS1965 de GEO.

Les cas 1 et 2 sont des cas d'utilisation de recherche d'information dans un ensemble de ressources hétérogènes. L'index est dans ce cas utilisé comme un outil d'intégration de données qui propose une interface unique pour interroger plusieurs ressources dont les modèles de représentation sont différents [Lenzerini,2002]. Ils ont pu remarquer qu'une recherche faite avec une conjonction de peu de concept (i.e. 2 ou 3) réduit suffisamment le nombre de résultats pour les rendre facilement consultable. Le troisième cas d'utilisation est plus atypique dans le sens où il informe sur le contenu d'un élément déjà identifié au sein d'une ressource.

Dans les trois cas d'utilisation, les résultats sont retournés aux utilisateurs sous formes d'un ensemble d'annotations. Les scores affectés aux annotations lors de l'indexation permettent de classer les résultats par ordre de pertinence. Il est possible d'obtenir directement de l'index des annotations détaillées, mais le plus souvent, il est plus intéressant d'obtenir des annotations agrégées et si nécessaire, obtenir le détail des annotations à l'origine de cette agrégation. Un

service Web est disponible pour chacun de ces trois cas d'utilisation. En conséquence, des utilisateurs peuvent composer ces services pour supporter des cas d'utilisation plus complexes. Par exemple, l'obtention, pour une ressource et un concept donné, des éléments annotés avec le meilleur score (cas 1), puis pour l'ensemble de ces éléments obtenir les concepts les plus représentatifs (cas 3).

Ces trois cas supportent différents scénarios de recherche, par exemples :

- Les chercheurs travaillant sur le projet Trialbank (<http://www.trialbank.org>) à l'Université de Californie à San Francisco, utilisent des annotations d'essais cliniques sur le VIH/sida de ClinicalTrials.gov afin de développer une application Web pour visualiser et comparer les essais. En particulier, ils exploitent l'origine des annotations (i.e. le champ de métadonnées) pour désigner leur interface.
- Des chercheurs de l'université de Washington utilisent l'index pour connecter de nombreuses ressources de données sur les nanoparticules. Ils exploitent les ontologies pour la recherche d'information sur ces ressources.
- Des chercheurs du Medical College of Wisconsin utilisent les annotations d'expérimentation d'expression génétique de GEO pour fouiller cette ressource et accélérer significativement la construction d'une base de données sur le génome du rat qui supporte leur recherche sur les associations gène-maladies et gène-phénotype. Ils ont développés une plateforme Web de curation et d'exploration des annotations de l'index (<http://gminer.mcw.edu>).
- A l'Université Stanford, ils utilisent l'index pour fouiller des données relatives au financement de projets scientifiques et à leur impact en termes de publications. Ils utilisent les concepts d'ontologies comme dénominateur commun entre une ressource de données financières (<http://www.researchcrossroads.org/>) et une ressource de publications (PubMed). Ils évaluent alors si les tendances des publications reflètent, dans le temps, celles des financements.

6. Comparaison entre les travaux

Onto pneumo : vise à construire une ontologie de la pneumologie pour l'aide au codage médical. Il propose un système avec lequel le pneumologue peut construire la représentation des connaissances du patient (codage médical);ils modélisent le thésaurus de spécialité de la pneumologie sous la forme de concept primitifs ainsi ils ont codes et identifier les pathologies listées dans les compte rendus d'hospitalisation. Ils utilisent le logiciel Syntex-Upery comme outil

d'analyse et de traitement automatique du langage et l'éditeur DOE qui permet de construire l'ontologie.

Pharmaco vigilance : l'objectif est de construire une ontologie des effets indésirables. L'idée est de construire des groupes de terme en fonction de la demande de l'utilisateur en s'appuyant sur la classification automatique des concepts à partir de leur définition formelle, ils considèrent que l'UMLS (unified Medical Language System) qui propose une fusion de nombreux produits terminologiques médicaux (Snomed-Ct, MedDra, Who-Art) est le cadre approprié pour extraire ces définitions formelles ; la conséquence de ces définitions formelles est de structurer la terminologie pour faciliter le regroupement sémantique des concepts qui sont en relation selon certains critères médicaux. Les définitions formelles des concepts de Who-Art peuvent être obtenues en ajoutant des nouvelles propriétés extraites d'un système terminologique.

Contribution à la construction d'ontologies et à la recherche d'information: application au domaine médical :

visent à permettre un accès efficace à des informations pertinentes sur la maladie d'Alzheimer malgré le volume croissant de la donnée. Ils utilisent des techniques de traitement de langue pour extraire les connaissances à partir du texte et une méthode d'alignement de terme Français Anglais pour l'enrichissement terminologique de l'ontologie, à partir de ça ils construisent une application appelée le portail SemBip. L'objectif du portail Sem Bip est de permettre à différents utilisateurs, tels que des étudiants, médecins d'accéder facilement à une synthèse critique de la littérature mondiale de référence sur la maladie d'Alzheimer et les syndromes apparentés.

Indexation et intégration de ressources textuelles à l'aide d'ontologies

application au domaine biomédical :

Ils nous présentent un mécanisme d'indexation de ressources de données textuelles dirigé par les ontologies. Ils détaillent la création et l'utilisation d'un index de ressources de données biomédicales qui fournit un accès uniforme à plus d'une vingtaine de ressources indexées avec plus de 200 ontologies. Le NCBO développe et maintient une application Web appelée Bio Portal. Bio Portal fournit un accès uniforme à un grand nombre d'ontologies. Ainsi, ils ont mis en œuvre le système d'indexation présenté précédemment avec une des plus grandes collections d'ontologies biomédicales publiques disponibles (215 ontologies). Comme outil de

reconnaissance de concept, le système utilise Mgrep et MetaMap, l'outil de reconnaissance de concept qui sert de référence dans le domaine biomédical.

6.1. Discussion des résultats:

Pour le premier travail il garantit que presque toutes les expressions sont modélisées et chacune a un lien avec les concepts de l'ontologie. 92% modélisation de 38 concepts et 16 relations avec un Rappel de 80% et Précision de 87%. Le travail concernant la Pharmaco vigilance a un taux de 76% sur 8 groupes de termes évalué. Le troisième travail, Pour mesurer la pertinence des services du portail Sem Bip, ils ont réalisé une évaluation système pour l'efficacité de la stratégie d'expansion de requête et une évaluation orientée utilisateur. L'évaluation est toujours en cours mais les premiers retours sont positifs. Dans le dernier travail, Ils comparent Mgrep avec MetaMap .Ils ont utilisé quatre ressources de données et quatre dictionnaires différents pour évaluer les deux outils en termes de précision, rapidité d'exécution, et de passage à l'échelle. Mgrep s'est montré très rapide avec une précision plus élevée pour. En outre, Mgrep est ouvert à tout type de dictionnaire et n'est pas limité, Ils ont donc pu utiliser Mgrep pour traiter les ontologies NCBO.

- 7. Conclusion

Le web sémantique réclame par définition de construire des ontologies permettant, par leur caractère formel, d'automatiser un certain nombre de tâches. Dans cet article, nous avons voulu montrer que la médecine a des spécificités qui contraignent l'usage des ontologies et leur construction : Les thésaurus ont permis de satisfaire les buts de conceptualisation liée à l'exploitation, recherche d'information, classification, et le partage des données. Une meilleure automatisation et prise en compte de la dimension des données médicales passe par l'usage des ontologies dans le contexte du web sémantique.

Dans ce contexte, la construction des ontologies peut prendre en compte l'expression linguistique des connaissances de deux façons : a) en tenant compte des thésaurus dont la dimension terminologique a été historiquement prise en compte à la création ou b) en utilisant les textes générés par les professionnels de santé durant leur activité qui fournissent une ressource traçant les conceptualisations d'un domaine.

•

CHAPITRE IV :

Proposition d'une
application médicale

1. Introduction :

Après avoir abordé dans la partie théorique, les concepts fondamentaux des ontologies, les systèmes de recherche d'information ainsi que l'utilisation des ontologies dans le processus de la recherche d'information et une citation de quelques travaux connexes on passe à la partie pratique où nous allons expliquer notre projet.

Dans le domaine médical le bon diagnostic, le bon suivi et avoir des informations détaillées sur les patients permet aux médecins spécialistes de bien choisir les médicaments et prescrire le meilleur traitement qui convient, même les pharmaciens peuvent conseiller un patient de prendre un médicament selon les symptômes ainsi que des informations médicales sur ce dernier. Les médicaments sont responsables d'effets indésirables ces effets peuvent être connus et observés pendant les phases expérimentales de leur développement, ou encore apparaître lors de leur commercialisation et de leur usage à grande échelle.

Notre objectif c'est d'aider le médecin ou le pharmacien de prendre une décision optimale concernant le médicament idéal à partir de connaître toutes les informations qu'ils ont besoin de savoir sur le médicament et le patient pour éviter les effets indésirables sur la santé de ce dernier. Les médicaments sont responsables d'effets indésirables ces effets peuvent être connus et observés pendant les phases expérimentales de leur développement, ou encore apparaître lors de leur commercialisation et de leur usage à grande échelle.

Dans ce contexte, nous tentons de proposer dans cette deuxième partie de notre mémoire une application qui a d'une part, un moteur de recherche pour la gestion des médicaments qui aide les médecins et beaucoup plus les pharmaciens de choisir idéalement le médicament. D'autre part une application pour la gestion du patient qui présente l'automatisation de la tâche médicale concerne le choix du médicament qui convient à partir des données de patient qu'on les déclare en entrée. Cette étude comporte deux phases importantes de représentations de connaissances qui sont : la construction d'une ontologie des médicaments (MED) et la conception du système de recherche d'information fondé sur l'ontologie construite.

2. Problématique :

Les ontologies ont connu un grand succès dans le domaine de la santé. Etant l'une des meilleurs moyens de la représentation des connaissances, elle permet de donner une description précise et exploitable par la machine d'un univers de discours, d'après ce qu'on a vu dans le chapitre précédent ce moyen est généralement utilisé dans le domaine de la santé comme une base pour la conception des applications de classification, stockage des ressources terminologiques et recherche d'information. L'idée d'utiliser l'ontologie comme une base de connaissance pour la recherche d'information médicale est inspiré par un travail (pharmacovigilance) réalisé par L'OMS (Organisation Mondiale de la Santé) qui définit un signal en pharmacovigilance comme toute information sur une relation causale possible entre un médicament et un effet indésirable, la pharmacovigilance a pour objet la détection, l'évaluation et la prévention des effets indésirables liés aux médicaments (EIM). C'est une application fondée sur une ontologie construite spécifiquement pour augmenter les signaux de détections de causalité possibles entre un médicament et un effet indésirable.

Le domaine d'étude qu'on a choisis à notre projet est dans le cadre de la pharmacie afin de faire la prescription des médicaments ayant le minimum des effets indésirables (EI) sur le patient en prenant compte ses caractéristiques et ses maladies, il s'agit d'un véritable enjeu de santé publique car les EIM causent aujourd'hui plus de 3 % des hospitalisations en France et sont une des dix principales causes de décès aux Etats-Unis.

Notre projet vise à développer une application médicale qui aide les médecins et les pharmaciens à éviter les effets indésirables liés aux médicaments (EIM), donc on va construire une ontologie (MED) comme une base de connaissance pour le stockage connecté avec l'application qui contient deux interfaces.

Les contributions principales présentées dans ce mémoire, s'articulent autour:

- La modélisation du domaine étudié à travers la conception d'une ontologie qui doit comprendre les concepts liés aux médicaments, et à la structure organisationnelle du cadre médical. Cette ontologie est le noyau de notre application et la base de connaissance du quelle on va extraire les résultats.

- Un processus de recherche d'information qui répond au besoin des utilisateurs et extrait les résultats pertinents.
- La construction d'une application qui contient deux interfaces: La 1^{ère} interface c'est du Gestion patient et la 2^{ème} Gestion médicament.

3. Approche proposée :

Notre approche est détaillée dans la figure suivante qui présente les trois étapes de notre travail :

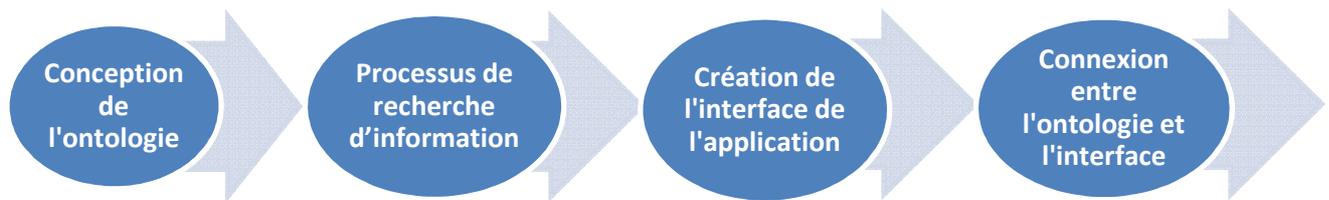


Figure IV.1. Le processus de développement de l'application proposée

3.1. Conception de l'ontologie du domaine :

Notre approche se base sur l'utilisation d'une ontologie, nous avons modélisé le domaine choisi dans le sens où elle répond aux objectifs déjà prévus, on définissant d'abord l'objectif de la conception ainsi que le scénario de l'utilisation, ensuite tous les concepts liés au domaine et les relations entre eux, cette modélisation a été basé sur les compétences et les validations d'un médecin du domaine et des pharmaciens.

La construction d'une ontologie d'un domaine à partir de zéro nécessite un professionnel informatique avec la validation d'un spécialiste du domaine étudié (un médecin ou un pharmacien). Les grandes étapes de ce processus sont inspirées de la méthodologie de construction d'ontologies « METHONTOLOGY ».

L'objectif principal de la construction de l'ontologie sera l'exploitation de son potentiel sémantique dans la recherche d'information qui sera concrétisé dans notre système de recherche d'information. Nous avons décidé de construire une ontologie qui va couvrir la branche étudiée et qui doit servir de base de connaissance pour la construction de l'application proposée. Le procédé de la construction de notre ontologie passe par quatre phases : spécification, conceptualisation, formalisation, implémentation.

- Phase 1 : Spécification : cette étape consiste à établir un document spécifiant des réponses à des questions fondamentales à la construction de l'ontologie. Les

connaissances qui seront représentées extraites principalement à partir d'un dictionnaire des médicaments VIDAL et validées par un médecin. Dans notre quête on collecte tous les informations et les aspects qui décrivent les médicaments pour bien modéliser le domaine étudié.

Ontologie	<ul style="list-style-type: none"> • Ontologie des médicaments
Objectifs	<ul style="list-style-type: none"> • Une base de connaissance pour faciliter la prise de décision dans la prescription des médicaments ayant le minimum des effets indésirables.
Domaine	<ul style="list-style-type: none"> • De la pharmacie (médicament).
Source	<ul style="list-style-type: none"> • Une étude de domaine en utilisant un dictionnaire des médicaments VIDAL et la validation d'un médecin et un pharmacien.
Utilisateurs	<ul style="list-style-type: none"> • Les médecins et les pharmaciens.
Scénario d'utilisation	<ul style="list-style-type: none"> • Interroger la base de connaissance (ontologie) pour extraire l'information médicale idéale (le médicament qui répond au besoin).

FigureIV .2. Spécification des besoins de l'ontologie

Phase 2 : Conceptualisation :

C'est l'étape la plus importante dans le processus de construction de l'ontologie, elle consiste à identifier et à structurer, à partir des sources d'informations, les connaissances du domaine c'est lors de cette étape qu'on définit les concepts, les relations, les attributs et les axiomes qui seront décrits dans l'ontologie. L'activité de conceptualisation va se dérouler sur les étapes suivantes :

- Construction du glossaire de termes (concepts) : A partir des sources de données que nous avons décrites précédemment dans l'étape spécification, nous avons construit un glossaire de termes qui regroupe les principes les plus importants dans le domaine de la pharmacie. Le tableau qui suit représente une partie des concepts regroupés dans le glossaire ainsi que leurs descriptions.

Nom du concept	Synonyme (anglais)	Description	Type
Patient	Patient	Le malade	classe
Nom commercial (NomC)	Commercial name	Est choisi par le producteur du médicament	classe
Nom pharmaceutique (NomP)	Pharmaceutic name	Est en fait la Dénomination commune internationale, la DCI. Celle-ci est attribuée par l'Organisation mondiale de la santé (OMS)	classe
Classe thérapeutique	Therapeutic class	Les médicaments sont classés en fonction de leurs effets, par exemple les anti-inflammatoires »	classe
Indication	Indication	Un signe ou une raison motivant la prescription d'un traitement ou d'un médicament pour une maladie	classe
Contre indication	Contraindication	Une condition qui interdit la prise d'un médicament.	classe
Forme	Shape	La forme d'un médicament : comprimé, sirop, suppositoire, gélule...	
Age	Age	L'âge qu'autorise le patient la consommation de tel médicament	
Effets indésirables (...)	Side effects (...)	<ul style="list-style-type: none"> • Réaction nocive et non voulue à un médicament • 	<ul style="list-style-type: none"> • •

Tableau IV.1. Extrait du glossaire des termes

- Construction du glossaire de sous-classes : Le tableau suivant montre certains sous-concepts présents dans notre ontologie.

Concept	sous-concept
Nom commercial (NomC)	ACTIFED ALVITYL ASPEGIC100mg LAROXYL50mg/2ml VENTOLINE100µg

Nom	paracétamol, pseudo éphédrine, diphénhydramine
pharmaceutique (NomP)	vitamines acétylsalicylate de DL-lysine amitriptyline salbutamol
Classe thérapeutique	Décongestionnant ORL vitamines associées Antalgique, antipyrétique, anti-inflammatoire Antidépresseur imipraminique Antiasthmatique
Indication	<ul style="list-style-type: none"> • nez bouché, écoulement nasal clair, maux de tête, fièvre • déficit lié à une alimentation insuffisante ou déséquilibrée • douleur d'intensité légère, rhumatisme inflammatoire chronique • énurésie nocturne, dépression, algie rebelle • risque d'asthme, exacerbations d'asthme, asthme d'effort
Contre indication	<ul style="list-style-type: none"> • glaucome à angle fermé, déconseillée pendant la grossesse, contre-indiquée pendant l'allaitement. • allergie au salicylé et paracétamol, Ulcère gastroduodéal en évolution, maladie hémorragique, Insuffisance hépatocellulaire. • allergie au paracétamol, Insuffisance hépatocellulaire. • allergie au tramadol et paracétamol, Intoxication aiguë par l'alcool. • Antécédents d'infarctus du myocarde, pathologie cardiaque ischémique, vasospasme coronarien, pathologie vasculaire
Forme	Gélule Solution injectable Crème Solution pour pulvérisation nasale Collyre, flacon
Age	Adulte Enfant à partir 1 mois à 12 ans Nourrisson de 1 à 4 mois De la naissance à 18 mois Nourrisson, enfant, adulte
Effets indésirables	Irritation oculaire Météorisme, Diarrhée

Risque de crise de glaucome aigu par fermeture de l'angle Eczéma allergique de contact, Photosensibilisation Constipation, flatulence, nausées, douleurs épigastriques, diarrhées • (...) (...)
--

Tableau IV .2 . Glossaire des sous-classes

- Construction de glossaire de relations : Les relations dans une ontologie relient le concept cible à un concept source et décrivent une interaction entre les notions d'un domaine. Le tableau suivant présente certaines des relations existantes dans notre ontologie.

Relation	Concept cible	Concept source
A	Nom commercial	Nom pharmaceutique
A	Nom commercial	Classe thérapeutique
A	Patient	Age
Dépend de	Indication	Symptome
Dépend de	Contre indication	Grossesse, Allaitement
Dépend de	Contre indication	Allergie
(...)		

Tableau IV.3. Glossaire de relation

- Construction du glossaire d'attributs : Nous avons défini des attributs essentiellement pour les concepts du patient. Le tableau suivant présente certains des attributs définis dans notre ontologie.

Attribut	Domains	Ranges
Nom	Patient	String
Prénom	Patient	String
Age	Patient	Int, String
Allergie	Patient	String
Sexe	Patient	String
Symptome	Patient	Int
(...)		

Tableau IV.4. Glossaire d'attributs

Le résultat de cette étape de conception est un glossaire des termes, sous termes et relations qui représente le domaine étudié, le but maintenant est de coder cette ontologie en OWL DL pour obtenir une ontologie opérationnelle ça ce que nous allons expliquer dans le chapitre suivant.

3.2. Un processus de recherche d'information :

Cette étape présente un système de recherche d'information qui se base sur notre ontologie Pour satisfaire les besoins de l'utilisateur et permettre de récupérer des documents pertinents et relatifs à sa recherche. Ce processus est relié avec l'interface qu'on va détailler son accès.

Dans ce système nous avons implémenté trois principaux modes de recherche :

- Mode de recherche par exploitation de l'arborescence de l'ontologie: Ce mode offre la possibilité d'accéder à l'ensemble des concepts constituant l'ontologie des médicaments.
- Mode de recherche par mots clés : dans ce mode nous avons implémenté une simple.
- Recherche en langage naturelle, qui permettra à un utilisateur non professionnel de recherche des informations sur un concept donné même s'il ignore l'appellation exacte du sous-concept nous ajoutons aussi la possibilité de préciser et bien choisir ce dernier dont l'utilisateur fait ça recherche.
- Mode de recherche Guidée: dans ce mode nous avons implémenté une recherche qui repose sur le langage d'interrogation qui nous donne l'avantage de faire la recherche guidé par quelques informations sur le patient que l'utilisateur doit les mentionner en entrée comme paramètre.

3.3. Création de l'interface de l'application :

Le rôle de ce module est de proposer une interface conviviale entre utilisateur et le système de recherche d'information basé sur l'ontologie. Il permet la collecte les informations nécessaires sur les requêtes des utilisateurs, et la restitution des résultats obtenus après la recherche. Les taches de module sont :

-Chargement de l'interface de choix : lorsque l'utilisateur se connecte au système, le module charge l'interface du choix, ensuite le chargement de l'interface choisit (1-Gestion patient, 2-Gestion médicament).

-Affichage du contenu de l'ontologie :

Dans le cas où l'utilisateur choisit (1-Gestion patient) le module affiche une fiche à remplir en entrée par les informations sur le patient selon les concepts suivants (symptôme, allergie, âge, sexe : homme/ femme (grossesse et allaitement)) afin de faire la recherche sur l'ontologie que nous avons déjà construire comme une base de connaissance MED pour nous extraire comme

résultat le médicament qui convient.

Au cas où l'utilisateur choisit (2-Gestion médicament) le module affiche une interface qui présente un moteur de recherche de médicament selon 3 concepts (le nom commercial, dénomination et la classe thérapeutique) relié avec notre ontologie MED. Ce moteur nous permet d'avoir des informations détaillés sur ce dernier (nom commercial, dénomination, classe thérapeutique, forme, indication, contre-indication, âge, effet indésirable).

3.4. Connexion entre l'ontologie et l'interface :

C'est la dernière étape dans notre approche, elle consiste à faire interagir notre interface graphique et la base de connaissance, pour interroger l'ontologie afin d'extraire la connaissance médicale. La base de connaissance est le noyau de notre application elle a permis de faire la capitalisation, l'organisation, et la représentation des connaissances de notre domaine d'étude, pour cela cette étape est très importante pour rendre l'application fonctionnelle, on va d'abord définir la liste des requêtes à suivre pour extraire l'informations de l'ontologie à l'interface-graphique, puis appliquer ces requêtes dans notre environnement du développement (Java) pour récupérer les résultats et les afficher dans notre interface-utilisateur.

4. Fonctionnement de l'application:

Le but de notre application est de donner un coup de main au utilisateur (un médecin ou un pharmacien) pour prendre une décision optimale concernant le médicament ayant le minimum des effets indésirables sur la santé des malades en prenant compte ses caractéristiques et ses maladies. On a utilisé l'ontologie comme une base de connaissance grâce à sa fonctionnalité de modéliser et stocker de manière informatique un ensemble de connaissances, idées, concepts du domaine ou données, plus l'avantage de l'accessibilité aux informations médicales et le pouvoir de les consulter et les récupérer.

En fonction des modèles de connaissances de notre domaine stockés dans l'ontologie et d'une approche guidée par la modélisation du processus.

Notre approche:

- Servir de faire une recherche libre et facile dans l'ontologie des médicaments même si l'utilisateur ignore l'appellation exacte du sous-concept nous ajoutons aussi la possibilité de le préciser.
- Faire une recherche concernant un médicament selon le concept choisi (par nom commercial, par dénomination, par classe thérapeutique) et extraire tous les informations relatifs à ce dernier de la base de connaissance vers une interface-utilisateur adaptée.

- Estimer l'information optimale (le médicament idéal) qui répond aux besoins des utilisateurs selon le profil de patient qui doit contenir tous les paramètres administratifs (nom, prénom, âge, sexe) et les paramètres vitaux (allergie, symptôme), la recherche d'information pertinente ici basée sur les paramètres vitaux.
- Stocker le profil des patients dans une base de données.

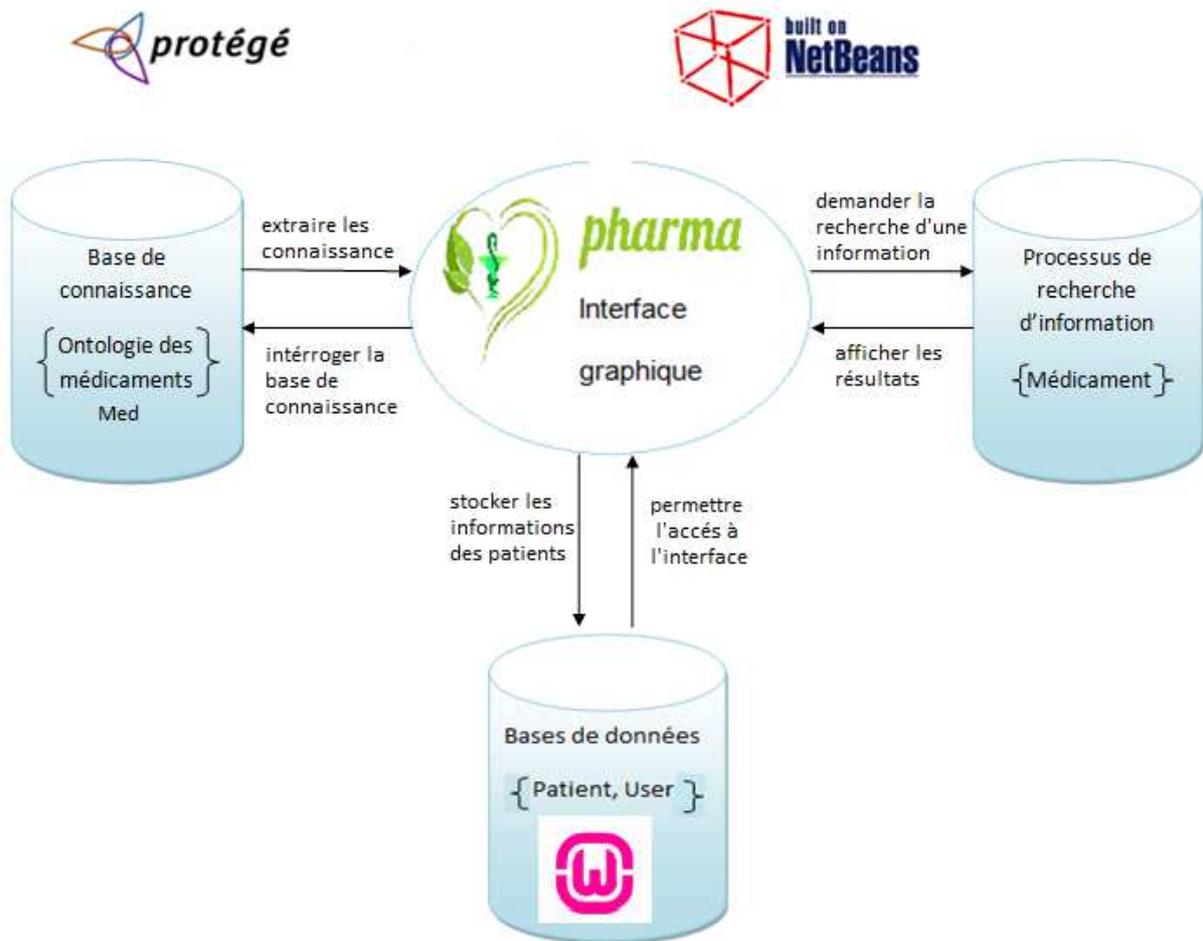


Figure IV.3. Architecture générale de notre approche

5. Conclusion :

Dans ce chapitre, nous avons présenté les étapes détaillées que nous avons suivi pour la Construction d'ontologie des médicaments, ensuite nous avons expliqué notre système de recherche d'information basé sur cette ontologie, cela nous permet de mettre en évidence les étapes nécessaires pour la création de l'ontologie ainsi que les différentes méthodes utilisées et les différents composants à savoir pour ça. Nous avons opté pour la méthode METHONTOLGY dans les étapes de conception de l'ontologie.

La prochaine étape consiste donc en la concrétisation du ce que nous avons proposé, en d'autres termes, la réalisation d'un système de recherche d'information basé sur notre ontologie.

CHAPITRE V :

Implémentation et résultat

1. Introduction

Pharma est une approche qui considère l'ontologie comme une base de connaissance pour établir la recherche d'information sur les médicaments et c'est un système d'aide à la décision médical.

Ce chapitre est consacré à l'implémentation de notre l'application pharma tout en présentant les langages, les outils et environnements de développements utilisés ainsi que les copies d'écran de chaque étape.

2. Implémentation de l'ontologie :

Une ontologie cherche à représenter le sens des concepts et des relations qui les lient. Elle comprend une partie terminologique, inventaire du vocabulaire pour les métadonnées et les concepts, la déclaration des instances (valeurs) et les propriétés particulières qui expriment relations entre concepts et instances.

La phase implémentation est la dernière étape du processus de création d'ontologie qui consiste à traduire l'ontologie réalisé par un Langage ontologique et le choix des outils d'implémentations de l'ontologie, la connexion à celles-ci et le langage du développement sont des factures importants pour le bon fonctionnement du système ainsi que pour la qualité des résultats de celui-ci.

Pour notre implémentation nous avons opté pour le langage OWL, et l'outil utilisé était Protégé.

2.1. OWL (Ontology Web Language):

RDF/RDF(S) ne permet pas d'exprimer certaines notions que l'on voudrait décrire avec les ontologies. Dans le but d'étendre l'expressivité de RDF Schema, le groupe de travail sur les ontologies du W3C a proposé le langage d'ontologie du Web (OWL), basé sur le langage DAML+OIL [McGuinness et al. 2002]. La sémantique du langage (ou plus exactement celle de OWL DL et OWL Lite) peut être dénie via une transformation en Logiques de Description (DL), dont ce langage est inspiré. On peut utiliser les outils de raisonnement développés pour les DL an d'effectuer des inférences sur OWL.

Le groupe de travail sur les ontologies du W3C, dans le souci de répondre aux critères exigés pour un langage de représentation d'ontologies, a déni trois différents sous-langages de OWL :

-OWL Lite : est le sous langage le plus simple, il répond à des besoins de hiérarchie de classification et de fonctionnalités de contraintes simples.

-OWL DL : est plus complexe que OWL Lite, il est fondé sur la logique de description SHIN(D). OWL DL garantit la complétude des raisonnements (calculabilité des inférences) et leur décidabilité (leur calcul se fait en une durée finie).

-OWL Full : est la version la plus complexe d'OWL, il se destine aux utilisateurs souhaitant une expressivité maximale. Il a l'avantage de la compatibilité complète avec RDF/RDFS, mais l'inconvénient est qu'il ne garantit pas la complétude et la décidabilité des calculs liés à l'ontologie.

2.2. Protégé:

Protégé [Gennari et al. 2003], développé à l'Université de Stanford¹⁵, est sans aucun doute l'environnement d'édition d'ontologies le plus utilisé aujourd'hui. Son noyau est basé sur le modèle des frames [Minsky, 1975] mais les ontologies développées avec cet environnement peuvent être exportées dans plusieurs formats (RDF(S), OWL, XML Schema).

Protégé s'enrichit régulièrement de l'apport de la communauté des utilisateurs et développeurs grâce au système de plugins qui permettent de rajouter de nouvelles fonctionnalités à l'outil.

Protégé a été choisi pour implémenter une ontologie interprétable par la machine ce qui doit permet de interroger cette base de connaissance pour les besoins de l'application. La Figure suivante montre la hiérarchie de l'ontologie implémentée via Protégé.

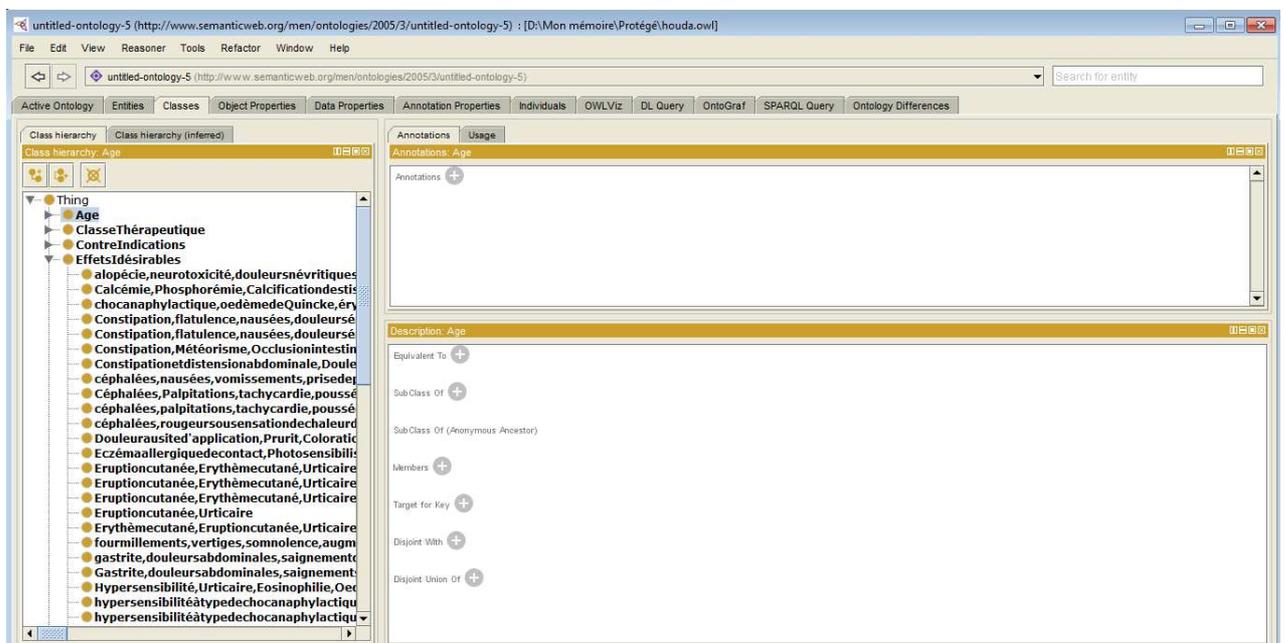


Figure V.1. Editeur d'Ontologie Protégé

3. implémentation de l'interface :

L'implémentation de l'interface consiste à rendre la maquette de l'interface graphique d'application exploitable par la machine et l'utilisateur l'outil choisi pour l'implémentation est NetBeans, une Platform de développement orientée Java.

3.1. Netbeans :

NetBeans est un environnement de développement intégré (EDI), placé en open source en juin 2000 sous licence CDDL (Common Développment and Distribution License). En plus de Java, NetBeans permet également de supporter différents autres langages, comme C, C++, JavaScript, XML, et HTML. Il comprend toutes les caractéristiques d'un IDE moderne (éditeur en couleur, projets multi-langage, éditeur graphique d'interfaces et de pages Web).

3.2. Construction de la base de données :

Notre applications consiste 2 basses de donnés, la première permis de modifier d'ajouter les patients et ces informations et de stocker les médicaments prescrit à ces derniers, et la deuxième est concerné sur le nom d'utilisateur et le mot de passe du pharma.

Ces bases de données sont construites par MySQL.

3.2.1. MySql :

MySQL est un système de gestion de bases de données relationnelles (SGBDR). Il est distribué sous une double licence GPL et propriétaire. Il fait partie des logiciels de gestion de base de données les plus utilisés au monde, autant par le grand public (applications web principalement) que par des professionnels, en concurrence avec Oracle, Informix et Microsoft SQL Server. Son nom vient du prénom de la fille du co-créateur Michael Widenius, MySQL fait référence au StructuredQuery Language, le langage de requête utilisé. MySQL AB a été acheté le 16 janvier 2008 par Sun Microsystems pour un milliard de dollars américains. En 2009, Sun Microsystems a été acquis par Oracle Corporation, mettant entre les mains d'une même société les deux produits concurrents que sont Oracle Database et MySQL. Ce rachat a été autorisé par la Commission européenne le 21 janvier 2010. Depuis mai 2009, son créateur Michael Widenius a créé MariaDB pour continuer son développement en tant que projet Open Source.

4. Connexion ontologie-interface :

Cette étape consiste à établir des requêtes orientées Java pour faire appel à la connaissance stockée dans l'ontologie pour cela on a utilisé l'API Jena.

4.1. Jena :

Jena est projet open source d'une API java permettant l'accès à des ontologies OWL. Comprend des raisonneurs intégrés mais permet également d'utiliser des raisonneurs externes. Jena comporte les outils suivantes :

1. Une API pour le langage RDF
2. Un module de lecture/ écriture sur les serveurs RDF : RDF/XML, N3 et NTRIPLE
3. Une API pour le langage OWL
4. RDQL : un langage d'interrogation des ontologies Implémenter en RDF et OWL

La page d'accueil de Jena, illustrée ci-dessous, est <http://jena.apache.org/>



Figure V.2. Page d'accueil Jena

5 .Architecture logicielle :

5.1. Ontologie :

L'environnement de développement Protégé nous aidons de créer une ontologie basée sur les informations des médicaments et des patients.

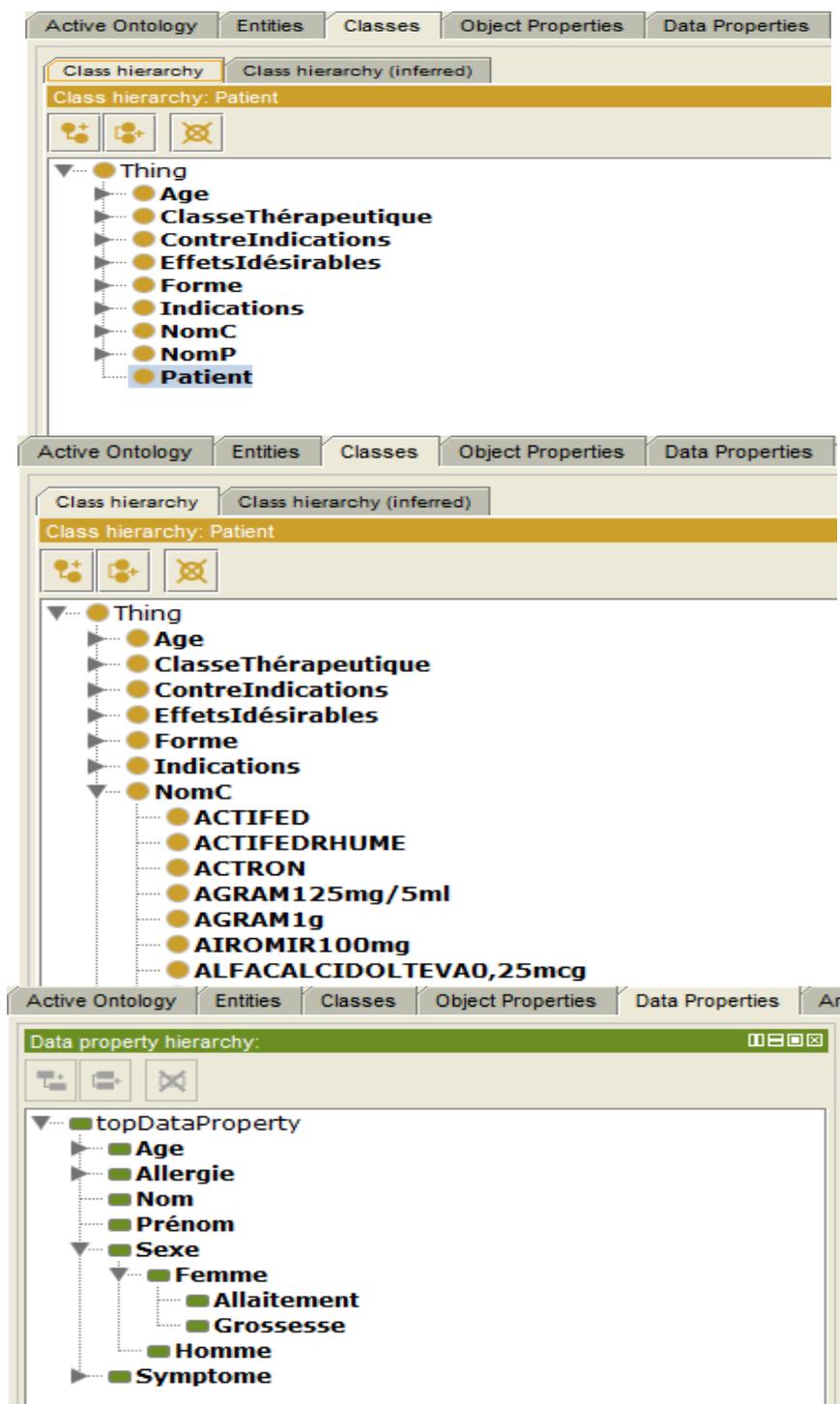


Figure V.3.Vue globale de nos ontologies

5.2. La base de données :

Nous avons créé une base de données qui sert à ajouter et stocker les informations et les modifications sur les patients et les médicaments prescrits. Cette base de données consiste en 8 attributs, les figures ci-dessus

ID	Nom	Prenom	Age	Sexe	Allergie	Symptome	Médicament
1	Belloula	Bouchra	enfant à partir de 8 ans	Femme	gluten	nez bouché	DOLI RHUME

Figure V.4. Vue globale de notre base de données

5.3. L'interface :

Dans ce qui suit nous allons présenter le prototype réalisé à travers des captures d'écran.

1-Fenêtre d'identification : cette fenêtre est importante car sans elle on ne peut pas accéder aux autres fenêtres de pharma dont où l'utilisateur peut gérer son nom d'utilisateur et son mot de passe et s'il trompe l'ouverture de l'application sera impossible la figure suivante représente une capture d'écran de cette fenêtre.



Figure. V.5. fenêtr principale d'application

Quand le mot de passe et le nom d'utilisateur est correcte la connexion au pharma sera établi



Figure. V.6. fenêtr de l'accé à PHARMA

2-Fenêtre d'accueil : C'est la fenêtre principale de notre application, l'implémentation de cette fenêtre sert à accéder autres fenêtres de l'application.



Figure. V.7.fenêtre d'accueil d'application

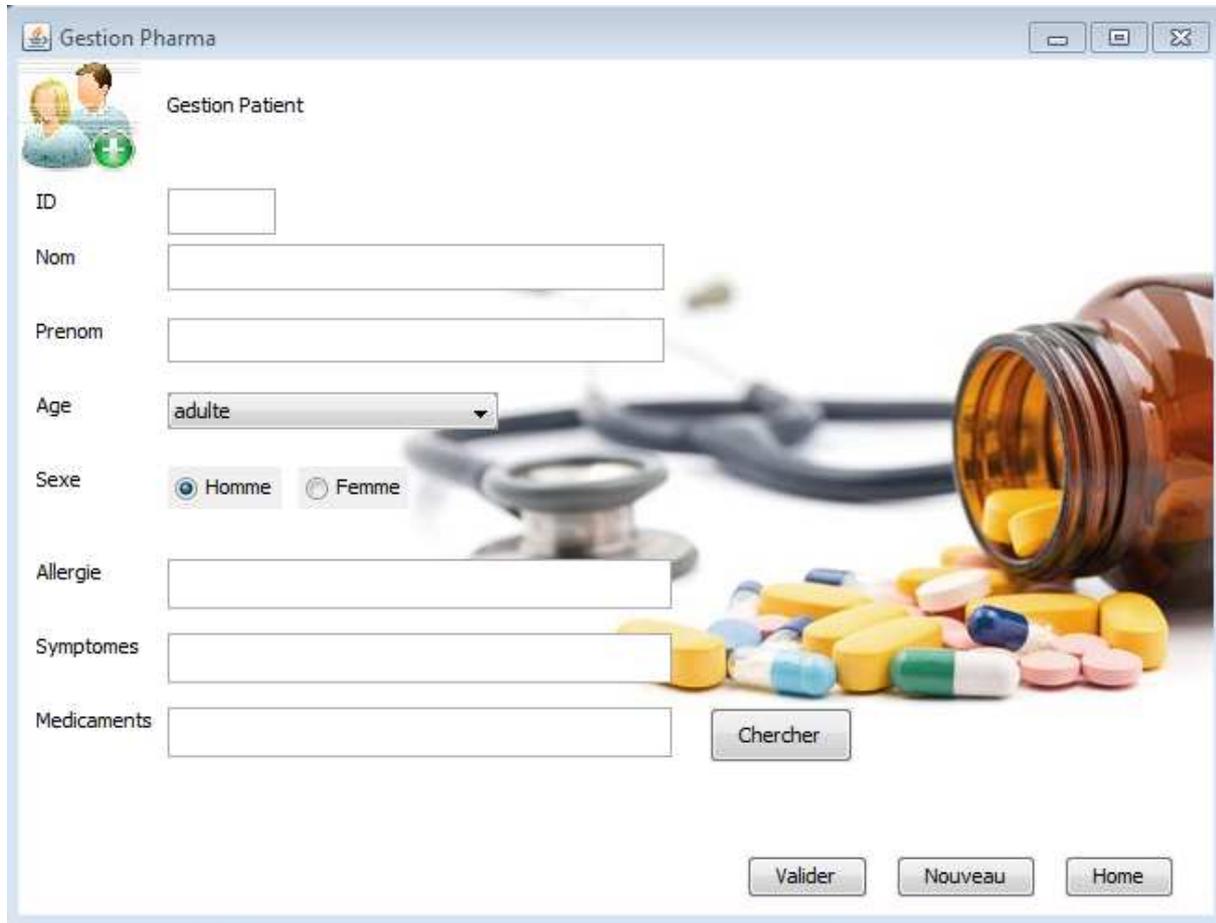
3. fenêtre gestion médicaments : cette partie est un moteur de recherches sur les médicaments et à l'utilisateur le choix de type de recherche par dénomination, classe thérapeutique ou nom commercial.



Figure. V.8.fenêtre gestion médicament

D'abord on choisissant le type de recherche puis on écrit dans le champ de recherche est le tableau au-dessous sera établie ; le bouton accueil retourne à la fenêtre d'accueil.

4. fenêtre gestion patient : cette fenêtre contient les informations des patients déjà enregistré et aussi l'utilisateur peut ajouter d'autre personne et de rechercher les médicaments à prescrire selon les indications du malade et ses allergies son âges et si c'est une femme il rendre compte si elle est dans la période d'allaitement ou enceinte.



The screenshot displays a web-based form titled "Gestion Patient" within a window labeled "Gestion Pharma". The form contains the following fields and controls:

- ID**: A text input field.
- Nom**: A text input field.
- Prenom**: A text input field.
- Age**: A dropdown menu with "adulte" selected.
- Sexe**: Radio buttons for "Homme" (selected) and "Femme".
- Allergie**: A text input field.
- Symptomes**: A text input field.
- Medicaments**: A text input field.

Buttons include "Chercher" (next to Medicaments), "Valider", "Nouveau", and "Home" at the bottom. The background features a stethoscope and a spilled pill bottle.

Figure. V.9.fenêtre gestion patient

L'utilisateur remplit les champs d'id (les numéros de patient) son nom et prénom l'âge (adulte, enfant). Et le sexe (homme ou femme (normal, période d'allaitement ou enceinte)) l'allergie du malade ces symptômes puis il clique sur chercher pharma nous résume les différents médicaments possible en tenant compte de ces concepts puis le bouton valider sert à sauvegarder le nouveau patient dont le bouton nouveau sert d'ajouter un nouveau patient et le bouton home retourne vers la fenêtre d'accueil.

6. Exemple :

6.1. Gestion patient :

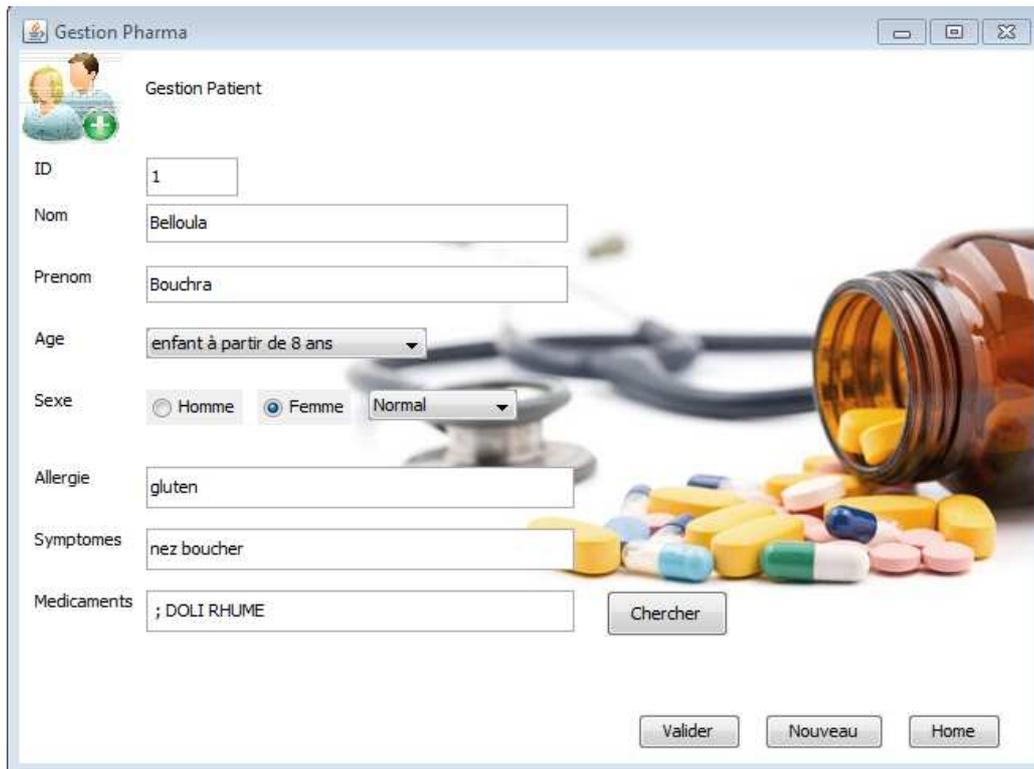


Figure. V.10.fenêtre gestion patient à l'étape de recherche dumédicament

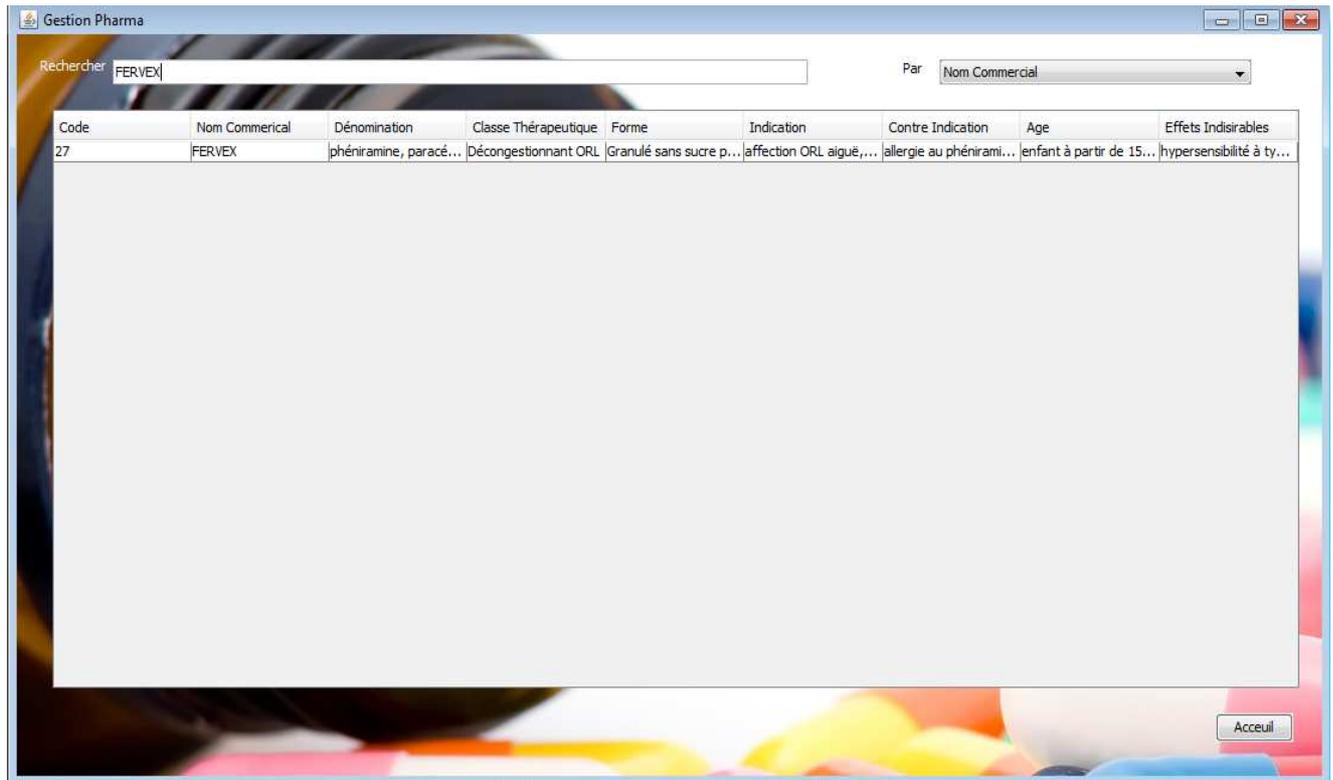
6.2. Gestion médicament

On recherche par classe thérapeutique :antalgique

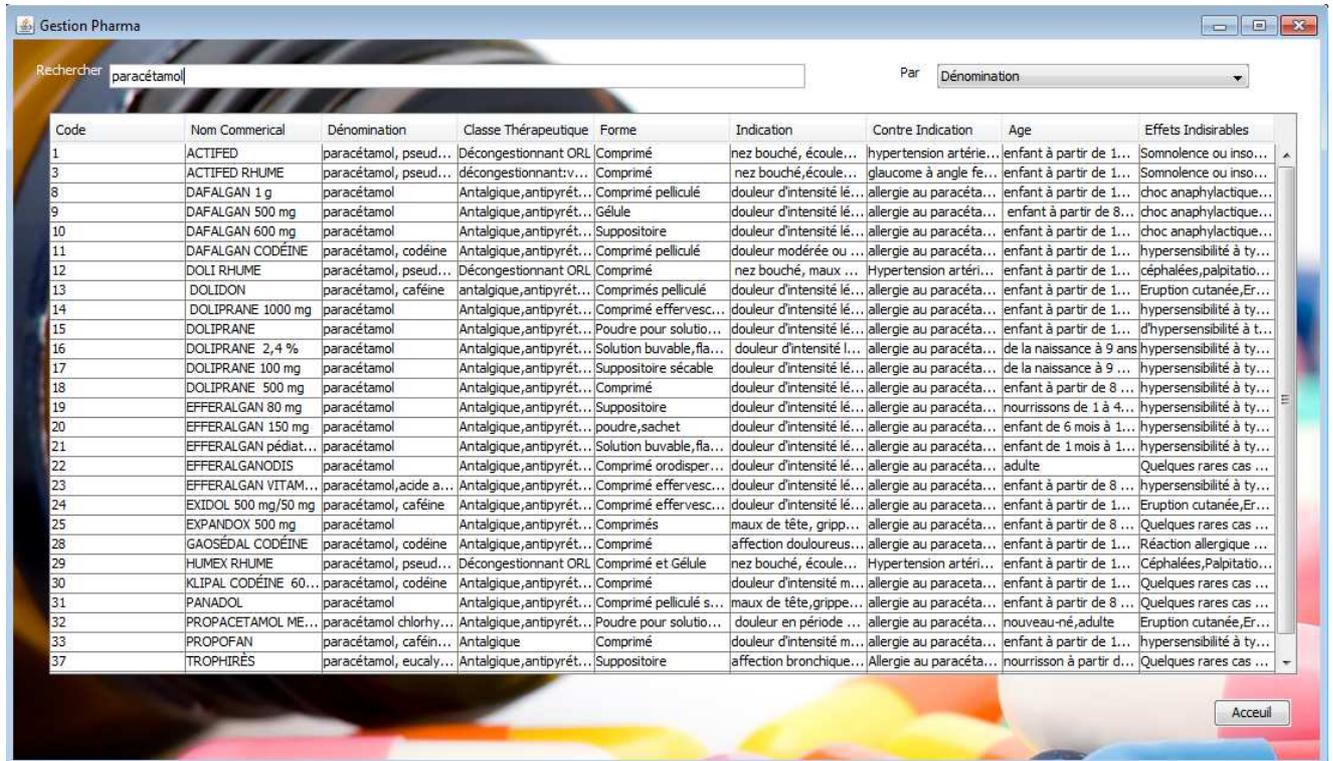


Figure. V.11.fenêtre gestion médicament à l'étape de recherche dumédicament

Puis on recherche par nom commercial :fervex



On recherche par dénomination : paracétamol



7. Analyse et résultats :

Notre application est très utiles dont nous avons atteint notre objectif de construire une ontologie contient une base de connaissance sur les médicaments ou on est besoin dans notre moteur de recherche quand a construit notre approche aide à la décision médical dont le médecin est besoin d'une liste de médicaments selon certain concept et critère. L'interface créée dans cette application répond aussi aux critères ergonomiques puisqu'elle permet de faire interagir son utilisateur qui n'est pas expert en informatique avec toute une application informatique et technique derrière mais de manière très simple.

8. Conclusion :

Nous avons présenté à travers ce chapitre l'implémentation de notre application on définissant tous les outils et langage de programmation utilisés et le fonctionnement de notre application pharma. Le travail réalisé est un système de recherche d'information basé sur l'ontologie des médicaments. Ce prototype a pour but de montrer l'utilité de l'utilisation des ontologies et leurs apports à la recherche d'information. Le système conçu ne propose pas un moteur de recherche mais un système qui permet l'exploitation de l'ontologie, il permet de voir le traitement sémantique et pertinent des requêtes et les résultats de ce traitement

:

CONCLUSION GENERALE ET PERSPECTIVES

Conçues comme réponse aux problèmes posés par l'intégration de connaissances au sein des systèmes de recherche d'information, les ontologies apparaissent désormais comme une clé pour la manipulation automatique de l'information au niveau sémantique. Au fur et à mesure des recherches et des idées se dégagent autour du contenu des ontologies, afin de créer une couche sémantique qui servira comme intermédiaire entre les bases documentaires et les systèmes de recherche d'information.

Dans cette thèse, nous avons proposé de construire une application pour le domaine médical, plus précisément pour le domaine pharmaceutique. Cette application a comme objectif d'aider le médecin ou le pharmacien dans les prescriptions des médicaments avec un minimum d'effets indésirables. De ce fait il nous a fallu d'abord développer une ontologie regroupant les concepts qui appartiennent au domaine choisi. En s'appuyant sur cette ontologie comme une base de connaissance, nous avons développé un moteur de recherches dans une interface-utilisateur avec l'objectif d'offrir une information plus détaillée aux utilisateurs. L'outil ainsi créé peut être utilisé pour améliorer la prise de décision médicale

Nous pouvons conclure que le travail que nous avons réalisé dans ce mémoire nous a permis d'explorer le domaine de l'ingénierie ontologique qui nous été méconnu, de plus nous avons eu la chance de manipuler et d'acquérir une certaine maîtrise des outils de création et de manipulation des ontologies, de plus il nous a permis de toucher au domaine des systèmes de recherche d'information basés sur les ontologies, qui sont présentés comme la nouvelle génération des systèmes de recherche d'information.

Comme perspectives à notre travail serait d'élargir encore plus les concepts de l'ontologie MED à une plus vaste gamme de médicament afin de rendre l'application plus exploitable par un large public de praticiens et Envisager une éventuelle intégration de l'ontologie MED dans un moteur de recherche dédié au domaine des pharmaciens .

L'ontologie pourra toujours être enrichie par plus de connaissances dans le domaine médical, ce qui nécessite une collaboration avec le personnel médical.

Il doit être intéressant d'enrichir les requêtes d'extraction de connaissance afin de clarifier la relation entre les différents concepts.

Transformé l'application construite à une application Androïde, pour qu'elle soit plus pratique

BIBLIOGRAPHIE

[Harris 68] Harris Z., *Mathematical Structures of Language*, John Wiley and Sons, New-York, USA, 1968.

[Minsky 75] Minsky, M. (1975). *A framework for representing knowledge*. *The Psychology of Computer Vision*, Winston, Patrick, ed. New York :McGraw Hill, pages 211-77.

[Neeches et ses collègues 93] Neeches, Finin T, Fikes R.E, Gruber T.R, Senator T et Swartou W.R. « Enabling technology for knowledge sharing » *AI Magazine*. Vol.12, no 3, 1993.

[GRUBER 93] Gruber T. (1993). *A Translation Approach to Portable Ontology Specifications*. *Knowledge Acquisition*, 5(2), 199-220.

[Schmid 94] Schmid, H., 1994. *Probabilistic Part-of-Speech Tagging Using Decision Trees*.
[Edwards et al, 94] Edwards I., Biriell C., « Harmonisation in pharmaco-vigilance », *Drug Safety*, vol. 10, n° 2, p. 93-102, 1994.

[GUARINO 95] Guarino N. (1995b). *Understanding, building and using ontologies*. *International J. Human-Computer Studies*, 46, 293-310.

[TRGU 95] THOMAS R. GRUBER « *A Translation Approach to Portable Ontology Specifications*. *Knowledge Acquisition*, Vol.5 » 1995

[PHMA 95] PHILIPPE MARTIN « *Using the WordNet Concept Catalog and a Relation Hierarchy for Knowledge Acquisition* » 18 août 1995 international Workshop on Peirce. Université de Californie, Santa Cruz, USA.

[UMKM 95] USCHOLD M. & KING M « *Towards a methodology for building ontologies*, in *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing* » 1995.

[USH 96] USCHOLD M. M., GRUNINGER M. “*Ontologies: principles, methods and applications*”. *Knowledge Engineering Review*.11(2).93-155. 1996.

[BORG 96] A. Borgida. “*On the relative expressiveness of description logics and predicate logics*”. *Artificial Intelligence*, volume 82, number 1-2, pages 353–367, 1996.

[BORST 97] Borst W. N. (1997). Construction of Engineering Ontologies. Center for Telematica and Information Technology, University of Twente, Enschede, NL

[GUARINO 97] Guarino N. (1997b). Understanding, building and using ontologies. International J. Human-Computer Studies, 46, 293-310.

[GUARI 97] Nicola Guarino. « Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. In Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology » 1997. SCIE 1997, M. T. Pazienza (Eds.), Springer Verlag, pp. 139-170.

[FER, GOM 97] FERNANDEZ M., A. GOMEZ-PEREZ et al. "METHONTOLOGY : from ontological art owards ontological engineering". In Proceedings of the Spring Symposium Series on Ontological Engineering (AAAI'97), AAAI Press, 1997.

[STU 98] STUDER R., BENJAMINS R., FENSELD. Knowledge Engineering: principales and methods. IEEE transactions on Data and Knowledge Engineering. 25(1-2): 161-197. 1998

[FDES 98] DIETER FENSEL, STEFAN DECKER, MICHAEL ERDMANN, AND RUDI STUDER. « Ontobroker: How to make the WWW Intelligent. Research report, in Proceedings KAW98» April 1998 the 11th Banff Knowledge Acquisition for Knowledge-Based System Workshopresearchreport.Institut AIFB, Banff, Canada

[MIZO 98] MIZOGUCHI R « A Step towards Ontological Engineering » Juin, 1998. Paper presented at the 12th National Conference on AI of JSAI

[BLAS et AL 98] BLAZQUEZ M. FERNANDEZ M., GARCIA-PINAR J. M. & GOMEZ-PEREZ A., Building Ontologies at the Knowledge Level using the Ontology Design Environment, in Proceedings of the Banff Workshop on Knowledge Acquisition for Knowledge-based Systems, 1998.

[MIZO 98] MIZOGUCHI R « A Step towards Ontological Engineering » Juin, 1998. Paper presented at the 12th National Conference on AI of JSAI

[MIZO 98] MIZOGUCHI R « A Step towards Ontological Engineering » Juin, 1998. Paper presented at the 12th National Conference on AI of JSAI

[BLAS et AL 98] BLAZQUEZ M. FERNANDEZ M., GARCIA-PINAR J. M. & GOMEZ-PEREZ A., Building Ontologies at the Knowledge Level using the Ontology Design Environment, in Proceedings of the Banff Workshop on Knowledge Acquisition for Knowledge-based Systems, 1998.

[Spackman et al 98] Spackman K., Campbell K., « Compositional Concept Representation using SNOMED: Towards Further Convergence of Clinical Terminologies », Journal of the American Medical Informatics Association, vol. 5, n° suppl., p. 740-744, 1998.

[WCR 99] WEIQIN CHEN AND RIIICHIRO MIZOGUCHI « Communication Content Ontology For Learner Model Agent in multi-Agent Architecture. Workshop on Ontologies for Intelligent Educational Systems » juillet 19- 23, 1999 Ninth International Conference on Artificial Intelligence in Education, AI-ED'99, Le Mans, France

[IHLB 99] M. IKEDA, Y. HAYASHI, J. LAI, W. CHEN, J. BOURDEAU, K. SETA AND R. MIZOGUCHI « An ontology more than a shared vocabulary. Workshop on Ontologies for Intelligent Educational Systems » Juillet 19-23, 1999 Ninth International Conference on Artificial Intelligence in Education, AI-ED'99. Le Mans, France

[GPBV 99] GOMEZ PEREZ A., BENJAMINS V.R. « Overview of knowledge sharing and components: Ontologies and problem Solving Methods. Workshop on Ontologies and problem-Solving Methods » 1999. Stockholm (Suède).

[NOYN, GUIN 01] NOY N. et McGuinness D. L. "Ontology Development 101: A Guide to Creating Your First Ontology". Technical Report KSL-01-05Stanford: Knowledge Systems Laboratory, mars 2001.

[Haarslev et al 01] Haarslev V., Möller R., « Description of the RACER System and its Applications », International Workshop on Description Logics proceedings, 2001.

[CHA 02] CHARLETJ., L'ingénierie des connaissances : développements, résultats et perspectives pour la gestion des connaissances médicales. Mémoire d'habilitation à diriger des recherches en Informatique de l'université de Pierre et Marie Curie. Déc. 2002.

[SUR, ERDM 02] SURE Y., ERDMANN M., ANGELE J., et al. "Onto Edit: Collaborative ontology development for the semantic web", presented at International Semantic Web Conference 2002 (ISWC 2002), Sardinia, Italy, 2002.

[Lenzerini 02] LENZERINI M. (2002). Data Integration: A Theoretical Perspective. In L. POPA, Ed., *21st ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, PODS'02*, p. 233–246, Madison, WI, USA.

[Bourigault et al 02] Bourigault D., Lame G., « Analyse distributionnelle et structuration de terminologie. Application à la construction d'une ontologie documentaire du Droit », *Traitement automatique des langues*, 2002.

[Bachimont et al 02] Bachimont B., Isaac A., Troncy R., « Semantic Commitment for Designing Ontologies: A Proposal», in A. Gomez-Pérez, V. Benjamins (eds), *13th International Conference on Knowledge Engineering and Knowledge Management (EKAW'02)*, vol. (2473) of *Lecture Notes in Artificial Intelligence*, Springer Verlag, Sigüenza, Espagne, p. 114-121, 1-4 October, 2002.

[Brown 02] Brown E. G., « Effects of coding dictionary on signal generation: a consideration of use of Med DRA compared with WHO-ART », *Drug Safety*, vol. 25, n° 6, p. 445-52, 2002.

[McGuinness et al 02] McGuinness, D. L., Fikes, R., Hendler, J. A. et Stein, L. A. (2002). *Daml+oil : An ontology language for the semantic web*. *IEEE Intelligent Systems*, 17(2)(5).

[Gennari et al 03] Gennari, J., Musen, M. A., Fergerson, R. W., Grosso, W. E., Cru-bezy, M., Eriksson, H., Noy, N. F., et Tu, S. W. (2003). The evolution of prot: An environment for knowledge-based systems development. *International Journal of Human-Computer Studies*.

[XuanET al 07] XUAN W., DAI M., MIREL B., ATHEY B., WATSON S. J. & MENG F. (2007). Interactive Medline Search Engine Utilizing Biomedical Concepts and Data Integration. In *BioLINK: Linking Literature, Information and Knowledge for Biology, SIG, ISMB'08*, p. 55–58, Vienna, Austria.

[Melle BOUARROUDJ.S 09] Melle BOUARROUDJ Samia, « Raisonement sur une ontologie enrichie par enrichie des règles SWRL pour la recherche sémantique d'images annotées », thèse de magister, (Ecole Doctorale en Informatique de l'Est -Pole ANNABA, 2009.