

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITÉ MOULOUD MAMMERI DE TIZI-OUZOU
FACULTÉ DE GÉNIE ÉLECTRIQUE ET INFORMATIQUE
DÉPARTEMENT D'INFORMATIQUE



THÈME

CONSTRUCTION D'UNE ONTOLOGIE DE DOMAINE MÉDICAL

Promoteurs : Mme D. TAOURI
Présidente : Mlle S. AIT ADDA
Examinatrice : Mlle S. ILTACHE
Examinatrice : Mlle Y. YESLI

Mémoire présenté en vue de l'obtention du
Master en Informatique, option :(CPI)

Réaliser par :

GHANINE Mohammed
MEDJBOUR Aghiles

Session 2014 - 2015

Sommaire

1	Ingénierie des ontologies	12
1.1	Définitions d'une ontologie	12
1.2	Intérêts de l'ontologie	14
1.3	Les composants d'une ontologie	15
1.3.1	Concepts	15
1.3.2	Relations	16
1.3.3	Fonctions	17
1.3.4	Axiomes	17
1.3.5	Les instances	17
1.4	La Typologie des ontologies	18
1.4.1	Typologie selon l'objet de conceptualisation	18
	Les ontologies de représentation des connaissances	18
	Les ontologies génériques	19
	Les Ontologies de domaine	20
	Les Ontologies de tâches	20
	Les Ontologies d'application	20
1.4.2	Selon le niveau de formalisme de représentation	20
	Ontologies Informelles	20
	Ontologies Semi-informelles	21
	Ontologies Semi-formelles	21
	Ontologies Formelles	21
1.4.3	La typologie selon le niveau de détail	21
	Granularité fine	21
	Granularité large	21
1.5	Méthodologie de construction d'une ontologie	21
1.5.1	La méthode ENTREPRISE	22
1.5.2	La Méthode TOVE	24
1.5.3	La Méthode METHONTOLOGY	24
1.6	Formalismes de représentation des connaissances	26

1.6.1	Frames	26
1.6.2	Logiques de descriptions	27
1.6.3	Réseaux sémantique	30
1.7	Les langages de représentation d'ontologies	30
1.7.1	XML (Extensible Markup Language)	30
1.7.2	RDF (Resource Description Framework)	30
1.7.3	RDFS (RDF Schéma)	34
1.7.4	OWL (Web Ontology Language)	34
	OWL Lite	35
	OWL DL	35
	OWL FULL	35
1.7.5	DAML-OIL :	35
1.8	Les outils d'édition	36
1.8.1	Ontolingua	36
1.8.2	WebOnto	37
1.8.3	Protégé	37
1.9	Domaines d'application d'ontologies	38
1.9.1	Les ontologies et la représentation des connaissances	38
1.9.2	Les ontologies et le web sémantique	38
1.10	Conclusion	39
2	Ressource terminologique et ontologique en médecine	40
2.1	Ressources terminologique et ontologique en médecine	40
2.1.1	CIM	40
2.1.2	CCAM	42
2.1.3	MeSH	43
2.1.4	CISMeF	44
2.1.5	SNOMED	45
2.1.6	UMLS	46
2.1.7	FMA	46
2.1.8	DOLCE	47
2.1.9	GALEN	47
2.1.10	MENELAS	48
2.2	Conclusion	49
3	Conception de l'ontologie	50
3.1	Processus de construction d'une ontologie de domaine	50
3.2	Construction d'une ontologie de domaine	51

3.2.1	Évaluation des besoins	51
3.2.2	Conceptualisation	51
	Construire le glossaire de termes	51
	Construction de diagramme des relations binaires et des attributs	53
	Construction d'un dictionnaire de concepts	55
	Déterminer la liste des relations binaires	55
	Construction de la table d'attributs	56
	Construction de la table des axiomes	57
	Construction de la table des instances	58
3.2.3	Formalisation	59
	Le niveau terminologique ou TBox	59
	Construction d'ABox	59
4	Implémentation et Réalisation	61
4.1	Étude de protégé	61
4.2	Les moteurs d'inférence	62
	4.2.1 Racer	62
	4.2.2 Pellet	62
4.3	Les Langage d'interrogation	63
	4.3.1 nRQL (new Racepro Query Language)	63
	4.3.2 RDQL (RDF Data Query Language)	64
	4.3.3 SPARQL	64
4.4	Les logiques de description	65
4.5	choix de langage	65
4.6	Choix de l'outil	65
4.7	Etapas de construction de l'ontologie	66
	4.7.1 Lancement de Protégé	66
	4.7.2 Création d'un nouveau projet	67
	4.7.3 Création des classes	68
	4.7.4 Création des classes disjointes	69
	4.7.5 Création des relations	70
	4.7.6 Création des Individus	71
	4.7.7 Création des Axiomes	72
	4.7.8 Génération du code RDF/XML	73
	4.7.9 Les classes et la hiérarchie des classes de notre ontologie	73
	4.7.10 Quelques requêtes SPARQL	75
	4.7.11 Type de Traitement :médicament parfois chirurgical	80

Liste des tableaux

1.1	Une base de connaissances composée d'une T-Box et d'une A-Box .	28
3.1	Glossaire de quelques termes	52
3.2	dictionnaire de quelques concepts	55
3.3	Table de quelques relations binaires	56
3.4	Table des attributs	56
3.5	Table des axiomes	57
3.6	Table des instances	59
3.7	Axiomes terminologiques (TBox)	60
3.8	Assertions sur les individus (ABox)	60
4.1	Avantages et inconvénients du moteur d'inférence Racer	62
4.2	Avantages et inconvénients du moteur d'inférence Pellet	63

Table des figures

1.1	schéma qui représente les différentes structures d'ontologie	18
1.2	Classification des Ontologies selon [N.Guarino, 1998]	19
1.3	méthode d'Uschold et King.	23
1.4	processus de développement et cycle de vie de METHONTOLOGY	26
1.5	Exemple de graphe RDF	32
1.6	Exemple de document RDF	34
1.7	Exemple de document RDFs	35
1.8	Hierarchie de langage OWL	36
3.1	Diagramme des relations binaires	54
4.1	Lancement de protégé	66
4.2	Création d'un nouveau projet	67
4.3	Création d'une classe	68
4.4	Classes disjointes	69
4.5	Ajout de la relation "Traite"	70
4.6	Création des individus	71
4.7	Création d'un axiome sur la classe Appendicite	72
4.8	extrait de code RDF/XML	73

Remerciements

Nous remercions le bon Dieu de nous avoir mis sur la route du savoir

Nous remercions infiniment notre promotrice : Mme D. TAOURI, pour nous avoir encadrés durant cette année, Nous tenons à lui exprimer notre profonde gratitude.

Nous remercions également les membres du jury pour nous avoir honorées avec leurs présences et pour avoir juger notre travail.

Nous tenons à saluer la peine et l'effort fournis par l'ensemble de nos professeurs afin d'assurer notre formation tout au long de notre cursus universitaire et leur disons de ce fait, merci. Nous exprimons notre infinie gratitude à nos chers parents en reconnaissance de leurs sacrifices, dévouement, soutien et encouragements.

Enfin, nous remercions tous ceux qui ont contribué de près ou de loin à réaliser ce modeste travail.

Ces quelques mots ne traduisent guère tout ce que nous avons pu recevoir de la part de chacun d'entre eux, mais nous souhaitons néanmoins qu'ils y trouvent l'expression de notre infinie reconnaissance.

Dedicaces

Je dédie ce modeste travail à :

Toutes ma familles qui m'a toujours soutenue et encouragé dans tous ce que j'entreprends en particulier ma mère et ma Fiancée.

Mon ami et binôme GHANINE Mohammed. Sans lui ce travail ne serait pas accompli.

Tous mes amis(es) qui se reconnaîtront.

Toute personne ayant contribué de près ou de loin au bon déroulement de ce projet.

Aghiles

Dedicaces

Je dédie ce modeste travail à :

Toutes ma familles qui m'a toujours soutenue et encouragé dans tous ce que j'entreprends en particulier mon père et ma mère ,mes grands parents et mes frères.

Mon ami et binôme MEDJBOUR Aghiles. Sans lui ce travail ne serait pas accompli.

Tous mes amis(es) qui se reconnaîtront.

Toute personne ayant contribué de près ou de loin au bon déroulement de ce projet.

Mohammed

Introduction Générale

Aujourd'hui une grande banque de connaissances distribuées sur toute la planète grâce à internet. Le web est la source d'information privilégiée dans tous les domaines de la vie. La majorité des pages web disponibles sont publiées dans un format qui ne facilite pas leur traitement automatique par des machines. Le langage HTML (Hyper Text Markup Language) ne décrit pas le contenu des documents ni la sémantique des documents, c'est un langage qui décrit comment les pages sont affichées. Alors que le traitement du document et son interprétation syntaxique et sémantique reviennent à l'utilisateur humain. La création des pages web destinées uniquement à un traitement par des utilisateurs humains, les applications destinées à faciliter l'utilisation et l'exploitation du web se trouvent limitées dans leurs fonctionnalités par leur incapacité d'interpréter l'aspect sémantique des pages web. Ce problème est à l'origine d'une nouvelle vision du web : « c'est le web sémantique ». Dans ce sens, les ontologies apparaissent aujourd'hui comme des composants logiciels avancés qui s'insèrent au centre des systèmes informatiques pour leur apporter une dimension sémantique qui, jusqu'à présent, leur faisait défaut. Un des plus importants projets d'Ingénierie ontologique à ce jour consiste à ajouter au web une sur-couche de connaissances qui permettrait de faire de la recherche d'informations au niveau sémantique et non plus seulement au niveau syntaxique. Le « Web sémantique » se veut un web dont le contenu peut être exploité et surtout appréhendé par des machines. Au même titre que l'ingénierie des connaissances, l'ingénierie ontologique est un domaine pluridisciplinaire puisque la construction d'ontologies qui est l'objet de travail de l'Ingénierie ontologique demande à la fois une analyse sémantique, et donc linguistique des informations, la mise en place d'outils de raisonnement et de calcul, la définition de langages de représentation et la réalisation de systèmes informatiques pour les utiliser. C'est un domaine qui a des répercussions sur les systèmes d'aide à la décision, les systèmes d'enseignement assisté par ordinateur, les systèmes de gestion de connaissances... Une véritable ingénierie s'est constituée autour du développement d'ontologies dans des domaines aussi variés que la médecine, le droit, la biochimie, l'indexation de séquences audiovisuelles, l'électronique... Cette ingénierie a pour objectifs

la construction d'ontologies, leur exploitation, leur maintenance et, de manière générale, leur gestion tout au long de leur cycle de vie

Le domaine médical est l'un des domaines qui joue un rôle primordial dans notre vie et l'accès aux connaissances médicales est un enjeu majeur pour les professions de santé comme pour le grand public.

Problématique

: Bien que l'information ne soit pas statique, et sujette à des modifications, des enrichissements, s'altère avec le temps, et parvient de différentes sources, nous avons besoin d'outils et de modèles qui permettent aux utilisateurs et experts du domaine de constituer, consulter et maintenir leurs connaissances du domaine. Ainsi, pour faire face à l'importante masse d'informations sur le web, il y a un besoin accru de développer des outils et des techniques qui servent à trouver les documents contenant l'information pertinente pour un besoin bien spécifique. Un moyen de parvenir à améliorer la recherche d'informations sur le web est le marquage sémantique des ressources en utilisant des ontologies qui définissent les concepts du domaine et leurs relations. Cela soulève plusieurs problèmes au centre de nos réflexions :

1. Quelles sont les connaissances les plus pertinentes d'un domaine et comment peut-on les trouver ?
2. Comment faut-il organiser ces connaissances de façon à obtenir le système le plus efficace et le plus informatif tout en leur conservant un maximum de richesse ?

Objectif de notre travail

- Le but de ce travail consiste à mettre en lumière : La construction d'une ontologie de domaine (Medical Domain ontology) dans le domaine de médecine, en utilisant Le " Ontology Web Language " (OWL). Le processus suivi pour la construction de cette ontologie est inspiré de la méthode METHONTOLOGY (Spécification, Conceptualisation, Formalisation, Opérationnalisation, Évaluation)
- Cette ontologie est utilisée par un réseau de santé qui regroupe tous les acteurs intervenant dans le processus de soin, des acteurs humains (médecins, patients, laborantins ...etc.)

Domaines concernés :

Ce mémoire situe à la croisée de plusieurs domaines de recherche tels que l'Ingénierie des connaissances, l'Ingénierie ontologique, le Web sémantique, la modélisation informatique, l'informatique médicale, la gestion et l'optimisation des coûts en médecine.

Nous sommes intéressés au Web sémantique et l'Ingénierie ontologique avec l'utilisation des techniques de cette dernière, dans le cadre de l'informatique médicale. Nous espérons que le travail réalisé sera profitable, et nous présentons brièvement ces quatre principaux champs de recherche auxquels ce travail apporte sa contribution la plus significative.

Organisation du mémoire :

Ce mémoire contient 4 chapitres :

Chapitre 1 : Qui s'intitule Ingénierie Ontologique, nous avons défini la notion d'ontologie ainsi que les composants de cette dernière, et nous avons examiné les différents types d'ontologies et les méthodes de construction, nous avons aussi consacré une partie dans ce chapitre sur les outils et langages d'ontologies et enfin quelques domaines dont les ontologies interviennent..

Chapitre 2 : présente les ressources terminologiques et ontologiques dans le domaine de médecin.

Chapitre 3 : présente la construction d'une ontologie de domaine partons de connaissances brutes et arrivant à une ontologie opérationnelle, ce processus est inspiré de la méthode ' METHONTOLOGY ' qui contient cinq étapes (Évaluation des besoins, Conceptualisation, Formalisation, Implémentation et Maintenance). Cette ontologie permet aux utilisateurs et experts humaines (Médecins, étudiants, Patients...) de constituer, consulter et maintenir leurs connaissances sur ce domaine, et d'avoir un langage médical commun et standardisé.

Chapitre 4 : Ce dernier s'intitule Implémentation et réalisation, nous avons détaillé encore plus sur le logiciel utilisé (Protégé 4.3), les moteurs d'inférences souvent utilisés Racer et Pellet et les langages d'interrogation utilisés par ces derniers comme SPARQL et nRDQL, nous avons mentionné également les deux langages assertionnel et terminologique, et enfin nous présenterons notre ontologie implémentée sous l'éditeur protégé qui sera interrogée par des requêtes afin de tester la consistance de cette dernière.

Chapitre 1

Ingénierie des ontologies

Introduction

Durant ces dernières années, les ontologies sont largement utilisées et ont prouvé leurs utilités dans de nombreux domaines tels que : l'ingénierie de connaissances, l'intelligence artificielle, l'intégration des sources de données, la recherche d'information, le commerce électronique et sont au coeur du Web Sémantique [**T. Berners-Lee, Hendler & O. Lassila, 2001**] Cet engouement est motivé par le fait que les ontologies sont un moyen efficace pour la gestion et le partage des connaissances d'un domaine particulier entre personnes et/ou systèmes. Ce chapitre a pour titre ingénierie ontologique qui est une discipline de l'informatique référant à l'ensemble des activités qui concernent le processus de développement d'ontologie, les méthodes et les cadres méthodologiques de sa construction, donc Nous allons présenter dans ce chapitre, en premier lieu la notion d'ontologie, en relevant quelques définitions qui lui ont été attribuées et quelques intérêt plus les différents éléments qui la constituent. Par la suite, nous présenterons ses différentes classifications en se concentrant sur la richesse de la structure interne de l'ontologie. Ainsi, on va désigner les différentes méthodologies de conception. En outre, nous faisons un survol des principaux formalismes de représentation de connaissances à savoir les frames, les logiques de description et les réseaux sémantiques qui sont à l'origine des langages permettant d'exprimer des ontologies.

1.1 Définitions d'une ontologie

Le terme «ontologie» vient du domaine de la philosophie au XIXème siècle, qui s'intéresse à l'étude de l'être ou de l'existence. En philosophie, l'ontologie est une branche de la métaphysique qui s'intéresse à la notion d'existence, son premier sens a trouvé son origine depuis ARISTOTE, où l'ontologie est l'étude des propriétés

générales de ce qui existe. Par analogie, le terme est repris en informatique et sciences de l'information dans les années 80, où l'ontologie est un terme technique désignant un artefact [Tom Gruber, 2009] qui est conçu pour un but, qui est de permettre la modélisation des connaissances sur un domaine quelconque, réel ou imaginaire. Les ontologies sont employées dans l'intelligence artificielle, le Web sémantique, le génie logiciel, l'informatique biomédicale ou encore l'architecture de l'information comme une forme de représentation de la connaissance au sujet d'un monde ou d'une certaine partie de ce monde. Les ontologies informatiques sont des outils qui permettent précisément de représenter un corpus de connaissances sous une forme utilisable par un ordinateur.

En informatique, plusieurs définitions ont été données pour l'ontologie :

En 1991, Neches et ses collègues ont été les premiers à en proposer une définition :

« Une ontologie définit les termes et les relations de base du vocabulaire d'un domaine ainsi que les règles qui indiquent comment combiner les termes et les relations de façon à pouvoir étendre le vocabulaire »

En 1993, Gruber a proposé une définition à cette notion qui est la plus célèbre couramment citée dans la littérature :

« Une ontologie est une spécification explicite d'une conceptualisation. »

En 1997, Borst a modifié légèrement la définition de Gruber en citant qu'une ontologie est définie comme étant :

« Une ontologie est une spécification formelle d'une conceptualisation partagée. »

En 1998, Studer et ses collègues ont rassemblé ces deux définitions (celles de Gruber et Borst) dans une seule qui est :

« Une ontologie est une spécification formelle et explicite d'une conceptualisation partagée. »

- **Spécification explicite** : signifie que les concepts, les propriétés, les relations, les fonctions, les restrictions et les axiomes de l'ontologie sont définis de façon déclarative.
- **Formelle** : réfère au fait qu'une ontologie doit être traduite dans un langage

interprétable par une machine.

- **Conceptualisation** : réfère à un modèle abstrait d'un phénomène du monde en identifiant les concepts appropriés à ce domaine.
- **Partagé** : réfère au fait qu'une ontologie capture la connaissance consensuelle c'est-à-dire non réservée à quelques individus, mais partagée par un groupe ou une communauté. [F. Amourache, 2008]

C'est une base de formalisation des connaissances. Elle se situe à un certain niveau d'abstraction et dans un contexte particulier.

C'est aussi une représentation d'une conceptualisation partagée et consensuelle, dans un domaine particulier et vers un objectif commun. Elle classe en catégories les relations entre les concepts.

1.2 Intérêts de l'ontologie

Quelques intérêts de l'utilisation des ontologies :

Vocabulaire unifié : un vocabulaire commun constitue le rôle d'une ontologie, qui permet l'unification des vocabulaires d'un domaine particulier.

Structuration des concepts : la structuration des données d'un domaine par une hiérarchie des concepts.

Interopérabilité : l'ontologie constitue une solution intéressante pour le problème d'interopérabilité entre les systèmes d'informations hétérogènes. Elle permet de définir clairement les différents concepts utilisés en vue d'assurer une communication sans ambiguïté entre les systèmes hétérogènes.

La communication : elle peut avoir lieu entre les hommes et/ou les systèmes. Les ontologies permettent alors le partage de la compréhension et la communication dans des contextes particuliers et selon les besoins. Ainsi, on peut utiliser l'ontologie pour créer un réseau de relations qui définit les connexions entre les composants du système. Cette caractéristique de communication est offerte grâce à la non-ambiguïté des termes utilisés et définis par l'ontologie dans le système.

Explication de l'implicite : l'ontologie explicite les connaissances implicite pour résoudre le problème d'ambiguïté entre concepts.

Proposition d'un méta- modèle : par définition, un modèle constitue une abstraction d'une réalité. Comme la notion d'ontologie se base sur la définition des concepts et des relations elle s'intègre parfaitement dans une solution pour création des modèles.

1.3 Les composants d'une ontologie

1.3.1 Concepts

Un concept peut représenter un objet matériel, une notion, une idée ; les concepts sont aussi appelés termes ou classe de l'ontologie, constituent les objets de base manipulés par les ontologies. Ils correspondent aux abstractions pertinentes du domaine du problème, retenues en fonction des objectifs qu'on se donne et de l'application envisagée pour l'ontologie. [M. Uschold & M. King ,1995].

Selon [Gomez-Perez, 1999], les concepts peuvent être classifiés selon plusieurs dimensions :

- Niveau d'abstraction : concret ou abstrait,
- Atomicité : élémentaire ou composé,
- Niveau de réalité : réel ou fictif,

Un concept peut aussi être divisé en trois parties :

- **Un Terme** : ce dernier permet de désigner un concept ou bien 'label de concept'. **Exemple** : Médecin
- **Une Intension** : ou notion est un ensemble de propriétés qualitatives ou fonctionnelles communes aux individus auxquels le concept s'applique, et permettant de définir le concept.

Exemple : Établissement de soin conçu et aménagé pour les soins d'un grand nombre de personnes malade.

- **Une Extension** : ou ensemble d'objets qui regroupe les objets manipulés à travers le concept, autrement dit instance du concept.

Exemple : La liste des établissements de soin

Par ailleurs, un terme peut ne pas avoir une extension comme par exemple le concept « Vérité » qui a le sens de « Ce Qui Est Vrai », il s'agit alors d'un concept générique qui correspond à une notion abstraite. [G.Falquet ,01]

Souligne qu'on peut trouver des concepts partageant la même extension mais pas leur intention et portent le même terme, ceci correspond à des points de vue différents sur un même concept.

Un Concept peut être caractérisé selon les propriétés qui lui sont associé, comme suit :

- **La généralité** : un concept est dit générique s'il n'admet pas d'extension, par **exemple** : La vérité est un concept générique.
- **L'identité** : un concept porte une propriété d'identité si cette dernière peut différencier deux instances de ce concept, par **exemple** : le concept « étudiant » porte une propriété d'identité liée au « numéro de l'étudiant », deux étudiants ne peuvent pas avoir le même numéro.
- **La rigidité** : un concept est dit rigide si toute extension de ce concept reste une extension de toutes les connaissances possibles, **Exemple** : « Humain » concept rigide. « Etudiant » concept non rigide.
- **L'anti-rigidité** : un concept est dit anti-rigide s'il peut être une instance pour d'autres concepts, par **exemple** : « Chercheur » est un concept anti-rigide car « Chercheur » est avant tout « Humain ».
- **L'Unité** : un concept est dit concept unité, si pour chacune des ses instances, les différentes parties de l'instance sont liés par une relation qui ne lie pas d'autres instances de concepts, par **exemple** : les deux parties d'un couteau sont reliés par une relation « Liée », « poignée » est liée à une « lame ».

Deux concepts portent les propriétés suivantes :

- **L'équivalence** : Deux concepts sont équivalents s'ils ont la même extension.
- **La disjonction** : Deux concepts sont disjoints si leurs extensions sont disjointes, **exemple** : « Lumière », « Obscurité ».
- **La dépendance** : Un concept C1 est dépendant d'un concept C2 si pour toute instance de C1 il existe une instance C2 qui ne soit ni partie ni constituant de l'instance C2, par exemple : « Père » est un concept dépendant de « Fils » et vice-versa.

1.3.2 Relations

Les relations traduisent les interactions existant entre les concepts présents dans le domaine d'analyse. Elles sont formellement définies comme tout sous ensemble d'un produit cartésien de n ensembles, c'est-à-dire $R : C1 \times C2 \times \dots \times Cn$

Les concepts peuvent être reliés entre eux par des relations au sein d'une ontologie. Par exemple : la relation « Écrit » lie une instance de concept « Personne » et une instance du concept « Article », dans cet ordre. Ces relations incluent les associations suivantes :

- Sous-classe-de/Is-a : Spécialisation, généralisation,
- Partie-de : agrégation ou composition,
- Associé-a, instance-de

Ces relations nous permettent de capturer la structuration ainsi que l'interaction entre les concepts, ce qui permet de représenter une grande partie de la sémantique via l'ontologie

1.3.3 Fonctions

Les fonctions sont des cas particuliers de relations dans lesquelles le nième élément de la relation est défini de manière unique à partir des n-1 éléments précédents. Formellement, les fonctions sont définies ainsi : $F : C_1 x C_2 \dots x C_{n-1} \rightarrow C_n$.

1.3.4 Axiomes

Permettent de modéliser des assertions toujours vraies, à propos des abstractions du domaine traduites par l'ontologie. Ils permettent de combiner des concepts, des relations et des fonctions pour définir des règles d'inférences et qui peuvent intervenir, par exemple, dans la déduction, la définition des concepts et des relations, ou alors pour restreindre les valeurs des propriétés ou les arguments d'une relation.

1.3.5 Les instances

Les instances ou individus constituent la définition extensionnelle de l'ontologie. Ils représentent des éléments singuliers véhiculant les connaissances à propos du domaine du problème.

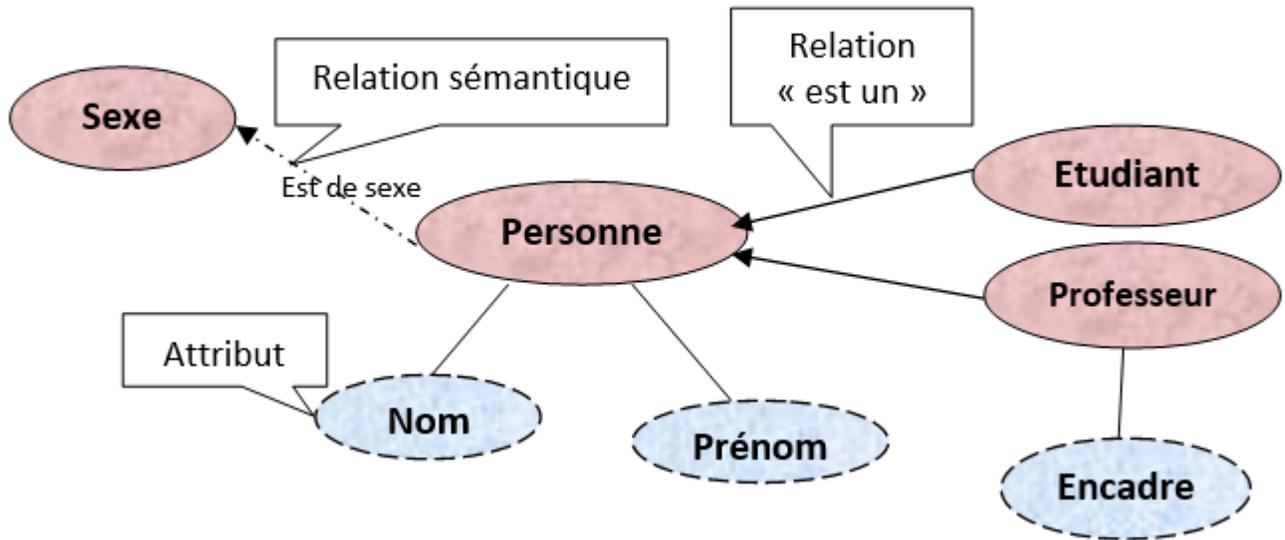


FIGURE 1.1 – schéma qui représente les différentes structures d'ontologie

1.4 La Typologie des ontologies

Les ontologies peuvent être classifiées selon plusieurs dimensions à savoir : objet de conceptualisation, niveau de formalisme de représentation, niveau de détail, et niveau de complétude. [Psyché, Mendes & Bourdeau, 2004] Comme le proposent [N. Guarino, 1998] la classification peut également se faire en fonction des objets que modélisent les ontologies pour répondre à un objectif précis :

1.4.1 Typologie selon l'objet de conceptualisation

Les ontologies de représentation des connaissances

Ce type d'ontologies regroupe les concepts impliqués dans la formalisation des connaissances. On peut citer l'exemple de l'ontologie de frames [Gomez- Perez, 1999], qui définit les primitives de représentation des langages à base de frames : Classes, Instances, facettes, Propriétés, relations.

Les ontologies de niveau supérieures (Haut niveau) Ce type d'ontologies vise à étudier les catégories des choses qui existent dans le monde, comme les concepts de niveau supérieur d'abstraction tels que les entités, les événements, les processus, les actions, le temps, les relations et les propriétés. L'ontologie de niveau

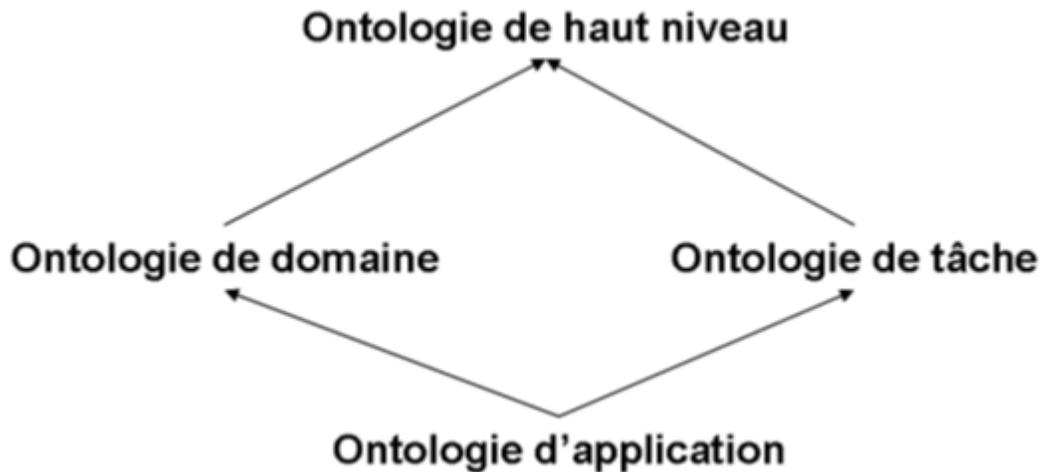


FIGURE 1.2 – Classification des Ontologies selon [N.Guarino, 1998]

supérieur est fondée sur la théorie de l'identité et la théorie de la dépendance. Des recherches ont été poursuivies sur la théorie de l'ontologie, ces dernières ont affirmé que les fondements philosophiques étaient des principes à suivre afin de concevoir l'ontologie de niveau supérieur. [Guarino, 1997]

Les ontologies génériques

Ce type d'ontologies aussi appelée méta-ontologies, véhiculent des connaissances génériques moins abstraites que celles véhiculées par l'ontologie de niveau supérieure, ces dernières peuvent être réutilisées à travers différents domaines. Elle peut adresser des connaissances réelles ou encore des connaissances visant à résoudre des problèmes génériques à travers différents domaines. [Gomez & Perez, 1999]

On prend deux exemples d'ontologies citées par [Borst,1997] :

1. l'ontologie méréologique qui est une branche de l'ontologie formelle et qui est une application de la logique des prédicats qui traite des relations entre la partie et le tout comme par **exemple** :

Le Toit d'une maison,
 Les organes du corps,
 Les cellules de l'organisme,

2. l'ontologie topologique tout comme la théorie de méréologie, l'ontologie to-

pologique contient des relations, associé-à.

Les Ontologies de domaine

Ce type d'ontologies est construit sur un domaine de connaissances bien spécifique ; elles fournissent le vocabulaire des concepts de ce domaine et les relations entre ces derniers, les activités de ce domaine ainsi que les théories et les principes de base de ce domaine. Ces ontologies de domaine constituent donc des méta-descriptions d'une représentation de connaissances du domaine. La plupart des ontologies existantes sont des ontologies du domaine. Elle caractérise la connaissance du domaine où la tâche est réalisée. Un domaine serait par exemple : la médecine, la recherche.

Les Ontologies de tâches

Ce type d'ontologies est utilisé pour gérer ou conceptualiser des tâches spécifiques dans les systèmes, telles que les tâches de diagnostique, de planification, de conception, de configuration et de tutorat, soit tout ce qui concerne la résolution de problèmes. Elle fournit un ensemble de vocabulaires et de concepts qui décrit une structure de résolution des problèmes inhérente aux tâches et indépendante du domaine.

Les Ontologies d'application

Ce sont les ontologies les plus spécifiques. Elles permettent de décrire des concepts dépendants à la fois d'un domaine et d'une tâche. Dans cette classification, la notion d'ontologie d'application définit le contexte d'une application qui décrit la sémantique des informations et des services manipulés par une ou un ensemble d'applications sur un même domaine.

1.4.2 Selon le niveau de formalisme de représentation

Ceci est lié au niveau du formalisme de représentation du langage utilisé pour rendre l'ontologie opérationnelle. [M.uschold & M.gruninger, 1996b]

Nous pouvons citer quatre catégories qui sont :

Ontologies Informelles

ontologies opérationnelles dans un langage naturel sans aucune restriction (sémantique ouverte), cela aide à rendre l'ontologie plus compréhensible pour l'utilisateur.

Ontologies Semi-informelles

utilisation d'un langage naturel structuré, cela permet d'augmenter la clarté de l'ontologie tout en réduisant l'ambiguïté.

Ontologies Semi-formelles

utilisation d'un langage artificiel défini formellement en utilisant des classes, des relations, des fonctions, objets et des axiomes pour décrire une ontologie.

Ontologies 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. [Gomez-Perez, 1999]

1.4.3 La typologie selon le niveau de détail

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 :

Granularité fine

correspondant à des ontologies très détaillées, possédant ainsi un vocabulaire plus riche capable d'assurer une description détaillée des concepts pertinents d'un domaine ou d'une tâche. Ce niveau de granularité est très utile afin d'établir un accord entre les agents qui utiliseront cette ontologie.

Granularité large

correspondant à un vocabulaire moins détaillé, par exemple l'ontologie de haut niveau possède une granularité large ; les concepts qu'elle traduit sont normalement détaillés dans d'autres ontologies de domaine ou d'application.

1.5 Méthodologie de construction d'une ontologie

Il existe plusieurs méthodes d'ingénierie ontologique mais jusqu'à présent aucune d'elles n'est proposée comme une méthodologie générale de construction d'ontologie ou bien normalisée, de ce fait certains auteurs ont tenus à proposer des méthodologies inspirées de leur expérience de construction d'ontologie.

On entend par méthodologie un ensemble de procédures de travail, d'étapes, un cycle de développement qui pourra être adopté lors de la construction d'ontologies. C'est dans les années 1995 que les premières ontologies commencent à naître et parmi ces ontologies on cite :

1.5.1 La méthode ENTREPRISE

Ils ont proposés une première méthode de construction d'ontologie inspirée de leur expérience acquise lors du développement des ontologies dans le domaine de la gestion [M. Uschold & M. King, 1995], Cette méthode repose sur les quatre étapes suivantes :

- A) identification du but et la portée de l'ontologie dont les raisons pour lesquelles l'ontologie en cours de construction est clarifiée ainsi que les utilisateurs potentiels de l'ontologie.
- B) Construire l'ontologie cette étape est divisée en trois activités :

- **Capture de l'ontologie** : Identifier les concepts et les relations fondamentaux, produire des définitions précises et non ambiguës à ces éléments en langage naturel, identifier les termes dénotant ces éléments et enfin essayer d'arriver à un argument.

Désormais, cette analyse est insuffisante afin de spécifier la sémantique de ce domaine, certaines notions doivent d'abord être lues par un expert ou un spécialiste du domaine. La sémantique doit alors être validée par les experts du domaine.

Approche descendante, à partir d'un nombre réduits de concepts qu'on veut spécialiser. **Approche ascendante**, consiste à considérer les termes spécifiques et de trouver des termes génériques associés.

Approche intermédiaire, dans laquelle les concepts se structurent autour des concepts importants du domaine ni trop généraux, ni trop spécifiques.

- **Codage de l'ontologie** : La représentation explicite de la conceptualisation dans un langage formel.
- **Réutiliser et intégrer** : Éventuellement des ontologies existantes, Cette activité peut être effectuée en parallèle avec l'activité de capture et/ou de codage.

- C) évaluer l'ontologie. Cette étape consiste à confronter l'ontologie aux objectifs, logiciels et utilisateurs pour lesquelles elle a été élaborée, évaluer l'ontologie en vérifiant la validité de la taxonomie et vérifier si toutes les instances d'une classe sont aussi des instances de la classe mère et qu'il n'y a pas de classe isolée.
- D) documenter l'ontologie. Cette étape consiste à établir des instructions de documentation d'ontologie qui diffèrent selon le type et le but.

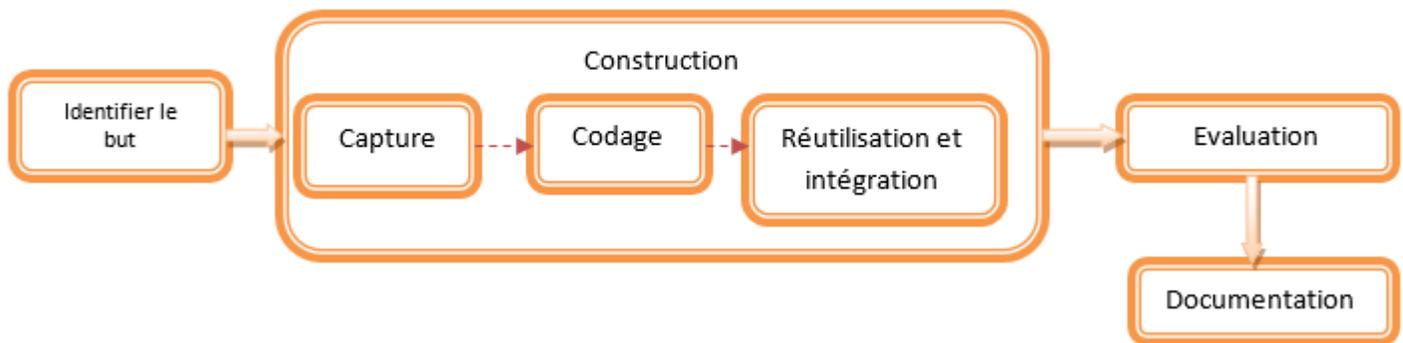


FIGURE 1.3 – méthode d'Uschold et King.

1.5.2 La Méthode TOVE

Cette méthode est basée sur l'expérience du développement de l'ontologie du projet TOVE (Toronto Virtual Enterprise). Elle aboutit à la construction d'un modèle logique de connaissance. [M.gruninger & M.fox, 1995] l'ontologie est développée selon les étapes suivantes :

- Identification des scénarios Cette étape consiste à identifier les applications possibles dans laquelle l'ontologie sera employée.
- Formulation de questions informelles Cette étape consiste à formuler un ensemble de questions basées sur des scénarios exprimées en langage naturel, afin de déterminer la portée de l'ontologie. Ces questions et réponses sont utiles afin d'extraire les concepts principaux, leurs propriétés et les relations qui existe entre ces concepts.
- Spécification Formelle Cette étape consiste à représenter des axiomes et des définitions pour les termes de la terminologie, en utilisant le formalisme de la logique du premier ordre. Les concepts seront représentés sous forme de constantes ou bien de variables. D'autre part, les propriétés et les relations seront représentées par des prédicats.
- Évaluation de la complétude de l'ontologie Cette étape consiste à effectuer deux tâches telles que la vérification et la validation, en assurant la conformité et la fidélité sémantique de l'ontologie au domaine de connaissances.

1.5.3 La Méthode METHONTOLOGY

Cette méthode est développée au Laboratoire d'Intelligence Artificielle de l'Université polytechnique de Madrid. Elle vise la construction d'ontologies au "niveau connaissance" [Fernandez-Lopez, 1999] et repose sur :

- **Spécification**
Cette étape a pour but de fournir une description claire du problème étudié ainsi que la façon de le résoudre. Elle permet de préciser l'objectif, la portée et le degré de granularité de l'ontologie qui sera construite.
- **Conceptualisation**

Cette étape consiste à organiser et structurer la connaissance acquise durant la phase spécification en utilisant des langages d'implémentation dans lesquelles l'ontologie va être formalisée. Les représentations intermédiaires utilisées sont : les taxonomies de concepts, les diagrammes des relations binaires, le glossaire des termes, le dictionnaire des concepts, le tableau des relations binaires, spécifier des contraintes sur les attributs dans une table d'attributs, spécifier des axiomes sur les concepts dans une table d'axiomes logiques, décrire les instances des concepts dans une table d'instances.

– **Formalisation**

Cette étape consiste à choisir ou à construire un langage formel capable d'intégrer toutes les propriétés de l'ontologie l'intérêt de l'utilisation d'un langage formel et de permettre de réduire les ambiguïtés du langage naturel en offrant une plus grande expressivité et d'autre part rendre l'ontologie compréhensible par les machines.

– **Implémentation**

Cette étape consiste à la codification de l'ontologie formelle dans un langage opérationnel du Web Sémantique.

– **Maintenance**

Cela peut s'agir d'une maintenance évolutive de l'ontologie (nouveaux besoins de l'utilisateur), ce qui permet la validation de celle-ci. Cette activité est généralement faite par le constructeur et des experts du domaine. La validation se base sur l'exploitation des services d'inférences associés à la logique de Description, et qui sont offerts par des raisonneurs.

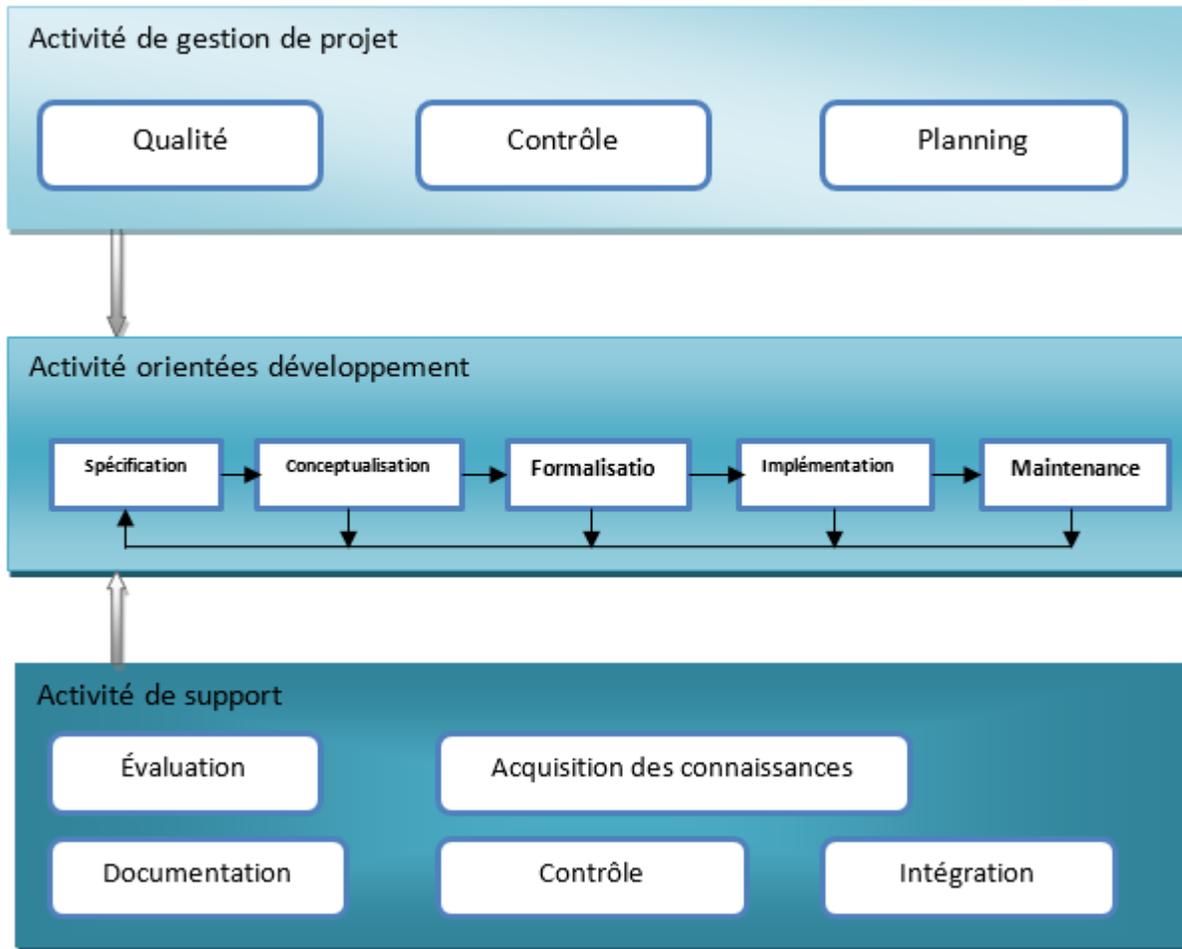


FIGURE 1.4 – processus de développement et cycle de vie de METHONTOLOGY

1.6 Formalismes de représentation des connaissances

1.6.1 Frames

GRUBER a initialement proposé Le modèle des Frames [M. Kifer, G. Lausen and J. Wu, 1995] comme langage de représentation d'ontologies, Le principe de ce modèle est de décomposer les connaissances en classes (ou frames) qui représentent les concepts du domaine. À un frame est rattaché un certain nombre d'attributs (slots), chaque attribut pouvant prendre ses valeurs parmi un ensemble de facettes (facets). Une autre façon de présenter ces attributs est de les considérer comme des relations binaires entre classes dont le premier argument est appelé domaine (domain) et la deuxième portée (range). Des instances des classes, correspondant à l'extension de chaque concept, peuvent être ajoutées, ainsi que des fonctions qui sont des types particuliers de relations liant un ensemble de classes

à une valeur calculée à partir des valeurs des attributs des classes. La spécification de propriétés conceptuelles des attributs (ou relations) recourt à des formules de la logique du premier ordre. [F. Amourache, 2008]

1.6.2 Logiques de descriptions

Les logiques de description (LDs) découlent directement des travaux fondateurs de Bachmann et de son système KL-ONE. Depuis le début des années 90, la recherche en logique de description s'est considérablement développée. Les logiques de description peuvent être considérées comme un fragment de la logique du premier ordre, dans lequel les formules ont une variable libre pour les descriptions de concept et deux variables libres pour les descriptions de relations. [F. Amourache, 2008]

Une LD est composée de deux parties :

- **Un langage terminologique TBOX** : définition des notions basiques ou dérivées et de comment elles sont reliées entre elles. Ces informations sont "génériques" ou "globales", vraies dans tous les modèles et pour tous les individus. La première colonne du tableau présente un exemple de T-Box, Les prochains paragraphes explicitent divers aspects des T-Box en se référant à cet exemple.

T-Box	A-Box
Femelle $\sqsubseteq \top$	Humain (Amel)
Male $\sqsubseteq \top$	Femelle(Amel)
Femelle $\sqsubseteq \top \sqcap \neg \text{Male}$	Femme (Lynda)
Male $\sqsubseteq \top \neg \text{Femelle}$	Humain (Sofiane)
Animal $\equiv \text{Male} \sqcup \text{Femelle}$	$\neg \text{Femelle}(\text{Sofiane})$
Humain $\sqsubseteq \text{Animal}$	Homme (Amar)
Femme $\equiv \text{Humain} \sqcap \text{Femelle}$	relationParentEnfant (Lynda, Amel)
Homme $\equiv \text{Humain} \sqcup \neg \text{Femelle}$	relaionParentEnfant (Sofiane, Amar)
Mère $\equiv \text{Femme} \sqcup \exists \text{ relationParentEnfant}$	
Père $\equiv \text{Homme} \sqcup \exists \text{ relationParentEnfant}$	
MèreSansFille $\equiv \text{Mère}$	
$\sqcap \forall \text{ relationParentEnfant.} \neg \text{Femme}$	
relationParentEnfant $\sqsubseteq \top_R$	

TABLE 1.1 – Une base de connaissances composée d’une T-Box et d’une A-Box

1. **Les entités atomiques** : Les concepts atomiques et les rôles atomiques constituent les entités élémentaires d’une T-Box. Les noms débutant par une lettre majuscule désignent les concepts, alors que ceux débutant par une lettre minuscule dénomment les rôles (par exemple : les concepts : Femelle, Mâle, Homme et Femme, et le rôle relation Parent Enfant).
2. **Les concepts et les rôles atomiques prédéfinis** : Les LDs prédéfinissent minimalement quatre concepts atomiques : le concept τ et le rôle \hat{R} , les plus généraux de leur catégorie respective, et le concept $\hat{}$ ainsi que le rôle \hat{R} les plus spécifiques (c’est-à-dire l’ensemble vide).
3. **Les entités composées** : Les concepts et les rôles atomiques peuvent être combinés au moyen de constructeurs pour former respectivement des concepts et des rôles composés. Par exemple, le concept composé Mâle \sqcap Femelle résulte de l’application du constructeur \sqcap aux concepts atomiques Mâle et Femelle. Le concept Mâle \sqcap Femelle s’interprète comme l’ensemble des individus qui appartiennent aux concepts Mâle et Femelle. Les différentes LDs se distinguent par les constructeurs qu’elles proposent. Plus les LDs sont expressives, plus les chances sont grandes que les problèmes d’inférence soient non décidables ou de complexité très élevée. Par contre, les LDs trop peu expressives démontrent une inapti-

tude à représenter des domaines complexes.

4. **La définition formelle de T-Box** : Une T-Box contient des axiomes terminologiques de la forme $C \Xi D$ ou $C \subseteq D$. La première sert à énoncer des relations d'équivalence entre concepts, alors que la seconde permet d'exprimer des relations d'inclusion. Une interprétation I satisfait un axiome $C \subseteq D$ si et seulement si $CI \hat{=} DI$. Une interprétation I satisfait un axiome $C \Xi D$ si et seulement si $CI = DI$. Une interprétation satisfait une T-Box (est un modèle de T-Box) si et seulement si l'interprétation satisfait tous les axiomes de la T-Box.

- **un langage assertiennel ABOX** : défini par un ensemble d'individus désignant des objets nécessairement différents dans toute interprétation et d'instances de concepts et de relations vérifiées par ces individus. Une A-Box contient un ensemble d'assertions sur les individus : (1) **des assertions d'appartenance** et (2) **des assertions de rôle**.

Chaque A-Box doit être associée à une T-Box, car les assertions s'expriment en termes de concepts et de rôles de la T-Box. La deuxième colonne du Tableau 1 illustre un exemple d'A-Box. Une A-Box désigne des individus dans ses assertions par des noms qu'elle leur donne. L'exemple du Tableau 1.1 comprend les individus nommés suivants : Amel, Sofiane, Amar et Lynda.

Dans ce qui suit nous représentons par les lettres a, b les individus nommés. Une fonction d'interprétation assigne à chacun de ces noms a , un individu aI tel que $aI \hat{=} DI$. Les moteurs d'inférence pour les logiques de descriptions adoptent généralement l'hypothèse de noms uniques, c'est-à-dire que pour tout individu nommé a et $b \Rightarrow aI^1bI$.

Chaque assertion d'appartenance d'une A-Box (notée $C(a)$ ou $a : C$), déclare que pour cette A-Box, il existe un individu nommé a , membre du concept C de la T-Box associée. Une interprétation satisfait une assertion d'appartenance $C(a)$ si et seulement si $aI \hat{=} CI$. Une assertion de rôle, de la forme $R(a, b)$ (ou $(a, b) : R$) indique que pour cette A-Box, il existe un individu nommé a qui est en relation avec un individu nommé b par le rôle R (défini dans la T-Box associée), tel que a fait partie du domaine de R et b fait partie de l'image (le Co-domaine) de R . Une interprétation satisfait une assertion de rôle $R(a, b)$ si et seulement si $(aI, bI) \hat{=} RI$.

1.6.3 Réseaux sémantique

Les réseaux sémantiques ont été proposés par Quillian en 1968 pour représenter explicitement un modèle psychologique de la mémoire associative humaine.

Un réseau sémantique est composé d'une part, de noeuds qui peuvent représenter indifféremment des objets, des concepts ou des évènements et d'autre part, de liens (arcs orientés étiquetés) entre les noeuds qui représentent leurs relations. Chaque réseau sémantique est donc une structure de graphe dédiée à la description d'objets et de leurs relations binaires.

De plus, ils peuvent aussi servir de représentation graphique de prédicats binaires de la logique des prédicats. Les symboles de variables, fonctions et prédicats sont alors représentés par un réseau de noeuds : les variables et les fonctions sont représentés par des noeuds et les prédicats binaires par des liens. [9]

1.7 Les langages de représentation d'ontologies

Pour le développement d'ontologies, il est nécessaire de choisir un langage ou un ensemble de langages dans lesquelles cette dernière est exprimée et utilisée, pour cela nous allons présenter les différents types de langages tel-que XML ; RDF/RDFs ; OWL et enfin DAML-OIL.

1.7.1 XML (Extensible Markup Language)

XML est un langage de description de documents utilisé pour décrire la structure des documents qui dérive de SGML (Standard Generalized Markup Language) et HTML (Hyper Text Markup Language). Il s'agit alors d'un langage formé de balises qui permet de structurer les documents ou bien qui décrit comment les balises sont utilisées pour diviser les documents de données structurées en différentes parties et la manière d'identifier ces parties. Il est utilisé dans la plupart des projets publication sur le web et dans les bases de données. Ce dernier est beaucoup plus utilisé pour stocker des documents que pour échanger des données. [01]

1.7.2 RDF (Resource Description Framework)

Développé [2] par le W3C (World Wide Web Consortium), RDF est le langage de base de web sémantique dont chaque ressource est pourvue d'un identifiant URI (Uniform ressource identifier). Tout document RDF est composé d'un ensemble de

triplets sujet, prédicat, objet. Un ensemble de tels triplets est appelé un graphe RDF. Ceci peut être illustré par un diagramme composé de noeuds et d'arcs orientés, dans lequel chaque triplet est représenté par un lien noeud-arc-noeud (d'où le terme de "graphe"), ou :

- **Sujet** : Représente la ressource décrite, pointée par une URI
- **Prédicat** : Représente la relation utilisée pour décrire une ressource
- **Objet** : Représente la valeur d'une propriété associée à une ressource spécifique.

Exemple



FIGURE 1.5 – Exemple de graphe RDF

L'exemple suivant est tiré du site du W3C qui décrit une ressource avec les déclarations "il y a une personne qui a comme identifiant `http://www.w3.org/People/EM/contact#me`, dont le nom est **Eric Miller**, dont l'adresse email est **em@w3.org**, et qui a le titre de **"Docteur"**".

La ressource "`http://www.w3.org/People/EM/contact#me`" est le **sujet**.

L'objet est : "Eric Miller" (avec le prédicat "quel est son nom"), `em@w3.org` (avec le prédicat "quel est son email"), et "Dr." (avec le prédicat "quel est son titre").

Le sujet est une URI.

Les prédicats sont aussi des URIs. Par exemple, l'URI pour chaque prédicat est :

"quel est son nom" est `http://www.w3.org/2000/10/swap/pim/contact#fullName`,
 "quel est son email" est `http://www.w3.org/2000/10/swap/pim/contact#mailbox`,
 "quel est son titre" est `http://www.w3.org/2000/10/swap/pim/contact#personalTitle`.
 De plus, le sujet a le type (avec le prédicat `http://www.w3.org/1999/02/22-rdf-syntax-ns#type`) `personne`
 (with URI `http://www.w3.org/2000/10/swap/pim/contact#Person`).

Par conséquent, les suivants "sujet, prédicat, objet", c'est-à-dire les suivants triplets RDF peuvent être exprimés :

`http://www.w3.org/People/EM/contact#me`,
`http://www.w3.org/2000/10/swap/pim/contact#fullName`, "Eric Miller"

`http://www.w3.org/People/EM/contact#me`,
`http://www.w3.org/2000/10/swap/pim/contact#personalTitle`, "Dr."

`http://www.w3.org/People/EM/contact#me`, `http://www.w3.org/1999/02/22-rdf-syntax-ns#type`, `http://www.w3.org/2000/10/swap/pim/contact#Person`

`http://www.w3.org/People/EM/contact#me`, `http://www.w3.org/2000/10/swap/pim/contact#mailbox`, `em@w3.org`

RDF peut aussi être exprimé en XML afin notamment d'échanger ses données avec les agents logiciels du Web ou autres. A propos du langage, un document RDF est délimité par l'élément `rdf:RDF` qui comporte un ou plusieurs éléments `rdf:Description` pour chacune des descriptions de ressources comprises dans le document. Chaque description comprend un attribut `rdf:about` qui pointe vers l'URI de la ressource à décrire et un à plusieurs éléments représentant chacun un prédicat. Lorsqu'un prédicat a pour valeur une autre ressource, l'attribut `rdf:resource` pointera vers son URI. Les autres caractéristiques de RDF permettent la composition des énoncés en structures plus complexes comme le groupement de sujets et/ou d'objets en listes énumérées ou bien la réification d'énoncés, i.e. la création de nouveaux énoncés à partir d'énoncés existants. RDF répond aux besoins de la plupart des outils d'annotation. En effet, les documents RDF sont des

documents XML valides, leur modélisation sous forme de réseau sémantique apporte une flexibilité nécessaire et il est possible de réutiliser des énoncés existants pour composer des documents RDF plus complexes.

Exemple : Amayas a 19 ans et habite à Tizi-ouzou

```

<rdf :RDF>
<rdf :Description about='Amayas'>
<rdf :Property about='ville'>
    Tizi-Ouzou
</rdf :Property>
<rdf :Property about='age'>
    19
</rdf :Property>
</rdf :Description>
</rdf :RDF>

```

FIGURE 1.6 – Exemple de document RDF

Par contre, RDF ne fournit pas de mécanisme de contrainte de classes ou de types pour les différentes parties du triplet. Il n'est donc pas assez puissant pour représenter de vraies ontologies avec un système de raisonnement approprié.

1.7.3 RDFS (RDF Schéma)

RDFS décrit les ressources avec des classes, les propriétés et les valeurs, autrement dit c'est un langage extensible de représentation des connaissances et qui permet aussi de définir le vocabulaire pour décrire des classes et des propriétés hiérarchisées en taxinomies. [03]

Sur l'exemple de Amayas, nous définissons le concept de personne, une taxinomie de concepts et l'instance Amayas.

1.7.4 OWL (Web Ontology Language)

Le langage d'ontologie Web OWL est conçu pour être utilisé par les applications qui ont besoin de traiter le contenu de l'information au lieu de simplement présenter des informations pour les humains. OWL est une extension de RDF et RDFS qui fournit un vocabulaire supplémentaire avec une sémantique formelle. OWL a trois sous-langages plus expressifs : OWL Lite, OWL DL et OWL Full.[04]

```

<rdf :RDF>
<rdfs :Class rdf :about='Personne'>
  <rdfs :subClassOf rdf :resource='Thing' />
</rdfs :Class>

<rdf :Property about='age'>
<rdfs :domain rdf :resource='Personne' />
<rdfs :range rdf :resource='xsd :integer' />
</rdf :Property>
<rdf :Property about='ville'>
<rdfs :domain rdf :resource='Personne' />
<rdfs :range rdf :resource='xsd :string' />
</rdf :Property>
</rdf :RDF>

<Personne rdf :ID='Amayas'>
<age rdf :resource='19' />
<ville rdf :resource='Tizi-Ouzou' />
</Personne>

```

FIGURE 1.7 – Exemple de document RDFs

OWL Lite

Moins Complexe, destiné aux utilisateurs qui ont besoin d'une hiérarchie de concepts simple.

OWL DL

Complexe, permet une expressivité plus importante et garantit la complétude des raisonnements.

OWL FULL

Plus complexe, permet le plus haut niveau d'expressivité.

1.7.5 DAML-OIL :

DAML + OIL est un langage de balisage sémantique pour les ressources web. Il s'appuie sur les standards du W3C antérieurs, tels que RDF et RDF schéma, et s'étend ces langues avec riches primitives de modélisation. DAML + OIL fournit des primitives de modélisation couramment trouvés dans les langues basé sur les images. Il a été construit à partir de la langue originale de l'ontologie DAML-ONT

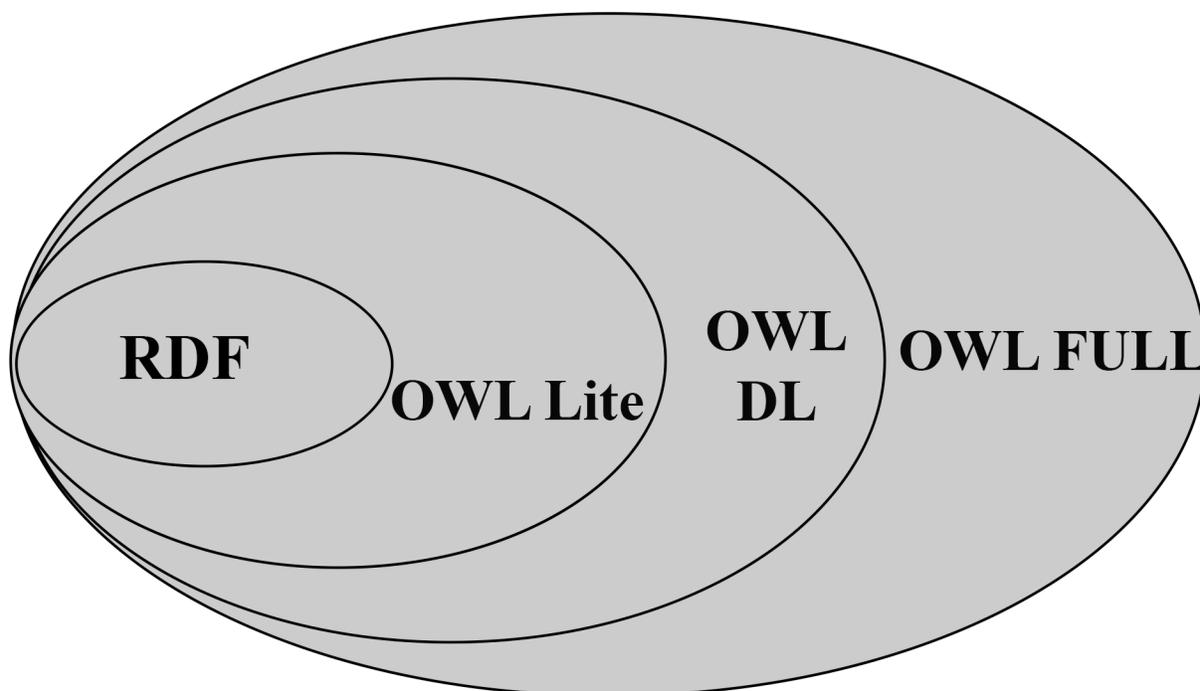


FIGURE 1.8 – Hiérarchie de langage OWL

dans le but de combiner de nombreux composants linguistiques. Le langage a une sémantique propre et bien défini [05]

1.8 Les outils d'édition

Il existe plusieurs éditeurs d'ontologies qui utilisent des formalismes variés et offrent différentes fonctionnalités. Dans ce qui suit nous allons définir quelques outils : **Ontolingua**, **WebOnto**, **Protégé**.

1.8.1 Ontolingua

[Farquhar & Fikes, 1996] Développé à l'université de Stanford, l'architecture de son serveur permet d'accéder à une bibliothèque d'ontologies et à des traducteurs de langages de programmation (Prolog, Lomm) et possède un éditeur de création et de parcours des ontologies. Il permet trois types d'interaction les collaborateurs qui souhaitent écrire et examiner des ontologies à distance, Applications éloignées susceptibles de vouloir interroger et modifier des ontologies sur

le serveur via l'Internet et les applications locales.

Pour intégrer les ontologies Ontolingua on distingue trois possibilités

Inclusion : en utilisant et incluent les définitions d'autres ontologies.

Restriction : en important les définitions depuis d'autres ontologies et les rendre plus spécifiques.

Raffinement polymorphe : on redéfinit une définition importée depuis n'importe quelle ontologie.

1.8.2 WebOnto

Développer au Knowledge Media Institute à L'Open University. C'est un outil accessible sur Internet et principalement graphique permettant de construire coopérativement des ontologies. Il permet une visualisation graphique et séparée des différents composants d'une ontologie (classes, instances, relation, règles, procédures) adaptée à la construction d'ontologies de grande taille. Et il offre éventuellement des services inférentiels, basés sur le langage OCML permettant de répondre à des requêtes et des vérifications de cohérence.[6]

1.8.3 Protégé

Protégé [Noy, et al., 2000] C'est un éditeur d'ontologie qui a été développé par le Département d'Informatique Médicale de l'Université de Stanford ; PROTÉGÉ est une plate-forme Open Source autonome, qui fournit un environnement graphique permettant l'édition, la visualisation et la vérification des contraintes (contrôle) de l'ontologie.

Le modèle de représentation de connaissances de PROTÉGÉ, est issu du modèle des frames. Ce dernier contient des classes (pour modéliser les concepts), des slots (pour modéliser les attributs des concepts) et des facettes (pour définir les valeurs des propriétés et des contraintes sur ces valeurs), ainsi que des instances des classes.

L'édition des listes de ces trois types d'objets se fait par l'intermédiaire de l'interface graphique, sans avoir besoin d'exprimer ce que l'on a à spécifier dans un langage formel : il suffit juste de remplir les différents formulaires correspondant à ce que l'on veut spécifier. L'interface très bien conçue, et l'architecture logicielle permettant l'insertion de plugins pouvant apporter de nouvelles fonctionnalités, notamment les plugins pour gérer les représentations sous forme graphique, par exemple OWLViz et la prise en charge de nouveaux langages ,ont participé à son

succès et le rendent l'éditeur d'ontologie jouissant de la plus grande renommée à l'heure actuelle, servant de référence pour une importante communauté d'utilisateurs.

1.9 Domaines d'application d'ontologies

1.9.1 Les ontologies et la représentation des connaissances

Les ontologies sont apparues en informatique, plus précisément en Ingénierie des Connaissances, dans le cadre des démarches d'acquisition des connaissances pour les systèmes à base de connaissances (SBC). Les SBC proposaient alors de spécifier, d'un côté des connaissances du domaine modélisé, et de l'autre, des connaissances de raisonnement qui manipule et utilise ces connaissances du domaine. L'idée de cette séparation modulaire était de construire mieux et plus rapidement des SBC en réutilisant le plus possible des composants génériques, que ce soit au niveau du raisonnement ou des connaissances du domaine.

En conclusion, l'objectif de l'ingénierie ontologique est de diversifier les applications des Systèmes à Base de Connaissances (SBC), et de permettre une représentation des connaissances indépendantes de ces diverses applications, de manière à assurer sa portabilité d'une application à l'autre.

[**Bachimont 2003**]

1.9.2 Les ontologies et le web sémantique

Le Web actuel est essentiellement syntaxique, la structure des ressources étant bien définie, mais leur contenu restant inaccessible aux traitements machines, seuls les humains étant capables de l'interpréter. [**Lee 2002**]

Le Web sémantique a alors l'ambition de lever cette difficulté en associant aux ressources du Web des entités ontologiques comme références sémantiques, ce qui permettra aux différents agents logiciels d'accéder et d'exploiter directement le contenu des ressources et de raisonner dessus. Ce référencement sémantique peut aussi résoudre les problèmes d'interprétation des ressources informationnelles provenant des applications hétérogènes et réparties et de permettre ainsi à ces applications d'être intégrées sémantiquement. [**Ushold 2002**]

L'architecture du Web sémantique repose sur une hiérarchie des langages d'assertion et de description d'ontologies ainsi que sur un ensemble de services pour

l'accès aux ressources au moyen de leurs références sémantiques, pour gérer l'évolution des ontologies, pour l'utilisation des moteurs d'inférences capables d'effectuer des raisonnements complexes ainsi que des services pour la vérification de la validité sémantique de ces raisonnements. [Oberle 04]

1.10 Conclusion

Nous avons détaillé dans ce chapitre la notion d'ontologie, en essayant d'éclaircir la notion en présentant certaines définitions. Nous avons montré aussi les différents intérêts des ontologies et sa structure, ainsi que les différentes classifications, les méthodologies les plus représentatives de leur construction. Ensuite nous avons présenté les principaux formalismes de représentation de connaissances. Finalement nous avons étudié les langages d'ontologies et les outils nécessaires aux développements des ontologies dans divers domaines d'applications. La construction des ontologies est un processus difficile et fastidieux. Les outils que nous avons présentés dans ce chapitre contribuent à amortir cet effort à travers les facilités et les multiples avantages qui portent les langages de développement d'ontologies dans de nombreux usages. Dans le chapitre suivant en mettre la lumière sur les différentes ressources terminologiques et ontologiques dans le domaine médical.

Chapitre 2

Ressource terminologique et ontologique en médecine

Introduction

Il existe dans le domaine médical un grand nombre de ressources terminologiques et ontologiques (RTO) construites pour répondre à des besoins précis et divers : la CIM et la CCAM sont des classifications utilisées pour le codage médico-économique des dossiers patients à des fins statistiques et budgétaires, le thésaurus MeSH vise à indexer les connaissances médicales pour la recherche d'information dans des bases documentaires, le catalogue CISMef utilise pour partie les termes du MeSH ainsi que d'autres, la SNOMED est une nomenclature dédiée au codage des dossiers électroniques des patients mais avec une granularité plus fine, l'UMLS a pour objectif d'améliorer l'accès à l'information médicale à partir de sources diverses, le FMA est une ontologie de l'anatomie, DOLCE est une ontologie de haut niveau construite pour être reliée à des ontologies de domaine, GALEN est une ontologie médicale à visée généraliste, MENELAS est une ontologie ciblée sur les pathologies coronariennes . . . Ces ressources font toutes partie, bien qu'à des degrés divers, de notre univers de recherche. Nous les décrivons dans les sections suivantes indépendamment les unes des autres.

2.1 Ressources terminologique et ontologique en médecine

2.1.1 CIM

L'appellation complète de la Classification internationale des maladies est «Classification statistique internationale des maladies et des problèmes de santé connexes» (en anglais : International Statistical Classification of Diseases and Related Health Problems). La désignation usuelle abrégée de « Classification internatio-

nale des maladies » est à l'origine du sigle couramment utilisé pour la désigner : « la CIM » (en anglais : ICD).

La CIM permet le codage des maladies, des traumatismes et de l'ensemble des motifs de recours aux services de santé. Elle est publiée par l'Organisation Mondiale de la Santé et est utilisée à travers le monde pour enregistrer les causes de morbidité et de mortalité, à des fins diverses, parmi lesquelles le financement et l'organisation des services de santé ont pris ces dernières années une part croissante. Elle bénéficie d'une remise à niveau régulière, la version la plus récente étant la 10e révision (CIM-10[10], publiée en 1993). Il s'agit d'une classification monoaxiale avec 21 chapitres principaux dont 17 concernent des maladies et 4 concernent les signes et résultats anormaux, les causes de traumatismes, d'empoisonnement ou de morbidité, l'état de santé et les facteurs de recours aux soins. Les catégories de maladies sont définies en fonction d'un caractère commun qui peut être l'étiologie (1 = maladies infectieuses, lettres A et B), la topographie (9 = maladies de l'appareil circulatoire, lettre I), la physiologie (15 = grossesse et accouchement, lettre O) ou la pathologie (II = tumeurs). Les affections (symptômes, maladies, lésions traumatiques, empoisonnements) et les autres motifs de recours aux soins sont répertoriés dans la CIM avec une précision qui dépend de leur importance, c'est-à-dire de leur fréquence et de l'intensité du problème de santé publique qu'ils posent. Par exemple, le chapitre des maladies infectieuses est le plus gros et le plus détaillé parce que ces maladies sont la première cause mondiale de morbidité et de mortalité.

La classification aboutit par subdivisions successives à un code à trois caractères (une lettre correspondant au chapitre puis deux chiffres) pour les maladies définies à un niveau général, décliné par l'ajout d'un quatrième chiffre (derrière un point) pour désigner les diagnostics précis et les formes cliniques ; le sous-code 9 désignant l'absence de précision (SAI = sans autre indication) et le sous-code 8 les autres formes non précédemment définies. Par exemple, le code B25.0 désigne une pneumopathie à cytomégalovirus, le code B25.8 désigne les autres maladies à cytomégalovirus et le code B25.9 désigne une maladie à cytomégalovirus, sans précision. Dans certains cas, un cinquième chiffre a été rajouté afin d'améliorer la finesse de description. Par exemple, le code M66.58 désigne une déchirure spontanée d'un tendon, sans précision, dont le siège n'est pas précisé. La CIM-10 a introduit la notion de troubles iatrogènes (Les troubles iatrogènes sont occasionnés par le traitement médical qu'il y ait ou non erreur dans le traitement prescrit). Elle compte au total 16 390 entrées (date de dernière actualisation, janvier 2006) et comprend trois volumes. La contrainte d'avoir un seul arbre hiérarchique implique

qu'une entité pathologique soit représentée une seule fois dans la classification, ce qui pose des difficultés. Ainsi les tumeurs sont extraites de leur chapitre d'appareil et regroupées dans un chapitre spécial. Parfois cependant, une même maladie peut apparaître en deux places distinctes (avec deux codes). C'est le cas lorsqu'une maladie appartient à un processus pathologique initial général (code associé à une dague), par exemple la tuberculose, et correspond à des manifestations localisées à un appareil (code associé à un astérisque), par exemple une tuberculose rachidienne. De plus, le principe de différenciation n'est pas constant. La classification de l'Organisation Mondiale de la Santé sert, en France, au codage des causes de décès ainsi qu'au regroupement des séjours hospitaliers en groupes homogènes de malades dans le cadre du PMSI. [A. Baneyx, 2007]

2.1.2 CCAM

Avant la Classification Commune des Actes Médicaux [11] (CCAM) existaient en France deux nomenclatures des actes médicaux : la Nomenclature Générale des Actes Professionnels (NGAP) et le Catalogue Des Actes Médicaux (CDAM). La NGAP est la nomenclature de la médecine ambulatoire. Elle permet la tarification des actes de médecine libérale. Il s'agit d'une liste de libellés d'actes assortis de cotations qui fixent les honoraires des professionnels : médecins, dentistes, sages-femmes et auxiliaires médicaux du secteur libéral. Le CDAM permet de décrire l'ensemble des actes réalisés lors de l'hospitalisation d'un patient. Plus récent que la NGAP, le CDAM a vu le jour lors de la mise en place du programme de médicalisation des systèmes d'information, en 1985. Il s'agit d'une nomenclature des actes médicaux comportant pour chaque acte un code, un libellé, un indice de coût relatif et, le cas échéant, la lettre Y indiquant le groupe homogène de malades.

La réglementation contraint les établissements de soins et les professionnels à utiliser simultanément ces deux nomenclatures conçues pour des objectifs différents. Dans ce contexte, un groupe de travail s'est réuni d'avril 1994 à fin 1995 pour étudier la faisabilité d'une classification commune des actes médicaux c'est-à-dire une liste unique de libellés et de codes dont le principe serait étendu, à terme, à l'ensemble des professions de santé. L'élaboration de la CCAM par le Pôle Nomenclature de la CNAMTS et par le Pôle d'Expertise et de Référence National des Nomenclatures de Santé (PERNNS) a été lancée en 1996. Elle remplace maintenant les deux nomenclatures NGAP et CDAM. Elle sert à la fois au programme de médicalisation des systèmes d'information dans tous les établissements publics et privés, et aux praticiens du secteur libéral (médecins et dentistes) pour leurs

honoraires. La CCAM est classée par grands appareils et non par spécialités. Elle comprend 17 chapitres, tels que : le système nerveux central, périphérique et autonome, les oreilles, le système cardiaque et vasculaire, le système immunitaire et hématopoïétique, le système respiratoire, le système digestif . . . Chaque libellé comporte la mention de deux axes obligatoires - l'action et la topographie - et de deux axes facultatifs - la voie d'abord et la technique utilisée, par exemple : Biopsie / du rein / par voie transcutanée / sans guidage. La CCAM a été construite avec l'ontologie GALEN. [A. Baneyx, 2007]

2.1.3 MeSH

Le MeSH est un thésaurus médical qui compte, dans sa version 2005, 22 995 mots clés (ou descripteurs), 83 qualificatifs et environ 57 000 synonymes. Il a été conçu à la National Library of Medicine aux Etats-Unis comme support de l'Index Medicus, répertoire des principales publications scientifiques, et est utilisé par les systèmes de recherche bibliographique Medlars et MEDLINE (Accessible grâce au moteur de recherche Pubmed sur le site). Il est traduit en français par l'INSERM [12] et sert aussi de thésaurus au site CISMef. Le MeSH est organisé en deux parties : une liste alphabétique de termes (lexique) et une structure multiaxiale. Les 200 000 termes du lexique sont distribués selon 15 axes, allant de l'anatomie à la géographie. Les termes équivalents sont rapportés à celui des 20 000 termes principaux (descripteurs) qui exprime le mieux le concept, termes auxquels sont associés un code alphanumérique. Les descripteurs s'organisent selon une structure hiérarchique et associative qui permet, par exemple, de répondre à une requête sur les virus en proposant aussi des documents sur les antiviraux ou les vaccins antiviraux. Le MeSH comprend jusqu'à neuf niveaux de profondeur. Ces principaux composants sont les Headings (MH pour Main Headings par la suite), les Subheading et les Supplementary Concept Records. En outre, des connecteurs permettant des références explicites entre termes expriment les relations de synonymie, de voisinage ou d'association tandis que des qualificatifs permettent de considérer les différentes facettes d'un concept (par exemple : tuberculose/traitement). Les MH respectent un certain nombre de propriétés [Nelson et al, 2001] : d'abord, ils couvrent tout le champ de la médecine qu'on veut bien leur faire couvrir et ne se recoupent pas les uns les autres. Ils forment une partition du domaine.

Les seuls recouvrements acceptés sont ceux de généralisation (broader-than) et de spécialisation (Narrower-than) mis en oeuvre dans leur structure hiérarchique. Ces arbres proposent des hiérarchies selon plusieurs points de vue et partagent les mêmes MH. Il est alors évident que ceux-ci ne peuvent être des concepts : ils

représentent un ou plusieurs concepts et constituent des classes de descripteurs (appelés par la suite, simplement, descripteurs). [A. Baneyx, 2007]

2.1.4 CISMeF

Le site du Catalogue et Index des Sites Médicaux Francophones [13], CISMeF a pour but « d’assister les professionnels de santé dans leur quête d’informations » [Thirion et al, 1999]. Il s’agit d’un catalogue et d’un index spécialisé référençant les sites médicaux francophones répondant à un critère de « qualité de l’information de santé sur l’Internet (NetScoring) ». Ce catalogue est le fruit d’un projet initié en février 1995 par le Centre Hospitalier Universitaire de Rouen.

Il existe d’autres catalogues médicaux, mais nous ne présenterons ici que CISMeF parce qu’il s’agit du catalogue médical francophone le plus fréquemment renvoyé lors de requêtes génériques portant sur le domaine médical, à partir de moteurs de requête eux aussi génériques. [Zweigenbaum et al, 2002]

Les bases documentaires de CISMeF sont indexées manuellement : quand une page web est cataloguée, elle est indexée pour pouvoir être retrouvée et reproposée aux utilisateurs. Le langage RDF est utilisé pour décrire des informations bibliographiques à l’aide des balises de métadonnées définies par le Dublin Core [14] [Darmoni et Thirion, 2000] : créateur de la ressource (`jdc:Creatori`), langue utilisée dans le document (`jdc:Languagei`), sujets abordés (`jdc:Subjecti`), type de lecteur souhaité (`jdc:Audiencei`) ... À chaque document est associé un certain nombre de descripteurs caractérisant son genre (cours, conférence, séminaire, etc.), son url d’accès et les sujets qu’il aborde. Ces derniers sont extraits du thésaurus MeSH, ou de sa version française traduite par l’INSERM. CISMeF utilise le MeSH mais propose des méta-termes. Les méta-termes correspondent à des spécialités biologiques ou médicales concernées par un ou plusieurs mots clés (ou arborescences de mots clés), qualificatifs, ou types de ressources. Par exemple pour le terme cancérologie, on trouve les arborescences, mots clés et qualificatifs du thésaurus MeSH concernant cette spécialité : antinéoplasiques (arb), marqueur biologique tumeur (arb), oncologie médicale (arb), secondaire (qualificatif), service oncologie hôpital (MeSH), service oncologie hôpital (type de ressource), tumeurs (arb). Le catalogue CISMeF se consulte par l’intermédiaire de l’interface [15] Doc’CISMeF. [A. Baneyx, 2007]

2.1.5 SNOMED

La SNOMED (Systematized Nomenclature of Medicine [16]) combine une nomenclature de plus de 50 000 termes et une classification multiaxiale et multi-domaines comportant à l'origine 7 axes : topographie, morphologie, étiologie, altération fonctionnelle, nosologie, actes médicaux (College of American Pathologists, 1993). La 3ème édition compte désormais 200 000 termes et 11 axes définis par une lettre (par exemple, T pour topographie, E pour étiologie). À l'intérieur de chaque axe, les éléments sont organisés suivant une structure hiérarchique. La classification d'un terme repose sur une décomposition de celui-ci en combinaison de termes appartenant à différents axes. Ainsi un diagnostic est traduit par plus d'un élément signifiant, mais chaque axe ne doit pas être obligatoirement validé. Par exemple, la juxtaposition : T2856 (lobe supérieur du poumon gauche) / M4100 (inflammation) / F0300 (fièvre) / E2012 (pneumocoque) correspond à la phrase « Pneumonie fébrile à pneumocoque du lobe supérieur gauche ». L'ajout de connecteurs concernant notamment les liens de causalité permet de décrire un fait complexe en plusieurs phrases. La SNOMED est une des classifications médicales les plus complètes mais un même concept peut y être décrit de différentes façons et rien n'empêche de créer par combinaison des concepts inconsistants [le Bozec, 2001]. Ce modèle pose encore des problèmes, par exemple : les termes des différents axes ne sont pas complètement indépendants entre eux, l'axe Maladie fait souvent double emploi, certains concepts peuvent apparaître dans plusieurs axes.

En raison de ce type de défaut, la SNOMED a évolué en SNOMED-RT (pour Root Procédure) puis en SNOMED-CT (pour Clinical terms), fusion de la SNOMED-RT et d'une terminologie britannique Clinical Terms de la NHS (services de santé britanniques). Dans cette nouvelle configuration, la SNOMED-RT respecte un certain nombre de principes : structure hiérarchique de concepts, définitions de types ou de rôles pour des concepts, consistance, exploitation dans le cadre d'une logique de description [Dolin et al, 2001] qui en font une ontologie formelle. La principale difficulté qui semble apparaître dans la transformation de la SNOMED en une ontologie est le choix qui a été fait de conserver les termes de la classification comme concepts et labels des concepts de l'ontologie [Spackman et al., 2002] : la volonté de transformer la SNOMED en une réelle ontologie formelle se heurte à la « nécessité » que s'imposent les auteurs, de conserver la classification presque telle quelle avec les souplesses et les imprécisions d'un paradigme de construction pour partie linguistique. En l'état, cette transformation n'est pas totalement assumée, en particulier au niveau de la normalisation.

La SNOMED-CT se veut une terminologie des soins de santé cliniques, dynamique et valide scientifiquement. Son objectif est de rendre les connaissances de soins de santé plus accessibles pour l'ensemble des spécialités médicales. La terminologie Core SNOMED-CT contient plus de 361 800 concepts de soins et comporte également 975 000 descriptions et près de 1,47 million de relations sémantiques. Pour des informations plus complètes sur le MeSH, la SNOMED et CISMef dans un cadre de recherche d'informations. [A. Baneyx, 2007]

2.1.6 UMLS

UMLS a été mis en place dans le but d'améliorer l'accès à l'information médicale à partir de sources diverses : bases de données bibliographiques, bases de données d'enregistrements cliniques et bases de connaissances médicales [Lindberg & Humphreys, 1990]. Un des moyens d'UMLS est alors de définir un vocabulaire médical de base, un « métathésaurus » qui reprend et dédouble les termes de l'ensemble des 95 ressources terminologiques qu'il inclut (MeSH, SNOMED...) [17]. Ce métathésaurus propose une description hiérarchique des connaissances médicales utilisées dans divers documents et systèmes à base de connaissances. De plus, un réseau sémantique de 134 types sémantiques environ permet de typer tous les termes du métathésaurus. L'intérêt d'UMLS réside dans sa grande couverture du domaine médical (1 276 301 concepts dans la version A du 1er trimestre 2006) et dans sa disponibilité. Le métathésaurus n'est pas une ontologie : il n'a pas été fait dans ce but et une tentative de réutilisation comme une ontologie (l'utilisation du métathésaurus pour construire l'ontologie de MENELAS) s'est soldée par un échec [Charlet et al, 1996]. Le réseau sémantique a une structure beaucoup plus proche d'une ontologie mais n'en a pas la précision : il recense 134 types plus ou moins généraux en médecine qui servent à typer chaque terme du métathésaurus. [A. Baneyx, 2007]

2.1.7 FMA

Le Foundational Model of Anatomy [21] (FMA) est une ontologie de référence pour le domaine de l'anatomie. C'est une représentation de toutes les entités anatomiques et les relations nécessaires pour la modélisation symbolique de la structure phénotypique du corps humain dans une forme qui soit compréhensible par l'homme et qui soit également traitable par une machine [Rosse & Mejino, 2003]. Cette ontologie permet de décrire les localisations des tumeurs ou des métastases, ce qui permet par exemple de déduire qu'une tumeur principale située dans le

lobe supérieur du poumon gauche est également située dans le poumon gauche (car le lobe supérieur est une partie du poumon). Le FMA est mis à la disposition d'utilisateurs qui peuvent récupérer des parties de la modélisation pour les intégrer dans leur propre ontologie. Cette ontologie suscite de réels espoirs dans les communautés de l'Ingénierie des connaissances et de l'Informatique médicale. [A. Baneyx, 2007]

2.1.8 DOLCE

L'ontologie DOLCE [22] (Descriptive Ontology for Linguistic and Cognitive Engineering), élaborée par l'équipe de Nicola Guarino (LOA, Trento, Italie), constitue un des résultats du projet européen WonderWeb Foundation Ontologies Library. DOLCE est une ontologie de haut niveau (niveau supérieure) dont la vocation est d'être utilisée pour concevoir des ontologies de domaine. Sa structure repose sur la distinction philosophique entre entités perdurantes et endurantes (Pour plus d'informations sur la distinction entre les notions d'entités perdurantes et endurantes dans les topontologies, le lecteur peut se reporter à l'article de I. Johansson 2005). [A. Baneyx, 2007]

2.1.9 GALEN

Le projet GALEN, développé à l'université de Manchester, vise à mettre en place un serveur de terminologie en médecine. Développé depuis 1992 au sein de projets européens successifs, il est centré sur un Common référence model, une ontologie de la médecine. Cette ontologie respecte une structure arborescente au niveau de ses types primitifs et est le coeur du système et des services qu'il propose [Rector, 1998]. GALEN utilise un formalisme appelé GRAIL (Galen Representation and Integration Language) qui permet de saisir la connaissance terminologique dans le domaine médical [Rector et al, 1992]. Ce formalisme est hautement génératif et permet de définir des concepts complexes (ou définis), composés de concepts plus élémentaires (ou primitifs). Tous les concepts, et les relations qui les lient, sont représentés indépendamment du langage dans lequel ils sont exprimés.

Nous n'allons pas décrire le système par le menu et nous allons plutôt nous intéresser au Common référence model et à sa mise en oeuvre. L'ontologie de la médecine ne couvre que les domaines dans lesquels le projet s'est développé, où des opportunités se sont créées. Ainsi, le département de santé publique et

d'informatique médicale de Saint-Étienne participe au développement de la classification commune des actes médicaux (CCAM) et justifie l'utilisation d'un tel outil pour le développement cohérent d'une terminologie médicale [**Rodrigues et al, 1998,1999**]. Les promoteurs du projet sont confrontés au problème de l'évolution rapide de la médecine et donc des ontologies attenantes, au point qu'il est difficile d'avoir une vue complète et cohérente d'une ontologie d'un domaine précis. Confronté à la difficulté de compréhension des ontologies, le projet GALEN y a répondu, même si ça n'était pas au début le but d'un tel module, par l'utilisation d'un générateur de langage naturel permettant de valider les représentations proposées avec les praticiens. Enfin, les promoteurs du projet, s'ils construisent des ontologies, ne proposent pas réellement de méthode argumentée et constructive. [**A. Baneyx, 2007**]

2.1.10 MENELAS

MENELAS [23] est un projet européen piloté de 1992 à 1995 par le DIAM/SIM/DSI/AP-HP. Le but du projet MENELAS était la conception et l'implémentation d'un système pilote capable d'accéder à des rapports médicaux rédigés en langage naturel dans trois langues : l'anglais, le français et le néerlandais. Ce système devait pouvoir analyser le contenu de rapports médicaux (comptes rendus d'hospitalisation ou CRH) et l'archiver dans une base de données sous la forme d'un ensemble de structures conceptuelles (graphes conceptuels de [**J. Sowa, 1984**]). Ces structures, qui constituent la représentation de chaque CRH, Une partie des informations était encodée à l'aide de nomenclatures internationales, ce qui permettait leur échange à partir de CRH écrits en différentes langues. Le projet a été confronté aux problèmes habituels de la compréhension de textes en langage naturel. Il s'agit bien sûr des problèmes inhérents au langage, comme la paraphrase, l'ambiguïté, la métonymie et de façon plus générale la description et la mise en oeuvre de connaissances syntaxiques et sémantiques adéquates et d'une couverture suffisante. Il s'agit aussi des problèmes généraux de représentation du sens des énoncés et des connaissances à fournir au système. On rejoint alors des problématiques classiques en Intelligence artificielle : acquisition, représentation, mise en ?uvre, validation de connaissances complexes. La question de la construction de l'ontologie²⁹ de MENELAS a été abordée de façon approfondie dans [**Bouaud et al, 1994 ; Charlet et al, 1996**]. Elle a amené la mise au point de la méthodologie ARCHONTE, principalement par B. Bachimont, et l'implémentation d'un système opérationnel, principalement par [**J. Bouaud, 1992**], se servant de cette ontologie. L'ontologie construite comporte plus de 1 800 types et 300 relations. [**A. Baneyx, 2007**]

2.2 Conclusion

Dans ce chapitre nous avons essayé de présenter de nombreuses ressources terminologiques utilisées dans le domaine médical en citant comme exemple DOLCE, GALAN, FMA qui sont des ontologies et SNOMED, MENALS, MeSH, CIM. Donnant à chacune de ces ressources une définition bien précise et leurs propriétés. Au cours du chapitre suivant, nous nous intéressons aux étapes de construction de notre ontologie de domaine médicale.

Chapitre 3

Conception de l'ontologie

Introduction

Ce chapitre présente notre contribution à la problématique posée dans ce mémoire, à savoir la construction d'une ontologie de domaine médicale. Cette ontologie est appelée une ontologie de domaine, d'abord en suivant un processus de construction d'une ontologie de domaine partant de connaissances brutes et arrivant à une ontologie opérationnelle. Par la suite, nous décrirons comment nous avons utilisé l'outil PROTÉGÉ pour éditer notre ontologie en langage OWL. Enfin, nous illustrerons la phase de vérification par le biais du système d'inférence et quelques requêtes avec SPARQL.

3.1 Processus de construction d'une ontologie de domaine

Nous avons suivi un processus de construction d'une ontologie de domaine partant de connaissances brutes et arrivant à une ontologie de domaine opérationnelle représentée par langage *OWL*. Ce processus est inspiré de la méthodologie *METHODOLOGIE* qu'on a bien détaillé dans le premier chapitre, et il est composé de cinq étapes :

1. Évaluation des besoins.
2. Conceptualisation.
3. Formalisation.
4. Implémentation.
5. Vérification et Évaluation.

3.2 Construction d'une ontologie de domaine

Pour réaliser cette ontologie de domaine nous suivons le processus décrit précédemment :

3.2.1 Évaluation des besoins

Pour réaliser cette ontologie de domaine nous suivons le processus décrit précédemment :

- **Le domaine de connaissance** : Le domaine medical.
- **L'objectif** : Partager de façon collaborative les connaissances médicales et faciliter la recherche et l'intégration d'informations provenant des multiples sources d'information médicales.
- **Les utilisateurs** : Les utilisateurs de cette ontologie sont : les *médecins*
- **Les sources d'informations** : L'encyclopédie Médicale, interviews avec les médecins, sites web médicaux (doctissimou, eureka santé, experts du domaine ...) Voir annexe.
- **La portée de l'ontologie** : Médecin, Médicament, maladie, diagnostique

3.2.2 Conceptualisation

Dans cette étape on distingue les principales tâches suivantes :

Construire le glossaire de termes

Construire un glossaire de termes est la première tâche à effectuer dans l'étape de conceptualisation. Il recueille et décrit tous les termes (concepts, instances, attributs, relations entre les concepts, etc.) qui sont utiles et potentiellement utilisables dans le domaine que nous allons représenter leurs descriptions détaillées et non ambiguës en langage naturel.

Le tableau suivant présente la liste de quelques termes utiliser dans notre ontologie :

Terme	Signification
Maladie	La maladie est une altération des fonctions ou de la santé d'un organisme. se traduisant par des symptômes et des signes. Elle concerne un ou plusieurs patients.
Médecin Spécialiste	Il est un professionnel de santé. Il peut exercer à l'hôpital ou avoir une activité libérale.
Médicament	Un médicament est une substance présentée comme possédant des propriétés curatives, préventives ou administrée en vue d'établir un diagnostic. Un médicament est le plus souvent destiné à guérir, à favoriser la guérison, à soulager ou à prévenir des maladies humaines ou animales.
Donne comme effet	Relation entre médicaments et effet secondaire.
Symptôme	Manifestation d'une maladie observée par le médecin lors de l'examen de son patient.
Diagnostic	le raisonnement menant à l'identification de la cause (l'origine) d'une défaillance, d'un problème ou d'une maladie, ou tout simplement à la détermination d'une espèce biologique par rapport à une autre (taxinomie), à partir des caractères ou symptômes relevés par des observations, des contrôles ou des tests.
Maladies endocriniennes	Les maladies endocriniennes sont causées par un dysfonctionnement des hormones sécrétées par les glandes endocrines ou due à un trouble métabolique
Maladie dermatologique	La dermatologie est la branche de la médecine qui s'occupe de la peau, des muqueuses et des phanères (ongles, cheveux, poils)

TABLE 3.1 – Glossaire de quelques termes

Construction de diagramme des relations binaires et des attributs

Une relation binaire permet de relier deux concepts entre eux (un concept source et un concept cible). « Si R est une relation entre deux concepts $C1$ et $C2$ alors pour tout couple d'instances des concepts $C1$ et $C2$, il existe une relation de type R qui lie deux instances de $C1$ et $C2$ ». Cette tâche permet de représenter d'une manière graphique les différentes relations qui existent entre les divers concepts que se soit de même ou de différentes hiérarchies.

Construction d'un dictionnaire de concepts

Une fois la taxonomie de concepts et le diagramme de relations binaires sont effectués, il faut donc spécifier les propriétés qui décrivent chaque concept de la hiérarchie dans un dictionnaire de concepts. Un dictionnaire de concepts contient tous les concepts du domaine, leurs synonymes, leurs acronymes, leurs attributs et leurs relations. Les relations spécifiées pour chaque concept sont celle où leur domaine est le concept lui-même.

Ce tableau représente le dictionnaire de quelques concepts utilisés dans notre ontologie.

Nom du concept	Attributs	Relations	Acronyme	synonymes
Maladie	/	Est_diagnostique_par Est_traitée_par Présente	Mal	Souffrance
Médecin Spécialiste	Numéro Agrément Nom Prénom Age Téléphone Sexe	Traite	DR	Docteur
Médicement	/	Donne_comme_Effet	Medoc	Traitement

TABLE 3.2 – dictionnaire de quelques concepts

Déterminer la liste des relations binaires

Le but de cette tâche consiste à construire une table de relations binaires décrites en détail. Pour chaque relation utilisé dans le diagramme des relations binaires, nous définissons le nom de la relation, le nom des concepts sources et cibles, le nom de la relation inverse et les cardinalités source et cible. Le tableau suivant présente quelques relations binaires utilisées dans notre ontologie.

Nom de relation	Concept source	Concept cible	Relation inverse	Cardinalité
Traite	Medecin Spécialiste	Maladies	/	1..n
Présente	Maladies	Symptômes	/	1..n
Est diagnostiquer par	Maladies	Diagnostique	/	1..n
Est traitée par	Maladies	Traitement	/	1..n
Donne comme effet	Médicaments	Effet secondaire	/	1..n

TABLE 3.3 – Table de quelques relations binaires

Construction de la table d'attributs

La table des attributs comporte une description détaillée des attributs inclus dans le dictionnaire de concepts, et l'ensemble de contraintes et de restrictions sur ces valeurs. Le tableau suivant présente quelques attributs utilisés dans notre ontologie.

Nom attribut	Type	Card (min/max)	Valeur par défaut	Domaine de valeur
Nom	String	1..1	-	-
Numéro d'agrément	String	1..1	-	-
Age	Integer	1..1	-	-
Type médicament	String	1..6	-	Comprimé, gélule, sirop, crème, injectable

TABLE 3.4 – Table des attributs

Construction de la table des axiomes

Dans cette étape, nous définissons les concepts au moyen d'expressions logiques possible. Pour chaque axiome, il faut spécifier la description de l'axiome en langage naturel, l'expression logique qui décrit formellement l'axiome en logique du premier ordre, les concepts, les relations et les variables utilisées. Le tableau suivant présente quelques axiomes utilisés dans notre ontologie.

Nom de concept	Description de l'axiome en langage naturel	expression logique
Ophthalmologue	Un Ophthalmologue traite les maladies de l'œil	$\forall x \text{ Ophthalmologue}(x) \wedge \exists y \text{ maladie de l'œil}(y) \wedge \text{Traite}(x,y)$
Medecin spécialiste	Chaque médecin spécialiste est soit un pneumologue, cardiologue, un pédiatre	$\forall x \text{ Medecin spécialiste}(x) \Rightarrow \text{Pneumologue}(x) \vee \text{Cardiologue}(x) \vee \text{Pédiatre}(x)$
Maladies	Chaque maladie est détectée par un ou plusieurs diagnostics	$\forall x \text{ maladie}(x) \wedge \exists y \text{ diagnostique}(y) \wedge \text{détecter par}(x, y)$
Traitement	Un Traitement est soit un Médicament, Thérapeutique, Chirurgie	$\forall x \text{ Traitement}(x) \Rightarrow \text{Médicament}(x) \vee \text{Thérapeutique}(x) \vee \text{Chirurgie}(x)$

TABLE 3.5 – Table des axiomes

Construction de la table des instances

La table des instances décrit les instances connues; qui sont déjà identifiées dans le dictionnaire de concepts. Pour chaque instance, il faut spécifier le nom de l'instance, le nom du concept où elle appartienne, ses attributs et les valeurs qui lui y sont associés. Le tableau ci-après illustre quelques instances créées.

Concept	Instance
Pneumologue	Dr Hameg
Diagnostique	Electrocardiogramme
Médicament	Antalgique
Symptôme	Constipation
Effet Secondaire	Fièvre
Antidiabétiques	Insulines
Maladies Endocrinienne	Goitre

TABLE 3.6 – Table des instances

3.2.3 Formalisation

Comme cité auparavant, dans cette étape, nous utilisons le formalisme de la logique de description pour formaliser le modèle conceptuel que nous avons obtenu dans l'étape de conceptualisation. Le résultat est une base de connaissances en logique de description composé de deux parties T-BOX et A-BOX.

Le niveau terminologique ou TBox

Voici quelques définitions et axiomes terminologiques représentées dans le tableau suivant :

Construction d'ABox

Voici quelques assertions sur les individus représentées dans le tableau suivant :

Axiome terminologique
Medecin Spécialiste \subseteq Thing
Effet Secondaire \subseteq Thing
Pneumologue \subseteq Medecin Specialiste
Traitement \subseteq Thing
Medicament \subseteq Traitement
Antibiotique \subseteq Medicament \subseteq Traitement
Diagnostique \subseteq Thing \wedge (\existsdéetecte. Maladie)
Endocrinologue \subseteq Medecin Specialiste \wedge (\exists NumAgrément.String) \wedge (\existsnom.String) \wedge (\existsprenom.String) \wedge (\exists Telephone.String) \wedge (\existstraite. Maladies Endocriniennes)
Maladies œil \subseteq maladie \wedge (\existsdiagnostiquer par .Ophtalmologue)
(Pneumologue \wedge Cardiologue \wedge Viscéralgie \wedge pédiatre) \subseteq Medecin Specialiste

TABLE 3.7 – Axiomes terminologiques (TBox)

Assertion sur les individus
<ul style="list-style-type: none"> • Pneumologue (Dr_ Hameg) • Diagnostiquer par (Maladies de l'œil, Ophtalmologue) • soigné par (Insuline) • Viséraliste (Dr_Hamrioui) • Cardiologue (Dr_Hemdani) • Médicament (Antalgique (panadole)) • Médicament (Antidiabétiques (Oraux(Amarel))) • Thérapeutique (chimiothérapie) • Diagnostique (Biopsie) • Traitement (Chirurgie) • Douleur (Arthralgie) • Maladies (Maladies Dermatologique (Acné)) • Symptôme (fatigue)

TABLE 3.8 – Assertions sur les individus (ABox)

Chapitre 4

Implémentation et Réalisation

L'ontologie que nous avons obtenue jusqu'à cette étape est une ontologie formelle, maintenant on doit implémenter cette ontologie pour qu'elle soit manipulable dans un système informatique. Pour cela on doit spécifier un langage de représentation de connaissance doté de capacités d'inférences.

4.1 Étude de protégé

Protégé est un éditeur d'ontologies hautement extensible, capable de manipuler des formats très divers, c'est aussi une librairie JAVA qui peut être étendue afin de créer des applications à base de connaissance en utilisant un moteur d'inférence pour raisonner et déduire de nouveaux faits. Incluant des plugins pour les langages RDF, DAML+OIL et OWL pour la manipulation d'ontologies dans différents formats, il aide ou permet de construire des ontologies pour le web sémantique. Protégé permet la création et l'édition des ontologies grâce à ces deux outils distincts :

- **Protégé-Frame** : Permet de créer facilement une interface graphique afin de bien gérer une ontologie, les formulaires se génèrent automatiquement en se basant sur le schéma d'ontologie créé. Il offre également la possibilité de personnaliser l'interface selon les besoins de l'utilisateur.
- **Protégé-Owl** : C'est une extension de protégé qui supporte le langage OWL. Il permet de créer des classes, propriétés, instances grâce aux nombreuses propriétés offertes par OWL. Il est aussi optimale d'interroger un raisonneur afin de contrôler l'intégrité du modèle et de créer un modèle d'inférences.

4.2 Les moteurs d'inférence

La sémantique formelle du langage OWL permet l'application des méthodes de raisonnement afin d'effectuer des dérivations logiques. Ces dérivations sont effectuées par des moteurs d'inférences ou bien des raisonneurs sémantiques, qui sont des programmes qui permettent de lire des ontologies à partir de fichiers OWL ou bien de serveurs distants. Autrement dit ce sont des systèmes capables de gérer et d'utiliser la sémantique du langage d'ontologies.

4.2.1 Racer

Racer [R.Moller & M.Wessel, 2001] est l'un des moteurs les plus connus et le plus utilisé grâce à ces performances et sa stabilité. Racer travaille sur les ontologies décrites par son langage mais celles aussi décrites par RDF ou OWL. Ce raisonneur possède également un langage de requête nRQL (new Racepro query language) pour interroger les ontologies sur la ABOX et la TBOX. Racer se présente sous la forme d'un serveur qui peut être accédé par le protocole TCP ou http. Voici un tableau résumant les avantages et inconvénients du moteur d'inférence Racer :

Avantages	Inconvénients
<ul style="list-style-type: none"> -Documentation disponible provenant des utilisateurs et concepteurs. -Racer permet l'ajout d'assertions et d'individus dans la ABOX après le chargement de l'ontologie. -Racer permet l'utilisation des règles SWRL. 	<ul style="list-style-type: none"> -Il ne permet pas l'utilisation de type de données utilisateur, car il possède ces propres types de données et il effectue une conversion avec les types de base. -Pas de valeurs multiples pour un datatype property, car il suppose que toutes les propriétés sur les datatypes sont fonctionnelles. -Il n'existe pas de version libre d'utilisation, produit commercial.

TABLE 4.1 – Avantages et inconvénients du moteur d'inférence Racer

4.2.2 Pellet

Le moteur Pellet [E.Sirin & B.Parsia, 2005] est l'un des projets d'un groupe de recherche sur le web sémantique de l'université du Maryland, disponible en Open-Source et offre des évolutions fréquentes. Pellet travaille sur des ontologies décrites en RDF ou OWL et permet les requêtes avec RDQL et SPARQL sur la ABOX et la TBOX. Voici un tableau résumant les avantages et inconvénients

du moteur d'inférence Pellet : Le moteur d'inférence Pellet utilise deux langages

Avantages	Inconvénients
-Pellet permet l'utilisation des types donnés utilisateur.	-Il possède une documentation pauvre en comparaison de celle de Racer.
-Pellet est Open-Source et développé en Java.	-Il ne permet pas l'utilisation de règles SWRL.
-Pellet est un raisonneur OWL DL complet.	-Il n'offre pas de système de souscription à un concept.

TABLE 4.2 – Avantages et inconvénients du moteur d'inférence Pellet

d'interrogation de requêtes qui sont RDQL et SPARQL, quant à Racer il utilise son propre langage de requêtes qui est nRQL. Dans ce qui suit nous allons aborder ces trois langages d'interrogation d'ontologies.

4.3 Les Langage d'interrogation

L'utilisation des ontologies permet de définir des types d'interrogation qui s'appuie sur les langages du web sémantique, ou plusieurs moteurs d'inférence sont intégrés au système pour interroger la base de connaissances. Racer et Pellet sont des exemples de moteurs d'inférence reposant sur la logique de description ; Plusieurs langages d'interrogation ont été définis afin d'interroger ces moteurs. Ces derniers fournissent des mécanismes permettant d'exprimer des requêtes complexes. En effet il existe trois langages d'interrogation RDQL, SPARQL, nRQL qui se basent sur la reconnaissance des graphes RDF. Pellet utilise RDQL et SPARQL comme langage d'interrogation, et Racer utilise nRQL

4.3.1 nRQL (new Racepro Query Language)

nRQL [V.Haarslev & M.Wessel, 04] est un langage d'interrogation de Racer, comme RDQL et SPARQL, nRQL est basé sur la recherche de graphes RDF, sa syntaxe est proche des deux autres, sauf sa notation préfixée des opérateurs. Nous allons présenter quelques exemples de requêtes nRQL :

- Pour savoir si un concept contient des individus ou non :

(Retrieve()(?x|Concept|))

Si le concept contient des individus alors le résultat retourné est T (True) sinon le résultat retourne NIL.

- Pour avoir toutes les instances d'un concept :

$$(\text{Retrieve}(?x)(?x|\text{Concept}|))$$

- Vérifier l'appartenance d'un individu à un concept :

$$(\text{Retrieve}(|\text{individu}||\text{Concept}|))$$

Si l'individu appartient au concept, le résultat retourné est T (True) sinon NIL.

4.3.2 RDQL (RDF Data Query Language)

RDQL [07] est un langage d'interrogation RDF, qui possède une syntaxe proche de SQL (Structured Query Language) utilisant les classes **SELECT**, **WHERE**, **USING**. RDQL est destiné à être utilisé dans des scripts et pour les expériences au niveau des langages de modélisation.

Exemple d'une requête RDQL :

$$SELECT * WHERE(?x?y?z)$$

.

4.3.3 SPARQL

SPARQL est un langage de requête qui peut être utilisé pour exprimer des interrogations à travers diverses sources de données, les résultats d'interrogation peuvent être des ensembles de résultats ou bien des graphes RDF. Car il est capable de rechercher des motifs de graphes ainsi que leur conjonction et leur disjonction. Une requête SPARQL est présentée sous la forme de sujet-prédicat-objet comme suit

$$SELECT ?sujet ?propriété ?valeur WHERE ?Sujet ?propriété ?valeur$$

4.4 Les logiques de description

4.5 choix de langage

Notre choix a été orienté vers OWL qui est le langage standard de représentation et de spécification de l'ontologie, comparé à RDFS qui est insuffisant pour codifier l'ontologie en termes de fonctionnalités sémantiques. Ces raisons suivantes illustrent notre choix de OWL DL :

- OWL DL permet une expressivité importante que OWL lite.
- OWL DL permet d'exprimer des cardinalités multiple.
- OWL DL offre un niveau d'expressivité suffisant tout en maintenant la complétude de calculs (toutes les inférences sont calculables) et la décidabilité (leurs calculs se fait en une durée finie).

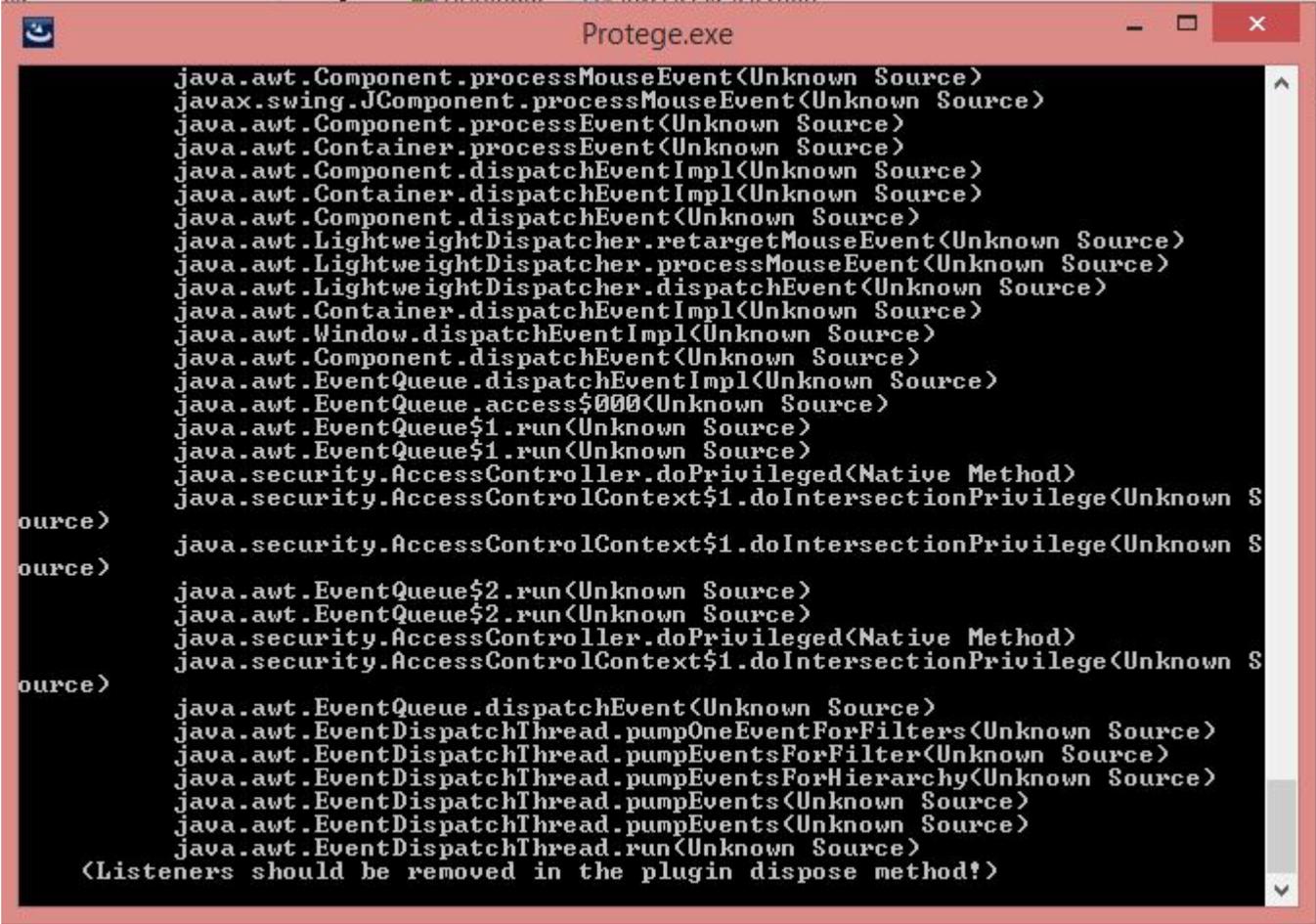
4.6 Choix de l'outil

Pour l'édition de notre ontologie nous avons choisi l'outil Protégé parce qu'il n'impose pas de méthodologie, c'est aussi un éditeur hautement extensible, capable de manipuler des formats très divers. Le support d'OWL, comme de nombreux autres formats, est possible dans Protégé grâce à un plugin dédié. Il a la particularité de maintenir un espace de nommage unique pour tous ces cadres. Mais il faut savoir que c'est un outil qui est sensible à la casse, ainsi *personne* et *Personne* sont deux classes différentes.

4.7 Etapes de construction de l'ontologie

A partir de la nous présentons les étapes essentielles de la construction de notre ontologie sous Protégé.

4.7.1 Lancement de Protégé



```

Protege.exe
java.awt.Component.processMouseEvent(Unknown Source)
javax.swing.JComponent.processMouseEvent(Unknown Source)
java.awt.Component.processEvent(Unknown Source)
java.awt.Container.dispatchEvent(Unknown Source)
java.awt.Container.dispatchEventImpl(Unknown Source)
java.awt.Component.dispatchEvent(Unknown Source)
java.awt.LightweightDispatcher.retargetMouseEvent(Unknown Source)
java.awt.LightweightDispatcher.processMouseEvent(Unknown Source)
java.awt.LightweightDispatcher.dispatchEvent(Unknown Source)
java.awt.Container.dispatchEventImpl(Unknown Source)
java.awt.Window.dispatchEventImpl(Unknown Source)
java.awt.Component.dispatchEvent(Unknown Source)
java.awt.EventQueue.dispatchEventImpl(Unknown Source)
java.awt.EventQueue.access$000(Unknown Source)
java.awt.EventQueue$1.run(Unknown Source)
java.awt.EventQueue$1.run(Unknown Source)
java.security.AccessController.doPrivileged(Native Method)
java.security.AccessControlContext$1.doIntersectionPrivilege(Unknown S
source)
java.security.AccessControlContext$1.doIntersectionPrivilege(Unknown S
source)
java.awt.EventQueue$2.run(Unknown Source)
java.awt.EventQueue$2.run(Unknown Source)
java.security.AccessController.doPrivileged(Native Method)
java.security.AccessControlContext$1.doIntersectionPrivilege(Unknown S
source)
java.awt.EventQueue.dispatchEvent(Unknown Source)
java.awt.EventDispatchThread.pumpOneEventForFilters(Unknown Source)
java.awt.EventDispatchThread.pumpEventsForFilter(Unknown Source)
java.awt.EventDispatchThread.pumpEventsForHierarchy(Unknown Source)
java.awt.EventDispatchThread.pumpEvents(Unknown Source)
java.awt.EventDispatchThread.pumpEvents(Unknown Source)
java.awt.EventDispatchThread.run(Unknown Source)
(Listeners should be removed in the plugin dispose method?)
  
```

FIGURE 4.1 – Lancement de protégé

4.7.2 Création d'un nouveau projet

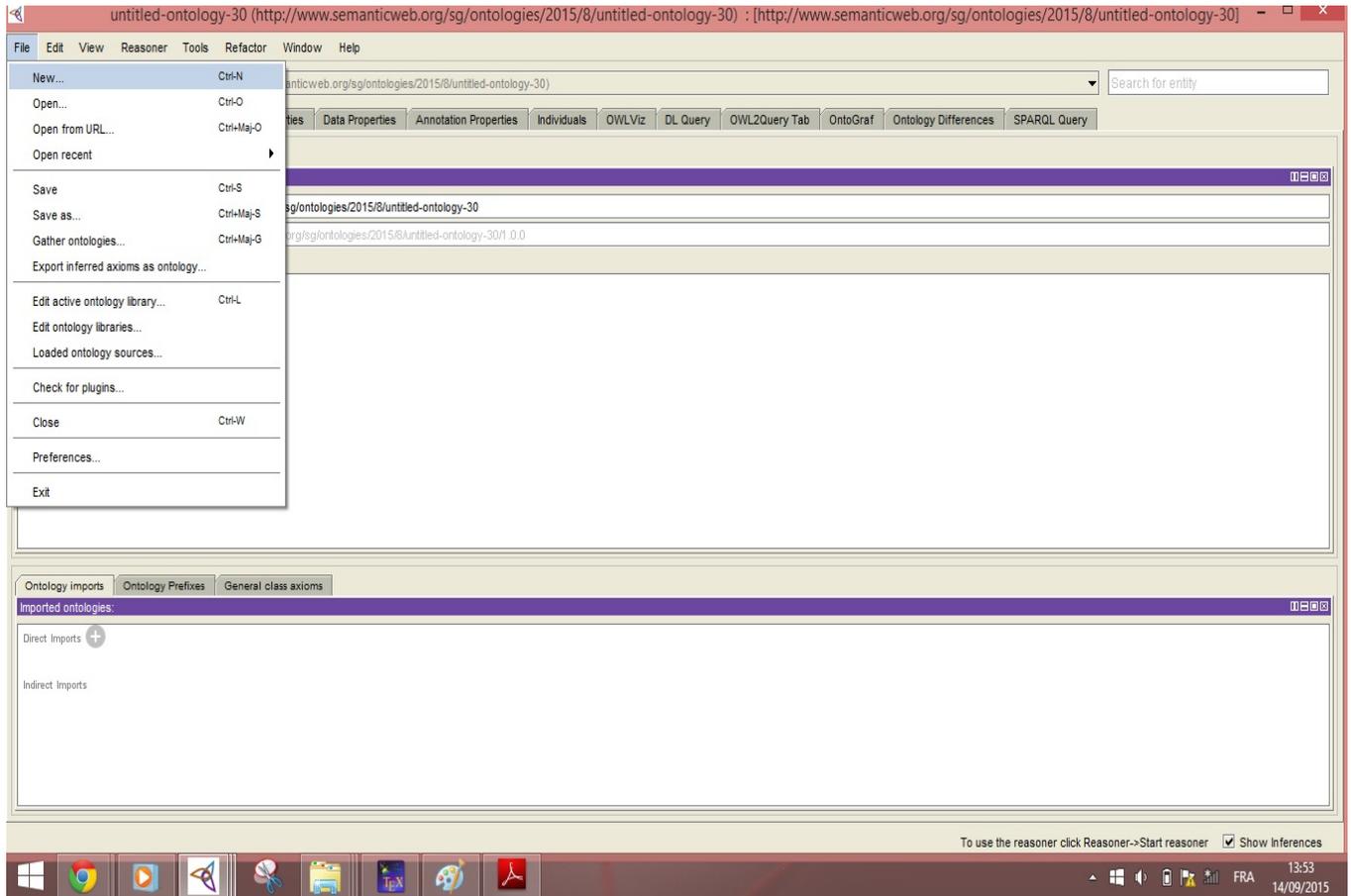


FIGURE 4.2 – Création d'un nouveau projet

4.7.3 Création des classes

Création de la classe Maladie

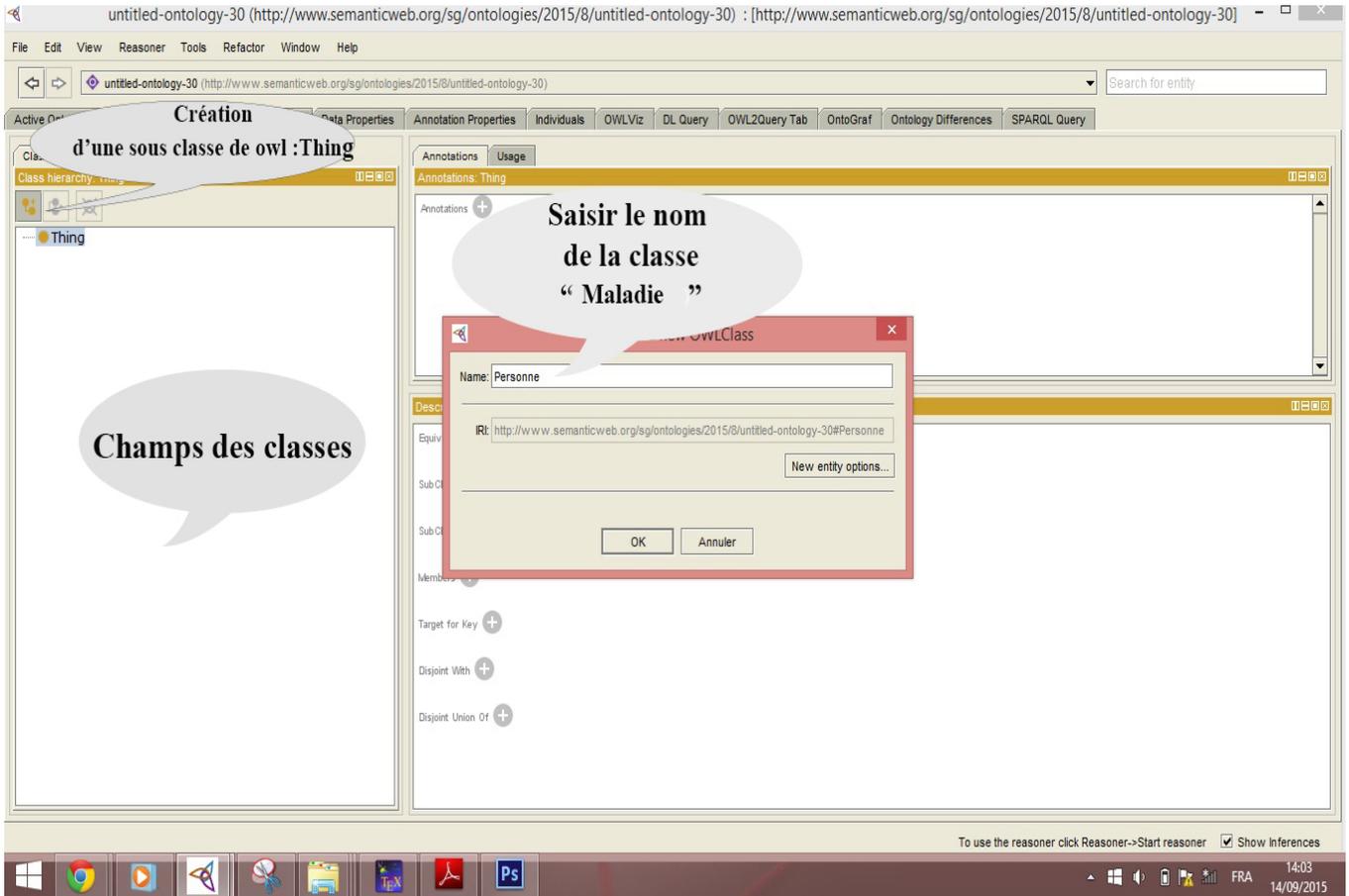


FIGURE 4.3 – Création d'une classe

4.7.4 Création des classes disjointes

Les classes ayant le même parent sont dites disjointes, car un individu ne peut pas être une instance de plus d'une même classe.

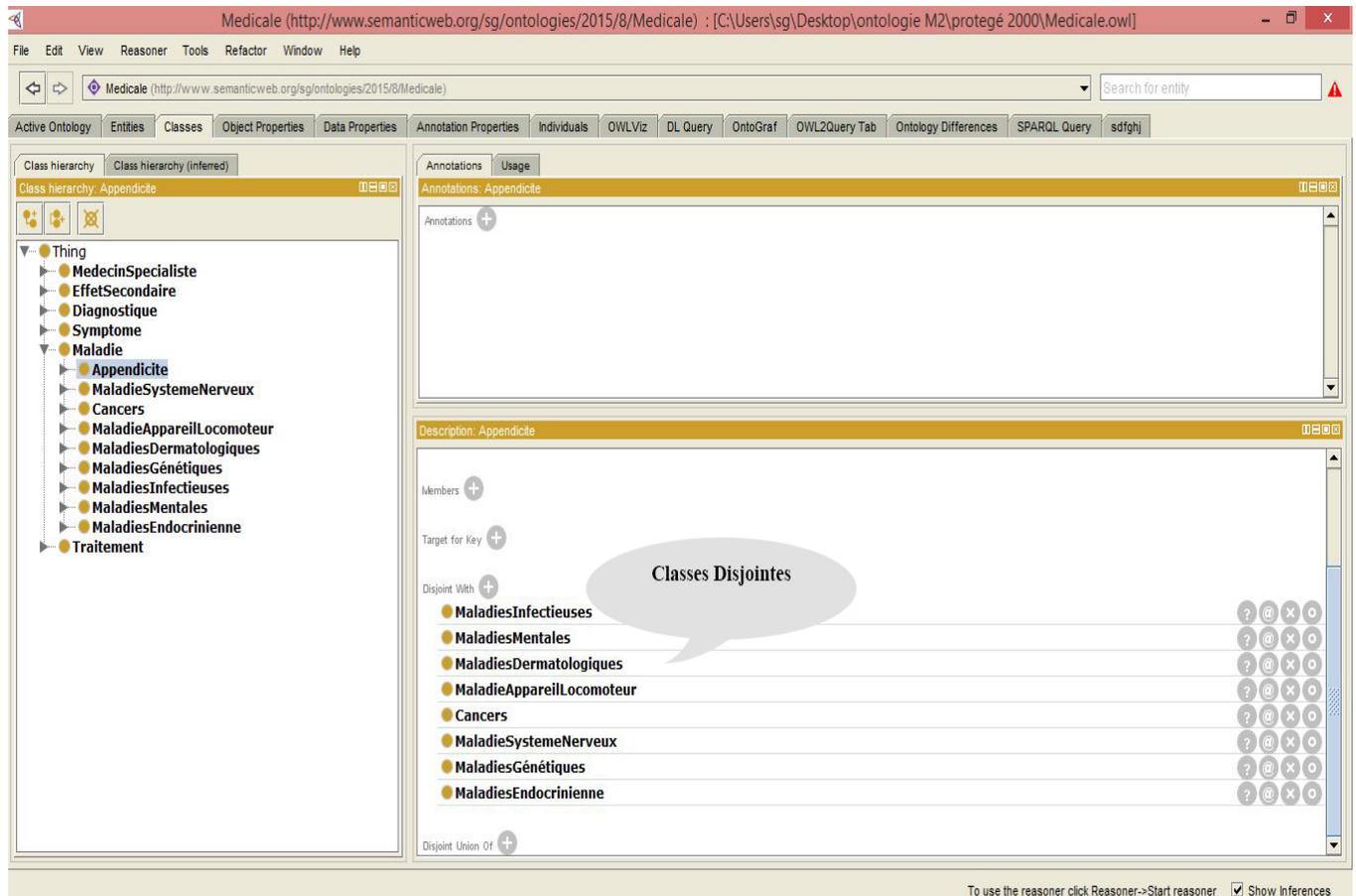


FIGURE 4.4 – Classes disjointes

4.7.5 Création des relations

On crée la relation est "Traite" qui a pour domaine la classe Médecin spécialiste et pour image la Maladie.

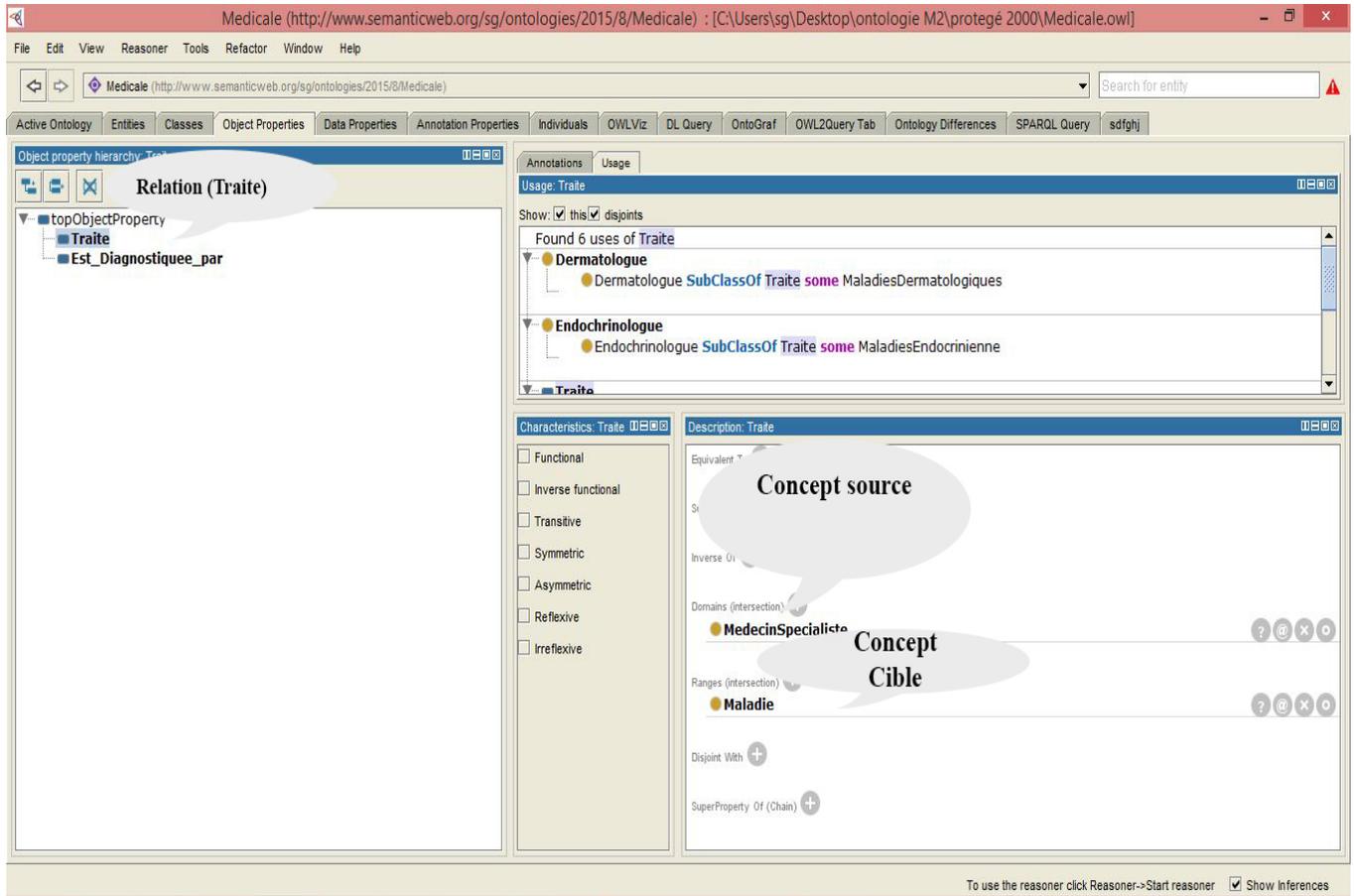


FIGURE 4.5 – Ajout de la relation "Traite"

4.7.6 Création des Individus

Pour chaque classe on peut créer un ou plusieurs individus on prend par exemple la classe médecin spécialiste ; on choisit Pneumologue on le renomme « Dr_Hameg » puis on remplit les champs (relations ou attributs) associés à ce médecin.

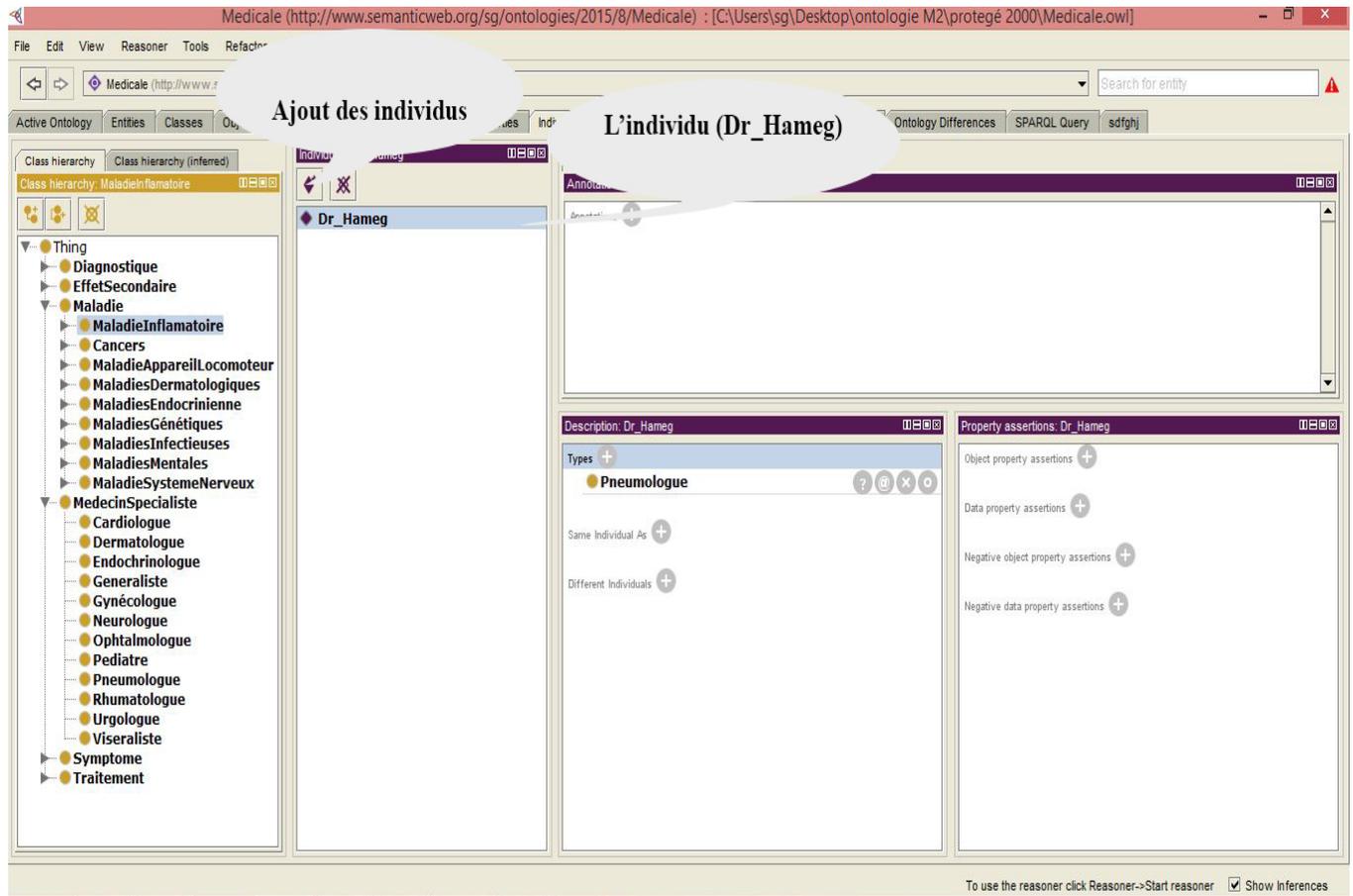


FIGURE 4.6 – Création des individus

4.7.7 Création des Axiomes

Les axiomes sont des restrictions sur les classes, les instances ainsi que les propriétés, ils démontrent comment les concepts interagissent les uns par rapport aux autres et permettent d'établir des inférences. On appelle la classe sur laquelle on restreint la propriété « filler ».

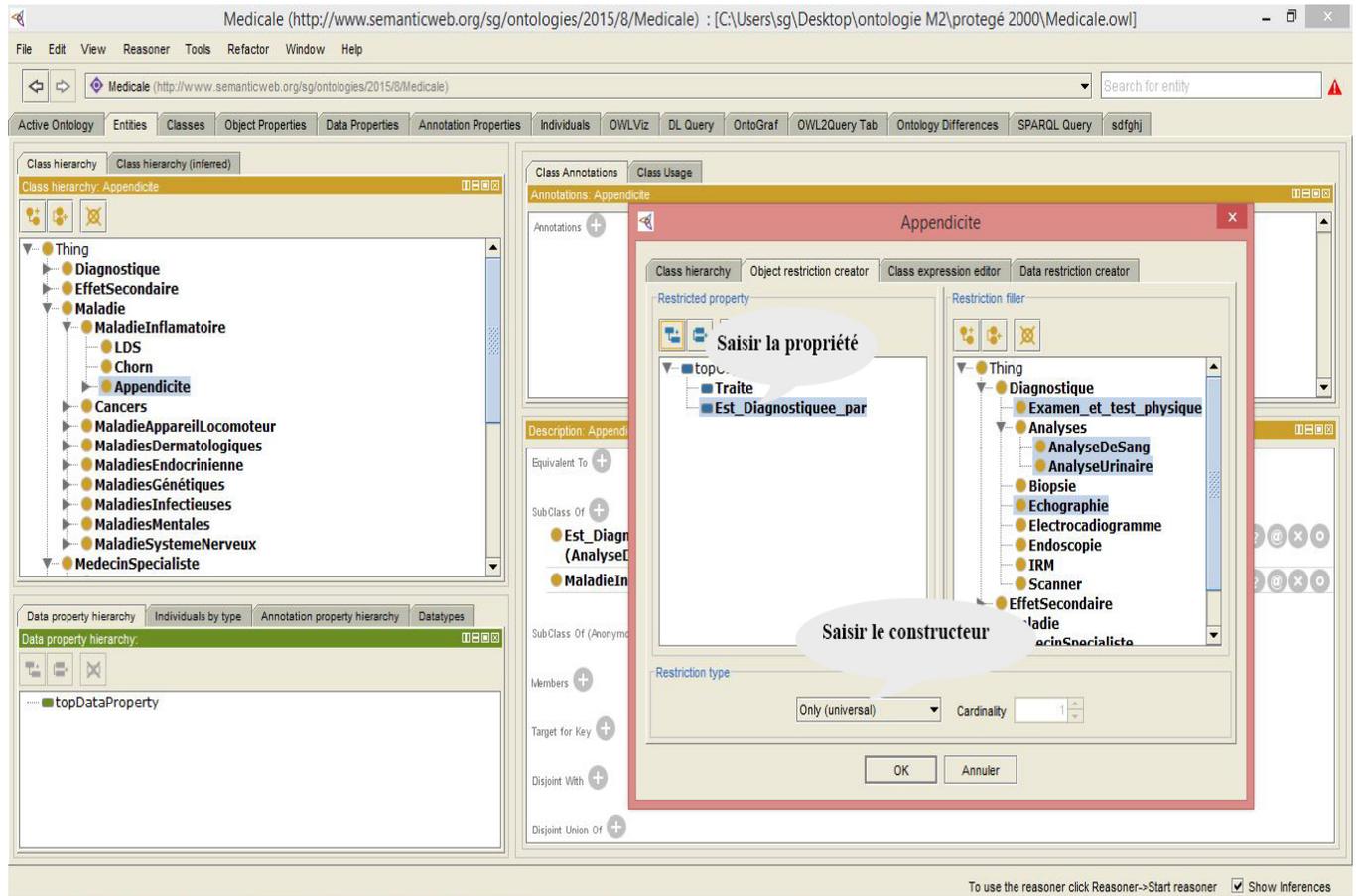


FIGURE 4.7 – Création d'un axiome sur la classe Appendicite

4.7.8 Génération du code RDF/XML

Sous protégé on peut générer plusieurs code dans la barre d'outil on choisie Window puis show Ontology views puis RDF/XML , la figure suivante présente le code généré.

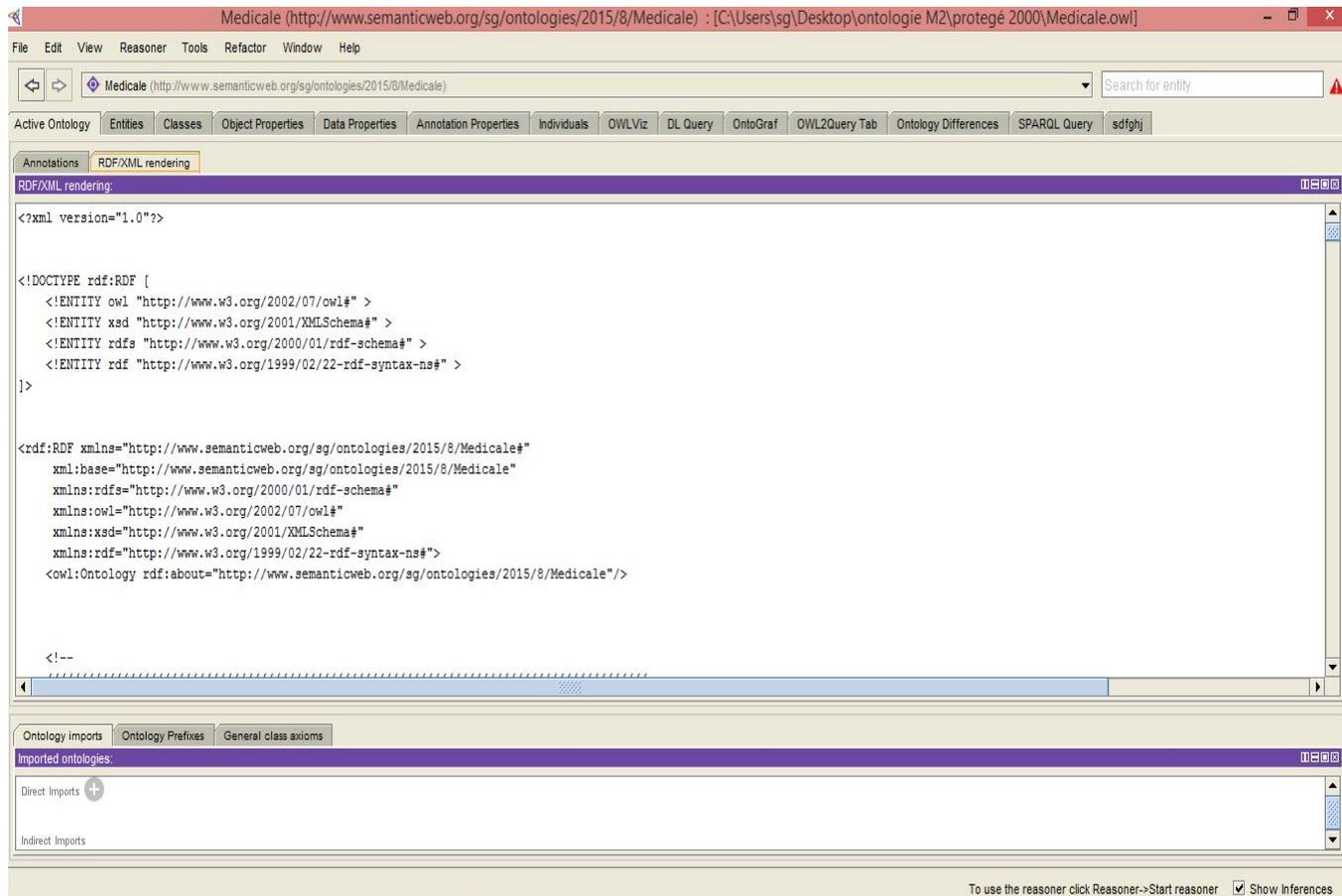
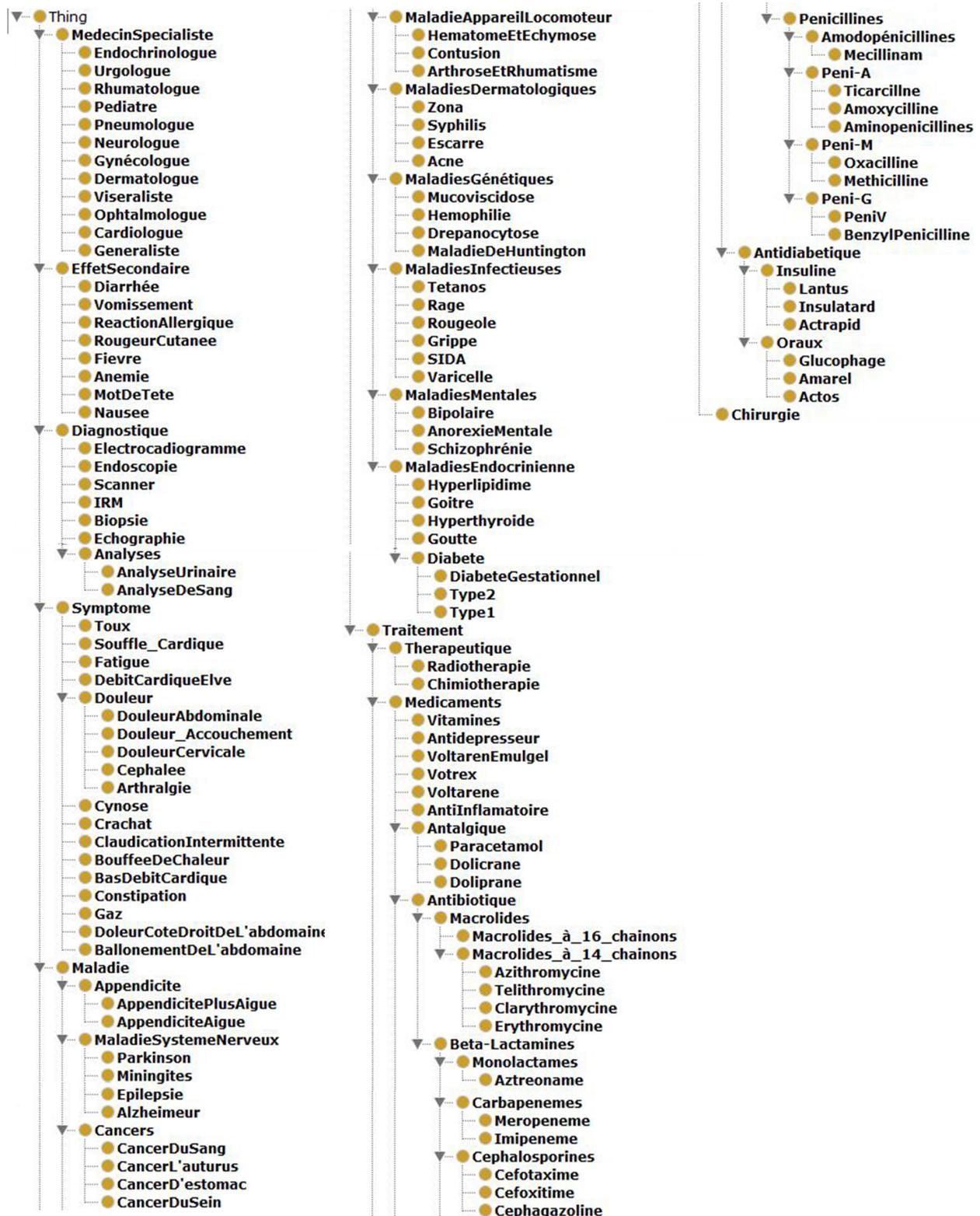


FIGURE 4.8 – extrait de code RDF/XML

4.7.9 Les classes et la hiérarchie des classes de notre ontologie



Les classes et la hiérarchie des classes de notre ontologie

4.7.10 Quelques requêtes SPARQL

SELECT ?subject WHERE ?subject rdf:type owl:DatatypeProperty

Elle permet de donner tous les attributs de notre ontologie.

SELECT ?subject WHERE ?subject rdf:type owl:ObjectProperty

Elle permet de donner toutes les relations de notre ontologie.

SELECT ?subject WHERE ?subject rdf:type owl:Class Elle permet de donner toutes les classes de notre ontologie.

SELECT ?subject ? Object WHERE ?subject rdfs:subClassOf? Object Elle permet de donner les sous classe dans subject et sa classe mère dans Object.

SELECT ?subject ? Object WHERE ?subject rdf:type ? Object Elle permet préciser c'est quoi les classes, les relations, les attributs.

Conclusion générale

On considère que nous avons réussi à réaliser une bonne partie de ces objectifs, dans la mesure où nous avons appliqué des méthodes d'ingénierie des connaissances et utilisé des outils et langages du web sémantique.

L'extraction manuelle de la connaissance à partir de corpus est un processus très coûteux en temps. Il n'a pas toujours été facile de déterminer si un terme devait être représenté en tant que classe ou attribut et certaines connaissances du domaine. Nous nous sommes basés sur les critères de T.Gruber pour évaluer notre ontologie. Les recherches dans le domaine médical ne cessent d'évoluer surtout en matière de nouvelles techniques d'imagerie et de nouveaux médicaments, c'est pour cela qu'il est nécessaire de prévoir une possibilité d'évolution pour l'ontologie. Notre ontologie évolue par l'ajout manuel de nouveaux concepts il suffit de déterminer à quel axe primitif appartiennent ces nouveaux concepts.

A travers ce modeste travail on a pu acquérir des compétences dans différents domaines nous citons :

- En ingénierie ontologique (qu'est-ce qu'une ontologie, méthodes de construction d'ontologies, processus de construction d'ontologies ...)
- La notion de web sémantique ainsi que ses langages et outils.
- Des connaissances en médecine.
- Des notions sur la logique de description et les moteurs d'inférences. Cependant il existe plusieurs perspectives envisageables à notre travail :
 - Utiliser cette ontologie pour faire de la recherche d'informations en médecine.
 - Étendre cette ontologie par l'ajout de nouveaux concepts, pour en faire une ontologie générale en médecine et l'exploiter dans une base de connaissance pour aider les experts dans la prise de décision.
- Tester encore le raisonnement, aux limites du langage OWL et du raisonneur Pellet.

Annexe

Les sources d'information

Entretien Réalisé avec Dr AMARI Mohamed, Résident en Viscérale au CHU NEDIR Mohamed de Tizi-Ouzou

Médecin spécialiste :

Viscéraliste

Maladie

Appendicite : on distingue deux cas ;

1. **Appendicite Aigue** : *Symptômes* :

- De la constipation,
- De la diarrhée
- Des gaz ;
- une douleur sourde démarrant au niveau de la région péri-ombilicale, se transformant en une douleur vive dans la partie inférieure droite de l'abdomen.
- une douleur qui se déplace.
- une perte d'appétit ;
- une faible fièvre (37 °C à 39 °C) ;
- des nausées et des vomissements ;
- une sensibilité à la pression de la région inférieure droite de l'abdomen ; la sensibilité au rebond est un bon indicateur d'une appendicite

2. **Appendicite plus avancée** : *Symptômes* :

- Ballonnement et une contracture de l'abdomen.
- Douleur du côté droit de l'abdomen.

Annexe

Diagnostique :

- Examen et tests physique.
- Analyse de sang.
- Prélèvement urinaire
- Echographie.

Type de Traitement :

Chirurgical L'intervention chirurgicale est l'unique traitement des appendicites aiguës

Annexe

Entretien Réalisé avec Dr DEBIANE, médecin Rhumatologue au CHU NEDIR Mohamed de BELLOUA (Sanataurium)

Médecin spécialiste :

Rhmatologue en 1er lieu

LUPUS ERYTHEMATEUX DISSEMINÉ

Symptômes :

- Articulations douloureuses et enflées.
- Éruption sur le visage en papillon avec une sensibilité accrue au soleil
- ulcérations cutanées voire gangrené
- Atteinte rénale a type d'insuffisance rénale.
- Atteinte cardiaque : pericardite, myocardite
- Atteinte pulmonaire : pneumopathie, lupique
- Atteinte neurologique : convulsions, AVC

Diagnostic :

- Examen clinique minutieux.
- Analyse de sang : VS,CRP,mais surtout bilan immunologique.
- Prélèvement urinaire
- Radiographies, échographie selon l'atteinte.

Type de Traitement :

Médicament

- Corticoides
- Aspirine,anti-inflammatoires non steroïdiens (AINS)
- Anti paludeens de synthese
- Immunosuppresseurs

Annexe

Entretien Réalisé avec *Dr REZGUI Melissa*, interne au CHU NEDIR Mohamed de TIZI-Ouzou

Médecin spécialiste :

Médecin interniste

Maladie de crohn

Symptômes :

- Articulations douloureuses et enflées.
- Éruption sur le visage en papillon avec une sensibilité accrue au soleil
- Douleurs abdominale.
- Diarrhées.
- Dysphagie
- NB :** ces signes évoluent par poussées séparées de remissions .
- Signes généraux : atteintes rénale, articulaire, dermatologique, oculaire

Diagnostic :

- Interrogatoire a la recherche d'autres pousseées.
- Examen clinique minutieux
- Analyse de sang : VS, CRP,
- Endoscopie

4.7.11 Type de Traitement : médicament parfois chirurgical

- Corticoïdes
- Aspirine, anti-inflammatoires non stéroïdiens (AINS)
- Si formes très avancées résection du segment atteint

Annexe

Liste des principales familles de maladies

Informations reçus de la part de Mlle KAREB Rezika Etudiante en 7^{ème} année medecine, département medecine ,Faculté Biomedicale, Université Mouloud Mammeri de Tizi-Ouzou

maladies dermatologiques

La dermatologie est la branche de la médecine qui s'occupe de la peau, des muqueuses et des phanères (ongles, cheveux, poils)

Pathologie dermatologique :

1. Acné
2. Urticaire
3. Escarre
4. Zona
5. Syphilis

Maladies génétiques (à gène identifié, à gène non identifié)

Une maladie génétique est une maladie due à une ou plusieurs anomalies sur un ou plusieurs chromosomes qui entraînent un défaut de fonctionnement de certaines cellules de l'organisme

les maladies génétiques

1. Maladie de Huntington
2. Drépanocytose
3. Hémophilie A
4. Mucoviscidose

Annexe

Maladies Infectieuses

1. Varicelle
2. SIDA
3. Grippe
4. Rougeole
5. Rage
6. Tétanos

Maladies mentales

1. Schizophrénie
2. Anorexie mentale
3. Hypersomnie
4. Bipolaire

Maladies endocriniennes

Les maladies endocriniennes sont causées par un dysfonctionnement des hormones sécrétées par les glandes endocrines ou due à un trouble métabolique

les maladies endocriniennes

1. Diabète (type 1, type 2, diabète gestationnel)
2. Hyperthyroïdie
3. Goutte
4. Goitre
5. Hyperlipidémie
6. Ostéoporose

Références Bibliographique

- [**A. Baneyx, 2007**] A. Baneyx ; Construire une ontologie de la Pneumologie ; Thèse de doctorat ; le 06 février 2007 ; Université Pierre et Marie Curie ; PARIS 6 ; France.
- [**Bachimont 2003**] B. Bachimont, J. Charlet & R. Troncy, Ontologies pour le Web Sémantique. Action spécifique 32 CNRS / STIC Web sémantique Rapport final. 2003.
- [**Borst, 1997**] Borst W. N. Construction of Engineering Ontologies. Center for Telematica and Information Technology, University of Twente, Enschede, NL. [M.uschold & M.gruninger, 1996b] Uschold M. ET Grüninger M. "Ontologies : Principles, Methods and Applications". (1996b).
- [**Bouaud et al, 1994 ; Charlet et al, 1996**] Bouaud J, Charlet J . (1994). Acquisition and structuring of an ontology within conceptual graphs. In Proceedings of ICCS'94 Workshop on Knowledge Acquisition using Conceptual Graph Theory, p. 1-25, University of Maryland, College Park, MD
- [**Darmoni et Thirion, 2000**] Darmoni S.-J. & Thirion B. (2000). A standard metadata scheme for health resources. J Am Med Inform Assoc, 7(1), 108.109.
- [**E.Sirin & B.Parsia, 05**] E. Sirin, B. Parsia, B.C. Grau, A. Kalyanpur and Y. Katz, Pellet : A Practical OWL-DL Reasoner, University of Maryland, 2005.
- [**Farquhar & Fikes, 1996**] Farquhar (A.), Fikes (R.), Rice (J.),1996 : «The Ontolingua Server : A Tool for Collaborative Ontology Construction», Proceedings of the 10th Knowledge Acquisition for Knowledge- Based Systems Workshop, Banff, Alberta, Canada, p. 44.1-44.19, 1996.

- [**Fernandez-Lopez, 1999**] Fernandez-Lopez, M., Gomez-Pérez, A., Pazos-Sierra, J. & Pazos-Sierra, A. Building a Chemical Ontology Using Methontology and the Ontology Design Environment. IEEE Intelligent Systems, January/February, pp. 37-46. (1999)
- [**F. Amourache, 2008**] F.Amourache "Construction d'une ontologie pour l'annotation des cvs/ offres d'emploi"; le 01/12/2008; Université Mentouri de Constantine; Algérie.
- [**Guarino, 1997a**] Guarino N. Some organizing principles for a unified top-level ontology. AAAI Spring Symposium on Ontological Engineering, 57-63. (1997a).
- [**Guarino ; 1997b**] Guarino N. (1997b). Understanding, building and using ontologies. International J. Human-Computer Studies, 46, 293-310, (1997b).
- [**Gomez-Perez, 1999a**] Gomez-Perez, Ontological Engineering : A state of the art. Expert, 1999.
- [**Gomez-Perez, 1999b**] Gomez-Perez, Tutorial on ontological Engineering, Paper presented at the proc, 1999.
- [**G.Falquet ,2001**] G. Falquet & C.L. Mottaz-Jiang, "Navigation hypertexte dans une ontologie multipoints de vue", NîmesTIC' . 2001.
- [**Lee 2002**] T.B. Lee et al., "The semantic Web". in Scientific American, May 2002.
- [**M. Kifer, G. Lausen and J. Wu, 1995**] M. Kifer, G. Lausen and J. Wu "Logical Foundations of Object- Oriented and Frame-Based Languages". Journal of the ACM (JACM), 42(4) :741 ? 843.
- [**M.gruninger & M.fox, 1995**] Grüninger, M. & Fox, M.S. Methodology for the Design and Evaluation of Ontologies. Proceedings of the IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing, 1995.
- [**M. Uschold & M. King, 1995**] M. Uschold& M. King, "Towards a me-

- thontology for building ontologies”, in Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing, IJCA, 1995.
- **Noy, N. F., Ferguson, R. W., & Musen, M. A. (2000)**. The knowledge model of Protege-2000 :Combining interoperability and flexibility. In R. Dieng, & O. Corby (Ed.), 12th International Conference in Knowledge Engineering and Knowledge Management (EKAW’00) (pp. 17 ?32). (Lecture Notes in Artificial Intelligence LNAI 1937) SpringerVerlag.
 - [**Psyché, Mendes & Bourdeau, 2004**] Valéry PSYCHÉ, Olavo MENDES, Jacqueline BOURDEAU, Apport de l’ingénierie ontologique aux environnements de formation à distance, 2004.
 - [**R.Moller & M.Wessel, 2001**] V. Haarslev, R. Moller and M. Wessel, RACER User’s Guide and Reference Manual, University of Hamburg, Computer Science Department, 2001.
 - [**Tom Gruber, 2009**] Tom Gruber dans l’Encyclopédie des systèmes de base de données, Ling Liu et M. Tamer Ozsü (Eds.), Springer-Verlag, 2009.
 - [**T. Berners-Lee, Hendler & O. Lassila, 2001**] T. Berners-Lee, J. Hendler and O. Lassila, (2001). ”The Semantic Web”. Scientific American, 284(5) :34 ?43.
 - [**Oberle 2004**] D. Oberle, R. Volz, B. Motik et S. Staab, ”An extensible ontology software environment”. In S.Staab et R. Studer (Eds.), Handbook on Ontologies (pp. 299-320) : Springer Verlag. 2004.
 - [**Uschold 2002**] M. Uschold and M.Gruninger, ”Creating semantically integrated communities on the World Wide Web”. Honolulu : Semantic Web Workshop, 2002.

Références Webiographique

- [01] : <http://www.lehtml.com/xml/syntaxe.html>
- [02] : http://www.w3schools.com/webservices/ws_rdf_intro
- [03] : http://www.w3schools.com/webservices/ws_rdf_schema
- [04] : <http://www.w3.org/2004/OWL/>
- [05] : <http://www.w3.org/2004/Daml-Oil/>
- [06] : <http://www.w3.org/2001/sw/WebOnt/>
- [07] : <http://jena.sourceforge.net/tutorial/RDQL/>
- [9] : <http://www.cartographie-semantique.fr/propositions/sndf-notre-formalisme-de-description-pour-la-cartographie-semantique/>
- [10] : <http://www.med.univ-rennes1.fr/noment/cim10/>
- [11] : <http://www.cnamts.fr/san/ccam/somccam.html>
- [12] : <http://dicdoc.kb.inserm.fr:2010/basismesh/mesh.html>
- [13] : <http://www.chu-rouen.fr/cismef/>
- [14] : <http://dublincore.org/>
- [15] : <http://doccismef.chu-rouen.fr/>
- [16] : <http://www.snomed.org/>
- [20] : <http://www.nlm.nih.gov/research/umls/>
- [21] : <http://sig.biostr.washington.edu/projects/fm/index.html>
- [22] : <http://www.loa-cnr.it/DOLCE.html>
- [23] : <http://estime.spim.jussieu.fr/Menelas/Ontologie/html/>