

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE**  
**MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE**  
**UNIVERSITE MOHAMED BOUDIAF - M'SILA**

**FACULTE** : Mathématique et Informatique

**DEPARTEMENT** : Informatique

**N°** : .....



**DOMAINE** : Informatique

**FILIERE** : INFORMATIQUE

**OPTION** : SIGL

**Mémoire présenté pour l'obtention**  
**Du diplôme de Master Académique**

**Par:** Oucief Affef

**Intitulé**

**Construction d'une Ontologie pour Représenter**  
**La Sémantique des Langues Naturelles**

**Soutenu devant le jury composé de :**

Bourahla Moustafa

Université de M'sila

Président

M<sup>me</sup> Bouzaroura Ahlam

Université de M'sila

Rapporteur

Kadri Said

Université de M'sila

Examineur

**Année universitaire : 2017 /2018**

# *Remerciements*

*Je remercie d'abord Dieu le tout puissant et miséricordieux, qui ma donné la santé, la force et la patience pour accomplir ce travail.*

*Au terme de ce travail, je tiens à exprimer ma profonde gratitude et mes sincères remerciements à mon encadreur:*

*M<sup>me</sup> Bouzaroura Ahlem, pour la qualité de son encadrement, sa patience et ses conseils judicieux et pertinents.*

*J'adresse mes vifs remerciements aux membres de jury pour l'intérêt qu'ils ont porté à cette recherche en acceptant d'examiner et juger ce travail.*

*Je tiens également à remercier tout les enseignants du département d'informatique.*

*Enfin, j'exprime mes remerciements à toutes les personnes qui ont contribué de près ou de loin pour la réaliser ce travail.*

# *Dédicaces*

*A ceux qui m'ont donné la vie, symbole de beauté, de fierté,  
de sagesse et de patience.*

*A ceux qui sont la source de mon inspiration et de mon  
courage à qui je dois l'amour et la reconnaissance.*

*Je dédie ce modeste travail à:*

*Mon père*

*Ma mère*

*Mes sœurs, Lemya et Racha*

*Mes deux familles, Oucief et Benkherbache*

*Mes amies, Ikram, Hannane et Abir*

*Affef*

# Sommaire

<b>Remerciements</b> .....	
<b>Introduction</b> .....	<b>i</b>

## Chapitre 1: Les Ontologies

1.1 Introduction.....	1
1.2 Ontologie .....	1
1.3 Les types d'ontologies .....	1
1.3.1 Les ontologies de haut niveau (top-level ontologies) : .....	1
1.3.2 Les ontologies de domaine (domaine ontologies) et les ontologies de tâche (task ontologies) .....	2
1.3.3 Les ontologies de tâches .....	2
1.3.4 Les ontologies d'application (application ontologies) .....	2
1.3.6 Les ontologies de représentation des connaissances (méta ontologie) .....	2
1.4 Les composantes d'une ontologie .....	2
a. Les concepts : .....	2
b. Les relations : .....	2
c. Les fonctions : .....	2
d. Les axiomes :.....	3
e. Les instances: .....	3
1.5 Les étapes de construction d'ontologies : .....	3
1.5.1 Methontology : .....	3
1.5.2 L'approche TOVE .....	4
1.5.3 L'approche SENSUS .....	4
1.5.4 Outils d'ingénierie ontologique à partir du texte .....	5
1.5.6 Outils d'ingénierie ontologique à partir des modèles conceptuels ou opérationnels .....	6
1.6 Les langages d'ontologies .....	7
1.6.1 KIF .....	7
1.6.2 KL-ONE .....	7
1.6.3 RDF et RDF Schéma .....	7
1.6.4 DAML + OIL .....	8
1.6.5 OWL .....	8
1.6.6 Le modèle orienté objet .....	8
1.7 Avantages .....	8
1.8 Les Editeurs .....	8
1.8.1 Ontorion Fluent Editor : .....	8
1.8.2 TopBraid Composer .....	8
1.8.3 OntoStudio .....	9
1.8.4 Protégé .....	9
1.9 Exemple d'ontologie .....	9
1.10. Conclusion .....	9

## Chapitre 2: La Sémantique

2.1 Introduction .....	11
2.3 La sémantique.....	11
2.3 Les domaines de l'informatique où la sémantique figure .....	11
2.3.1 Traitement automatique de langage naturel .....	11
2.3.2 Analyse sémantique .....	12
2.3.3 Web sémantique .....	12
2.3.4 La fouille du Web(Web Mining) .....	12
2.3.5 Fouille de textes(Text Mining) .....	13
2.3.6 Analyse des Sentiments et l'opinion mining .....	13
2.4 Modèles de présentation des connaissances sémantiques .....	14
2.4.1 Chomsky .....	14
2.4.2 Corpus .....	14
2.4.3 Thesaurus .....	14
2.4.4 Wordnet .....	15
2.4.5 Ontologie .....	15
2.5 Approche de l'analyse sémantique .....	15
2.5.1 Approche basée sur un dictionnaire .....	15
2.5.2 Approche basée sur le corpus .....	15
2.5.3 Sac de mot (bag of word) .....	17
2.5.4 L'Analyse Sémantique Latente (LSA) .....	17
2.3 Conclusion .....	17

## Chapitre 3: Implémentation

3.1 Introduction.....	18
3.2 Construction d'une ontologie.....	18
3.3 Ontologie être_vivant utilisée: .....	17
3.4 Description de notre approche .....	21
3.5 Implémentation .....	22
3.6 Fonctionnement de l'application.....	24
3.7 Conclusion .....	27
<b>Conclusion générale.....</b>	<b>30</b>
<b>Références.....</b>	<b>31</b>

## Liste des Figures

<b>Figure 1.1</b> : processus de développement et le cycle de vie de Methontology .....	4
<b>Figure 1. 2:</b> Exemple d'une ontologie de pizza .....	10
<b>Figure 3.1</b> Document de spécifications de l'ontologie.....	19
<b>Figure 3.2:</b> Hiérarchies de concepts.....	20
<b>Figure 3.3:</b> Diagramme de relations binaires.....	21
<b>Figure 3.4:</b> Exemple de regroupement des classes.....	23
<b>Figure 3.5</b> : l'ontologie être-vivant utilisé.....	25
<b>Figure 3.6</b> : Synonyme de mots en jwnl.....	26
<b>Figure 3.7:</b> implémentation requette SPARQL (subclass).....	27
<b>Figure 3.8</b> : implémentation requette SPARQL (individuals) .....	28
<b>Figure 3.9</b> :l'interface de l'application (the odd word).....	28

## Listes des tableaux

<b>Tableau 3.1:</b> Glossaire de termes.....	20
<b>Tableau 3.2 :</b> Dictionnaire de concepts.....	21
<b>Tableau 3.3:</b> Table des relations .....	21
<b>Tableau 3.4:</b> La table des instances de concepts.....	22
<b>Tableau 3.5:</b> Axiomes de classes et de rôles.....	22
<b>Tableau 3.6:</b> Assertions d'individus.....	22

## Introduction Générale

De nos jours, l'informatique a pris une place considérable dans la vie humaine, en s'intégrant dans tous les domaines : domaine enseignement, domaine médical, domaine économie, Les nombreuses utilisations du web ont conduit à une explosion des données stockées. Ainsi, le développement de Internet et la démocratisation des supports multimédias qui permettent de créer et stocker de plus en plus des documents multimédias qui ont rendu difficile l'accès à l'information.

Par ailleurs, des techniques ont été développées pour accéder automatiquement aux informations pertinentes, pour les exploiter de manière qualitative et supérieure cet information aussi bien par les humains que par la machine, il a été proposé d'étendre le web classique vers le web sémantique. la représentation sémantique du contenu de l'information aide l'utilisateur et lui facilite ses différentes tâches de recherche, (commerce électronique..).

Ainsi, ces nouvelles technologies ont été combiné avec la linguistique appelée aussi science du langage ou traitement automatique du langage naturel qui étudie les langues naturelles de l'espèce humaine. le TALN a pour but de traiter automatiquement la langue naturel par les ordinateurs, on parle actuellement des langues naturelles contrairement aux langues "artificielles ou formelles" qui sont les langages de programmation informatique ou la logique informatique. Cependant, l'analyse du langage nécessite une connaissance de sa structure sur de nombreux niveaux: que sont les mots? Que signifient-ils? Comment se combinent-ils pour former la phrase? Comment contribuent-ils au sens de la phrase? Et par ailleurs, comment fonctionnent le monde et le raisonnement de l'humain dans le monde?

Comme en linguistique, la sémantique désigne le lien entre le programme et l'objet mathématique qui dépendra des propriétés que l'on souhaite connaître du programme. La sémantique traite l'organisation des significations et leurs relations dans le monde des signes observables. La conceptualisation humaine du monde n'est pas très bien comprise et il n'est pas évident de bien décrire les phénomènes d'organisation de la pensée humaine. Néanmoins, de bons modèles sémantiques ont été proposés, en supposant que des constituants sémantiques basiques soient organisés au sein des structures conceptuelles.

Depuis son émergence au début des années 1990, dans les recherches en modélisation de connaissances, la notion d'ontologie s'est rapidement diffusée dans un grand nombre de domaines de recherche en informatique. Définie comme la représentation formelle et

consensuelle au sein d'une communauté d'utilisateurs, des concepts propres à un domaine et des relations qui les relient, la notion d'ontologie apparaît comme un moyen de représenter explicitement et de partager des objets d'un domaine ainsi que leur sémantique. Par ailleurs, de nombreux travaux portent sur l'utilisation des ontologies dans des domaines aussi divers que le traitement automatique de la langue naturelle, la recherche d'information, le commerce électronique, le web sémantique, la spécification des composants logiciels, l'intégration de systèmes d'information, etc. Le nombre d'ontologies qui sont développées et utilisées pour différentes applications ne cesse d'augmenter. Mais un des problèmes majeurs avec les ontologies est leur changement leur évolution. Les ontologies peuvent changer pour des raisons diverses, par exemple lorsque les besoins des utilisateurs changent, nécessitant une conceptualisation différente. Mais aussi un défaut de conception peut avoir été remarqué dans la conceptualisation d'origine, en appliquant des corrections de modélisation, ou bien encore il peut s'avérer nécessaire d'élargir la représentation du domaine. Les changements dans une ontologie peuvent causer des problèmes sérieux pour les données instanciées, les applications et les services qui sont dépendants de l'ontologie, ainsi que pour toutes les ontologies qui importent cette ontologie modifiée.

**Objectifs:**

L'objectif principal de notre travail est la construction d'une ontologie pour représenter la sémantique des langues naturelles. La simulation de cette tâche par ordinateur s'avère difficile à réaliser.

**Organisation du manuscrit:**

Ce mémoire est structuré comme suit:

Le premier chapitre, présente la notion d'ontologie : ses types, ses langages, ses composants ainsi que les étapes de construction d'ontologie.

Le deuxième chapitre, décrit la notion de sémantique qui est une notion très importante en relation avec les connaissances.

Le troisième chapitre, est consacré à la présentation de l'approche adoptée pour la construction d'ontologie. Ainsi nous définissons les principes de base nécessaires à la compréhension de l'approche adoptée et nous passons par la suite à la présentation de l'implémentation réalisée et des résultats obtenus.

Enfin, nous terminons par une conclusion générale dans laquelle nous exposons le jugement final de notre approche est son efficacité.

## CHAPITRE 1

### LES ONTOLOGIES

#### 1. Introduction

Le terme « ontologie », construit à partir des racines grecques *ontos* (ce que existe,l'existant) et *logos* (le discours, l'étude), est un mot que l'informatique a emprunté à la philosophie au début des années 1990. En philosophie, l'Ontologie est une branche fondamentale de la Métaphysique, qui s'intéresse à la notion d'existence, aux catégories fondamentales de l'existant et étudie les propriétés les plus générales de l'être.

Par analogie, le terme est repris en informatique et en science de l'information, où une ontologie est l'ensemble structuré des termes et concepts représentant le sens d'un champ d'informations, que ce soit par les métadonnées d'un espace de noms, ou les éléments d'un domaine de connaissances.

#### 1.2 Ontologie

L'ontologie est un terme emprunté à la philosophie qui est une branche de la philosophie qui traite la nature et l'organisation de la réalité. Plusieurs définitions d'ontologies sont données dans la dernière décade, mais celle qui caractérise l'essentiel d'une ontologie est fondée sur la définition reliée dans [1] : « Une ontologie est une spécification formelle, et explicite d'une conceptualisation partagée ». *Conceptualisation* réfère à un modèle abstrait de certains phénomènes dans le monde qui identifie les concepts appropriés de ce phénomène. *Explicite* signifie que le type de concepts utilisés et les contraintes sur leur utilisation doivent être explicitement définis. *Formelle* réfère au fait qu'une ontologie doit être compréhensible par la machine, c'est-à-dire cette dernière soit capable d'interpréter la sémantique de l'information fournie. *Partagée* indique que l'ontologie supporte la connaissance consensuelle, et elle n'est pas restreinte à certains individus mais acceptée par un groupe [2].

## 1.3 Les types d'ontologies

### 1.3.1 Les ontologies de haut niveau (top-level ontologies) :

Elles décrivent des concepts très généraux comme l'espace, le temps, la matière, les objets, les événements, les actions, etc.

### 1.3.2 Les ontologies de domaine (domaine ontologies) et les ontologies de tâche (task ontologies)

Elles décrivent le vocabulaire lié à un domaine générique (comme la médecine, ou les automobiles) ou une tâche ou une activité générique (comme le diagnostic ou la vente)..

### 1.3.3 Les ontologies de tâches

Ce type d'ontologies décrit le vocabulaire concernant une tâche générique (ex. : enseigner, diagnostiquer...), notamment en spécialisant les concepts d'une ontologie de haut niveau. Certains auteurs emploient le nom « ontologie du domaine de la tâche » pour faire référence à ce type d'ontologie.

### 1.3.4 Les ontologies d'application (application ontologies)

Elles décrivent des concepts dépendant à la fois d'un domaine et d'une tâche particuliers dans ce domaine. Ces concepts correspondent souvent aux rôles joués par des entités.

### 1.3.6 Les ontologies de représentation des connaissances (méta ontologie)

Elles décrivent les concepts utilisés par les langages de représentation des ontologies.[3]

## 1.4 Les composantes d'une ontologie

Les ontologies fournissent un vocabulaire commun d'un domaine et définissent la signification des termes et des relations entre elles. La connaissance dans les ontologies est principalement formalisée en utilisant les cinq types de composants (Gomez et Benjamins 1999) à savoir : concepts (ou classes), relations (ou propriétés), fonctions, axiomes (ou règles) et instances (ou individus).

### a. Les concepts :

Aussi appelés termes ou classes 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, par

---

## CHAPITRE 1 – LES ONTOLOGIES

---

exemple, la description d'un ensemble d'objets, d'une tâche, d'une fonction, d'une stratégie, d'un processus de raisonnement, etc. [3]

### **b. Les relations :**

Traduisent les interactions existant entre les concepts présents dans le domaine ciblé. Ces relations sont formellement définies comme tout sous ensemble d'un produit cartésien de  $n$  ensembles, c'est à dire  $R : C_1 \times C_2 \times \dots \times C_n$

- la relation de spécialisation (subsumption),
- la relation de composition (méronymie),
- la relation d'instanciation, etc. 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 de l'ontologie. [5]

### **c. Les fonctions :**

Sont des cas particuliers de relations dans lesquelles le  $n$ ème élément (extrant) de la relation est défini de manière unique à partir des  $n-1$  éléments précédents (intrants). Formellement, les fonctions sont définies ainsi :  $F : C_1 \times C_2 \times \dots \times C_{n-1} \rightarrow C_n$ . Comme exemple de fonctions binaires, nous pouvons citer la fonction « mère-de » ; tandis que, « *prix-de voiture-usagée* » qui calcule le prix d'une voiture de seconde main en fonction du modèle de voiture, de la date de fabrication et du nombre de kilomètres est une fonction ternaire..

### **d. Les 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.

### **e. Les instances:**

Où individus constituent la définition extensionnelle de l'ontologie. Ils représentent des éléments singuliers véhiculant les connaissances (statiques, factuelles) à propos du domaine du problème.

## **1.5 Les étapes de construction d'ontologies :**

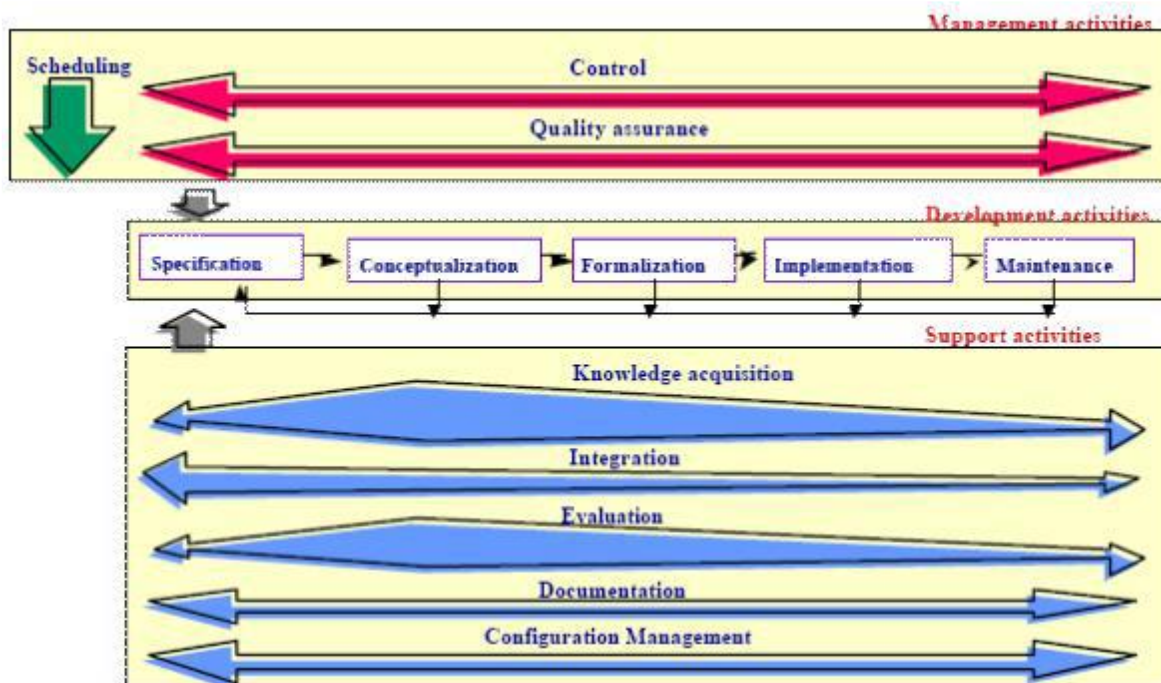
Les ontologies réalisées sont très différentes les unes des autres. Dans cette section nous citons quelques méthodologies et outils existants.

# CHAPITRE 1 – LES ONTOLOGIES

## 1.5.1 Methontology :

La methontology se veut être une approche intégrée s'inscrivant dans le cadre d'un processus de gestion de projet complet, allant de la spécification des besoins jusqu'à la phase de réalisation et maintenance. Ce processus est composé des étapes suivantes:

- Spécification : déterminer l'utilisation future de l'ontologie.
- Conceptualisation : obtenir un modèle du domaine au niveau des connaissances.
- Formalisation : transformation du modèle conceptuel en modèle formel.
- Intégration : réutilisation d'autres ontologies.
- Implémentation : construction d'un modèle opératoire utilisable par un ordinateur.
- Maintenance : mise à jour de l'ontologie en cas de besoin.[4]



**Figure 1 :** processus de développement et le cycle de vie de Methontology [4]

## 1.5.2 L'approche TOVE

Cette méthode tente de construire un modèle logique de connaissance où les ontologies sont construites en respectant les étapes suivantes [5]:

- Identification de scénarii (problèmes) dépendants d'une application ;

# CHAPITRE 1 – LES ONTOLOGIES

---

- Formulation de questions informelles (basées sur les scénarii) auxquelles l'ontologie doit permettre de répondre ;
- Spécification d'une terminologie à partir des termes apparaissant dans les questions.
- Spécification formelle (en KIF) des axiomes et des définitions pour les termes de la terminologie.
- Evaluation de la complétude de l'ontologie.

## 1.5.3 L'approche SENSUS

Avec cette approche, les ontologies de domaine à développer sont dérivées d'une ontologie générale qui est l'ontologie SENSUS. Cela est fait en reliant les termes spécifiques du domaine à cette ontologie et d'enlever dans SENSUS, les termes qui ne sont pas pertinents dans la nouvelle ontologie. Ainsi, le squelette de cette nouvelle ontologie est générée automatiquement avec l'outil OntoSaurus.

La construction d'une ontologie dans un domaine spécifique suit les étapes suivantes :

- *Première étape* : identifier les termes clés du domaine.
- *Deuxième étape* : relier manuellement les termes clés à SENSUS
- *Troisième étape* : inclure tous les concepts qui se trouvent sur le chemin depuis le terme clé jusqu'à la racine de SENSUS.
- *Quatrième étape* : ajouter les nouveaux termes de domaine. Dans cette étape on ajoute manuellement tous les termes qui sont pertinentes pour le domaine et qui ne sont pas encore apparue. Ainsi les étapes 2 et 3 sont répétées pour inclure les concepts sur le chemin, des nouveaux concepts jusqu'à la racine de SENSUS.
- *Cinquième étape* : Ajouter le sous arbre entier. Dans cette étape, celui qui fait l'ontologie doit faire attention aux noeuds qui ont un grand nombre de chemins qui passe à travers le nouvel arbre généré.

## 1.6 Outils d'ingénierie ontologique à partir du texte

### 1.6.1 Terminal

La méthodologie TERMINAE [6] propose une approche pour sélectionner les concepts, leurs propriétés, les relations et leur regroupement. Cette méthodologie repose sur l'utilisation d'outils de traitement automatique des langues analysant les termes de textes et les relations

---

## CHAPITRE 1 – LES ONTOLOGIES

---

lexicales. Les termes sont regroupés suivant leur contexte et facilitent la création de concepts et de relations sémantiques. Les concepts et relations sont ensuite formalisés dans un modèle.

Cette méthodologie est composée de plusieurs étapes :

- La première consiste en la description des besoins (utilisation de l'ontologie, connaissance à représenter...).
- L'étape suivante conduit à construire un corpus sur lequel les outils de traitement automatique de langues se reposent.
- La troisième étape correspond à l'étude linguistique. Des outils sont appliqués au corpus afin d'extraire les termes et leurs relations lexicales et syntaxiques. Une application de la méthodologie est proposée par les auteurs à partir des outils LEXTER.
- La phase suivante, appelée phase de normalisation, vise à conceptualiser les résultats de l'étape précédente. Les termes à conserver sont sélectionnés en fonction de leur contexte et définis à partir d'une définition en langage naturel. Les concepts sont ensuite identifiés ainsi que les relations sémantiques entre eux. Ils sont représentés sous forme d'un réseau sémantique.
- La dernière étape est celle de la formalisation. Le réseau sémantique précédemment obtenu est traduit et enrichi dans un langage formel.

TERMINAE a l'avantage de répondre à certaines questions et d'axer le choix des concepts et des relations de l'ontologie sur l'extraction de termes d'un corpus de référence

### 1.6.2 KAON

KAON [7] est un environnement de construction d'ontologie qui associe différents modules, correspondant à autant de phases identifiées de l'ingénierie ontologique : création, stockage, raffinement, exploitation, maintenance et application d'ontologies. Dans l'ensemble de ce processus, le module de « création ontologique » : TextToOnto applique des « stratégies de fouille de texte à des corpus textuels pour la création semi-automatique d'ontologies ».

### 1.6.3 Acquisition des termes à partir d'une analyse linguistique

Les avantages d'une analyse de corpus sont nombreux [8]:

- Les textes contiennent des connaissances explicites par opposition aux connaissances des experts souvent tacites, difficilement exprimables et difficilement accessibles.
- L'analyse automatisée de corpus est un gain de temps par rapport aux entretiens avec des experts du domaine.

## CHAPITRE 1 – LES ONTOLOGIES

---

- Les textes sont des ressources fiables et stables puisqu'ils fixent par écrit, à un instant t, un certain nombre de connaissances sur le domaine.
- Les modèles sont a priori plus facilement compréhensibles et donc maintenables car le retour aux textes est possible.

### 1.6.4 LEXTER & SYNTAX

Lexter extrait dans un premier temps des syntagmes nominaux maximaux, en s'arrêtant aux unités linguistiques constituant des frontières des syntagmes; ces frontières des syntagmes sont constituées par une ponctuation forte, un verbe conjugué, un pronom ou déterminées selon des informations contextuelles complexes. Ensuite, « Le module acquiert par lui-même, à l'aide d'une méthode d'apprentissage endogène sur corpus, les informations de sous-catégorisation des noms et des adjectifs, propres aux corpus, dont il a besoin pour résoudre les cas d'ambiguïté de rattachement prépositionnel. Et enfin, un dernier module calcule les dépendances syntaxiques au sein des syntagmes nominaux maximaux et construit un réseau de têtes et d'expansions de candidats termes.

### 1.6.5 FASTR

FASTR, est un outil d'acquisition terminologique qui repère les variantes en corpus d'une même unité lexicale. Ce regroupement permet d'associer différents libellés à un seul terme.

Le système calcule des variantes syntaxiques, morpho-syntaxiques et sémantico-syntaxiques.

### 1.6.7 Outils d'ingénierie ontologique à partir des modèles conceptuels ou opérationnels

Il existe plusieurs travaux sur l'extraction automatique des termes à partir des schémas de bases de données conceptuel (E/A ou UML) ou relationnelle pour générer les ontologies.

L'objectif principal de ces travaux est l'annotation des bases de données. Pour ce faire il n'est pas nécessaire d'aller construire l'ontologie à partir de zéro mais ils ont pensé de la développer soit à partir du modèle conceptuel ou à partir du schéma relationnel. Z. Xu et al. [9] ont présenté ER2WO comme un outil d'extraction automatique basé sur une approche formelle de transformation du modèle conceptuel Entité/ Association vers une ontologie. ER2WO produit une ontologie codée en OWL DL à partir d'un fichier XML contenant le modèle E/A obtenu par l'outil PowerDesigner 4, ER2WO valide ses résultats par l'OWL Ontology Validator 5. Afin de réaliser cet outil, les auteurs ont défini un module de translation basé sur un ensemble de règles pour générer une ontologie dans la syntaxe abstraite du OWL DL. Cette ontologie constituera

l'entrée d'un autre module de transformation pour générer l'ontologie finale dans la syntaxe RDF/XML.

### 1.7 Les langages d'ontologies

#### 1.7.1 KIF

KIF [10], est un langage basé sur les prédicats du premier ordre avec des extensions pour représenter des définitions et des méta-connaissances. Tant que la logique du premier ordre est un langage de bas niveau pour l'expression d'ontologies, l'outil *Ontolingua* [11], permet aux utilisateurs de construire des ontologies KIF à un niveau plus élevé de la description par l'importation des définitions des ontologies prédéfinies.

#### 1.7.2 KL-ONE

KL-ONE [12], est un langage basé sur la logique de description [13]. Il est une formalisation de représentation de la connaissance à base de cadres 5 (« frame-based »). Ce système maintient la définition des concepts par un simple nommage, et l'indication de la correspondance des concepts dans une hiérarchie de généralisation/spécialisation.

#### 1.7.3 RDF et RDF Schéma

RDF [14] « *Resource Description Framework* » est un formalisme graphique pour représenter des méta-données. Il est basé sur la notion de triplet (sujet, prédicat, objet).

Le sujet et l'objet sont des ressources liées par le prédicat. RDF utilise la syntaxe XML, mais il ne donne aucune signification spécifique pour le vocabulaire comme sous classe de, ou le type. RDFS est reconnu comme un langage d'ontologie qui définit:

- Des classes et des propriétés.
- Les sous-classes, les super-classes, les sous-propriétés, et les super-propriétés.
- Le domaine de définition et le domaine image des propriétés.

#### 1.7.4 DAML + OIL

DAML + OIL est un langage construit sur des normes précédentes du W3C telles que RDF et RDF Schéma, et étend ces langages avec des primitives de modélisation plus riches. DAML+OIL a été conçu à partir du langage d'ontologie DAML-ONT (DARPA Agent Modelling Language-Ontology, Octobre 2000) en vue de combiner plusieurs composants du langage OIL [15]. OIL « *Ontology Inference Language* » est une représentation basée sur le Web, et une couche d'inférence pour des ontologies. Il combine les primitives de modélisation des

---

## CHAPITRE 1 – LES ONTOLOGIES

---

langages à base de cadres (frames) avec la sémantique formelle et le raisonnement fournis par la logique de description.

### 1.7.5 OWL

Développé par le groupe de travail sur le Web Sémantique du W3C, OWL peut être utilisé pour représenter explicitement les sens des termes des vocabulaires et les relations entre ces termes. OWL vise également à rendre les ressources sur le Web aisément accessibles aux processus automatisés [16], d'une part en les structurant d'une façon compréhensible et standardisée, et d'autre part en leur ajoutant des méta-informations. Pour cela, OWL a des moyens plus puissants pour exprimer la signification et la sémantique que XML, RDF, et RDF-S [17].

### 1.7.6 Le modèle orienté objet

Les différentes recherches, au sein de l'intelligence artificielle, pour la représentation de la connaissance ont abouti au développement de divers formalismes pour définir des ontologies, notamment le langage KIF [10] (*Knowledge Interchange Format*) et les langages de représentation de la connaissance dérivés de KL-ONE [17]. Cependant, ces langages de représentation sont peu connus hors des laboratoires de recherche en Intelligence Artificielle.

## 1.8 Avantages

Parmi les avantages de la méthode des ontologies la compréhension commune de la structure de l'information entre les personnes ou les fabricants de logiciels, fournir une description formelle du modèle qui permettent de raisonner, supporter les requêtes communes : questions sur la véracité des déclarations, requêtes attendants un objet à envoyer, assurer l'interopérabilité entre systèmes, permettre l'échange de connaissances entre Systèmes.

## 1.9. Les Editeurs

Il existe plusieurs éditeurs pour les ontologies, on cite parmi eux :

### 1.9.1 Ontorion Fluent Editor :

Un nouvel éditeur d'ontologie basé sur le langage CNL (controlled Natural Language) très compréhensible.

### 1.9.2 TopBraid Composer

Un environnement de modélisation visuelle pour les ontologies basé sur la plate-forme Eclipse et l'API Jena.

### 1.9.3 OntoStudio



# CHAPITRE 2

## LA SEMANTIQUE

### 2.1 Introduction

La sémantique domine un intérêt économique réel, bien que et l'informatique est utilisée de plus en plus fréquemment dans tous les appareils de la vie moderne elle requiert un niveau de qualité bien plus important que les applications traditionnelles de l'informatique. Ainsi, la manifestation évidente de l'intelligence artificielle est la possibilité d'un ordinateur ou d'un robot de comprendre notre langue et communiquer sans aucun obstacle avec nous. La grande société essaye d'intégrer ce type de programme dans notre vie quotidienne c'est le cas des assistants virtuels comme Siri (Apple), Cortana (Microsoft), Bixby (Samsung) ou Assistant (Google).

### 2.2 La sémantique

La sémantique est une branche de la linguistique qui étudie le sens des mots.

Selon la définition du dictionnaire (Larousse), elle est considérée comme l'étude des unités linguistiques et leurs combinaisons. La combinaison entre les unités linguistiques est fortement liée à la structure syntaxique de la langue. Par ailleurs, en informatique, c'est l'étude de la signification des programmes informatiques vus en tant qu'objets mathématiques. Comme en linguistique, ici la sémantique désigne le lien entre le programme, et l'objet mathématique qui dépendra des propriétés que l'on souhaite connaître du programme. En effet, c'est un travail d'actualité dans le domaine scientifique. Cependant, les linguistes informaticiens tentent de modéliser le langage (sous forme de graphes, d'espace vectoriel ou de réseaux neuronaux) afin que la machine puisse attribuer des sens aux mots qui, pour eux, ne présentent qu'une suite de caractères.

### 2.3 Les domaines de l'informatique où la sémantique figure

L'utilisation de la sémantique est utile dans divers domaines de l'informatique:

#### 2.3.1 Traitement automatique de langage naturel

Le traitement automatique de la langue naturelle (TALN) ou des langues (TAL) est une discipline à la frontière de la linguistique, de l'informatique et de l'intelligence artificielle.

---

## CHAPITRE 2-LA SEMANTIQUE

---

Elle concerne la conception de systèmes et techniques informatiques permettant de manipuler le langage humain. Historiquement la première application a débuté dès les années 1950.

### 2. 3.2 Analyse sémantique

L'analyse sémantique tente de découvrir de façon plus générale le sens des phrases ou de textes entiers. C'est une technique qui se fait sur la sémantique des phrases pour déterminer le sens des écrits. L'analyse sémantique se base sur deux analyses:

➤ Analyse lexicale

Elle assure la décomposition du texte source en ses différents constituants qu'on appelle entités lexicales ou tokens ,en lisant le texte caractère par caractère. Toute suite de caractères située entre des séparateurs est considérée comme un token.

➤ Analyse syntaxique

C'est la phase qui permet de vérifier si le texte source respecte la grammaire du langage .On distingue deux grandes catégories d'analyseurs syntaxiques ceux de l'analyse descendante et ceux de l'analyse ascendante.

### 2.3.3 Web sémantique

Le web sémantique a attiré depuis ces dernières années l'attention de nombreux chercheurs, dont le but d'arriver à un web intelligent, où les informations ne seraient plus stockées mais comprises par les ordinateurs afin d'apporter à l'utilisateur ce qu'il cherche vraiment.

La notion de web sémantique fait référence du web de demain dans lequel les utilisateurs devraient être déchargés d'une bonne partie de leurs tâches de recherche. Le web sémantique peut fournir aux utilisateurs, par moyen d'agents logiciels, des services automatiques et avancés Laublet et al.[17]. Berners-Lee et al.[18], l'écrivent en substance «Le web sémantique est une extension du web courant, dans lequel on donne une information un sens bien défini pour permettre aux ordinateurs et aux personnes de travailler en coopération».

### 2.3.4 La fouille du Web(Web Mining)

La fouille du Web (Web mining) est l'utilisation de techniques d'exploration de données pour automatiser, découvrir et extraire des informations à partir de documents Web et services. Ce domaine de recherche est si énorme aujourd'hui en raison des intérêts de diverses communautés de recherche, l'énorme croissance des sources d'information disponibles sur le Web et l'intérêt récent pour le commerce électronique[19].

---

## CHAPITRE 2-LA SEMANTIQUE

---

Il consiste à utiliser l'ensemble des techniques de la fouille de données afin de développer des outils permettant l'extraction d'informations pertinentes à partir de données du Web (documents, traces d'interactions, structure des liens, etc.

### **2.3.5 Fouille de textes(Text Mining)**

Le terme "fouille de texte" distingue dans la littérature,des méthodologies et des outils très différents. Selon la culture scientifique des chercheurs qui s'intéressent à la fouille de textes, ce terme recouvre des travaux en recherche d'information, en extraction d'information, en extraction de terminologies, en structuration d'ontologies, pour les systèmes questions, réponses...etc[20].

La fouille de textes est définie par comme l'ensemble des tâches qui, par analyse de grandes quantités de textes et la détection de modèles fréquents, essaie d'extraire de l'information probablement utile[21].

L'objectif de la fouille de textes est d'exploiter l'information contenue dans les documents textuels de différentes manières,incluant les analyses classiquement faites en fouille de données: découvrir des patrons et des tendances dans les données trouver des associations entre les notions,construire des règles de prédiction, etc [22].

### **2.3.6 Analyse des Sentiments et l'opinion mining**

Dans la littérature,sentiments analysis,opinion mining, opinion extraction, sentiments mining,affect analysis, emotion analysis,review minig, appraisal extraction, sont des termes utilisés pour désigner des technologies d'analyse automatique des discours écrits ou parlés, afin d'extraire des informations subjectives ou des émotions.

L'origine de la discipline d'analyse des sentiments se réfère aux sciences de la psychologie, la sociologie et d'anthropologie[23].

Le terme analyse sentimentale se réfère à l'extraction automatique de texte évaluatif qui aide à produire des résultats prédictifs. L'analyse des sentiments est le domaine d'étude qui analyse les opinions , les sentiments, les évaluations des attitudes et les émotions des gens vers des entités telles que des produits, des services,des organisations, des particuliers, des problèmes, des événements, des sujets et leur attributs. Il présente un grand espace extrêmement actif en traitement automatique des langues[24].

## **2.4 Modèles de présentation des connaissances sémantiques**

Plusieurs types de modèles de présentation des connaissances sémantiques ont été proposés afin de qualifier l'approche la plus abstraite.

---

## CHAPITRE 2-LA SEMANTIQUE

---

### 2.4.1 Chomsky

Noam Chomsky est le linguiste vivant le plus important et le plus connu. Son influence sur la linguistique contemporaine est déterminante. Né à Philadelphie en 1928, formé à la linguistique par Z. S. Harris – un des principaux théoriciens du distributionnalisme – au début des années 1950, il élabore dans sa thèse *The Logical Structure of Linguistic Theory*, écrite en 1955, les instruments formels qui permettent pour la première fois à la syntaxe et à la sémantique de rejoindre la phonologie dans le domaine des études rigoureuses et empiriquement motivées : y est développée l'idée fondamentale que la maîtrise d'une langue peut être simulée par le biais d'un ensemble de règles et de principes explicites, qui constituent une grammaire « générative », c'est-à-dire une procédure qui énumère et analyse mécaniquement tous les énoncés bien formés de la langue étudiée et rien qu'eux.

### 2.4.2 Corpus

Un corpus est un regroupement structuré de textes intégraux, documentés, éventuellement enrichis par des étiquetages, et rassemblés de manière théorique réflexive en tenant compte des discours et des genres, et de manière pratique en vue d'une gamme d'applications.

Tout corpus suppose en effet une préconception des applications, fussent-elles simplement documentaires, en vue desquelles il est rassemblé : elle détermine le choix des textes, mais aussi leur mode de « nettoyage », leur codage, leur étiquetage ; enfin, la structuration même du corpus. Ensemble homogène et significatif de données linguistiques observées et à partir desquelles pourra s'élaborer la théorie. La notion de corpus est évidemment fondamentale dans la linguistique structurale : désireuse de substituer à la normativité de la grammaire ou aux fondements pseudo-logiques de la signification un enregistrement de l'usage, écrit ou oral, ainsi qu'un relevé des différences en fonction des différents contextes, elle pose en principe l'absolue nécessité de s'appuyer sur des productions de parole. Les caractéristiques d'un corpus significatif sont : l'homogénéité (le groupe qui le produit est socialement défini), la synchronie (on ne peut travailler sur des énoncés trop éloignés chronologiquement les uns des autres), la moindre redondance possible (on supprimera des énoncés représentatifs d'un phénomène identique).

### 2.4.3 Thesaurus

Selon la norme internationale ISO 2788 (1986), les thésaurus nés dans les années 50, sont le «vocabulaire d'un langage d'indexation contrôlé organisé formellement de façon à expliciter les relations a priori entre les notions (par exemple relation génériquespécifique)». Selon la même norme un langage d'indexation est un «ensemble contrôlé de termes choisis dans une langue naturelle et utilisés pour représenter sous forme condensée, le contenu des documents» [25].

---

## CHAPITRE 2-LA SEMANTIQUE

---

Le thésaurus contient un lexique (l'ensemble de termes d'une langue de spécialité) définissant les descripteurs et les non descripteurs (termes interdits), un sous-ensemble de définitions et de notes d'application pratique et une structure classificatoire exprimée par des relations sémantiques entre les termes du lexique: relation d'équivalence intra linguistique (synonymie), relation d'équivalence inter linguistique (traduction), relation hiérarchique, relation d'association. Alors que les classifications organisent les sujets des documents, les termes des thésaurus visent à décrire des concepts.

### 2.4.4 Wordnet

Wordnet est une base de données de mots anglais qui sont liés par leurs relations sémantiques. C'est comme un dictionnaire/Thésaurus suralimenté avec une structure de graphe.

### 2.4.5 Ontologie

Une ontologie représente un modèle conceptuel sur lequel il est possible de développer des systèmes de connaissances qui soient partageable et réutilisables permettant ainsi, l'interopérabilité entre plusieurs sources d'information et de connaissances.

## 2.5 Approche de l'analyse sémantique

En se basant sur l'étude menée par les chercheurs cités dans la référence [26], concernant des récentes mises à jour des algorithmes et des applications de l'analyse des sentiments. Ces derniers ont présentés des différentes approches de l'analyse sémantique. Parmi ces approches on a :

### 2.5.1 Approche basée sur un dictionnaire

C'est un petit ensemble d'opinions de mots collecté manuellement avec des orientations connues. Ensuite, cet ensemble est développé en recherchant dans les corpus bien connus WordNet ou Thésaurus pour leurs synonymes et antonymes. Les mots nouvellement trouvés sont ajoutés à la liste de départ, alors l'itération suivante démarre.

Le processus itératif s'arrête lorsqu'aucun nouveau mot n'est trouvé. Une fois le processus terminé, une inspection manuelle peut être effectuée pour supprimer ou corriger les erreurs.

### 2.5.2 Approche basée sur le corpus

L'approche basée sur le corpus aide à résoudre le problème de trouver des mots d'opinion avec des orientations spécifiques au contexte. Ses méthodes dépendent des modèles syntaxiques ou des modèles qui se produisent ensemble avec une liste de graines de mots d'opinion pour trouver d'autres mots d'opinion dans un grand corpus. L'une de ces méthodes était représentée par Hatzivassiloglou et McKeown [27], ont commencé avec une liste d'adjectifs d'opinion, et les ont utilisés avec un ensemble de contraintes linguistiques pour identifier d'autres mots d'opinion

---

## CHAPITRE 2-LA SEMANTIQUE

---

adjectifs et leurs orientations. Les contraintes sont pour les connecteurs comme ET, OU, MAIS, SOIT OU SOIT. . ... . la conjonction ET par exemple dit que les adjectifs conjoints ont généralement la même orientation. Cette idée est appelée cohérence du sentiment, ce qui n'est pas toujours cohérent dans la pratique.

Il y a aussi des expressions adversatives telles que mais, cependant, qui sont indiqués comme des changements d'opinion. Afin de déterminer si deux adjectifs conjoints ont des orientations identiques ou différentes, l'apprentissage est appliqué à un grand corpus. Ensuite, les liens entre adjectifs forment un graphe et un grappage est effectué sur le graphe pour produire deux ensembles de mots: positif et négatif.

L'approche basée sur le corpus est effectuée utilisant une approche statistique ou une approche sémantique comme illustré dans les sous-sections suivantes :

- **Approche statistique:**

Trouver des modèles de cooccurrence ou Les mots d'opinion sur les semences peuvent être obtenus en utilisant des techniques statistiques. Cela pourrait être fait en dérivant des polarités postérieures en utilisant la co-occurrence d'adjectifs dans un corpus, comme proposé par Fahrni et Klenner . Il est possible d'utiliser l'ensemble des indexés documents sur le web comme corpus pour la construction du dictionnaire. Cela permet de s'affranchir du problème de l'indisponibilité de certains mots si le corpus utilisé n'est pas assez important

- **Approche sémantique :**

Donne un sentiment valeurs directement et repose sur des principes différents pour l'informatique la similitude entre les mots. Ce principe donne des résultats similaires valeurs de sentiment pour fermer sémantiquement les mots. WordNet pour exemple fournit différents types de relations sémantiques entre les mots utilisés pour calculer les polarités du sentiment. WordNet pourrait être utilisé aussi pour obtenir une liste de mots de sentiments itérativement et élargir l'ensemble initial avec des synonymes et des antonymes puis déterminer la polarité du sentiment pour un mot inconnu par le nombre relatif de synonymes positifs et négatifs de ce mot.

### 2.5.3 Sac de mot (bag of word)

Sac de mot est un algorithme qui compte combien de fois un mot apparaît dans un document. Ces chefs d'accusation mot nous permettent de comparer les documents et jauger leurs similitudes pour des applications comme la recherche, la classification des documents et modélisation de la rubrique.

---

## CHAPITRE 2-LA SEMANTIQUE

---

Avant la liste des mots avec leurs comptes de mot par document. Dans la table où les mots et les documents effectivement devient vecteurs est stockée, chaque ligne est un mot, chaque colonne est un document et chaque cellule est un nombre de mots. Chacun des documents du corpus est représenté par des colonnes de longueur égale. Ceux sont des vecteurs de wordcount, une sortie dépouillée du contexte.

### **2.5.4 L'Analyse Sémantique Latente (LSA)**

La LSA fut brevetée en 1988 et publiée en 1990. est une approche statistique qui est utilisée pour analyser les relations entre un ensemble de documents et les termes mentionnés dans ces documents afin de produire un ensemble de modèles significatifs liés aux documents et aux termes. LSA se base sur svm ou machine support vectore qui est un classificateur discriminant formellement défini par un hyperplan de séparation. En d'autres termes, étant donné les données d'apprentissage étiquetées (apprentissage supervisé), l'algorithme génère un hyperplan optimal qui catégorise les nouveaux exemples. Dans un espace dimensionnel, cet hyperplan est une ligne qui divise un plan en deux parties où, dans chaque classe, se trouve l'un ou l'autre côté.

### **2.6 Conclusion**

La sémantique, reste une tâche un peu complexe à implémenter. Cependant, l'évolution technologique au niveau d'architecture des ordinateurs avec l'apparition processeur multicœur nous oblige à envisager des solutions pour rendre cette tâche intellectuelle accessible et exécutée par les ordinateurs. Dans le prochain chapitre, nous allons présenter notre travail qui sert essentiellement à présenter des connaissances dans un domaine et d'essayer d'envisager les relations sémantiques entre ses concepts par le biais d'une ontologie.

## CHAPITRE 3

# IMPLEMENTATION

### 3.1 Introduction

La sémantique dans son sens large est difficile à manipuler par les ordinateurs, acquérir le sens est un processus complexe qui implique plusieurs étapes de raisonnement chez les êtres humains. Pour simuler le processus, on a besoin d'un modèle qui représente les connaissances et les relations qui les relient. D'où vient le rôle des ontologies qui sont devenues un excellent moyen pour distinguer les concepts, les relations et les méta-données supplémentaires.

Notre travail consiste à essayer de trouver la relation sémantique dans un groupe de mots. Pour la réalisation, on a opté pour utiliser la langue anglaise car il existe plusieurs bibliothèques, corpus et dictionnaires qui offrent des outils de traitement et qui facilitent l'implémentation. Le plus connu est WordNet qui est un dictionnaire de la langue anglaise dont le but est de classer et mettre en relation de diverses le contenu sémantique et lexical de la langue.

### 3.2 Construction d'une ontologie

Les ontologies représentent un modèle conceptuel sur lequel il est possible de développer des systèmes de connaissances qui soient partageables et réutilisables permettant ainsi l'interopérabilité. La définition de la démarche de construction d'ontologie reste un problème ouvert qui continue à susciter beaucoup d'intérêt dans la communauté du web sémantique. La majorité des approches existantes commencent la construction des ontologies par l'identification, puis l'organisation et la structuration des concepts et des relations à représenter.

L'objectif de notre travail est de trouver une relation sémantique pour les langues naturelles. Cependant, les langues naturelles sont trop riches en vocabulaire, on a choisi de cerner le domaine d'analyse sémantique. Nous avons décidé de nous focaliser sur le domaine de la science naturelle où on trouve un vocabulaire riche mais un peu réduit, malheureusement nous n'avons pas pu trouver une ontologie prête à être utilisée. Donc nous avons créé notre propre ontologie du domaine avec un nombre de concepts et individus limités mais qui peuvent être enrichis facilement depuis l'éditeur protégé.

Pour construire une ontologie, il faut passer par quelques techniques. Dans notre travail, on utilise la méthode Methontology qui s'inscrit dans le cadre d'un processus de gestion de projet complet, ce processus est composé des étapes suivantes : Spécification, Conceptualisation,

## CHAPITRE 3- IMPLIMENTATION

---

Formalisation , Intégration, Implémentation et Maintenance. Nous avons défini chaque étapes dans le chapitre 1 (les ontologies).

### 3.3 Ontologie être\_vivant utilisée:

Pour faciliter notre tâche et puisque notre objectif est de trouver une relation sémantique pour les langages naturelles,on cherche le mot étrangere dans une ontologie et le comparer avec d'autres mots on a opté pour construire une ontologie.

- **Domaine de connaissance**

Le domaine que va couvrir l'ontologie est d'analyse sémantique.

- **Objectif**

L'objectif de notre travail est de trouver une relation sémantique pour les langages naturelles,on cherche le mot étrangere dans une ontologie et le comparer avec d'autres mots

- **Utilisateurs**

Dans notre cas, les utilisateurs de l'ontologie sont des gens qui sont intéressé par les langues naturelles et linguistique

- **Les sources d'informations :**

Les sources d'informations sur lesquelles nous nous sommes basés pour arriver à la construction de l'ontologie d'annotation sont les experts du domaine (science naturelle) et encyclopedie.

**Portée de l'ontologie:**Science naturelle

**Figure 3.1** Document de spécifications de l'ontologie

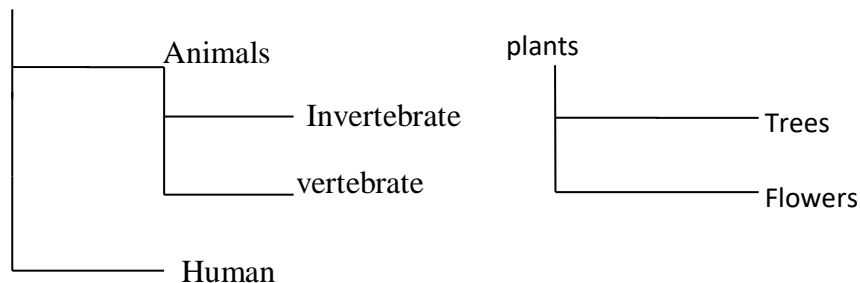
## CHAPITRE 3- IMPLIMENTATION

Nom de terme	Description
Living_beings	Sont des organisme dote de vies
Animals	un être vivanthétérotrophe
Human	Etre vivant est different par son intelligence
Plants	Etre vivant
Food	toute substance nutritive que les gens ou les animaux mangent ou boivent, ou que les plantes absorbent
Vertebrate	un animal d'un grand groupe se distinguant par la possession d'une colonne vertébrale ou d'une colonne vertébrale
Invertebrate	un animal dépourvu d'épine dorsale
Color	Les couleurs
Flower	Les fleurs
Trees	Les arbres
Classification	est un processus lié à la catégorisation, processus dans lequel les idées et les objets sont reconnus, différenciés et compris.
Environment	le monde naturel, dans son ensemble ou dans une zone géographique particulière
Carnivorous	Les animaux qui mange la viande
Herbevorous	Les animaux qui mange les herbs

**Tableau 3.1:** Glossaire de termes

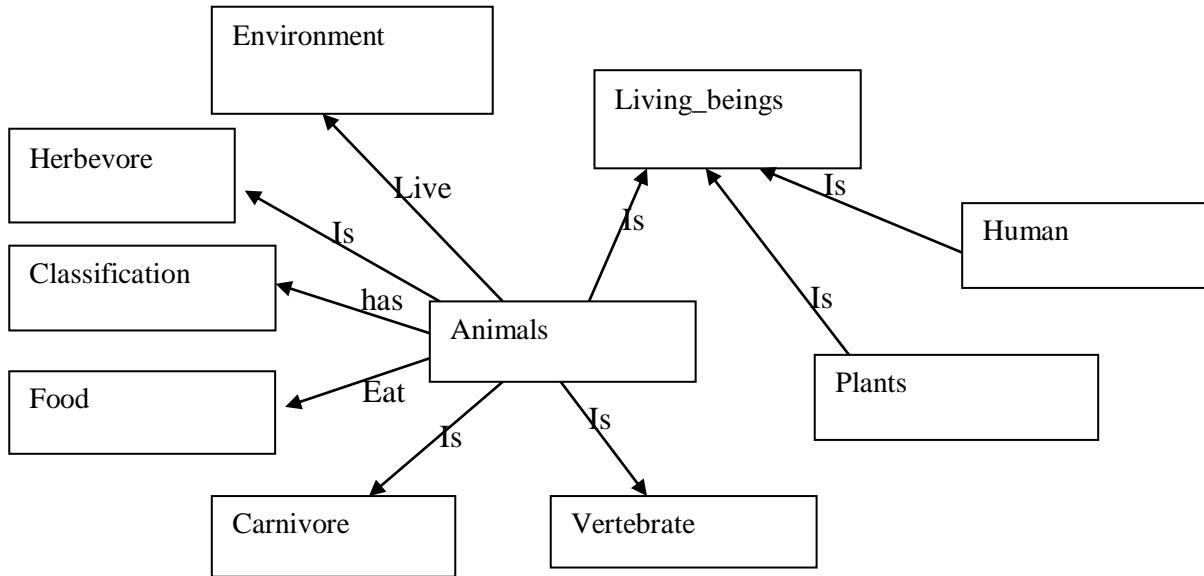
- Une partition d'un concept C est l'ensemble de sous classes de C qui ne partagent pas des instances en commun et couvrent C

Living-Being



**Figure 3.2:** Hiérarchies de concepts

## CHAPITRE 3- IMPLIMENTATION



**Figure 3.3:** Diagramme de relations binaires

Nom de concept	Attribut	Les relations
Animals		Is
Vertebrate		Is
Invertebrate		Is
Human		
Plants		
Color		
Flower		Is
Trees		Is

**Tableau 3.2 :**Dictionnaire de concepts

Nom de la relation	Concept source	Concept cible	Card source	Card cible	Relation inverse
Is	Animals	Vertebrate	1..1	1..n	Isnot
Eat	Vertebrate	Food	1..1	1..n	
Live	Vertebrate	Classification	1..1	1..n	

**Tableau 3.3:**Table des relations

## CHAPITRE 3- IMPLIMENTATION

Concept	Instance	Attribut
classification	Mammals	Nom
classification	Brids	Nom
Classification	Fish	Nom
Environment	Desert	Nom
Environment	Sea	Nom
Food	Herbs	Nom
Invertebrate	Molluscs	Nom
Invertebrate	Worms	Nom
Carnivore	Crow	Nom
(...)	(...)	(...)

**Tableau 3.4:** La table des instances de concepts

$\begin{aligned} \text{Living\_beings} &\subseteq \text{Thing} \\ \text{animals} &\subseteq \text{living\_beings} \\ \text{human} &\subseteq \text{living\_beings} \cap (\neg \text{animals.string}) \\ \text{plants} &\subseteq \text{living\_beings} \cap (\neg \text{invertebrate.string}) \cap (\neg \text{human.string}) \\ \text{invertebrate} &\subseteq \text{animals} \cap (\neg \text{invertebrate.string}) \end{aligned}$
--

**Tableau 3.5:** Axiomes de classes et de rôles

$\begin{aligned} \text{animals}(\text{cat}) \\ \text{flower}(\text{jasmine\_rose}) \end{aligned}$
---

**Tableau 3.6:** Assertions d'individus

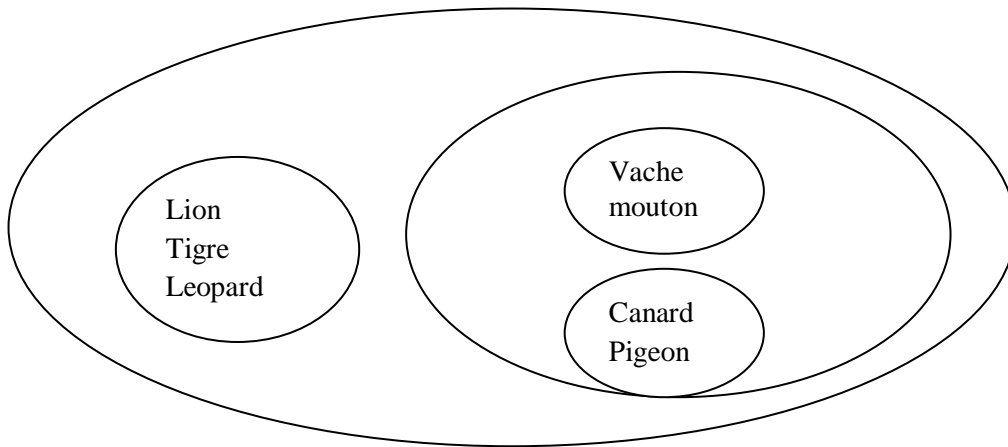
### 3.4 Description de notre approche

Trouver le sens d'une phrase est une tâche intellectuelle très compliquée, simuler cette tâche par les ordinateurs est difficile à réaliser. Pour cela, on a simplifié la tâche de notre application en essayant de trouver la relation sémantique entre un groupe de mots et trouver si un mot ou plus sont étrangères à l'ensemble (odd word en anglais).

## CHAPITRE 3- IMPLIMENTATION

---

Dans notre application l'utilisateur final est invité à donner des mots (dans le domaine de la science naturelle spécifiquement) puis le système essaye de trouver la relation sémantique en correspondance avec notre ontologie être\_vivant en simulant le processus de raisonnement des humains le plus possible. Sachant que les humains et d'après la psychologie cognitive tentent à regrouper les classes qui sont similaires :



**Figure 3.4:** Exemple de regroupement des classes

- L'illustration de cette phase est la suivante :

Ensemble de mots :

Lion –Desert-Chaise-Arbre-Poisson .

Mot étrangère :

Chaise .

Donc le processus de détection des relations sémantiques est comme dit :

- Comparer l'ensemble des mots donnés avec les concepts et les individus de l'ontologie. Si un mot ou plus n'existe pas dans notre ontologie on utilise WORDNET pour trouver un synonyme.
- Le synonyme est recherché dans l'ontologie. S'il est trouvé le mot est comparé par l'ensemble mot à mot .
- La comparaison est faite par l'interrogation de l'ontologie par le biais du SPARQL.

### 3.5 Implémentation

#### 3.5.1 Les outils de programmation

##### 3.5.1.1 Langage JAVA

Pour le choix de programmation de notre système nous avons opté pour le langage JAVA et cela pour de nombreuses raisons :

- C'est un langage orienté objet simple, qui réduit le risque des erreurs d'incohérence.
- Il permet d'accéder d'une manière simple aux fichiers et aux réseaux (notamment Internet).
- Il est caractérisé aussi par la réutilisation de son code ainsi que la simplicité de sa mise en œuvre.
- Il possède une riche bibliothèque de classes comprenant des fonctions diverses telles que les fonctions standards, le système de gestion de fichiers ainsi que beaucoup de fonctionnalités qui peuvent être utilisées pour développer des applications diverses.

##### 3.5.1.2 Jena API

Le framework Jena est une API java permettant de lire et de manipuler des ontologies décrites en document OWL, RDF (S), DAML+OIL et d'y appliquer certains mécanismes d'inférences. Jena est diffusé en code source libre. Les primitives pour lire et parcourir des ontologies OWL constituent l'outil le plus utile pour l'enrichissement. Nous allons esquisser les éléments les plus importants de ces primitives.

##### 3.5.1.3 Protégé :

Protégé 4.3 est un éditeur d'ontologies. Il permet de construire et d'éditer des ontologies sous divers formalismes RDF, DAML+OIL, OWL. De plus, Protégé est extensible en utilisant des plug-ins et des APIs java pour le développement d'applications particulières. L'interface de Protégé est assez simple, l'ensemble des fonctionnalités de l'éditeur étant regroupé en onglets.

Le premier onglet présente les classes de l'ontologie. Le deuxième onglet permet de créer des instances et de leur affecter des propriétés. L'onglet des requêtes permet d'effectuer des requêtes sur l'ontologie en cours d'édition.[25].

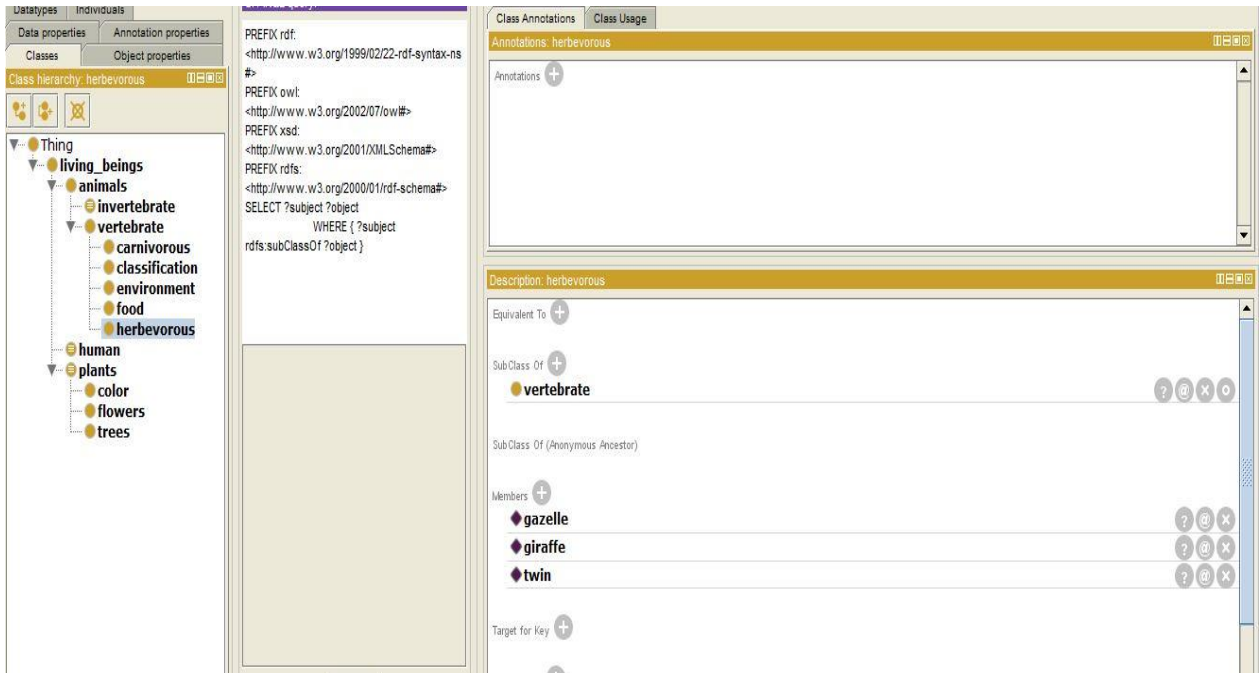


Figure 3.5 : l'ontologie être-vivant utilisée

### 3.5.1.4 SPARQL

Langage d'interrogation SPARQL (*Protocol and RDF QueryLanguage*) pour RDF Recommendation du W3C du 15 janvier 2008. SPARQL peut être utilisé pour exprimer des interrogations à travers diverses sources de données, que les données soient stockées nativement comme RDF ou vues comme du RDF via un logiciel médiateur (middleware). SPARQL est capable de rechercher des motifs de graphe (graph patterns) obligatoires et optionnels ainsi que leurs conjonctions et leurs disjonctions.

### 3.5.1.5 Wordnet

WordNet de version 2.0 qui est une base de données lexicographique. Le choix de WordNet est cause de diverses raisons :

- C'est la base la plus riche et la plus générale qui contient tous les domaines
- Il utilise la langue anglaise qui est la langue la plus utilisée dans le monde.

Il est vu comme une ontologie linguistique. Dans WordNet, un concept appelé un sens, est défini par un seul ensemble de *synonymes*, appelés *synset*. Les concepts dans le thesaurus

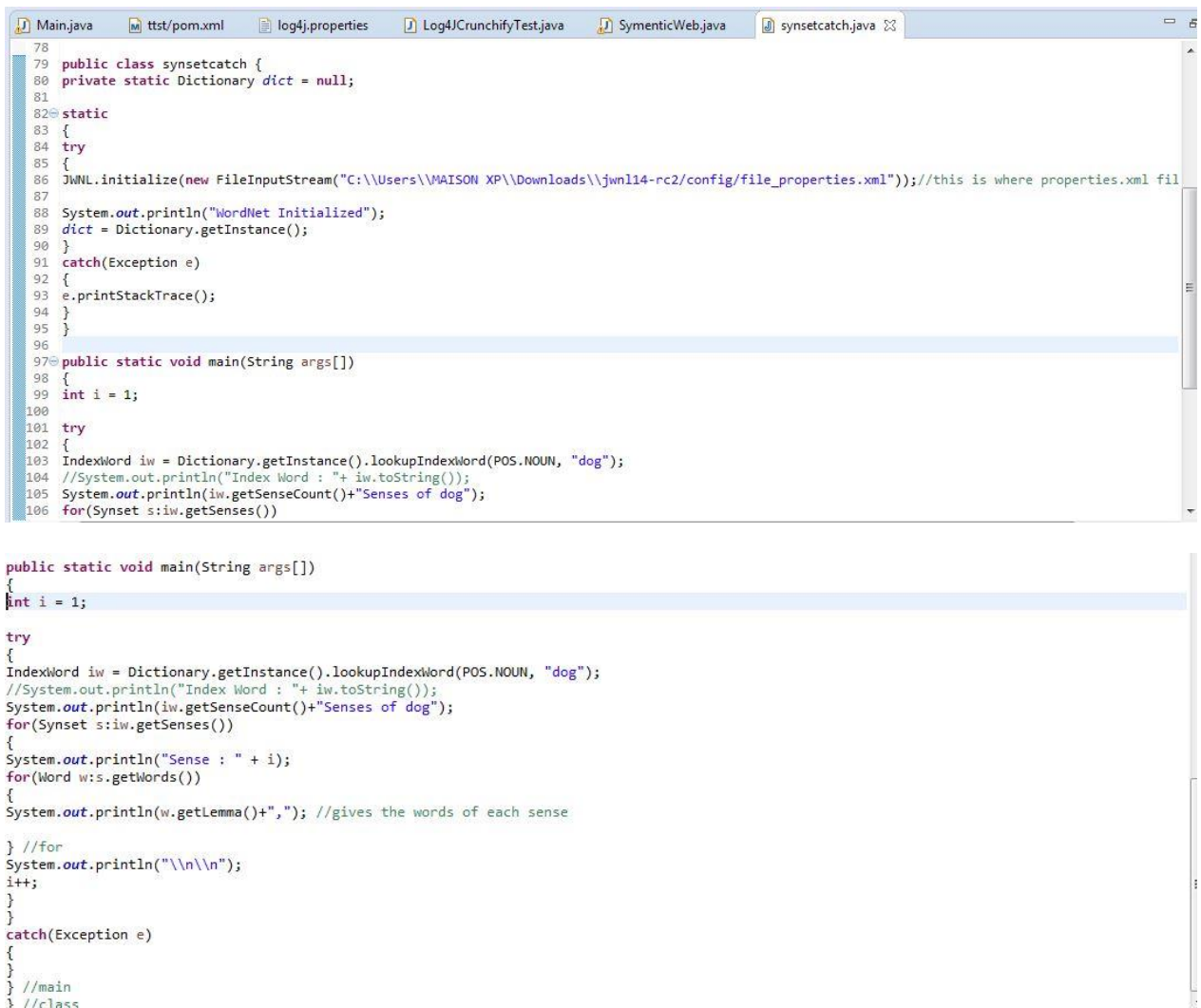
## CHAPITRE 3- IMPLIMENTATION

sont ambiguës (il est associé à plusieurs termes), car pour chaque terme, le numéro propre du sens est donné après le symbole .

On peut accéder à WordNet par le biais de son API, JWNL (un Java API pour accéder au dictionnaire apparenté de WordNet).

Nous avons utilisé le wordnet pour vérifier le sens de mot et le comparer avec l'ontologie .

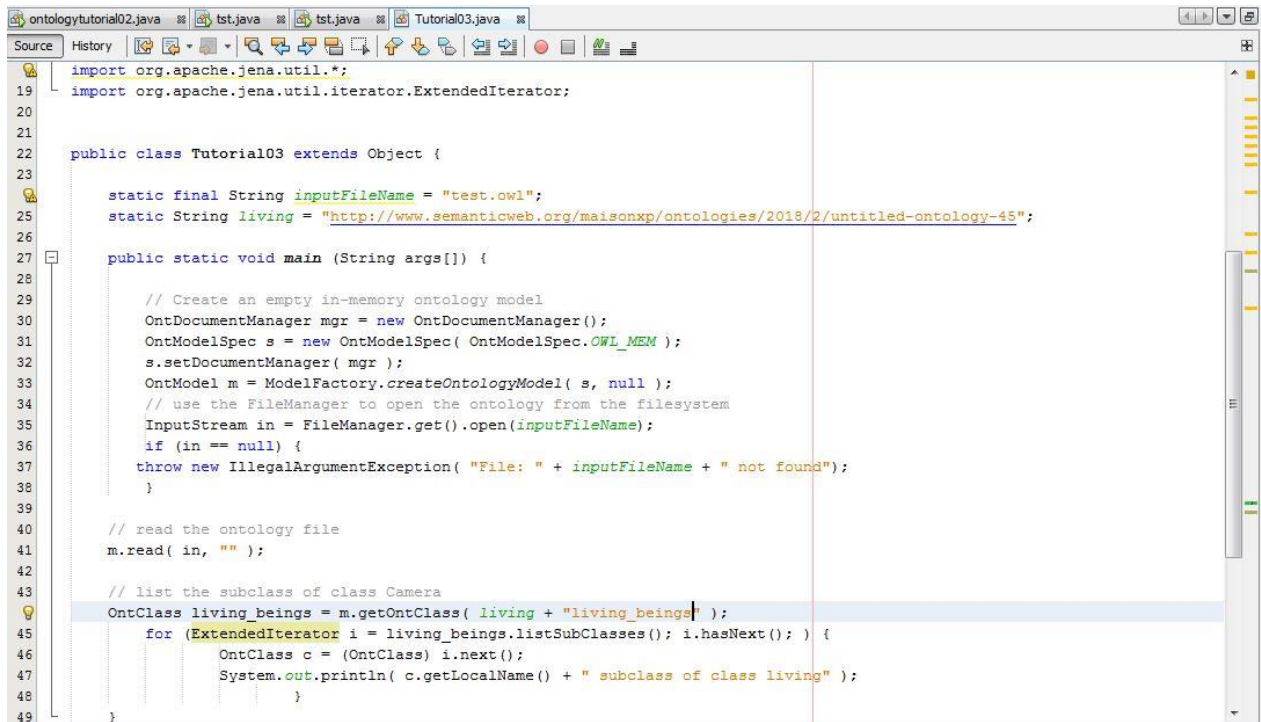
### 3.6 Fonctionnement de l'application



```
78
79 public class synsetcatch {
80 private static Dictionary dict = null;
81
82 static
83 {
84 try
85 {
86 JNWL.initialize(new FileInputStream("C:\\Users\\MAISON XP\\Downloads\\jwnl14-rc2/config/file_properties.xml")); //this is where properties.xml fil
87
88 System.out.println("WordNet Initialized");
89 dict = Dictionary.getInstance();
90 }
91 catch(Exception e)
92 {
93 e.printStackTrace();
94 }
95 }
96
97 public static void main(String args[])
98 {
99 int i = 1;
100
101 try
102 {
103 IndexWord iw = Dictionary.getInstance().lookupIndexWord(POS.NOUN, "dog");
104 //System.out.println("Index Word : "+ iw.toString());
105 System.out.println(iw.getSenseCount()+"Senses of dog");
106 for(Synset s:iw.getSenses())
107 {
108 System.out.println("Sense : " + i);
109 for(Word w:s.getWords())
110 {
111 System.out.println(w.getLemma()+","); //gives the words of each sense
112 }
113 } //for
114 System.out.println("\n\n");
115 i++;
116 }
117 catch(Exception e)
118 {
119 }
120 //main
121 } //class
```

Figure 3.6 : Synonyme de mots en jwnl

## CHAPITRE 3- IMPLIMENTATION



```
ontologytutorial02.java  tst.java  tst.java  Tutorial03.java
Source  History  [Icons]
19  import org.apache.jena.util.*;
20  import org.apache.jena.util.iterator.ExtendedIterator;
21
22  public class Tutorial03 extends Object {
23
24      static final String inputFileNames = "test.owl";
25      static String living = "http://www.semanticweb.org/maisonxp/ontologies/2018/2/untitled-ontology-45";
26
27      public static void main (String args[]) {
28
29          // Create an empty in-memory ontology model
30          OntDocumentManager mgr = new OntDocumentManager();
31          OntModelSpec s = new OntModelSpec( OntModelSpec.OWL_MEM );
32          s.setDocumentManager( mgr );
33          OntModel m = ModelFactory.createOntologyModel( s, null );
34          // use the FileManager to open the ontology from the filesystem
35          InputStream in = FileManager.get().open(inputFileNames);
36          if (in == null) {
37              throw new IllegalArgumentException( "File: " + inputFileNames + " not found");
38          }
39
40          // read the ontology file
41          m.read( in, "" );
42
43          // list the subclass of class Camera
44          OntClass living_beings = m.getOntClass( living + "living_beings" );
45          for (ExtendedIterator i = living_beings.listSubClasses(); i.hasNext(); ) {
46              OntClass c = (OntClass) i.next();
47              System.out.println( c.getLocalName() + " subclass of class living" );
48          }
49      }
```

Figure 3.7: implémentation requête SPARQL (subclass)

## CHAPITRE 3- IMPLIMENTATION

```
ontologytutorial02.java  tst.java  tst.java  Tutorial03.java
Source  History  [Icons]
20  *
21  * @author MAISON XP
22  */
23  public class tst {
24      static final String inputFileName = "test.owl";
25      static String living = "http://www.semanticweb.org/maisonxp/ontologies/2018/2/untitled-ontology-45"; /*get Expert
26
27      public static void main (String args[]) {
28          OntDocumentManager mgr = new OntDocumentManager();
29          OntModelSpec s = new OntModelSpec( OntModelSpec.OWL_MEM );
30          s.setDocumentManager( mgr );
31          OntModel m = ModelFactory.createOntologyModel( s, null );
32          InputStream in = FileManager.get().open(inputFileName);
33          if (in == null) {
34              throw new IllegalArgumentException( "File: " + inputFileName + " not found");
35          }
36          m.read( in, "" );
37          OntClass living_beings= m.getOntClass( living + "living_beings" );
38          //print out the name of the Expert class
39          System.out.println(living_beings.getLocalName());
40
41          //get instances of the Expert class
42          Iterator it = living_beings.listInstances();
43          //print out the instances of the Expert class
44          while(it.hasNext()){
45              Individual oi = (Individual)it.next();
46              System.out.println(oi.getLocalName());
47          }
48
49      }
```

Figure 3.8 : implémentation requette SPARQL (individuals)

Interface

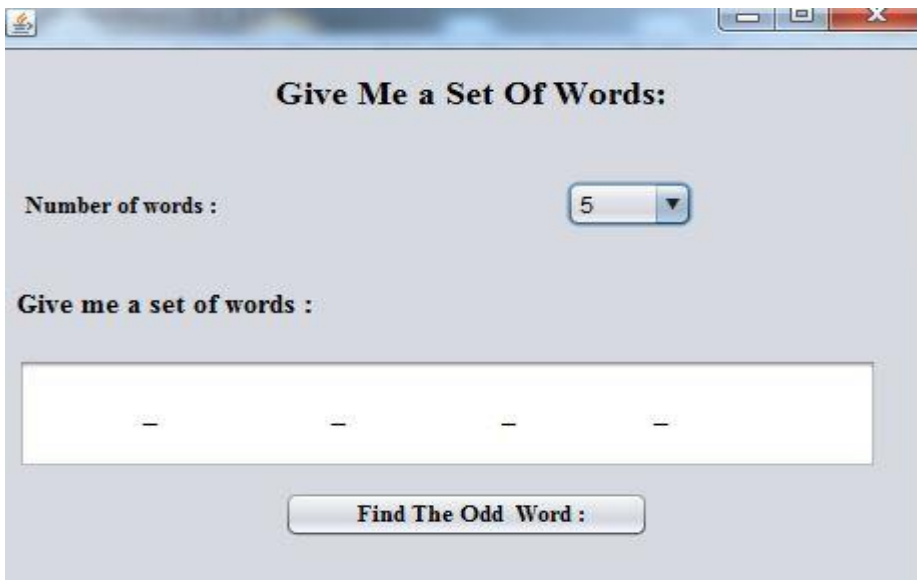


Figure 3.9 : l'interface de l'application (the odd word)

### **3.7 Conclusion**

Ce chapitre présente la description et la mise en œuvre des étapes d'implémentation pour notre approche de détection des relations sémantiques dans les langues naturelles. Puisque le vocabulaire est trop riche de termes on a choisi de réduire le domaine pour le présenter sous forme d'ontologie. Les relations sémantiques sont ensuite déduites à l'aide du langage SPARQL et le Framework Jena. Les résultats des requêtes sont ensuite manipulés pour avoir une relation sémantique entre un ensemble de terme pour isoler le terme étranger.

## Conclusion générale

Avec l'arrivée de web sémantique et ses promesses d'automatiser et d'unifier l'utilisation des données pour que les ordinateurs puissent les utiliser et les comprendre, l'emploi des ontologies pour présenter les connaissances est devenu une tendance. L'ontologie présente les connaissances ainsi les relations qu'elles les relie. La sémantique est exprimée d'une façon moins complète mais plus simple sous forme d'une représentation formelle.

L'acquisition du sens d'une langue naturelle devient une inspection et une recherche dans l'ontologie. A condition, que l'ontologie présente tous les concepts liés à cette langue. Cela est malheureusement difficile voire impossible, à cause de la richesse des langues naturelles en termes de vocabulaire et relations sémantiques.

De ce fait, dans notre travail nous proposons de cerner le domaine d'étude, on a choisi d'utiliser le domaine des sciences naturelles. Ce domaine est moins complexe, ses concepts sont connus et les relations sémantiques sont visibles et distinctes.

Donc, nous proposons une ontologie des êtres vivants, pour la création de cette ontologie nous avons adopté la méthode METHONTOLOGY pour la conceptualisation, afin d'atteindre un ensemble de représentations intermédiaires qui facilitent sa formalisation. En se basant sur cette formalisation, nous avons choisi le langage OWL, pour codifier l'ontologie formelle, et nous avons utilisé l'éditeur graphique PROTEGE-OWL, pour l'implémentation.

Pour tester notre approche de détection des relations sémantiques, nous avons proposé un système qui trouve le mot étranger dans un ensemble de mots. La réalisation est faite au biais du langage JAVA et la bibliothèque Jena. Les mots sont cherchés un par un dans l'ontologie cible, après un certain traitement, notre système distingue le mot étranger.

Notre contribution est essentiellement l'utilisation du dictionnaire WORDNET pour trouver un synonyme dans le cas où un mot ou plus n'existe pas dans notre ontologie. Une fois le synonyme est trouvé, ce dernier est comparé mot à mot par le langage d'interrogation d'ontologie SPARQL. Comme perspectives nous envisageons de pouvoir faire des requêtes sur plusieurs ontologies simultanément, aussi enrichir notre ontologie par d'autres concepts et de relations, utiliser d'autres techniques pour chercher des relations sémantiques entre les concepts et individus de l'ontologie.

---

---

## Références

- [1] Tom Gruber. **A translation approach to portable ontology specifications.** *Knowledge Acquisition*, 5(2):199–220, 1993.
- [2] J. Broekstra, M. Klein, S. Decker, D. Fensel, F. van Harmelen, and I. Horrocks. **Enabling Knowledge Representation on the Web by extending RDF Schema.** *In Proceedings of the tenth World Wide Web conference WWW'10*, Hong Kong, May 2001.
- [3] Djida BAHLOUL (Une approche hybride de gestion des connaissances basée sur les ontologies: *application aux incidents informatiques*) *thèse de doctorat* L'Institut National des Sciences Appliquées de Lyon.
- [4] Fernandez, M., Gomez-Pérez, A., Juristo, N. (1997) METHONTOLOGY: From Ontological Art Towards Ontological Engineering. Proceedings of the AAI-97 Spring. HDR Jean Charlet : <http://www-test.biomath.jussieu.fr/~jc/>.
- [5] Gruninger M., Fox. M., The logic of enterprise modeling. In Brown, J. & O'Sullivan, D., (Eds.), *Reengineering the Enterprise*, Chapman and Hall, pages 83–98, 1995.
- [6] Nathalie H., Josiane M. TtoO: une méthodologie de construction d'ontologie de domaine à partir d'un thésaurus et d'un corpus de référence, IRIT, Toulouse- France 2006.
- [7] Malaisé V, Méthodologie linguistique et terminologique pour la structuration d'ontologies différentielles à partir de corpus textuels , PhD thesis, Université Paris 7 - Denis Diderot, 2005.
- [8] Baneyx A., Construire une ontologie de pneumologie. Thèse de doctorat, Université Paris6 page 61 ; 85. Février 2007.
- [9] Xu Z., Cao X., Dong Y., and Su W., «Formal Approach and Automated Tool for Translating ER Schemata into OWL Ontologies», *Advances in Knowledge Discovery and Data Mining*, Springer-Verlag Berlin Heidelberg, 2004, pp. 464-475.
- [10] National Committee for Information Technology Standards, Technical Committee T2 (Information Interchange and Interpretation). **Draft proposed American national standard for Knowledge Interchange Format.** <http://logic.stanford.edu/kif/dpans.html>, 1998.
- Bibliographie Laboratoire LIRIS 2003 – 2004*

## REFERENCES

---

- [11] A. Farquhar, R. Fikes, and J. Rice. **The Ontolingua Server**: a tool for collaborative ontology construction. In *Proceedings of the 10th Knowledge Acquisition for Knowledge-Based Systems Workshop (KAW'96)*, 1996.
- [12] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, Peter Patel-Schneider, editors. **The Description Logic Handbook**. *Cambridge University Press*, 2003;
- [13] R. J. Brachman and J. G. Schmolze. **An overview of the KL-ONE knowledgeRepresentation System**. *Cognitive Science*, 9(2):171–216, April 1985.
- [14] Dan Brickley and R. V. Guha, Editors. **RDF Vocabulary Description Language 1.0:RDF Schema**, *W3C Recommendation*, 10 February 2004.<http://www.w3.org/TR/rdf-schema/>
- [15] I. Horrocks, D. Fensel, J. Broekstra, S. Decker, M. Erdmann, Goble, F. van Harmelen, M. Klein, S. Staab, R. Studer, and E. Motta. **OIL: The Ontology Inference Layer**. *Technical Report IR-479*, Vrije Universiteit Amsterdam, Faculty of Sciences, Sept. 2000.<http://www.ontoknowledge.org/oil/>.
- [16] Deborah L. McGuinness and Frank van Harmelen, Editors. **OWL Web Ontology Language Overview**, *W3C Recommendation*, 10 February 2004.<http://www.w3.org/TR/owl-features/>.
- [17] Tom Gruber, article « Ontology » dans l'Encyclopedia of Data Base Systems, Springer-Verlag, 2009.
- [17]: Laublet, P., Reynaud, C., Charlet, J. Sur quelques aspects du Web sémantique. Journées Scientifiques Web sémantique, Paris 2002.
- [18]: Berners-Lee, T., Hendler, J., Lasilla, O. "The semantic web". Scientific American. 2001.
- [19] O. Etzioni. The world wide web: Quagmire or goldmine. *Communications of the ACM*, 39(11):65–68, 1996
- [20] Hacène Cherfi, Etude et réalisation d'un système d'extraction de connaissances à partir de textes. Thèse de doctorat (informatique) Université Henri Pointcaré- Nancy 1. 2004.
- [21] Sebastiani F., « Machine learning in automated text categorization », *ACM Computing Surveys*, 34(1), 2002, p. 1-47.
- [22] M. Grobelnik, D. Mladenic et N. Milic-Frayling. Introduction. Dans Proc. Of the workshop on Text Mining, Boston, 2000. In Conjunction with the 6<sup>th</sup> Int 1 Conf. on knowledge Discovery and Data Mining kdd'00.
- [23] Meena Rambocas, Jo. Gama. Marketing research. The role of sentiment analysis, FEP Economic and Management. 2013

## REFERENCES

---

- [24] Bing, Liu, Opinion, Sentiment and Emotion in Text.. Cambridge University Press. 2015
- [25] Saadani, L., Bertrand, Gastaldy, S., (2000). Cartes conceptuelles et thésaurus : essai de comparaison entre deux modèles de représentation issus de différentes traditions disciplinaires. In: Kublik, A., ed. Les dimensions d'une science de l'information globale ACSICAIS [http://www.caisacasi.ca/2007call\\_fr.htm](http://www.caisacasi.ca/2007call_fr.htm)
- [26] Walaa Medhat, Ahmed Hassan, Hoda Korashy Sentiment analysis algorithms and applications: A survey Received 8 September 2013; revised 8 April 2014; accepted 19 April 2014 Available online 27 May 2014.
- [27] Hatzivassiloglou V, McKeown K. Predicting the semantic orientation of adjectives. In: Proceedings of annual meeting of the Association for Computational Linguistics (ACL'97); 1997.

## الملخص:

أصبح استخدام الأنطولوجيات لتمثيل المعرفة ضروريًا بسبب إمكانية التعبير عن العلاقات الدلالية في مجال المعالجة الآلية للغات الحية. الهدف من هذا العمل هو اكتشاف العلاقات الدلالية للغات الطبيعية. يستند النظام أساسًا إلى التمثيل للغات الطبيعية، ثم يتم التمييز بين العلاقات الدلالية عن طريق البحث في الأنطولوجيا وهذا من خلال المنصة Jena ولغة الاستجواب SPARQL. بما أن اللغة الطبيعية هي موضوع هذه الدراسة، فإن استخدام قاموس ضروري، استخدمنا WORDNET لإزالة الغموض.

الكلمات المفتاحية: OWL، أنطولوجيا، علم الدلالة، Jena، SPARQL، WORDNET

## Résumé :

L'utilisation des ontologies pour la représentation des connaissances est devenue une nécessité, la possibilité d'exprimer les relations sémantiques sous forme formelles les rendre plus importantes surtout dans le domaine du traitement automatique des langues naturelles. L'objectif de ce travail est de détecter les relations sémantiques des langues naturelles. La réalisation du système repose essentiellement sur la représentation formelle des langues naturelles puis la distinction des relations sémantique est réalisée par la manipulation et l'interrogation de l'ontologie par la plateforme Jena et le langage de requêtes SPARQL. Puisque le langage naturel est l'objet de cette étude, l'emploi d'un dictionnaire est indispensable, on a choisi d'utiliser le WORDNET pour enlever les ambiguïtés.

**Mot clés :** ontologie, OWL, sémantique, Jena, SPARQL, WORDNET

## Abstract :

the use of ontologies for the representation of knowledge has become a necessity, the possibility of expressing semantic relations in formal form making them more important especially in the field of automatic processing of natural languages. The objective of this work is to detect the semantic relations of natural languages. The realization of the system is based essentially on the formal representation of natural languages, then the distinction between semantic relations is made by the manipulation and interrogation of the ontology by the Jena platform and the SPARQL query language. Since natural language is the object of this study, the use of a dictionary is essential, we chose to use the WORDNET to remove ambiguities.

**Key words:** ontology, OWL, semantics, Jena, SPARQL, WORDNET