

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Université de Djilali BOUNAAMA Khemis Miliana



**Faculté des Sciences et de la Technologie**  
**Département de Mathématiques et d'Informatique**

Mémoire Présenté

Pour l'obtention du diplôme de

**Master** en Informatique

**Option** : « Génie Logiciel et systèmes distribués »

Titre :

**Alignement d'ontologie Biomédicale**

**Encadré par :**

**Réalisé par :**

**Mr. ILYES MOHAMMED BOUZIANE**

- DAHMANI Naziha

- ELMOKADDEM Karima

Soutenu publiquement le : 21 /09/2019

Année Universitaire 2019-2020

## *Remerciement*

*Nous exprimons notre profonde gratitude à Allah en tout lieu.*

*Nous remercions vivement notre encadrant Mr. Mohammed Bouziane Ilyes, d'avoir accepté de diriger ce travail et pour son aide précieux pour le temps qu'il a pris pour échanger et pour nous corriger*

*Ainsi qu'à tous les membres de jury Mr. F. Haniche, Mme. F. Boudali et Mr. A. Khalfi, qui nous ont honorés par leur présence, sans oublier tous les enseignants de notre cursus d'étude.*

*Nous tenons à remercier également nos familles pour leurs soutiens et encouragements, ainsi que toute autre personne qui a contribué de près ou de loin au bon déroulement de ce projet.*

*Merci*

## Dédicace

*Je dédie cet humble travail avec grand amour,  
sincérité et fierté :*

*Pour ces deux grands cœurs qui m'entourent  
toujours avec tendresse et de bonté, A ma très chers  
mère fadila et A mon très cher père djalloul, pour  
tous les sacrifices qu'ils ont fait à mon égard.  
Que ceci leur soit une récompense et un témoignage  
de ma profonde gratitude.*

*A mes sœurs Saïda, Zehor, souhaite tout le bonheur  
et le succès dans la vie.*

*Pour tous les membres de ma famille.*

*Et toutes les personnes que j'aime.*

*A ma chère binôme Karima avec qui j'ai partagé ce  
travail avec ses moments agréables et désagréables.*

**NAZIHA**

## *Dédicace*

*Je dédie ce travail*

*À mon père Ali*

*L'épaule solide, l'œil attentif compréhensif et la personne la plus digne de mon estime et de mon respect.*

*Aucune dédicace ne saurait exprimer mes sentiments, que Dieu te préserve et te procure santé et longue vie.*

*À ma mère Fatma*

*Tu m'as donné la vie, la tendresse et le courage pour réussir, tous ce que je peux t'offrir ne pourra exprimer l'amour et la reconnaissance que je porte.*

*En témoignage, je t'offre ce modeste travail pour te remercier pour tes sacrifices pour l'affection dont tu m'as toujours entourée.*

*À mon mari Oussama*

*Pour ta compréhension, ta confiance, ta patience et ta tendresse, tu m'as toujours soutenu et réconforté. Tu es tu resteras toujours ma source d'encouragement tu as supporté mes caprices pendant certaines périodes de ce parcours.*

*À mes frères*

*À mes sœurs*

*À ma belle famille*

*À mes chères amies*

*Karima*

## *Table des matières*

Liste des tableaux.....	III
Liste des figures .....	IV
Résumé .....	V
Introduction Générale .....	VII
<b>Chapitre1 : Généralité sur les ontologies</b>	
I.1. Introduction.....	4
I.2. Naissance de l'ontologie .....	4
I.3. La Notion d'ontologie .....	4
I.4. Rôle des ontologies .....	5
I.5. Types d'ontologies .....	6
I.5.1. Ontologie de domaine .....	6
I.5.2. Les ontologies de niveau supérieur ou de haut niveau .....	6
I.5.3. Ontologie générique .....	6
I.5.4. Ontologie de tâche .....	7
I.6. Composantes d'une ontologie .....	7
I.6.1. les concepts .....	7
I.6.2. Les relations .....	7
I.6.3. Les fonctions.....	8
I.6.4. Les axiomes .....	8

I.6.5. Les instances .....	8
<b>I.7. Domaine d'application .....</b>	<b>8</b>
I.7.1. Ontologies dans les systèmes d'informations .....	8
I.7.2. Ontologies dans le Web sémantique .....	9
<b>I.8. Classification des ontologies .....</b>	<b>9</b>
I.8.1. Classification selon l'objet de conceptualisation.....	10
I.8.2. Classification selon le niveau de granularité .....	10
I.8.3. Classification selon le niveau de formalisme de représentation .....	10
I.8.4. Classification selon le poids de l'ontologie .....	11
<b>I.9. L'ingénierie ontologique .....</b>	<b>11</b>
I.9.1. Processus de construction d'une ontologie .....	11
I.9.2. Méthodologies de construction d'ontologies .....	12
I.9.2.1. La méthodologie d'Uschold et King:.....	12
I.9.2.2. La méthodologie de Gruninger et Fox .....	13
I.9.2.3. La méthodologie Methontology.....	13
<b>I.9.3. Cycle de vie d'une ontologie.....</b>	<b>14</b>
I.9.3.1. Conception et évolution .....	15
I.9.3.2. Diffusion .....	16
I.9.3.3. Utilisation.....	16
I.9.3.4. Gestion .....	16
<b>I.10. Langage de représentation et de manipulation d'ontologie.....</b>	<b>17</b>
I.10.1. Le langage Rdf.....	17
I.10.2. Le langage RDFS (Resource Description Framework Schéma).....	17
I.10.3. Le langage Owl .....	17

<b>I.11. Conclusion .....</b>	<b>18</b>
-------------------------------	-----------

## **Chapitre II : Alignement d'ontologie**

<b>II.1. Introduction .....</b>	<b>21</b>
---------------------------------	-----------

<b>II.2. La nécessité de l'alignement des ontologies :.....</b>	<b>22</b>
-----------------------------------------------------------------	-----------

II.2.1. Problème issu de l'hétérogénéité .....	23
------------------------------------------------	----

II.2.1.1. L'hétérogénéité syntaxique .....	23
--------------------------------------------	----

II.2.1.2. L'hétérogénéité terminologique .....	24
------------------------------------------------	----

II.2.1.3. L'hétérogénéité conceptuelle .....	24
----------------------------------------------	----

II.2.1.4. L'hétérogénéité sémantique .....	25
--------------------------------------------	----

<b>II.3. Définition de l'alignement d'ontologies .....</b>	<b>26</b>
------------------------------------------------------------	-----------

II.3.1 le processus d'alignement .....	26
----------------------------------------	----

<b>II.4. Quelques domaine d'application de l'alignement d'ontologies .....</b>	<b>27</b>
--------------------------------------------------------------------------------	-----------

II.4.1. Ingénierie d'ontologie .....	27
--------------------------------------	----

II.4.2. Le web sémantique .....	27
---------------------------------	----

II.4.3. Le Web Biomédical .....	28
---------------------------------	----

<b>II.5. Les Méthodes d'alignement .....</b>	<b>29</b>
----------------------------------------------	-----------

II.5.1. La relation entre l'alignement et la similarité .....	29
---------------------------------------------------------------	----

II.5.2. La similarité .....	29
-----------------------------	----

<b>II.6. Les techniques d'alignement d'ontologie .....</b>	<b>31</b>
------------------------------------------------------------	-----------

II.6.1. Les méthodes terminologiques .....	31
--------------------------------------------	----

II.6.1.1. Les méthodes syntaxiques .....	31
------------------------------------------	----

II.6.1.2. Les méthodes linguistiques .....	32
--------------------------------------------	----

<b>II.6.2. Les méthodes structurelles :.....</b>	<b>33</b>
--------------------------------------------------	-----------

II.6.2.1. Les méthodes structurelles internes .....	34
-----------------------------------------------------	----

II.6.2.2. Les méthodes structurelles externes .....	34
-----------------------------------------------------	----

II.6.3. Les méthodes extensionnelles .....	35
II.6.4. Les méthodes Sémantiques .....	35
II.6.4.1. Les techniques basées sur les ontologies externes .....	36
II.6.4.2. Les techniques déductives .....	36
II.6.5. Les méthodes d'alignement combinées .....	36
<b>II.7. Approches d'alignement d'ontologies .....</b>	<b>38</b>
<b>II.8. Etude comparative des méthodes d'alignement .....</b>	<b>42</b>
II.8.1. Comparaison des caractéristiques externes .....	44
II.8.2. Comparaison des techniques de base .....	47
II.8.3. Comparaison des stratégies d'alignement .....	48
<b>II.9. Conclusion .....</b>	<b>49</b>
 <b>Chapitre III :Algorithme d'alignement</b>	
<b>III.1. Introduction .....</b>	<b>52</b>
<b>III.2. Méthode d'alignement.....</b>	<b>52</b>
III.2.1. L'UMLS et ontologie.....	53
III.2.2. Formats en entrée et en sortie.....	54
III.2.3. Le problème de calcul des similarités .....	54
<b>III.3. La méthode utilisée pour le calcul de la distance .....</b>	<b>55</b>
<b>III.3.1. Description de la méthode levenshtein pour calculer la similarité .....</b>	<b>56</b>
<b>III.4. Conclusion .....</b>	<b>57</b>
 <b>Chapitre IV :Réalisation et expérimentations</b>	
<b>IV.1. Introduction .....</b>	<b>59</b>
<b>IV.2. Environnement de réalisation .....</b>	<b>59</b>



<b>IV.2.1. présentation des outils de programmation .....</b>	<b>59</b>
<b>IV.2.2. présentation de l'interface du système .....</b>	<b>60</b>
<b>IV.3. Les ontologies de tests .....</b>	<b>62</b>
<b>IV.4. Résultats expérimentaux.....</b>	<b>62</b>
<b>IV.5. Conclusion .....</b>	<b>62</b>
<b>Conclusion Générale .....</b>	<b>64</b>
<b>Bibliographie .....</b>	<b>65</b>

## Liste des abréviations

**Asmov:** Automated Semantic Mapping of Ontology with Validation

**DS :** Distance.

**IA :** Intelligence Artificielle

**QOM:** Quick Ontology Matching

**NOM:** Naive Ontology Matching.

**OWL :** Ontology Web Language.

**Rdf :** Ressource Description Framework.

**Rdfs:** Ressource Description Framework/Schéma

**ROMIE:** Resource based Ontology Mapping within and Interactive and Extensible environment,

**SAT :** Propositional Satisfiability

**SBC :** Systèmes à base de connaissance

**URI :** Uniform resource identifié.

**UMLS:** Unified Medical Language System

**XML :** EXtensible Markup Language.

## **Liste des tableaux**

<b>Tableau 1: Quelques techniques et méthodes utilisé par les approches d'alignement d'ontologie.....</b>	<b>43</b>
<b>Tableau 2: Caractéristiques externes des principales méthodes d'alignement.....</b>	<b>45</b>
<b>Tableau 3 : Comparaison des techniques de composition.....</b>	<b>46</b>
<b>Tableau 4: Comparaison des techniques de combinaison.....</b>	<b>48</b>
<b>Tableau 5: Comparaison des techniques de combinaison et de sélection des principales méthodes d'alignement .....</b>	<b>49</b>

# Liste des figures

## Chapitre I

Figure 1: Le triangle sémantique.....	7
Figure 2: Processus de construction d'ontologie.....	12
Figure 3: La méthodologie de d'Uschold et King, (Entreprise).....	13
Figure 4: Les composantes de Methontology .....	14
Figure 5: Cycle de développement d'ontologie.....	15
Figure 6: Relation entre Unicode, XML, RDF, et OWL .....	18

## Chapitre II

Figure 1:Exemple de Deux Ontologies Hétérogènes.....	23
Figure 2:Etablissement des Correspondances Sémantiques.....	23
Figure 3:Les trois dimensions de l'hétérogénéité au niveau conceptuel.....	25
Figure 4: Schéma résumant le processus d'alignement.....	27
Figure 5: Composition d'un service web.....	28
Figure 6: Le graphe RDF et le graphe RDF biparti d'une ontologie OA ...	43

## Chapitre III

Figure 1: Méthode d'alignement.....	53
-------------------------------------	----

## Chapitre IV

Figure 1: L'interface principale de l'application .....	60
Figure 2: chargement d'une ontologie.....	61
Figure 3: Résultats des correspondances obtenus par l'algorithme proposé .....	62

## **Résumé:**

Le web sémantique a un but de rendre le contenu des documents web accessible afin d'assurer un échange d'information et une interopérabilité entre les différents systèmes ; Les ontologies sont proposées comme élément central du web sémantique.

L'hétérogénéité entre les connaissances décrites dans les ontologies dans le web ont fait surgir plusieurs questions de recherche intéressantes telles que l'intégration, l'évolution et l'alignement des ontologies. Donc notre projet était de développer un outil informatique qui permet d'aligner deux ontologies.

Pour réaliser l'application; nous avons exploité la base de données lexical « Wordnet » et « UMLS » pour découvrir si le concept concerné possède au moins un synonyme dans cette dernière, après nous avons utilisé la méthode Levenshtein qui calcule la distance entre les entités à aligner, qui est basée sur les chaînes de caractères, Puis trouver les valeurs des similarités. A partir de ces valeurs de similarités, les correspondances entre les concepts de deux ontologies sont identifiées, donc l'alignement est identifié.

On a testé notre algorithme avec deux ontologies de domaine biomédicale, l'un sur la maladie asthme et l'autre sur la maladie Autisme, et on a obtenu des résultats expérimentaux de mesure des similarités.

## ***Summary***

The semantic web has a purpose to make the content of web documents accessible to ensure an exchange of information and interoperability between different systems; Ontologies are proposed as central element of the semantic web.

The heterogeneity between the knowledge described in ontologies in the web has led to several interesting research questions such as the integration, evolution and alignment of ontologies. So our project was to develop a computer tool that aligns two ontologies.

To realize the application; we have exploited the lexical database "Wordnet" and "UMLS" to find out if the concept concerned has at least one synonym in the latter, after we have used the Levenshtein method which calculates the distance between the entities to align, is based on strings, then find the values of the similarities. From these similarity values the correspondences between the concepts of two ontologies are identified so the alignment is identified.

We tested our algorithm with two biomedical domain ontologies, one on asthma and the other on Autism disease, and experimental results of similarity measurements were obtained.

#### التلخيص:

الغرض من الويب الدلالي هو جعل محتوى مستندات الويب في متناول اليد لضمان تبادل المعلومات وقابلية التشغيل البيئي بين الأنظمة المختلفة ؛ يتم اقتراح الأنطولوجيا كعنصر أساسي في الشبكة الدلالية. أدى عدم التجانس بين المعرفة الموصوفة في الأنطولوجيات على شبكة الإنترنت إلى العديد من الأسئلة البحثية المهمة مثل تكامل وتطور ومواءمة الأنطولوجيا. لذلك كان مشروعنا هو تطوير أداة كمبيوتر تعمل على محاذاة الأنطولوجيين. لتحقيق التطبيق؛ لقد استغلنا قاعدة البيانات المعجمية "Wordnet" و "UMLS" لمعرفة ما إذا كان المفهوم المعني له مرادف واحد على الأقل في الأخير، بعد أن استخدمنا طريقة Levenshtein التي تحسب المسافة بين الكيانات لمحاذاة، يعتمد على سلاسل، ثم ابحث عن قيم التشابه. من قيم التشابه هذه ، يتم تحديد المراسلات بين مفهومي علم الوجود بحيث يتم تحديد المحاذاة.

اختبرنا خوارزمية لدينا اثنين من الأنطولوجيا الطبية الحيوية ، واحدة على الربو والآخر على مرض التوحد ، وتم الحصول على النتائج التجريبية لقياسات التشابه.

# ***Introduction Générale***

# Introduction Générale

De nos jours, les ontologies sont devenues l'une des plus importantes orientations de recherche avec l'avènement du Web Sémantique. Elles jouent un rôle primordial dans l'annotation de pages ou de services web puisqu'elles modélisent les concepts, attributs et relations utilisées pour annoter le contenu des ressources.

Dans de nombreux contextes applicatifs plusieurs ontologies couvrant un même domaine ou des domaines connexes sont développées indépendamment les unes des autres. L'hétérogénéité entre les connaissances exprimées au sein de chacune d'entre elles doit être résolue. C'est la problématique de l'interopérabilité. Cette dernière a pour objectif de permettre à des systèmes hétérogènes qui s'appuient sur une des ontologies de pouvoir communiquer et coopérer dans le but d'atteindre leurs objectifs. À cette fin, les liens sémantiques entre entités appartenant à deux ontologies différentes doivent être établis d'où l'alignement des ontologies.

Cependant, il n'existe pas d'ontologie universelle partagée, adoptée par tous les utilisateurs d'un domaine donné. Les problématiques et les tentatives d'amélioration de l'interopérabilité du système comptent donc sur la réconciliation des différentes ontologies utilisées dans un domaine par les différents systèmes. Cette réconciliation consiste à trouver les liens de correspondances entre ces ontologies, on parle donc de l'**alignement d'ontologies**.

C'est dans ce cadre que s'introduit ce projet de fin d'études, intitulé « Alignement des ontologies biomédicale ». En effet, il s'agit tout d'abord de faire une étude de l'état de l'art sur le sujet de l'alignement des ontologies.

Dans ce mémoire, nous allons dans un premier lieu donner une définition détaillée de la notion d'ontologie, ainsi les différents composants d'ontologie et leurs types, etc. Ensuite, dans le deuxième chapitre, nous présentons la notion alignement d'ontologies basée principalement sur la mesure de similarité. Nous présentons les différentes techniques utilisées pour mesure de la similarité. Ainsi, dans ce chapitre nous illustrons les domaines d'applications d'alignement d'ontologies.

Puis, dans le troisième chapitre nous présenterons la conception de notre système, nous avons présenter l'architecture générale pour le fonctionnement de notre application, puis nous allons le problème de calcul des similarités utilisant l'algorithme levenshtein, ainsi exploitant deux méthodes lexicales pour calculer la similarité « wordnet » et « UMLS ».



Enfin, le dernier chapitre est consacré à l'implémentation de notre système. La représentation des différents outils utilisés, ainsi que les ontologies de test et les résultats expérimentaux obtenus.

***Chapitre 1***

***Généralité***

***Sur les ontologies***

## **I.1. Introduction :**

Pour supporter l'interopérabilité, le partage et la réutilisation des connaissances formellement représentées parmi les différents systèmes, il est utile de définir un vocabulaire commun dans lequel les connaissances partagées seront représentées. La spécification de ce vocabulaire de représentation est appelée Ontologie [1].

Les ontologies sont aujourd'hui au cœur du Web sémantique. Elles proposent des représentations sémantiques des connaissances, susceptibles d'être manipulées par les machines. Elles sont utilisées pour formaliser les connaissances d'un domaine et ainsi ajouter une couche sémantique aux systèmes et applications informatiques. Les ontologies permettent de représenter de manière explicite les connaissances d'un domaine au moyen d'un langage formel afin qu'elles puissent être manipulées automatiquement et partagées aisément [2].

Dans ce chapitre, nous allons présenter les ontologies et leur application nous présentons aussi les différents composants d'ontologie et nous décrivons les différents types d'ontologies ainsi que leur cycle de vie, Les méthodologies et les outils de construction d'ontologies et les langages de représentation et de manipulation d'ontologie.

## **I.2. Naissance de l'ontologie :**

À l'origine, l'ontologie est une notion philosophique, et on considérait que l'étude de l'ontologie était une partie de la métaphysique, qui s'intéresse à l'étude des propriétés de l'être, et par extension de l'existence.

John McCarthy a introduit l'ontologie en intelligence artificielle en 1980 [3], par le principe que les concepteurs des systèmes intelligents fondés sur la logique devraient d'abord établir une base de données complète de ce qui existe, et ensuite utiliser ces données. Les ontologies sont apparues en intelligence artificielle, comme réponses aux problématiques de représentation et de manipulation des connaissances au sein des systèmes informatiques [2].

## **I.3. La Notion d'ontologie :**

Ce terme est construit à partir des racines grecques « onto » qui veut dire ce qui existe, l'être, l'existant, et « logos » qui veut dire l'étude, le discours, d'où sa traduction par « l'étude de l'être » et par extension de l'existence , nous pouvons dire que le mot « ontologie » provient du domaine de la philosophie, qui sert à désigner une théorie basée sur l'étude de l'être, d'une coté, et il peut aussi être interprété comme l'ensemble de ce qui existe avec ses relations, restrictions, axiomes et vocabulaires dans le domaine de l'Intelligence Artificiel [4].

Dans le cadre de l'intelligence artificielle ; « une ontologie définit les termes et les relations de base de vocabulaire d'un domaine, ainsi que les règles qui indiquent comment combiner les termes et les relations de façon à pouvoir étendre le vocabulaire» [5].

En fait, plusieurs définitions d'ontologies ont été proposées, mais celle qui caractérise l'essentiel d'une ontologie est fondée sur la définition de personne qui a introduit cette définition dans :

« Une ontologie est une spécification formelle et explicite d'une conceptualisation partagée », avec la signification des termes suivants :

**.Formelle** : réfère au fait qu'une ontologie doit être compréhensible par la machine, c'est-à-dire que cette dernière doit être capable d'interpréter la sémantique de l'information fournie.

• **Explicite** : signifie que le type de concepts utilisés et les contraintes sur leur utilisation doivent être explicitement définis.

• **Conceptualisation** : se réfère à un modèle abstrait de certains phénomènes dans le monde qui identifie les concepts appropriés de ce phénomène.

• **Partagée** : indique que l'ontologie supporte la connaissance consensuelle, et elle n'est pas restreinte à certains individus mais est acceptée par un groupe [3].

#### **I.4. Rôle des ontologies :**

Les ontologies qui sont actuellement au centre de la recherche dans le domaine du Web sémantique ont été développées dans le but de représenter des connaissances, que les machines peuvent comprendre, et du fait peuvent manipuler la sémantique des informations [5]. La construction des ontologies nécessite à la fois une étude des connaissances humaines et la définition de langages de représentation, ainsi que la réalisation de systèmes pour les manipuler. Les ontologies participent donc pleinement aux dimensions scientifiques et techniques de l'Intelligence Artificielle : scientifiques comme étude des connaissances humaines et plus largement de l'esprit humain, ce qui rattache l'Intelligence Artificielle aux sciences humaines, et techniques comme création d'artefacts possédant certaines propriétés et capacités en vue d'un certain usage.

Partant d'une pratique limitée voir artisanale, les méthodologies de construction d'ontologies et des outils de développement adéquats sont apparues. Émergeant des pratiques artisanales initiales, une véritable ingénierie se constitue autour des ontologies, ayant pour but leur construction mais plus largement leur gestion tout au long d'un cycle de vie.

Les ontologies apparaissent ainsi comme des composants logiciels s'insérant dans les systèmes d'information et leur apportant une dimension sémantique qui leur faisait défaut jusqu'ici.

Les ontologies représentent la composante logicielle qui s'intègre dans les systèmes d'information en leur fournissant un aspect sémantique, et de ce fait, leur champ d'application s'est considérablement élargie et couvre les systèmes conseillers comme les systèmes d'aide à la décision, les systèmes d'enseignement assisté par ordinateur " e-Learning ", etc., les systèmes de résolution de problèmes et les systèmes de gestion de connaissances.

Un des plus grands projets basé sur l'utilisation des ontologies consiste à ajouter au Web une véritable couche de connaissances permettant, dans un premier temps, la recherche d'information aussi bien au niveau syntaxique qu'au niveau sémantique.

L'enjeu de l'effort engagé est de rendre les machines suffisamment sophistiquées pour qu'elles puissent intégrer le sens des informations, qu'à l'heure actuelle, elles ne font que manipuler formellement. Mais en attendant que des ordinateurs chargés d'ontologies et de connaissances nous soulagent en partie du travail de plus en plus lourd de gestion des informations dont le flot a tendance à nous submerger, de nombreux problèmes théoriques et pratiques restent à résoudre [1].

### **I.5. Types d'ontologies :**

Les méthodes en ingénierie des connaissances ont répertorié plusieurs types d'ontologie liés à l'ensemble des objets conceptualisés et manipulés au sein d'un SBC .Parmi ces types d'ontologies, nous citons :

**I.5.1. Ontologie de domaine :** Cette ontologie exprime des conceptualisations spécifiques à un domaine Elles sont réutilisables pour plusieurs applications de ce domaine. Elle fournit les concepts et les relations permettant de couvrir les vocabulaires, activités et théories de ces domaines.

L'ontologie du domaine caractérise la connaissance du domaine ou la tâche à réalisée.

**I.5.2. Les ontologies de niveau supérieur ou de haut niveau :** Les ontologies de niveau supérieur modélisent des concepts de haut niveau auxquels ces derniers doivent être reliés au sommet des ontologies de plus bas niveaux. Cependant, il existe plusieurs ontologies de haut niveau qui se différencient par le critère utilisé pour classifier les concepts généraux de la taxonomie.

**I.5.3. Ontologie générique :** Elle est appelée également noyau ontologique, modélise des connaissances moins abstraites que celles véhiculées par l'ontologie de haut

niveau mais assez générales néanmoins pour être réutilisées à travers différents domaines. Cette ontologie inclut un vocabulaire relatif aux choses, évènements, temps, espace, causalité, comportement, fonction, etc.

**I.5.4. Ontologie de tâche :** c'est une ontologie spécifique pour résoudre un problème bien défini comme les problèmes de planification et de diagnostic. Elle offre un ensemble de termes avec nous pouvons décrire généralement comment résoudre un type de problèmes.

Elle contient des noms, des verbes et des adjectifs pour les descriptions des tâches [6].

## I.6. Composantes d'une ontologie :

La connaissance dans les ontologies est principalement formalisée en utilisant les cinq types de composants à savoir : concepts (ou classes), relations (ou propriétés), fonctions, axiomes (ou règles) et instances (ou individus).

**I.6.1. les concepts :** aussi appelés termes ou classes sont des notions (ou objets) permettant la description d'un concept, d'une tâche, d'une fonction, d'une action, d'une stratégie ou d'un processus de raisonnement, etc. Ils peuvent être abstraits ou concrets, élémentaires ou composés, réels ou fictifs. Habituellement, les concepts sont organisés en taxonomie. Une taxonomie est une hiérarchie de concepts (ou objets) reliés entre eux en fonction de critères sémantiques particuliers.

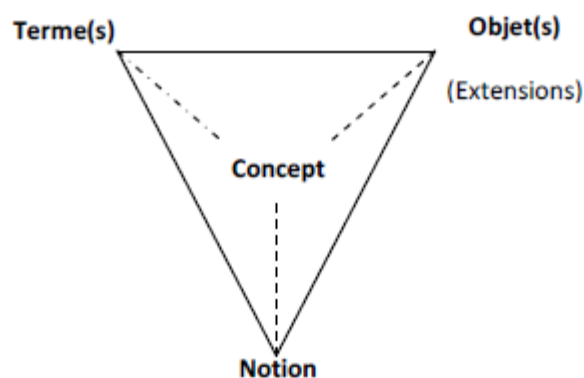


Figure I.1 : Le triangle sémantique [6]

**I.6.2. Les relations :** Elles représentent des interactions entre concepts permettant de construire des représentations complexes de la connaissance du domaine. Elles établissent des liens sémantiques binaires, organisables hiérarchiquement. Par exemple, les concepts

« Personnalité » et « Film » sont reliés entre eux par la relation sémantique « réalise (Personnalité, Film) ».

**I.6.3. Les fonctions :** Elles présentent des cas particuliers de relations dans lesquelles le nième élément de la relation est unique pour les n-1 éléments précédents

**I.6.4. Les axiomes :** Les axiomes désignent des vérités indémonstrables qui doivent être admises. Ce sont des affirmations considérées comme évidentes sans preuve. Ils permettent de contraindre les valeurs de classes ou d'instances.

**I.6.5. Les instances :** Les instances représentent les éléments des concepts et des relations dans un domaine donné [7] [8].

## **I.7. Domaine d'application :**

Le domaine d'application représenter on deux partie :

### **I.7.1. Ontologies dans les systèmes d'informations :**

Les ontologies sont développées et utilisées pour faciliter, la mise en œuvre et l'exploitation de certaines techniques et mécanismes nécessaires pour la gestion des connaissances et l'échange d'informations dans les différents systèmes d'informations et/ou applications logicielles.

Les auteurs affirment que les ontologies sont développées pour offrir une aide dans au moins trois domaines d'application à savoir la communication entre les êtres humains, l'interopérabilité entre les Systèmes d'Informations hétérogènes, ainsi que la réutilisabilité et le partage d'informations. Nous pouvons donc résumer le rôle des ontologies dans le domaine des systèmes d'informations comme suit:

#### ✓ **La communication :**

Une ontologie est un modèle standard qui décrit et spécifie explicitement son domaine, constituant ainsi un espace partageable entre les (sous) systèmes et/ou individus, où ils partagent leurs points de vue et compréhensions perspectives, éliminant ainsi tout risque de confusion ou d'incompréhension, et favorisant tout effort de collaboration et de partenariat.

#### ✓ **L'interopérabilité entre les systèmes d'informations hétérogènes:**

Pour assurer une meilleure interopérabilité entre des systèmes d'informations hétérogènes, une ontologie permet de modéliser les connaissances (les classes ainsi que les règles qui les régissent) qu'ils (les systèmes d'informations) doivent partager. En d'autres termes, en jouant le rôle d'un format d'échange, où elle répertorie toutes les classes qui doivent être échangées par les applications, l'ontologie facilite la communication, la

collaboration ainsi que la coopération entre différents systèmes d'informations. Cette coopération est principalement, dépendante de la flexibilité de communication entre les différents systèmes, car leurs actions et réactions résultent des messages envoyés et reçus.

✓ **Soutenir la spécification et la conceptualisation de systèmes:**

Une ontologie est un concept réutilisable et/ou partageable par plusieurs (sous) systèmes, ce qui permet de faciliter l'acquisition d'informations, l'analyse de données ainsi que la spécification des besoins. En plus, sa structuration plus lisible et compréhensible permet de faciliter la structuration de la documentation d'un logiciel, ce qui permet d'éviter tout risque d'ambiguïté ou de confusion lors de la spécification des besoins. Finalement, elle soutient l'automatisation du processus de vérification de cohérence réduisant ainsi les coûts de maintenance.

### **I.7.2. Ontologies dans le Web sémantique :**

Avec l'apparition du web sémantique, les ontologies ont joué un rôle technologique clé pour représenter et stocker les informations à travers des modèles qui préservent les sémantiques adressées par les domaines des applications porteuses.

Ceci facilite d'une part, l'indexation et la recherche d'information sur le web sémantique et améliore considérablement la qualité des résultats obtenus, à travers deux points majeurs:

- Une ontologie permet à l'utilisateur d'utiliser un autre terme que celui présent dans les documents.
- Une ontologie permet au système de recherche d'information d'aider l'utilisateur reformuler ses requêtes sur la base d'une proximité sémantique.

D'autre part, elle permet de faire des traitements automatiques à des modules logiciels au sein du web sémantique, que ce soit pour faire interagir et interopérer des machines entre elles ou faire interagir des machines avec des humains.

D'un autre côté, grâce au Web sémantique, l'ontologie a trouvé un jeu de formalismes standards à l'échelle mondiale, et s'intègre dans de plus en plus d'applications Web.

Cela se fait au profit des logiciels qui, à travers les ontologies et les descriptions qu'elles permettent, peuvent proposer de nouvelles fonctionnalités pour en améliorer les effets [4].

### **I.8. Classification des ontologies :**

Plusieurs critères peuvent être utilisés pour classer les ontologies. Dans cette section nous allons présenter les critères les plus cités, en l'occurrence: l'objet de la



conceptualisation, le niveau de granularité, le niveau de formalisme de représentation, et le poids de l'ontologie [4].

Une classification des ontologies selon les quatre critères ainsi cités:

### **I.8.1. Classification selon l'objet de conceptualisation:**

Selon l'objet de conceptualisation, les ontologies sont classifiées en six catégories :

❖ **Ontologie de représentation des connaissances:** Elle se focalise sur la représentation des classes impliquées dans la formalisation des connaissances.

❖ **Ontologie supérieure:** Son objectif est l'étude des catégories des choses qui existent dans le monde tels que: les entités, les relations, les propriétés, le temps, l'espace.

❖ **Ontologie Générique :** Elle représente des connaissances génériques mais assez générales pour être réutilisées à travers différents domaines.

❖ **Ontologie du Domaine :** Elle caractérise la connaissance du domaine où la tâche est réalisée.

❖ **Ontologie de Tâches :** Ce type d'ontologies est utilisé pour conceptualiser des tâches spécifiques dans les systèmes, telles que les tâches de diagnostic, de planification, de conception, de configuration, etc.

❖ **Ontologie d'Application :** Il s'agit de l'ontologie la plus spécifique où les classes correspondent souvent aux rôles joués par les entités du domaine tout en exécutant une certaine activité.

### **I.8.2. Classification selon le niveau de granularité :**

Selon le niveau de granularité (niveau de détail), les ontologies sont classifiées en deux catégories :

❖ **Ontologies de fine granularité :** Ce niveau correspond 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.

❖ **Ontologies de large granularité :** Ce niveau correspond à des vocabulaires moins détaillés. Par exemple les scénarios d'utilisation spécifique ou les utilisateurs sont déjà préalablement d'accord à propos d'une conceptualisation sous-jacente. Les ontologies de haut niveau possèdent une granularité large, compte tenu que les concepts qu'elles traduisent sont normalement raffinés ultérieurement dans d'autres ontologies de domaine ou d'application.

### **I.8.3. Classification selon le niveau de formalisme de représentation :**

On peut distinguer les ontologies selon le formalisme utilisé pour les exprimer.

❖ **Informelle** : l'ontologie est exprimée en langage naturelle. Cela peut permettre de rendre plus compréhensible l'ontologie pour l'utilisateur, mais cela peut rendre plus difficile la vérification de l'absence de redondances ou de contradiction ;

❖ **Semi-informelle** : l'ontologie est exprimée dans une forme restreinte et structurée de la langue naturelle ; cela permet d'augmenter la clarté de l'ontologie tout en réduisant l'ambiguïté ;

❖ **Semi-formelle** : l'ontologie est exprimée dans un langage artificiel défini formellement ;

❖ **Formelle** : l'ontologie est exprimée dans un langage artificiel disposant d'une sémantique formelle, permettant de prouver des propriétés de cette ontologie. L'intérêt d'une ontologie formelle est la possibilité d'effectuer des vérifications sur l'ontologie : complétude, non redondance, consistance, cohérence, etc.

#### **I.8.4. Classification selon le poids de l'ontologie :**

Selon leur poids, les ontologies sont classifiées en deux types :

- ❖ Ontologies de poids léger (classes et relations)
- ❖ Ontologies de poids lourd (classes, propriétés et relations)

### **I.9. L'ingénierie ontologique :**

#### **I.9.1. Processus de construction d'une ontologie :**

Le processus de construction d'ontologie est une collaboration qui réunit des experts du domaine de connaissance, des ingénieurs de la connaissance, voire les futurs utilisateurs de l'ontologie.

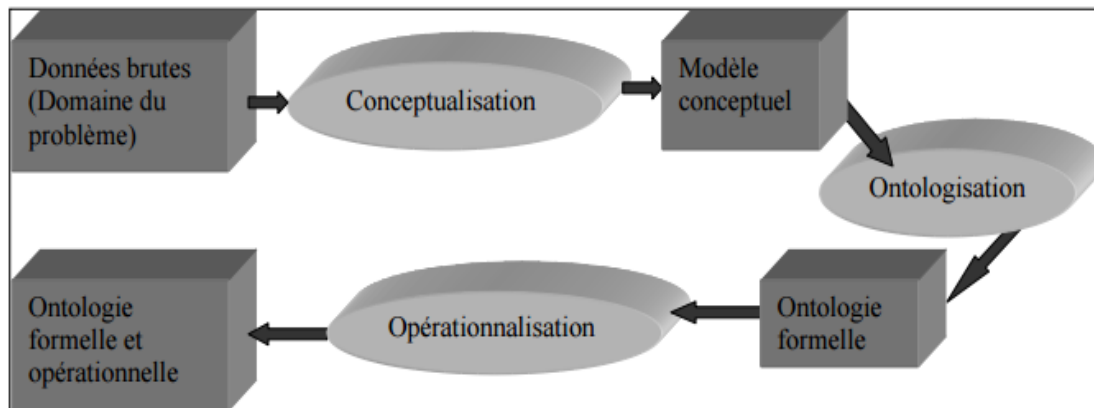
Cette collaboration ne peut être fructueuse que si les objectifs du processus ont été clairement définis, ainsi que les besoins qui en découlent. La construction d'ontologie s'effectue en trois étapes : [9]

❖ **La conceptualisation** (informelle ou semi formelle): En langage naturel ou semi-structuré, ne disposant pas de sémantique claire ou tout au moins d'une sémantique fixée à priori.

❖ **L'ontologisation** (formelle): Spécifie la syntaxe et la sémantique, Ce travail doit être mené par l'ingénieur de la connaissance, assisté de l'expert du domaine.

❖ **L'opérationnalisation** (formelle et opérationnelle) : transcription de l'ontologie dans un langage formel et opérationnel de représentation de connaissances. Ce travail doit être mené par l'ingénieur de la connaissance ; Il est à noter que ce processus n'est pas linéaire et

que de nombreux aller-retour sont à priori nécessaires pour bâtir une ontologie opérationnelle adaptée aux besoins.



**Figure I.2:** Processus de construction d'ontologie [9]

## I.9.2. Méthodologies de construction d'ontologies :

La construction des ontologies, nécessite le passage d'une étape indispensable, qui est l'analyse des méthodologies et des outils de construction d'ontologies qui existent, dans la perspective de pouvoir identifier une méthodologie et/ou un outil qui permet de construire une ontologie qui répond aux maximum aux besoins des utilisateurs :[4][9]

### I.9.2.1. La méthodologie d'Uschold et King:

Connue aussi par la méthodologie Entreprise, Elle s'est inspirée des constructions modélisant une entreprise [9].

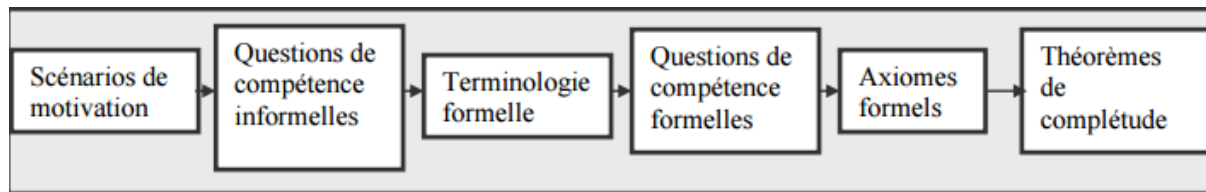
Elle comprend l'identification de l'objectif de l'ontologie, sa construction, son évaluation, et finalement sa documentation. Lors du processus de construction, les auteurs proposent de capturer les connaissances, les coder, et finalement intégrer si nécessaire des ontologies préexistantes dans l'ontologie courante. Les trois stratégies suivantes sont proposées par les auteurs pour identifier les classes :

- Une stratégie descendante (top down strategy) où les classes les plus générales sont identifiées puis sont spécialisées,

- Une stratégie ascendante (down top strategy) où les classes les plus spécifiques sont identifiées puis sont généralisées,

- Une stratégie mixte (midle out strategy), où les classes les plus importantes sont identifiées puis est généralisée et/ou spécialisées.

Cette méthodologie est indépendante du système qui l'exploite et son processus de construction d'ontologie est indépendant de l'objectif de l'ontologie.



**Figure I.3:** La méthodologie de d'Uschold et King, (Entreprise) [9]

### **I.9.2.2. La méthodologie de Gruninger et Fox:**

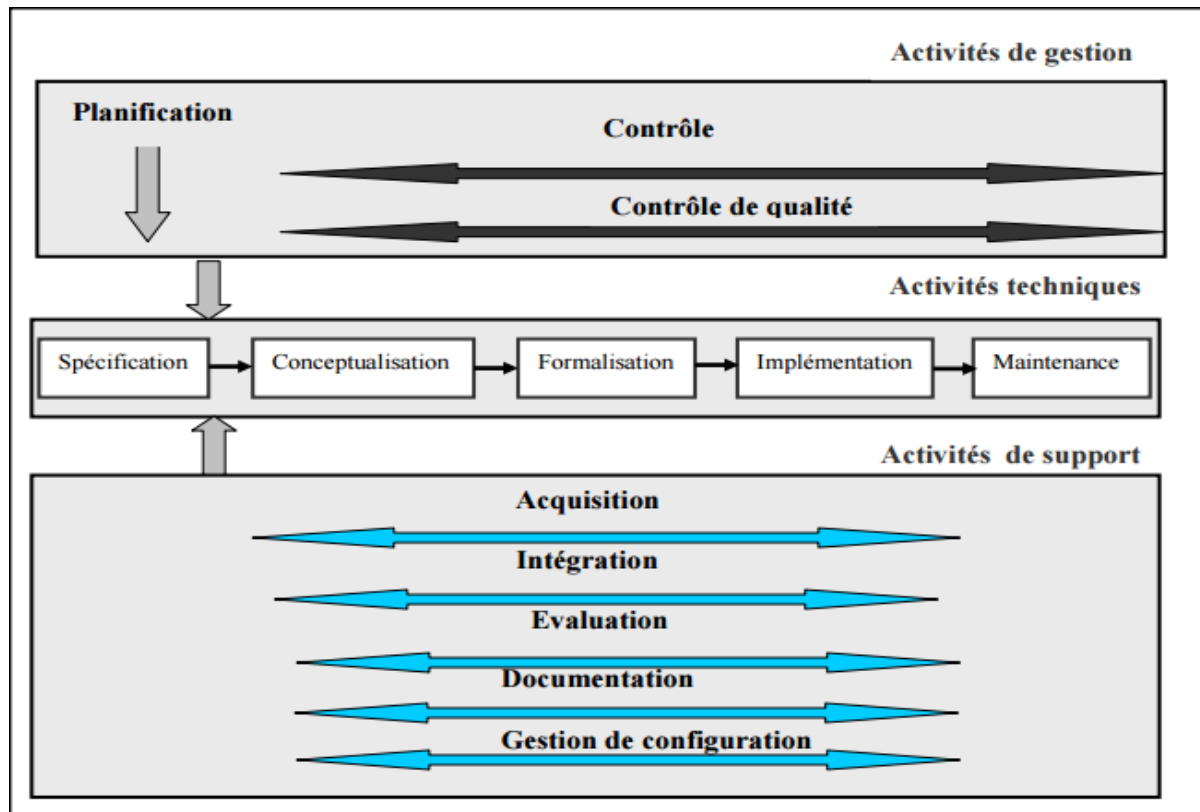
Appelée aussi la méthodologie TOVE elle est inspirée du développement des SBC utilisant la logique du premier ordre. Elle comprend: L'identification des principaux scénarios (les applications de l'ontologie), l'identification des questions de compétence (que le système est censé pouvoir répondre) et alors identifier les concepts et les axiomes de l'ontologie en utilisant une stratégie mixte (middle out). Dans cette méthodologie qui est semi dépendante du système qui l'exploite, une description informelle des spécifications de l'ontologie sont formalisés (logique de premier ordre), donc elle peut être utilisée comme un guide de transformation de scénarios informels en modèles formels.

### **I.9.2.3. La méthodologie Methontology :**

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

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

Un schéma illustratif de la méthodologie Methontology est présenté dans la figure I.4



**Figure I.4** : Les composantes de Methontology [9]

### I.9.3. Cycle de vie d'une ontologie :

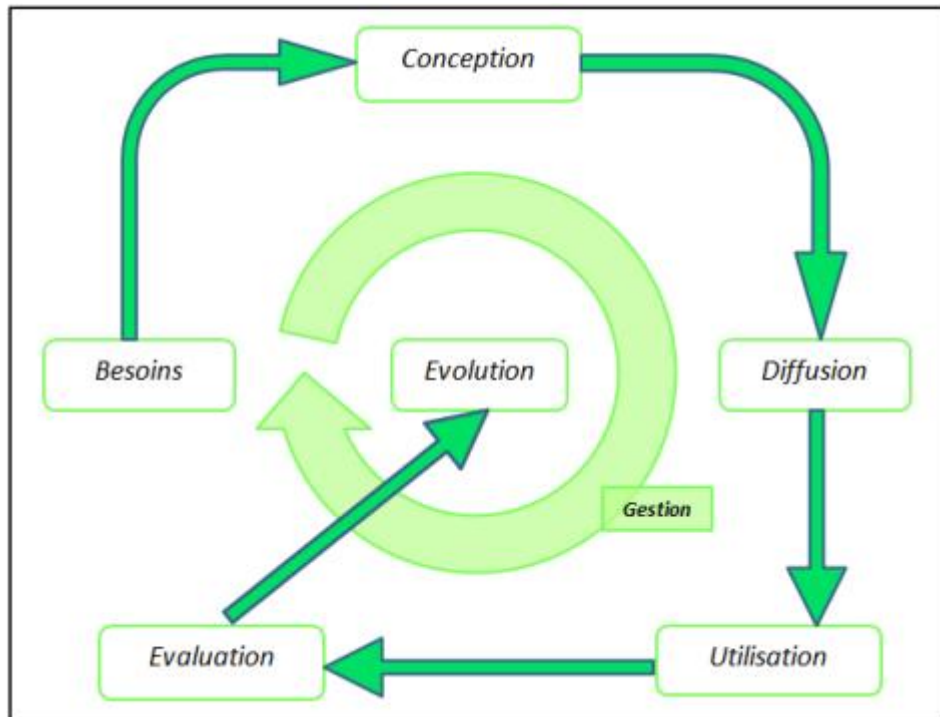
Les ontologies étant destinées à être utilisées comme des composants logiciels dans des systèmes répondant à des objectifs opérationnels différents, leur développement doit s'appuyer sur les mêmes principes que ceux appliqués en génie logiciel. Ainsi, les ontologies doivent être considérées comme des objets techniques évolutifs et possédants un cycle de vie qui nécessite d'être spécifié. Dans ce contexte, les activités liées aux ontologies sont d'une part des activités de gestion incluant la planification, le contrôle, et la garantie de qualité, et d'autre part des activités orientées développement regroupant les activités de pré développement, de développement et de post-développement, s'y ajoutent des activités transversales de support telles que l'évaluation, la documentation et la gestion de version.

Un cycle de vie inspiré du génie logiciel a été proposé. Il comprend une étape initiale d'évaluation des besoins, une étape de construction, une étape de diffusion, et une étape d'utilisation. Après chaque utilisation significative, l'ontologie et les besoins sont réévalués et l'ontologie peut être étendue et, si nécessaire, en partie reconstruite.

L'étape de construction peut être décomposée en trois principales étapes qui sont:

La conceptualisation, la formalisation appelée également ontologisation et l'opérationnalisation.

Le processus de construction peut être intégré au cycle de vie d'une ontologie comme indiqué en (figure I.5) [10] [11].



**Figure I.5 :** Cycle de développement d'ontologie [7]

### I.9.3.1. Conception et évolution :

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

- spécification des solutions ;
- conceptualisation et modélisation ;
- formalisation (logiques de description, graphes conceptuels, formalismes du web sémantique RDF, RDF(S) et OWL) ;
- intégration de ressources existantes ;
- implantation (graphes conceptuels, logiques de description).

Le choix de représentation et de conceptualisation faits dans l'ontologie représente un problème de conception et d'évolution. Notons aussi que l'évolution pose le problème de la maintenance de ce qui repose déjà sur l'ontologie. L'ontologie est à la fois un ensemble

évolutif et un ensemble de primitives pour décrire des faits et des algorithmes sur ces faits. Ses changements donc, ont un impact direct sur tout ce qui a été construit sur la base de cette ontologie. Le maintien de la cohérence dans une ontologie est un des points clés dans son utilisation.

La maintenance de l'ontologie soulève donc des problèmes d'intégration technique et des problèmes d'intégration aux usages. Les domaines de la maintenance comme l'historique et la gestion des versions, la propagation des changements après modification, sont des questions importantes de la recherche actuelle.

### **I.9.3.2. Diffusion :**

Le déploiement et à la mise en place de l'ontologie interviennent dans la phase de diffusion, où les problèmes sont fortement contraints par l'architecture des solutions. Pour l'application Web on utilisera des technologies adéquates et pour le partage de fichiers, nous pouvons utiliser des architectures distribuées ou l'architectures Peer to Peer. Dans toutes les architectures dans les serveurs web, services web, Peer to Peer, etc. la distribution des ressources et leur hétérogénéité du point de vue syntaxiques, sémantiques, protocolaires, contextuelles, ou autres posent des problèmes de recherche sur l'interopérabilité et le passage à l'échelle (larges bases, optimisation d'inférences, propagation de requêtes, etc.).

### **I.9.3.3. Utilisation :**

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

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

### **I.9.3.4. Gestion :**

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

## I.10. Langage de représentation et de manipulation d'ontologie :

Dans cette section, nous définissons quelques langages de représentation des ontologies les plus connus et les plus utilisés : [10]

### I.10.1. Le langage Rdf :

**RDF** (Resource Description Framework) est un modèle de données décrivant les ressources web à travers un ensemble de nœuds représentant les objets ou les ressources, et un ensemble d'arcs représentant les propriétés formant ainsi un graphe de relations entre les différentes ressources, où chaque ressource se voit associée un identifiant unique, URI (Unique Resource Identifier), ces ressources sont donc liées entre elles ou à un littéral par le biais d'un triplet, « sujet, prédicat, objet », mais comme RDF seul ne définit aucune primitive pour créer des ontologies, il donne une base pour plusieurs langages de définition d'ontologie tel que RDFS.

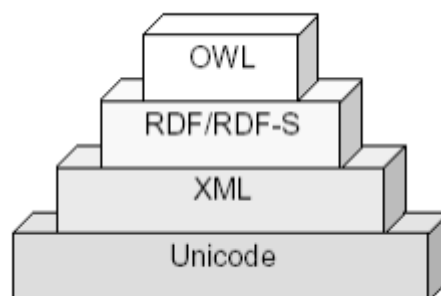
### I.10.2. Le langage RDFS (Resource Description Framework Schéma) :

RDFS offre des primitives pour organiser les objets du web sous forme d'une hiérarchie. Il était introduit comme étant une couche au dessus de RDF (basé sur RDF).

### I.10.3. Le langage Owl :

OWL (Ontology Web Language) : Comme son nom l'indique OWL est un langage de représentation d'ontologies destinées à être publiées et partagées sur le web.

Il s'agit de la dernière recommandation du W3C dans ce sens et constitue une surcouche de RDF(S) qui est lui-même une surcouche de XML. Voir la figure ci dessous.



**Figure I.6 :** Relation entre Unicode, XML, RDF, et OWL [10]

En OWL on distingue :

- les propriétés d'objet (Object Properties) : les relations, qui relient des instances de classes à d'autres instances de classes. C'est l'équivalent des triplets RDF dont l'objet est une ressource.



- les propriétés type de données (data type property) : les attributs, qui relient des instances de classes à des valeurs de types de données (nombres, chaînes de caractères,...). C'est l'équivalent des triplets RDF dont l'objet est une valeur littérale.

-Les axiomes fournissent de l'information au sujet des classes et des propriétés, spécifiant par exemple l'équivalence entre deux classes. Donc OWL permet de définir des ontologies comme un jeu de définition de classes, de propriétés et de contraintes [12].

Toute classe définie dans une ontologie OWL est une sous-classe d'owl Thing.

Pour répondre aux besoins d'expressivité ontologique croissante OWL s'est évolué en mettant en œuvre trois sous langages en l'occurrence OWL-Lite, OWL-DL, et OWL-Full:

➤ **OWL Lite** : Convient aux utilisateurs qui ont principalement besoin d'une hiérarchie de classification et de contraintes simples. Ce sous langage reprend tous les constructeurs de RDF (c'est-à-dire fournit des mécanismes permettant de définir un individu comme instance d'une classe, et de mettre des individus en relation).

➤ **OWL DL** : Convient aux utilisateurs qui veulent le maximum d'expressivité, ce sous langage reprend tous les constructeurs d'OWL LITE, il permet tout entier positif dans les contraintes de cardinalité, et le tire son nom de sa correspondance avec les logiques de descriptions.

➤ **OWL FULL** : ce sous langage reprend tous les constructeurs d'OWL DL, et tous les constructeurs de RDF Schéma. Il permet d'utiliser une classe en position d'individu dans les constructeurs.

Donc en fonction de leurs besoins, les développeurs d'ontologies choisissent le sous langage qui leur convient le plus, qu'il s'agisse de la simplicité et de la facilité d'implémentation (OWL-Lite), de la richesse d'expression (OWL-DL), ou d'éviter les restrictions pouvant leur y imposées (OWL-Full).

### **I.11. Conclusion :**

une ontologie est une conceptualisation formelle d'un domaine particulier. Elle regroupe les connaissances d'un domaine, à savoir les concepts, les propriétés et les relations entre ces derniers.

L'ontologie est un élément pivot pour l'avenir du web sémantique, en ajoutant des métas données sur les ressources du web afin de préciser leurs sémantiques et leurs données.

La croissance continue du nombre d'ontologies dans le web et la façon dont elles sont conçues, à savoir, chacun développe son ontologie selon sa vision et ses besoins, ont ramené à avoir de multiple ontologies dans le même domaine.

---

Le chapitre suivant met l'accent sur les travaux effectués sur l'alignement d'ontologies et les techniques adoptées pour élaborer des outils automatiques d'alignement.

# **Chapitre II**

## **Alignement d'ontologie**

## II.1. Introduction :

Le terme ontologie a été utilisé initialement en philosophie depuis le 19<sup>ème</sup> siècle. Dans ce domaine, il désigne l'étude de ce qui existe, l'ensemble des connaissances sur le monde. Les techniques d'alignement entre deux structures ont été développées dans de nombreux domaines [9].

L'alignement d'ontologies est une tâche complexe reposant sur la définition de mesures de similarité. De nombreuses études ont été effectuées. Un seul type de mesure se révélant la plupart du temps insuffisant pour détecter une similarité. Différents outils combinant plusieurs mesures de façon séquentielle ont été proposés : cette combinaison est effectuée a priori et n'est pas modifiable. Les Framework devraient être plus prometteurs [13].

Participer au web de données suppose d'être capable de s'interconnecter avec les données et ontologies déjà présentes et disponibles. Publier une nouvelle ontologie nécessite donc au préalable de la relier avec les "bonnes" ontologies publiées sur le web de données comportant des concepts similaires dans des domaines similaires

Relier deux ontologies consiste en fait à les aligner, trouver des correspondances entre les entités (concepts, propriétés ou instances) des deux ontologies [14].

Les techniques d'alignement représentent un cadre général, dans lequel plusieurs ontologies peuvent être exploitées [15].

Dans le domaine de la représentation des connaissances, les ontologies ne sont considérées que relatives aux différents domaines de connaissances. Elles répondent aux problèmes de représentation et de manipulation des connaissances. Les ontologies sont très utilisées dans la représentation des connaissances sur le Web [16].

Dans ce chapitre, nous présentons un état de l'art sur l'alignement des ontologies dans le domaine biomédical qui constitue la base de notre recherche. Et une fois que nous clarifions : la nécessité enchaînons plus en profondeur avec les méthodes de base utilisées pour réaliser l'alignement, ces dernières sont combinées pour construire des systèmes automatiques d'alignement qui à leurs tours sont classés selon différents critères, et nous présentons quelque approches d'alignement d'ontologies, enfin nous effectuons une comparaison entre les méthodes d'alignements.

## II.2. La nécessité de l'alignement des ontologies :

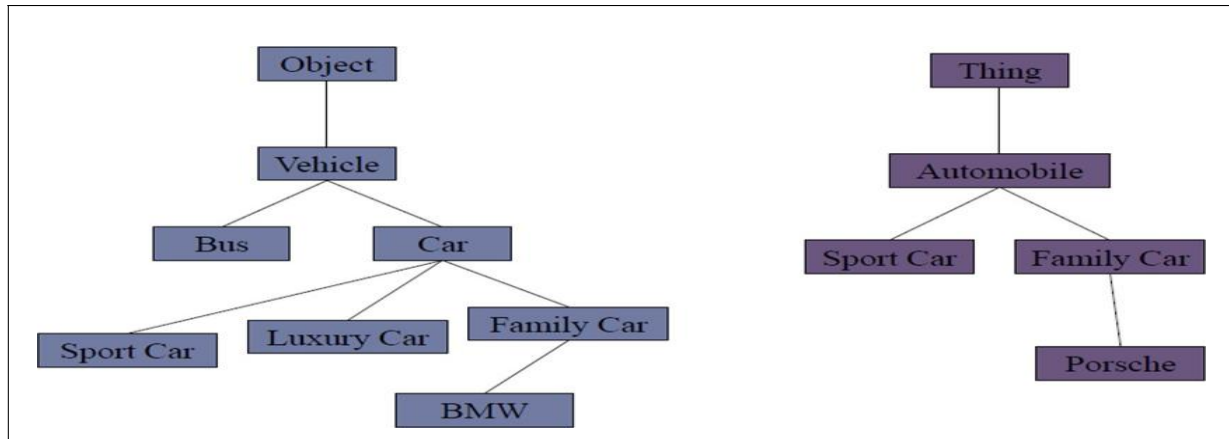
L'ontologie peut être utilisée dans de nombreux domaines d'application pour décrire et stocker les connaissances.

Cependant, puisqu'il n'existe pas une ontologie universelle à cause des divergences conceptuelles et pour les raisons que nous avons citées, ces connaissances sont exprimées par des ontologies hétérogènes et distribuées. Cette hétérogénéité peut se situer au niveau syntaxique, terminologique, conceptuel ou sémiotique. Nous présentons ci-dessus un exemple de deux ontologies hétérogènes qui décrivent le même domaine mais utilisent différents termes (Figure II.1) [17].

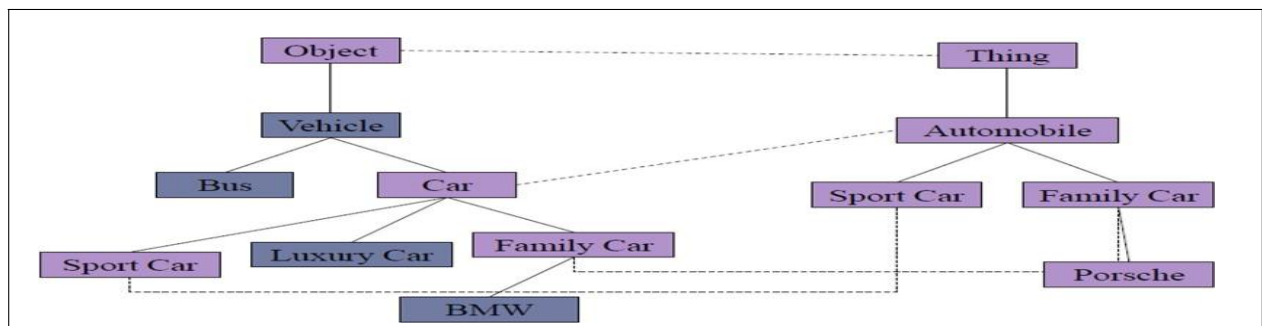
La solution proposée pour surmonter cette hétérogénéité consiste à utiliser l'alignement d'ontologies pour combler le fossé sémantique entre les ontologies hétérogènes. Le système d'alignement d'ontologies reçoit en entrée deux ou plusieurs ontologies et génère en sortie un ensemble de correspondances sémantiques entre les entités de ces ontologies. En effet, ces correspondances sémantiques sont les ponts qui maintiennent les ontologies hétérogènes ensemble (Figure II.2) et assurent leur interopérabilité sémantique [15].

En outre, avec l'énorme volume des ontologies (leur taille et leur nombre) déjà disponibles sur le web et leur constante évolution, l'identification manuelle des correspondances sémantiques est une tâche fastidieuse et coûteuse voire n'est pas réalisable. Par conséquent, les outils d'alignement d'ontologies sont tenus d'avoir la capacité d'identifier les correspondances sémantiques entre les entités de différentes ontologies de manière automatisée ou semi-automatisée afin de réduire la charge de la création et de la maintenance manuelle de l'alignement d'ontologies. Toute fois, l'identification automatique des correspondances sémantiques n'est pas une tâche triviale en raison de la diversité conceptuelle entre les ontologies. Plusieurs méthodes et systèmes d'alignement d'ontologies ont été élaborés pour aligner sémantiquement leurs entités. Ces méthodes sont basées généralement soit sur le calcul de la similitude des concepts, des relations et des instances ; soit sur des techniques de raisonnement logique. Elles peuvent être du type terminologique, structurel, sémantique et extensionnelle et sont souvent combinées [16].

Comme nous l'avons cité auparavant, l'alignement d'ontologies peut être utilisé dans différentes applications autres que le web sémantique. Nous présentons dans la section suivante quelques unes de telles applications



**Figure II.1 :** Exemple de Deux Ontologies Hétérogènes [17]



**Figure II.2:** Etablissement des Correspondances Sémantiques [15]

### II.2.1. Problème issu de l'hétérogénéité :

L'hétérogénéité n'est pas seulement due à la divergence des domaines que peuvent couvrir les ontologies mais aussi aux formalismes requis pour leur développement. Dans la littérature, plusieurs classifications des types d'hétérogénéité sont recensées. Quelques classifications se basent sur l'étude du décalage sémantique et structurel qui peut exister entre les ontologies. D'autres classifications élucident le degré d'hétérogénéité selon les niveaux d'interopérabilité sémantique [6] [18]. La littérature recense quatre types d'hétérogénéités, à savoir : l'hétérogénéité syntaxique, l'hétérogénéité terminologique, l'hétérogénéité conceptuelle et l'hétérogénéité sémantique.

#### II.2.1.1. L'hétérogénéité syntaxique :

L'hétérogénéité se produit quand deux ontologies sont décrites avec deux langages ontologiques différents, elle est dépendante du choix du format de représentation. Celui-ci consiste à décrire et à coder les entités d'un domaine donné de manière à ce qu'une machine puisse les manipuler afin de raisonner ou de résoudre des problèmes, Cette hétérogénéité se produit aussi quand deux ontologies sont modélisées en utilisant des

formalismes différents. Il est possible dans certains cas de traduire les ontologies dans différents langages ontologiques à condition de préserver la signification [6].

Dans la littérature, il existe de nombreux langages informatiques spécialisés dans la création et la manipulation des ontologies comme (XML, RDF, rdfs et OWL etc..) et chacun d'entre eux est basé sur une syntaxe propre et parfois ils sont utilisés pour représenter la même ontologie.

### **II.2.1.2. L'hétérogénéité terminologique :**

L'hétérogénéité est liée au processus de nommage des entités (classes, propriétés, relations) qui constituent une ontologie à partir d'un langage public, alors qu'elles désignent le même objet.

La cause d'une telle hétérogénéité revient à l'utilisation de différents langages naturels, ou des sous-langages techniques spécifiques à un domaine de connaissances bien déterminé. Dans ce qui suit, nous citons des exemples de ce type d'hétérogénéité :

**Synonymie** : Différents noms utilisés pour désigner une même entité.

- Polysémie : Le même mot désigne plusieurs entités.
- l'homonymie : Au contraire, le même terme peut représenter différents concepts ; l'homonymie est un problème qui nécessite bien souvent l'intervention humaine.
- Langage : Des mots provenant de différentes langues (Français, Anglais, Italien, etc.) utilisés pour désigner une même entité.
- Variations syntaxiques : variations syntaxiques du même mot (différentes prononciations, abréviations, utilisation des préfixes et des suffixes, etc.).

Finalement, l'encodage des données au sein de l'ontologie diffère bien souvent, que ce soit pour les dates, les unités (monnaie, distances, etc.) [19].

### **II.2.1.3. L'hétérogénéité conceptuelle :**

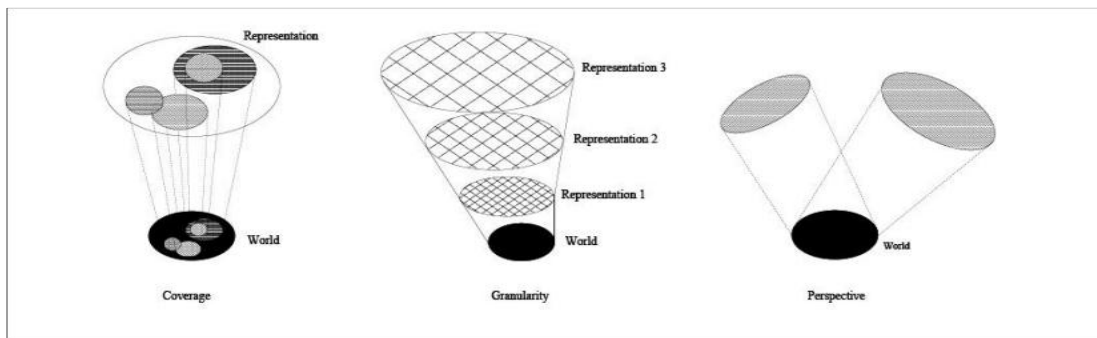
Appelée aussi hétérogénéité sémantique ou la différence logique [20] .Elle concerne la diversité des modélisations d'un même domaine de connaissances. Elle découle principalement de l'utilisation de différents (ou équivalents) axiomes décrivant les concepts ontologiques. Elle se manifeste aussi lors de l'utilisation de concepts totalement différents. Respectivement et évoquent la différence de conceptualisation et la différence de l'explication. La différence de conceptualisation se manifeste à travers la différence entre les concepts inclus dans la modélisation. La différence des explications se base sur la manière avec laquelle les concepts sont exprimés. proposent une classification précise de ces

différences. Par ailleurs présentent les trois principales raisons de la différence de conceptualisation : [21]

- Différence de convergence qui survient lorsque deux ontologies décrivent différentes connaissances avec le même niveau de détail pour une unique perspective.

- Différence de granularité qui se produit quand deux ontologies décrivent le même domaine avec une même perspective, mais avec différents degrés d'expression des détails.

- Différence de perspectives qui se manifeste quand deux ontologies décrivent un même domaine, avec un même degré d'expression des détails, mais avec des points de vue et des perspectives différents.



**Figure II.3** : Les trois dimensions de l'hétérogénéité au niveau conceptuel [6]

#### II.2.1.4. L'hétérogénéité sémantique :

Appelée aussi hétérogénéité pragmatique, elle s'intéresse à la manière dont les entités ontologiques sont interprétées par leurs utilisateurs. Ainsi, les entités ayant les mêmes interprétations sémantiques peuvent être interprétées de différentes manières par l'homme [14].

Ces différences d'interprétation sont dues principalement à la diversité des contextes et des domaines d'application des ontologies. Par conséquent, la manière de mettre en œuvre les entités ontologiques influence leurs interprétations. De plus, ce type d'hétérogénéité reste difficile à détecter par la machine.

Mais grâce à l'alignement des ontologies, on peut lier entre les ontologies du web sémantique même si l'hétérogénéité existe. Cette liaison se fait en cherchant les correspondances entre les ontologies. Afin de pouvoir réaliser quelques tâches comme l'échange, l'intégration... etc. [4]



### II.3. Définition de l'alignement d'ontologies :

L'alignement dans son sens général, désigne l'ajustement des objets par rapport aux autres selon une orientation voulue. Dans le domaine d'informatique, plus particulièrement dans le Web Sémantique, on parle de l'alignement des ontologies qui représente une tâche cruciale dans plusieurs sous-domaines d'application : la communication dans les systèmes multi-agents, data warehouse, etc. Et comme nous avons mentionné dans un premier temps que le problème actuel, c'est que dans un même domaine ou des domaines connexes, on risque d'avoir plusieurs ontologies différentes (hétérogénéité). C'est pour cela qu'on a introduit la notion de l'alignement d'ontologies qui sert à comparer deux ontologies et de gérer le plus automatiquement possible, des appariements sur les ontologies, et qui consiste à trouver des correspondances chacune liant deux entités (par exemple, des concepts, des instances, des propriétés, des termes, etc.) par une relation (équivalence, incompatibilité, etc.), éventuellement munie d'un degré de confiance. L'ensemble de ces correspondances est ce qu'on appelle alignement, et qu'on exploite après conjointement dans le même système [1] [4].

#### II.3.1 le processus d'alignement :

Le processus d'alignement regroupe trois dimensions : l'input, le processus d'alignement et l'output.

✓ **L'input** : ou bien l'entrée qui se constitue des structures destinées à être alignées, elles peuvent être des schémas XML, schémas relationnels, des ontologies. Dans notre cas, ce sont les dernières.

**Remarque** : l'input peut être enrichi par un alignement en entrée (qui aurait besoin d'être complété par une nouvelle itération d'alignement).

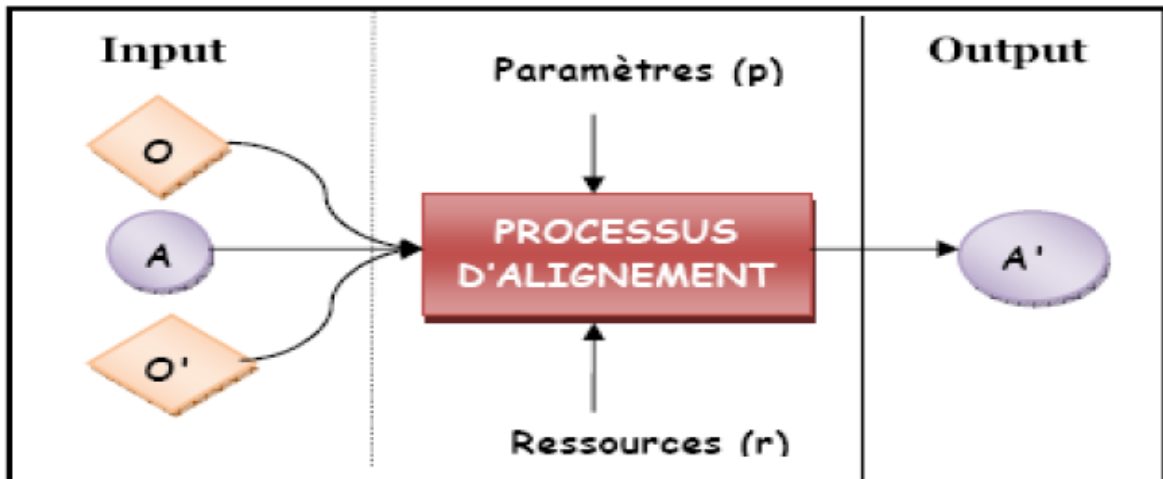
✓ **Le processus d'alignement** : est une tâche pendant laquelle à partir d'une ontologie  $O$  et une autre  $O'$ , il détermine un alignement  $A$  entre ces deux ontologies, cette tâche est réalisée en utilisant une stratégie ou une combinaison de techniques d'alignement de bases. On verra dans la suite de ce chapitre ces techniques en détail.

✓ **L'output** : ensemble d'alignement reliant les entités qui constituent les deux ontologies qui comporte :

- Id : un identifiant de l'alignement
- $e$  : une entité à aligner et qui appartient à  $O$
- $e'$  : une entité à aligner appartenant à l'ontologie  $O'$

- $r$  : la relation qui lie  $e$  à  $e'$  ( $=, \subseteq, \supseteq, \sqcap, \sqcup$ )
- $n$  : la mesure de confiance de  $r$  comprise dans  $[0,1]$ , 1 étant une relation forte [1].

La figure qui suit est une représentation schématique du processus d'alignement :



**Figure II.4:** Schéma résumant le processus d'alignement [1].

## II.4. Quelques domaines d'application de l'alignement d'ontologies :

L'alignement est devenu un concept incontournable de nos jours et ceci dans plusieurs domaines d'application. Nous en citons quelques exemples [18].

### II.4.1. Ingénierie d'ontologie :

Est un contexte où les concepteurs d'ontologies sont confrontés à l'hétérogénéité de ces dernières, plus précisément, par rapport aux applications suivantes : La construction d'ontologies : ces dernières années, le maître mot dans la démarche de construction des ontologies est la réutilisation d'ontologies déjà existantes, car la construction d'ontologies à partir de zéro est un processus long, coûteux et très laborieux, parallèlement, elle accentue le phénomène de l'hétérogénéité des ontologies, multipliant le nombre d'ontologies décrivant le même domaine (surtout lorsqu'on sait que l'objectif ultime du web sémantique est d'arriver à instaurer une ontologie de référence pour chaque domaine). Dans ce contexte, l'alignement des ontologies est la solution pour réaliser l'intégration et le rapprochement de ces différentes structures [2].

### II.4.2. Le web sémantique :

Le web sémantique a été créé afin d'assurer l'interopérabilité sémantique entre différentes sources d'informations exprimées par des ontologies. Cependant, ces ontologies sont hétérogènes et distribuées. L'alignement d'ontologies est la clé pour réaliser cette interopérabilité sémantique, il permet d'établir un ensemble de correspondances sémantiques entre les ontologies hétérogènes. Ces correspondances peuvent être utilisées pour diverses tâches, telles que la fusion des ontologies, la traduction des données, etc. le web sémantique est considéré comme une tâche complexe ou toutes les applications de l'alignement peuvent être rencontrées [15].

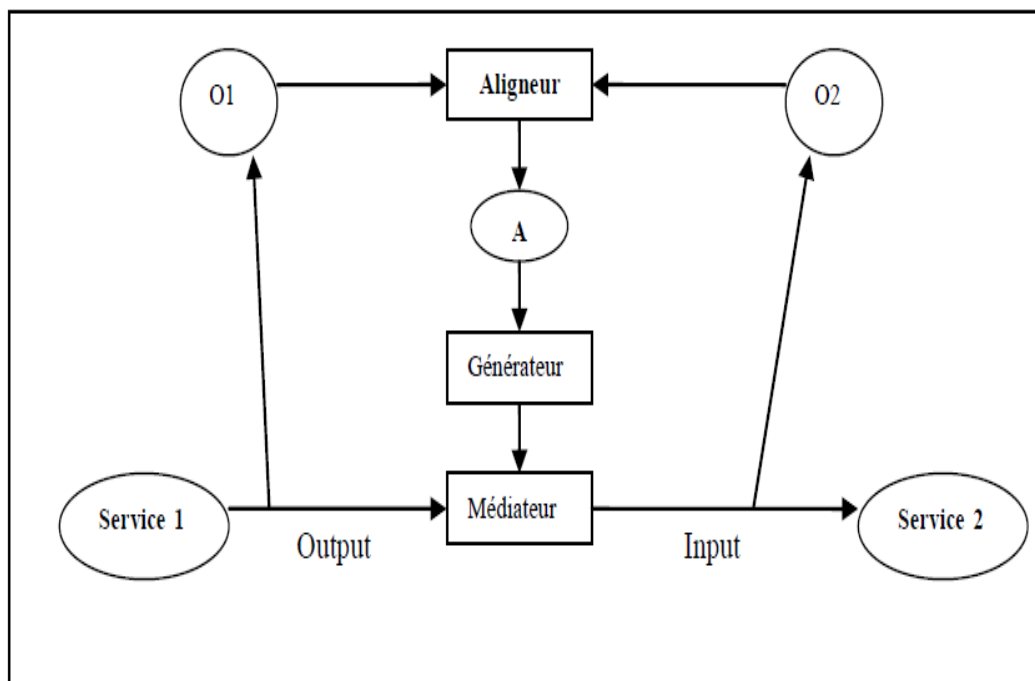


Figure II.5 : Composition d'un service web [15]

### II.4.3. Le Web Biomédical :

Le domaine biomédical est par ailleurs caractérisé par l'existence de nombreux standards terminologiques, thesaurus, et langages, partagés par les communautés biomédicales, qu'ils soient généralistes ou dédiés à un domaine de spécialité, ainsi que de riches banques de documents généralistes ou plus spécifiques (par exemple, sur les maladies rares ORPHANET, RARE DISEASE), qui représentent un acquis important mais aussi une contrainte forte puisqu'il n'est pas envisageable de les ignorer. Les ontologies doivent fournir les concepts et les relations utilisés pour le marquage sémantique des données en vue du Web Sémantique avec une signification partagée et réutilisable pour différentes applications et différents usagers.

L'alignement des ontologies aide à trouver rapidement sur le Web, avec le minimum de bruit possible, une information scientifique récente, a un intérêt non seulement pour le chercheur qui doit accéder à des bases hétérogènes et réitérer régulièrement les interrogations sur ces bases, pour les patients à la recherche d'informations, mais aussi dans la pratique médicale quotidienne où médecins et industriels pharmaceutiques sont amenés à rechercher de l'information. De plus, l'alignement des ontologies joue un rôle déterminant dans la mise en place d'une recherche biomédicale associant un plus grand nombre d'acteurs ; en facilitant la constitution d'entrepôts de données ou d'entrepôts d'informations fédérés, articulés autour d'ontologies communes [15].

## **II.5. Les Méthodes d'alignement :**

Les méthodes d'alignement d'ontologies se basent sur les méthodes de calcul de similarité. Dans cette partie, nous allons voir la relation entre l'alignement et la similarité, la définition de la similarité et les différentes méthodes de calcul de similarité (terminologiques, structurelles, extensionnelles et sémantiques).

### **II.5.1. La relation entre l'alignement et la similarité :**

L'alignement des ontologies consiste à chercher la similarité entre les ontologies à aligner, donc pour chercher des correspondances entre les entités des ontologies il faut analyser les hiérarchies des entités (la structure locale et globale des entités), cela en utilisant les techniques de l'alignement qui sont les méthodes de calcul de similarité (terminologiques, structurelles, extensionnelles et sémantiques) [4]

### **II.5.2. La similarité :**

La notion de similarité dans notre contexte n'est pas celle que l'on peut trouver en psychologie ou en mathématiques. En psychologie sociale, la similarité se rapporte à comment les attitudes, les valeurs, les intérêts et la personnalité correspondent entre les personnes. En mathématiques, plusieurs relations d'équivalence (qui sont des relations binaires réflexives, symétriques et transitives) sont appelées similarité. Ces relations existent par exemple :

- En géométrie : Deux objets géométriques sont similaires si l'un est isométrique avec le résultat d'un agrandissement ou rétrécissement uniforme de l'autre.

L'un peut être obtenu à partir de l'autre uniformément élargi, rétréci, avec une rotation éventuelle (tous les deux ont la même forme), ou en plus, en appliquant un effet du miroir (l'un a la même forme que l'image de miroir de l'autre). Par exemple, tous les cercles sont similaires entre eux, tous les carrés sont similaires entre eux, et toutes les paraboles sont

similaires entre elles. D'autre part, ni les ellipses, ni les hyperboles ne sont similaires entre elles. Deux triangles sont semblables si et seulement s'ils ont les mêmes 3 angles.

- En algèbre linéaire, deux matrices A et B de taille  $n \times n$  sont dites similaires s'il existe une matrice inversible P de même taille  $n \times n$  satisfaisant  $A P = B$ .

- En topologie, la similarité est une fonction telle que sa valeur est plus grande quand deux points sont plus proches (contrairement à la distance, qui est une mesure de dissimilarité : plus les points sont proches, plus la distance est petite).

Dans notre contexte, la notion de similarité sémantique est vue comme celle de la similarité topologique en mathématiques, où on l'associe à une fonction, appelée fonction de la similarité. La définition celle-ci peut changer selon les approches selon les approches, selon les propriétés souhaitées. sa valeur est souvent comprise entre 0 et 1, ce qui permet des possibilités d'interprétation probabiliste de la similarité. Des propriétés ou des caractéristiques communes possibles de la fonction sont des caractéristiques positives, auto similaires ou maximales, symétriques ou réflexives. On peut aussi trouver d'autres caractéristiques telles que la finitude ou la transitivité [18].

**Définition 1** (Similarité). La similarité  $S:O \times O \rightarrow R$  est une fonction d'une paire d'entités à un nombre réel exprimant la similarité entre ces deux entités telle que : [4]

- $\forall a, b \in O, S(a, b) \geq 0$  (positivité)
- $\forall a, b, c \in O, S(a, a) \geq S(b, c)$  et  $S(a, a) = S(a, b) \Leftrightarrow a = b$  (autosimilarité ou maximalité)
- $\forall a, b \in O, S(a, b) = S(b, a)$  (symétrie)
- $\forall a, b, c \in O, S(a, b) = S(b, c) \rightarrow S(a, b) = S(a, c)$  (transitivité)
- $\forall a, b \in O, S(a, b) \leq \infty$  (finitude)

La dissimilarité (dissimilitude) est parfois utilisée au lieu de la similarité. Elle est définie de manière analogue à la similarité, sauf qu'elle n'est pas transitive : [20]

**Définition 2** (Dissimilarité). La dissimilarité  $DS:O \times O \rightarrow R$  est une fonction d'une paire d'entités à un nombre réel exprimant la dissimilitude entre ces deux entités telle que :

- $\forall a, b \in O, DS a, b \geq 0$  (positivité)
- $\forall a, b, c \in O, DS a, a \leq DS b, c$  et  $DS a, a = 0$  (minimalité)
- $\forall a, b \in O, DS a, b = DS b, a$  (symétrie)
- $\forall a, b \in O, DS a, b \leq \infty$  (finitude)

**Définition 3** (Distance). La distance  $D:O \times O \rightarrow R$  est une fonction de la dissimilitude satisfaisant la définitivité et l'inégalité triangulaire : [20]

- $\forall a, b \in O, D a, b = 0 \Leftrightarrow a=b$  (définitivité)
- $\forall a, b, c \in O, a, + D b, \geq D a,$  (inégalité triangulaire)

Les valeurs de similarité sont souvent normalisées pour pouvoir être combinées dans des formules plus complexes. Si la valeur de similarité et la valeur de dissimilitude entre deux entités sont normalisées, notées  $S$  et  $DS$ , alors on a  $S + DS = 1$  [20].

**Définition 4** (Normalisation). Une mesure est une mesure normalisée si les valeurs calculées par cette mesure ne peuvent varier que dans un intervalle de 0 à 1. Ces valeurs calculées sont appelées valeurs normalisées. Les fonctions du calcul sont appelées fonctions normalisées et notées  $f$  [14].

## II.6. Les techniques d'alignement d'ontologie :

### II.6.1. Les méthodes terminologiques :

Les méthodes terminologiques comparent les chaînes de caractères afin d'en déduire la similarité (ou dissimilitude) en exploitant les relations d'hyponymie ou d'hyponymie. Certaines méthodes terminologiques se basent sur la comparaison des chaînes de caractères et sont appelées les méthodes syntaxiques. Ces méthodes syntaxiques comparent les structures des deux chaînes. Deux chaînes qui partagent des caractères ou des mots en commun seront considérées comme similaires. D'autres méthodes terminologiques qui ont recours à une base de données lexicale, sous forme de réseau sémantique, sont les méthodes linguistiques. Ces méthodes calculent la similarité entre deux chaînes en fonction des relations qu'elles entretiennent [4].

Ces méthodes peuvent encore être divisées en deux sous-catégories : les méthodes syntaxiques et les méthodes linguistiques.

#### II.6.1.1. Les méthodes syntaxiques:

Les méthodes syntaxiques comparent, les termes ou les chaînes de caractères ou bien les textes, des entités à aligner. Ces méthodes permettent de calculer la valeur de la similarité des entités textuelles. Ces entités sont caractérisées par des noms, des étiquettes, des commentaires, des descriptions, ...etc. [4].

Ces méthodes se déclinent en deux sous-catégories : La première sous-catégorie englobe des méthodes qui comparent des termes en se basant sur les caractères contenus dans ces termes et La deuxième sous-catégorie utilise les distances basées sur les tokens [18].

### ❖ Les méthodes se basant sur les chaînes de caractères :

Ces méthodes analysent la structure des chaînes de caractères, l'ordre des caractères dans la chaîne, le nombre d'apparitions d'une lettre dans une chaîne pour concevoir des mesures de la similarité. Par contre, elles n'exploitent pas la signification des termes [4].

Les résultats de la comparaison des chaînes seront améliorés si chaînes passent une phase qu'on appelle la phase de normalisation, pendant laquelle ces chaînes sont « nettoyées » ou « traitées », avant de les fournir aux formules calculant la similarité [21].

Par exemple, les mesures dans cette catégorie retournent une grande valeur de similarité (jusqu'à 1) si elles comparent les termes « Voiture » et « voitures », mais une petite valeur, voire la valeur 0, si elles comparent les termes « voiture » et « bagnole ».

### ❖ Les distances basées sur les tokens :

Les mesures présentées ci-dessus s'adaptent bien lorsque l'on veut comparer deux termes ou deux courtes chaînes de caractères. Il existe aussi des cas où l'on a besoin de comparer des textes longs ou bien des documents textuels. Dans ces cas, ces entités sont découpées en plusieurs morceaux, appelés tokens. Elles deviennent des ensembles des tokens, et la similarité entre elles est produite grâce aux mesures de similarité basées sur des tokens [21].

#### II.6.1.2. Les méthodes linguistiques :

Les méthodes linguistiques utilisant des ressources externes (dictionnaires, taxonomies, vocabulaires,...). La similarité entre deux entités représentées par des termes peut aussi être déduite en les analysants à l'aide des méthodes linguistiques. Les méthodes linguistiques permettent de déterminer la similarité entre deux entités. Ces entités sont représentées par des termes (ou mots). Ces méthodes prennent en charge les propriétés expressives et productives du langage naturel qui peuvent être intrinsèques ou extrinsèques. [20].

### ❖ Les méthodes intrinsèques :

Les méthodes intrinsèques fonctionnent avec le principe de chercher la forme canonique ou représentative d'un mot ou d'un terme (lemme) à partir de ses variantes linguistiques (lexème).

Une même entité ou un même concept peut être référencé par plusieurs termes (synonymie) ou par plusieurs variantes d'un même terme [18].

Par exemple, le résultat de la mesure de similarité exacte de deux mots « ran » et « running » sera égal à 0 (c.-à-d. ils sont différents), alors que le résultat de la même mesure pour les lemmes de ces mots sera égal à 1, ce qui indique que « ran » et « running » sont similaires.

La similarité entre deux termes est mesurée en comparant leurs lemmes. La recherche du lemme d'un mot peut être effectuée à l'aide d'un dictionnaire. Une approche automatique utilise les *stemmers*. Le stemmer est un algorithme qui permet de déterminer la forme radicale d'un terme. Cette forme est déduite à partir d'une forme infléchie ou dérivée d'un mot donné. Les radicaux trouvés par les *stemmers* n'ont pas besoin d'être identiques à la racine morphologique du mot. Il suffit que les mots similaires soient associés à un même radical, même si ce radical n'est pas une racine de mot valide. La lemmatisation est une approche qui détermine le radical exact d'un mot. Ce processus extrait la catégorie lexicologique d'un mot et applique des règles de normalisation différentes pour chaque partie du discours. Cette approche nécessite la connaissance de la grammaire ainsi que les différentes règles d'un langage [4].

#### ❖ Les méthodes extrinsèques :

Ces méthodes calculent la valeur de similarité entre deux termes en employant des ressources externes telles que des dictionnaires, des lexiques ou des vocabulaires. La similarité est décidée grâce aux liens sémantiques déjà existants dans ces ressources externes tels que des liens synonymes (pour l'équivalence), des liens hyponymes/ hypernymes (pour la subsomption) [20] Par exemple, à l'aide des ressources des synonymes, « voiture » et « bagnole » sont dites similaires.

Typiquement, l'API WordNet est un système lexicologique, qui a été employé pour chercher des relations telles que la synonymie entre les termes. WordNet est aussi exploité pour calculer la distance sémantique entre les termes. Les ressources externes utilisées dans les méthodes extrinsèques peuvent aussi être des vocabulaires ou des dictionnaires multilingues, ou d'autres systèmes tels qu'EuroWordNet et Polylex. [4].

#### II.6.2. Les méthodes structurelles :

Ce sont des méthodes qui déduisent la similarité de deux entités en exploitant des informations structurelles lorsque les entités en question sont reliées aux autres par des liens sémantiques ou syntaxiques, formant ainsi une hiérarchie ou un graphe des entités. Nous appelons méthodes structurelles internes les méthodes qui n'exploitent que des informations concernant des attributs d'entité, et méthodes structurelles externes les autres qui considèrent des relations entre des entités [21].



### II.6.2.1. Les méthodes structurelles internes :

Ces méthodes calculent la similarité entre deux entités en exploitant des informations des structures internes de ces entités. Dans la plupart des cas, ce sont des informations concernant des attributs de l'entité, telles que des informations du Co-domaine des attributs, celles de la cardinalité des attributs, celles des caractéristiques des attributs (la transitivité, la symétrie), ou celles des autres types de restriction sur des attributs [20].

Par exemple, deux entités sont considérées comme similaires, si elles possèdent, une proximité dans les valeurs de données ou la cardinalité.

### II.6.2.2. Les méthodes structurelles externes :

Contrairement aux méthodes décrites dans les méthodes structurelles internes, qui exploitent des informations des attributs d'entité, les méthodes structurelles externes exploitent des relations entre des entités elles-mêmes, qui sont souvent des relations spécifique-générique ou hyponyme-hyperonyme (is-a), et la relation de composition meronymie-holonymie (partie-tout). Avec ces relations, les entités sont considérées dans des hiérarchies et la similarité entre elles est déduite de l'analyse de leurs positions dans ces hiérarchies. L'idée de base est que si deux entités sont similaires, leurs voisines pourraient également être d'une façon ou d'une autre similaires.

Cette observation peut être exploitée de plusieurs manières différentes en regardant des relations avec d'autres entités dans des hiérarchies. Deux entités peuvent être considérées similaires si :

- Leurs super-entités directes (ou toutes leurs super-entités) sont similaires.
- Leurs sœurs (ou toutes leurs sœurs, qui sont les entités ayant la même super-entité directe avec les entités en question) sont déjà similaires.
- Leurs sous-entités directes (ou toutes leurs sous-entités) sont déjà similaires.

Leurs descendants (entités dans le sous-arbre ayant pour racine l'entité en question) sont déjà similaires.

- Toutes (ou presque toutes) leurs feuilles (les entités de même type, qui n'ont aucune sous-entité, dans le sous-arbre ayant pour racine l'entité en question) sont déjà similaires.
- Toutes (ou presque toutes) les entités dans les chemins de la racine aux entités en question sont déjà similaires.

Des combinaisons de ces différents cas sont aussi possibles. Les approches externes rencontrent des difficultés dans le cas où les hiérarchies à aligner possèdent des niveaux différents de granularité [18].

### II.6.3. Les méthodes extensionnelles :

Les méthodes extensionnelles se basent sur l'analyse des extensions des entités (concepts ou classes), donc pour déduire la similarité entre deux entités on va traiter leurs ensembles des instances qui représentent leurs extensions.

On aura deux cas vu qu'on va comparer les ensembles des instances, le premier cas c'est où nous aurions une partie commune entre les deux ensembles, dans cette situation il existe des mesures qu'on peut appliquer pour qu'on puisse faire des opérations sur ces ensembles.

Il y a la mesure de Hamming adapté pour les mesures extensionnelles qui en général compte le nombre d'éléments différents entre deux ensembles à comparer [4].

**Définition 1** (Distance de Hamming , version adaptée pour les ensembles des instances). Soit S et T deux ensembles. La distance de Hamming (appelée aussi la différence symétrique) entre S et T est une fonction de la dissimilarité. [18]

$$\overline{DS}_{Hamming}(S, T) = \frac{|S \cup T - S \cap T|}{|S \cup T|}$$

**Définition 2** (Distance de Jaccard, version adaptée pour les ensembles des instances). Soit S et T deux ensembles. Soit P(x) la probabilité d'une instance aléatoire être dans l'ensemble X. La distance de Jaccard est une fonction de la dissimilarité

$DS_{jaccard} 2^E \times 2^E \rightarrow [0,1]$  Telle que : [18]

$$\overline{DS}_{Jaccard}(s, t) = 1 - \frac{P(S \cap T)}{P(S \cup T)}$$

Les mesures ci-dessus produisent la similarité de deux entités qui est en fait la similarité entre les deux ensembles de leurs instances en se basant sur la comparaison exacte des éléments dans deux ensembles. Dans le cas où les ensembles des instances ne partagent aucune partie commune, ces mesures ne sont plus applicables (le résultat retourné sera toujours égal à 1, c.à.d. les entités à comparer sont toujours différentes).

### II.6.4. Les méthodes Sémantiques :

Les méthodes sémantiques disposent d'un modèle théorique utilisé pour justifier leurs résultats. Par conséquent, elles sont considérées deux approches. La première approche repose

sur le raisonnement dans les logiques de descriptions tandis que la deuxième approche regroupe les méthodes de déduction afin de déduire la similarité entre deux entités [3].

#### **II.6.4.1. Les techniques basées sur les ontologies externes :**

Lorsque deux ontologies doivent être alignées, il est préférable que les comparaisons se fassent selon un capital de connaissances commun. Ce type de techniques s'intéresse à l'utilisation d'ontologie formelle intermédiaire pour répondre à ce besoin. Cette ontologie va définir un contexte commun pour les deux ontologies à aligner, et sera utilisée dans notre étude afin de compléter les techniques classiques d'alignement cités précédemment qui se basent sur la structure ou la richesse du langage de représentation des ontologies, et ne donnent pas un bon résultat quand les ontologies à apparier sont faiblement structurées ou se limitent à de simples hiérarchies de classification.

L'idée est que cette ontologie, avec une couverture appréciable du domaine d'intérêt des ontologies (ou une ontologie encore plus générale comme une ontologie de haut niveau), va permettre de lever le voile sur les ambiguïtés concernant les différentes significations possibles des termes. Des exemples d'ontologies intermédiaires: FMA<sup>1</sup> « the Foundational Model of Anatomy » et SUMO<sup>2</sup> « the Suggested Upper Merged Ontology » [20]

#### **II.6.4.2. Les techniques déductives :**

Les méthodes sémantiques se basent sur des modèles de logique (tels que la satisfiabilité propositionnelle (SAT), la SAT modale ou les logiques de descriptions) et sur des méthodes de déduction pour déduire la similarité entre deux entités. Les techniques des logiques de description (telles que le test de subsomption) peuvent être employées pour vérifier des relations sémantiques entre des entités telles que l'équivalence (la similarité est égale à 1), la subsomption (la similarité est de 0 à 1) ou l'exclusion (la similarité est égale à 0), et permettent donc de déduire la similarité de deux entités [18].

#### **II.6.5. Les méthodes d'alignement combinées :**

Une entité peut être considérée sous plusieurs différents aspects, soit en s'appuyant sur son nom, soit sur ses attributs, ou soit sur ses relations avec d'autres entités. La similarité entre deux entités peut donc être calculée en se basant sur plusieurs aspects ; les caractéristiques d'une entité sont comparées avec les caractéristiques correspondantes d'une autre par une des mesures de similarité de base présentées dans la section II.5.2, cela retourne une valeur de la similarité (ou de la dissimilitude /distance) [20].

<sup>1</sup> <http://sig.biostr.washington.edu/projects/fm/>

<sup>2</sup> <http://www.ontologyportal.com>

Il faut donc un moyen pour combiner toutes les valeurs de similarité calculées de chaque aspect pour produire une seule valeur de similarité représentative pour deux entités à comparer. Cette partie analyse quelques approches existantes dans la littérature.

**Définition1 (Distance de Minkowski):** Soit  $O$  l'ensemble d'objets qui peuvent être analysés dans  $n$  dimensions. Soit  $x$  et  $y$  deux objets dans  $O$ . La distance de Minkowski entre  $x$  et  $y$  est une fonction de la dissimilitude  $DS_{Minkowski} : O \times O \rightarrow R$  telle que : [22]

$$DS_{Minkowski}(x, y) = \sqrt[p]{\sum_{i=1}^n DS(x_i, y_i)^p}$$

Cette distance est une mesure généralisée avec différentes valeurs de  $p$ ,  $p \geq 1$ .

Quand  $p$  est égale à  $1$ , elle devient la distance de « city block » et quand  $p = 2$  elle devient la distance euclidienne. La distance de chebychev (appelée aussi la distance de valeur maximum) est un cas spécial de la distance de minkowski avec  $p = \infty$  :

$$DS_{chebychev}(x, y) = \max_i DS(x_i, y_i)$$

Cette mesure n'est une fonction linéaire que quand  $p = 1$  ou  $p = \infty$ . Dans le cas où  $p = 1$ , une variante de cette mesure avec des poids est souvent utilisée. Le bon côté de cette variante est que nous pouvons contrôler l'influence (ou l'importance) de chaque dimension sur la valeur finale de la distance. Les dimensions plus importantes seront associées avec les poids plus élevés, donc les valeurs de ces dimensions influenceront mieux à la valeur agrégée finale [21].

**Définition2 (Somme pondérée) :** Soit  $O$  l'ensemble d'objets qui peuvent être analysés dans  $n$  dimensions. Soit  $x$  et  $y$  deux objets dans  $O$ . Soit  $w_i$  le poids de la dimension  $i$ . Soit  $DS(x_i, y_i)$  la dissimilitude de la paire des objets à la dimension  $i$ . La somme pondérée entre  $x$  et  $y$  est une fonction de la dissimilitude  $DS_{sp} : O \times O \rightarrow R$  telle que : [4]

$$DS_{sp}(x, y) = \sum_{i=1}^n w_i * DS(x_i, y_i)$$

En général,  $\sum_{i=1}^n W_i = 1 \rightarrow$  Version normalisé .

Une autre mesure similaire à la somme pondérée est le produit pondéré. Cependant, un inconvénient de cette mesure est que le résultat sera égal à  $0$  si une des dimensions est égale à  $0$ .

**Définition3 (Produit pondéré) :** Soit  $O$  l'ensemble d'objets qui peuvent être analysés dans  $n$  dimensions. Soit  $x$  et  $y$  deux objets dans  $O$ . Soit  $w_i$  le poids de la dimension  $i$ . Soit  $DS(x_i, y_i)$  la dissimilarité de la paire des objets à la dimension  $i$ . Le produit pondéré entre  $x$  et  $y$  est une fonction de la dissimilarité  $DS_{sp} : O \times O \rightarrow \mathbb{R}$  telle que : [4]

$$DS_{sp}(x, y) = \prod_{i=1}^n DS(x_i, y_i)^{w_i}$$

Toutes les approches, qui combinent des valeurs de similarité calculées par différentes mesures, emploient la méthode de la somme pondérée (lien vers le paragraphe). Cependant, certaines approches (par exemple Anchor-PROMPT) [1] déduisent des alignements en examinant des critères heuristiques sans utiliser des méthodes de combinaison des similarités.

## II.7. approches d'alignement d'ontologies :

Pour aligner deux ontologies, on doit chercher des correspondances entre eux en utilisant les méthodes de calcul de similarité. Il existe plusieurs méthodes et techniques pour calculer la similarité, aussi différentes façon pour combiner les résultats de similarité. Donc la diversité d'approches vient du fait que chaque approche utilise les méthodes adéquates selon les types d'ontologies à aligner. Nous allons présenter quelques approches :

### ✓ Une approche utilisant WordNet et une nouvelle mesure structurelle :

Cette technique commence par un prétraitement sur les termes de chaque ontologie, ce qui est important pour améliorer les résultats d'alignement, selon les réalisateurs de cette approche. Cette approche se compose de trois phases principales, la première est basée sur les chaînes de caractères plus particulièrement sur la mesure de Levenshtein qui sert à déterminer les relations d'équivalence et l'heuristique d'inclusion des labels afin de déterminer les relations spécialisations entre les concepts. La deuxième phase se focalise sur l'exploitation d'une ressource externe, le WordNet, ainsi que l'utilisation de la synonymie, pour qu'on puisse déduire les relations d'équivalence entre les concepts dont les labels ne sont forcément pas similaires, syntaxiquement parlant. La dernière phase exploite les résultats obtenus dans les deux phases précédentes en utilisant une technique structurelle, et elle combine la structure des deux ontologies ainsi que WordNet pour trouver d'autres possibles alignements. Lors de la phase structurelle, pour générer des correspondances, on a proposé une nouvelle mesure structurelle entre deux ontologies qui n'est qu'une généralisation de la mesure de Wu & Palmer [23].

### ✓ **ASMOV :**

Asmov (Automated Semantic Mapping of Ontologies with Validation) comme son nom l'indique est un nouvel algorithme qui automatise le processus d'alignement ainsi que l'intégration d'un unique processus de validation sémantique afin de s'assurer que l'alignement obtenue ne contient pas d'incohérences sémantiques.

L'algorithme ASMOV calcule la similarité entre les entités pour une paire d'ontologies selon quatre caractéristiques : éléments lexicaux (id, label, commentaires), structure relationnelle (hiérarchie : ascendant-descendant), structure interne (restrictions des propriétés des concepts ; types, domaines, ranges ; data-values pour les individus), extension (les instances des classes et les valeurs des propriétés). Et enfin, les différentes mesures obtenues en comparants ces quatre caractéristiques sont fusionnées en une valeur unique utilisant une somme pondérée.

Pour le calcul de la similarité lexicale, ASMOV, utilise par défaut WordNet. Cependant, autres dictionnaires peuvent être mieux appropriés pour d'autres domaines ; par exemple, UMLS est mieux convenable pour les ontologies médicales que le WordNet. Par conséquent, ASMOV définit les interfaces qui permettent aux utilisateurs experts d'implémenter des adaptateurs pour d'autres sources de connaissances. ASMOV inclut les adaptateurs WordNet et UMLS.

Et les résultats des différents benchmarks qui ont été réalisés montrent qu'ASMOV est un outil pratique pour les applications du monde réel qui demandent des alignements d'ontologies à la volée [4].

### ✓ **ROMIE :**

ROMIE (Resource based Ontology Mapping within and Interactive and Extensible environment), est une approche qui regroupe les méthodes d'alignement syntaxiques, linguistiques, structurelles et sémantiques pour offrir un alignement semi-automatique.

Cette approche se divise en deux phases : la phase d'enrichissement sémantique et la phase d'alignement. La première est basée sur l'analyse des informations développées par les ontologies à comparer (ressources web, données, documents, ...) et qui sont associées aux concepts de l'ontologie. A la fin de la phase d'enrichissement, une ontologie contient plus de relations sémantiques entre les concepts qui seront exploitées dans la deuxième phase. La phase d'alignement, prend deux ontologies et calcule la similarité entre les couples de chaque concept d'ontologies en employant plusieurs méthodes syntaxiques qui calculent par exemple la distance d'édition entre deux concepts, et/ou encore des méthodes linguistiques basées sur

le dictionnaire WordNet, par exemple. Puis selon les valeurs de similarité retournées, on génère un ensemble d'alignements candidats et en utilisant des méthodes de filtrage, on élimine les alignements les moins pertinents. Ces méthodes sont basées sur des relations structurelles et sémantiques entre les concepts. En fin l'ensemble des hypothèses d'alignements fournis par l'étape précédente est ordonné selon les valeurs de similarité [20].

✓ **OLA :**

OLA pour OWL-Lite Alignement est un algorithme pour aligner des ontologies représentées en OWL-Lite. Cet algorithme utilise les différentes méthodes de calcul de similarité pour trouver les correspondances entre les entités de deux ontologies en se basant sur leurs caractéristiques et leurs rapports avec d'autres entités et pour combiner ces valeurs de similarités calculées pour chaque paire d'entités il utilise la somme pondérée des valeurs de similarité de chaque caractéristique [4].

✓ **QOM :**

L'approche QOM (Quick Ontology Matching) est basée sur l'approche NOM (Naive Ontology Matching) cette dernière a montré qu'elle est effective, mais inefficace, donc on a pensé à l'optimiser, la chose qui a donné naissance à l'approche QOM. Cette approche comme son prédécesseur exploite RDF-triples, et pour mesurer la similarité entre les ontologies à aligner elle utilise la similarité terminologique et structurelle. Et, après la phase de l'analyse de la similarité et ses interprétations, de nouvelles décisions doivent être prises, comme : quels alignements candidats doit-on ajouter à l'agenda pour l'itération suivante.

Cette approche a montré de très bons résultats en termes de temps d'exécution ( $n \cdot \log(n)$  au lieu de  $n^2$ ,  $n$  étant le nombre d'entités dans les ontologies) en comparaison avec d'autres approches de la même classe de complexité [4]

✓ **ASCO2 :**

L'algorithme ASCO2 propose un modèle de calcul de similarité sur deux étapes, la similarité partielle et la similarité finale. La similarité partielle entre deux entités des deux ontologies est déduite entre les composantes correspondantes aux entités en question. Ces composantes sont des pièces de connaissance contenues dans les définitions de l'entité en employant des primitives du langage OWL. Les valeurs de similarité partielle sont ensuite agrégées dans un schéma de pondération variable pour obtenir une meilleure valeur de similarité finale de ces deux entités [14].

**✓ Anchor-PROMPT:**

Anchor-PROMPT est un algorithme qui permet de trouver des concepts qui sont sémantiquement similaires dans des ontologies différentes, en se basant sur : une liste des paires d'ancres comme entrée de l'algorithme, ces paires d'ancres sont les paires de concepts similaires définis par les utilisateurs ou trouvés en utilisant des méthodes de correspondance lexicologique. Aussi sur le graphe qui représente l'ontologie, où les nœuds sont des concepts (classes) et les arcs représentent les relations (slots) entre les concepts, cet algorithme fait une différence entre les relations de type «is-a » et les autres relations car les classes liées par « is-a » constituent des groupes d'équivalences qui sont considérés comme un seul nœud. L'algorithme analyse les chemins dans les sous-graphes limités par les ancres pour déterminer quels concepts apparaissent fréquemment en positions similaires sur les chemins similaires. Ces classes sont susceptibles de représenter les concepts qui sont sémantiquement similaires. Les résultats de cet algorithme sont liés à la taille de la liste des paires d'ancre initiale et la longueur maximale du chemin. Cet algorithme souffre ou rencontre des problèmes si les structures des ontologies sont différentes car il ne trouvera pas les chemins similaires avec la même longueur et si les noms de relations sont différents [21].

**✓ EDOLA :**

Transforme les ontologies sous forme d'un graphe OWL-Graphe. Les nœuds du graphe représentent les six types d'entités qui existent dans une ontologie OWL-Lite : les concepts, les instances des concepts, les types de données, les valeurs des types de données et les propriétés des classes. Les relations entre les entités au niveau de l'ontologies OWL-Lite sont les arcs entre les nœuds [24].

Le graphe OWL-Graphe permet de représenter quatre catégories de liens de spécialisation, d'attribution, d'instanciation et d'équivalence [25].

**✓ SODA :**

Transforme les ontologies sous forme d'un graphe DL-Graphe. Les nœuds du graphe sont les entités de l'ontologie : les classes, les propriétés et les instances. Les arcs du graphe décrivent les relations qui existent entre ces différentes entités [26]

Structural Ontology OWL-DL Alignement), implémente un nouvel algorithme d'alignement d'ontologies OWL-DL. La nouvelle méthode d'alignement repose sur le calcul des mesures de similarité. Il définit deux modèles de calcul de similarité (locale et globale).



La méthode combine les mesures de similarité locale (terminologique et structurelle) pour l'évaluation de la similarité globale. Elle permet de générer un alignement exploitant l'aspect structurel du voisinage des entités à appairer.

La méthode SODA est performante sur les petites et moyennes ontologies et n'est pas encore très bien adaptée pour les ontologies de grande taille [21].

✓ **FALCON :**

utilise les graphes RDF directs bipartis pour représenter les ontologies. Ces graphes sont dérivés des graphes RDF bipartis .La figure II.5 montre le graphe RDF et le graphe RDF biparti d'une ontologie  $O_A$ . [24]

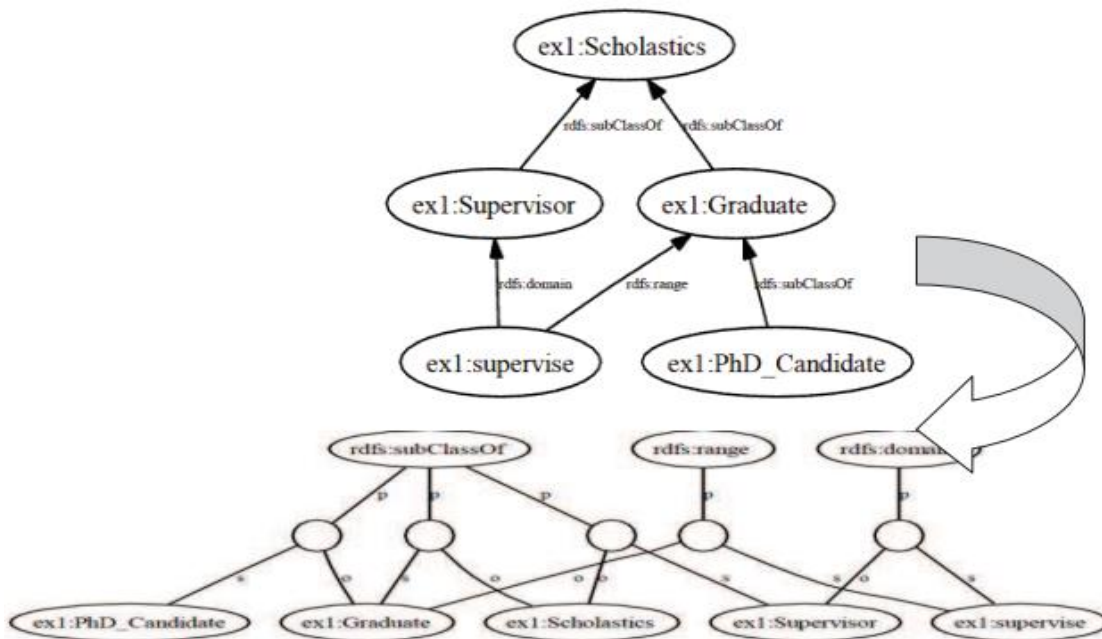


Figure II.6 : Le graphe RDF et le graphe RDF biparti d'une ontologie  $O_A$  [24]

**II.8. étude comparative des méthodes d'alignement :**

Les approches	Méthodes terminologiques	Méthodes structurelles	Méthodes extensionnelles	Méthodes sémantiques	Méthodes combinées
ASMOV	comparaison lexicale des étiquettes, descriptions, tokenisation,	comparaison des nœuds parents/enfants. Comparaison des types,	comparaison des instances classes/propriétés		somme pondérée

	WordNet	domaines, ranges. Mesure de Wu et Palmer			
OLA	termes comparaison des URIs, étiquettes, noms comparaison des types	comparaison des nœuds voisins			somme pondérée
QOM	String- comparaison des URIs et étiquettes	comparaison des ensembles des voisines directes des concepts / relations	comparaison des ensembles d'instances concepts / relations		Somme pondérée, Fonction de Sigmoid
Anchor- PROMPT	String- comparaison des classes/reliations	comparaison des positions			
SODA	Jaro-Winkler pour les noms et labels, Monge- Elkan pour les commentaires	comparaison des nœuds voisins			somme pondérée

**Tableau II.1 :** Quelques techniques et méthodes utilisé par les approches d'alignement d'ontologie [25]

La comparaison des méthodes d'alignement peut se faire sur plusieurs volets. Le premier volet permet de les comparer en se basant sur leurs caractéristiques externes. Ces caractéristiques regroupent principalement les entrées et les sorties des méthodes. Le second volet étudie les techniques de base utilisées par les méthodes d'alignement pour le calcul de la similarité. Les méthodes de similarité utilisées renseignent sur les éléments exploités au

niveau de l'ontologie afin de produire l'alignement. Le troisième volet indique les stratégies d'alignement exploitées par les méthodes d'alignement. Ces stratégies décrivent éventuellement la démarche exploitée par la méthode pour déterminer les correspondances entre les différents constituants des deux ontologies à aligner [11].

### II.8.1. Comparaison des caractéristiques externes :

Les méthodes que nous avons choisies dans notre comparaison, sont souvent citées dans la littérature comme des méthodes d'alignement d'ontologie dans le domaine du web sémantique. La plupart de ces méthodes ont été conçues pour aligner des ontologies RDFS/OWL et elles produisent en sortie des alignements au format RDF/XML. La table I présente leur comparaison selon les dimensions externes [26].

La table II.2 regroupe les principales méthodes d'alignement citées dans la littérature.

Méthode	Types d'ontologies	Relation	Cardinalité
Anchor-Prompt	RDF(S), OWL	↔	(0, n)-(0, n)
ASCO1	RDF(S)	↔	(0, n)-(0, n)
AUTOMS	RDF(S), OWL	↔	
COMA	SR, XML	↔	(0,1)-(0,1), (0,1)-(0, n), (0, n)-(0, n)
Cupid	XML, SOO	↔	(0, n)-(0, n)
GLUE	HT	↔	
GMO	RDF(S), OWL	↔	
H-match	RDF(S), OWL	↔	0,1)-(0,1), (0,1)-(0,n)
OLA	RDF(S), OWL	↔	(0,1)-(0,1)
oMAP	HT	↔	(0,1)-(0,n)
OpIMap	RDF(S), OWL	↔	(0,n)-(0,1)
QOM	HT, RDF(S), OWL	↔	(0,1)-(0,1)
RiMOM	HT	↔	(0,1)-(0,1), (0,1)-(0,n)
SBI-Hical	HT, RDF(S), OWL	↔	(0,n)-(0,n)
SCM	SR, SOO, XML, RDF	↔	(0,1)-(0,n)

SF	HT, RDF(S), OWL	$\Leftrightarrow$	(0,1)-(0,1)
S-match,	HT, RDF(S), OWL	$\Leftrightarrow$	

**Tableau II.2 :** Caractéristiques externes des principales méthodes d'alignement [26]

Elle décrit pour chaque méthode ses caractéristiques générales, c.-à-d le type de l'ontologie à aligner, le type des relations pris en considération et éventuellement les cardinalités des alignements produits.

Plusieurs méthodes d'alignement ont été proposées. Ces méthodes permettent l'alignement des hiérarchies textuelles (notées HT) et des ontologies. Les ontologies, à aligner, peuvent être décrites avec plusieurs langages de description. Ces langages regroupent le RDF(S) ou les différentes déclinaisons de l'OWL (OWL-Lite et OWL-DL).

D'autres méthodes d'alignement assurent l'alignement des structures. Ces structures décrivent des schémas de bases de données (schémas relationnels, notés SR, et les schémas orientés objets, notés SOO) ou même des documents XML. En effet, les méthodes Cupid et SIMILARITY FLOODING (notée SF dans la table II.2) sont considérées comme des méthodes d'alignement d'ontologies. Elles sont aussi considérées comme des méthodes représentatives des approches structurelles d'alignement d'ontologies. La méthode COMA a été développée pour obtenir un maximum de flexibilité. En effet, elle est adaptée à de nombreux types de schémas.

La table II.3 illustre, pour chaque méthode le type de composition utilisée, la technique de sélection exploitée et éventuellement les post-traitements effectués. La composition permet la combinaison de plusieurs techniques d'alignement ou plusieurs niveaux d'information. La technique de sélection regroupe le seuillage ou l'optimisation (maximisation ou minimisation) locale de l'appariement.

Les méthodes d'alignement d'ontologies reposant seulement sur des approches structurelles (voir table II.3) utilisent généralement une composition linéaire. La méthode oMap est basée sur un ordre de préférence entre les différentes techniques utilisées.

Méthode	Composition	Sélection	Post-traitement
Anchor-Prompt	Linéaire (exploite un alignement d'entrée)	Seuil	-
ASCO1	Moyenne pondérée (deux niveaux : terminologique et	Seuil	-

	structurel)		
AUTOMS	-	Maximisation locale	-
CAIMAN	-	Maximisation locale	-
COMA	Statistique (maximum, moyenne, etc.)	Seuil ou maximisation Locale	-
Cupid	Moyenne pondérée	Seuil	-
GLUE	Moyenne pondérée	-	Relaxation des labels
H-match	Moyenne pondérée	Seuil, ou maximisation Locale	-
OLA	Moyenne pondérée	Seuil, maximisation Globale	-
Omap	Priorités entre méthodes	Maximisation locale	-
OplMap	Moyenne pondérée	Seuil	Consistance
QOM	Moyenne pondérée, moyenne sigmoïde	Seuil et maximisation Locale	-
RiMOM	Moyenne pondérée, moyenne sigmoïde	Seuil et minimisation Locale	Réduction de la cardinalité, consistance
SBI-Hical	-	Seuil	
SCM	-	Maximisation locale	Consistance
SF	Linéaire (exploite un alignement d'entrée)	Seuil	Stabilisation
S-match, Ctx-match	Union ensembliste	Seuil, résolveur SAT	-

**Tableau II.3 :** Comparaison des techniques de composition **et** de sélection des principales méthodes d'alignement [11]

Les méthodes d'alignement reposent, dans la majorité des cas, sur une sélection exploitant un seuil ou sur une optimisation locale de l'appariement. La méthode OLA utilise une approche globale de maximisation de la similarité de l'alignement. La méthode RiMOM fonctionne avec une minimisation locale du risque.

### II.8.2. Comparaison des techniques de base :

La table II.4 présente une étude comparative qui montre les principales techniques de base utilisées par chaque méthode d'alignement. Ces techniques regroupent les méthodes terminologiques, les méthodes structurales, les méthodes extensionnelles et les méthodes sémantiques. Parmi les méthodes présentées dans la table II.4, la majorité sont des méthodes se basant sur les techniques de base. Les méthodes GLUE, V-Doc, oMap, OplMap, QOM et RiMOM utilisent également, en plus des techniques de bases, des stratégies d'alignement [11].

Méthode	Techniques terminologiques		Techniques structurales		
	Syntaxique	Linguistique	Attribut	Relation d'ordre	Relation transversale
Anchor-Prompt					Similarité des chemins
ASCO1	Jaro-Winkler (identi_cateurs et labels), TF-IDF (commentaires	Similarité synatxique (Jaro-Winkler) entre synset WordNet, tokenisation	Proposition d'entités du voisinage (subsumés, subsumantes, soeurs) qui sont similaires (à partir des similarités terminologiques)		
AUTOMS		LSA sur thésaurus WordNet			
COMA	A_xe, distance d'édition, n-gram, soundex	Synonymie (WordNet), Tokenisation	Table de compatibilit	Moyenne sur les subsumés et les feuilles	
Cupid		Basé sur un thésaurus	Table de compatibilit	Proportion similaire	d'attributs
GLUE				Prise en compte des	

				subsumants et descendants par la relaxation des labels	
QOM	Égalité, distance d'édition		Égalité des types de Données	SimSet sur subsumants, subsumés et sœurs	SimSet sur relations Directes

**Tableau II.4 :** Comparaison des techniques de combinaison et de sélection des principales méthodes d'alignement [11]

### II.8.3. Comparaison des stratégies d'alignement :

La table II.5 donne pour chaque méthode exploitant une stratégie d'alignement recensée, le type de prétraitement qu'elle réalise. La table II.4 fournit aussi la technique utilisée et la mesure sur laquelle s'appuie la comparaison des extensions. Les méthodes OplMap et OMap n'utilisent pas de mesures de comparaison étant donné que le processus de classification est implicite. En effet, avec ces méthodes, les valeurs des prédictions sont combinées (par une moyenne pondérée).

Au niveau des prétraitements utilisés, les méthodes recensées dans la table II.4 utilisent des prétraitements, qui assurent la classification des instances par l'intermédiaire d'un classifieur bayésien naïf (ou une amélioration), ou par la réindexation vectorielle. La méthode QOM n'utilise pas de prétraitements étant donné que la comparaison des extensions repose sur l'agrégation des valeurs de similarités individuelles [11].

Méthode	Prétraitement	Technique	Extensions
CAIMAN	Réindexation vectorielle	TF IDF – Rocchio	Cosinus
GLUE	Classification	Réseau bayésien naïf sur le contenu et chemin de noms	Jaccard
V-Doc	Réindexation vectorielle	TF IDF	Cosinus
Omap	Classification	Réseau bayésien naïf	

OpIMap	Classification	Réseau bayésien naïf et k-plus proches voisins	Consistance
QOM			SimSet sur les instances directes et celles subsumés directs
RiMOM	Classification	Réseau bayésien naïf	
SBI-Hical	Classification	Adaptation du réseau bayésien naïf aux hiérarchies	Kappa
SCM	Réindexation vectorielle	Kullback-Leibler avec espace vectoriel non euclidien	Cosinus

**Tableau II.5 :** Comparaison des techniques de combinaison et de sélection des principales méthodes d'alignement [11]

## II.9. Conclusion :

Dans ce chapitre, nous avons présenté l'état de l'art sur l'alignement des ontologies; quelques domaines d'application de l'alignement des ontologies et les dimensions de l'alignement, puis nous avons présenté les techniques et les méthodes utilisées dans la littérature qui visent le problème de recherche de la similarité, de la dissimilitude ou de la correspondance entre deux entités en général, qui apparaissent dans des ontologies. L'alignement d'ontologies revêt toute son importance dans des applications nécessitant la prise en compte d'une interopérabilité sémantique. Plusieurs approches d'alignement d'ontologies existent dans la littérature. Elles sont basées sur les mesures de similarités.

Les techniques d'alignement présentées sont les outils dont disposent les concepteurs de systèmes d'alignement pour trouver des solutions aux problèmes de l'hétérogénéité des ontologies.

Dans le chapitre suivant, nous présentons notre méthode d'alignement d'ontologies structurelle et lexicale.



**Chapitre III**  
**Algorithme**  
**d'alignement**

### III.1. Introduction :

Le domaine biomédical est caractérisé par l'existence de nombreux standards terminologiques, thesaurus, et langages, partagés par les communautés biomédicales, qu'ils soient généralistes ou dédiés à un domaine de spécialité. Les ontologies doivent fournir les concepts et les relations utilisés pour le marquage sémantique des données en vue du Web sémantique avec une signification partagée et réutilisable pour différentes applications et différents usagers.

Dans ce chapitre, nous présentons notre méthode d'alignement d'ontologies, qui exploite les caractéristiques du langage OWL ; ce langage décrit et structure les deux ontologies existante dans le domaine biomédicale que nous avons choisi pour tester notre système ainsi que déduire les valeurs de similarités entre les entités de deux ontologies avec quelque d'exemples de calcul.

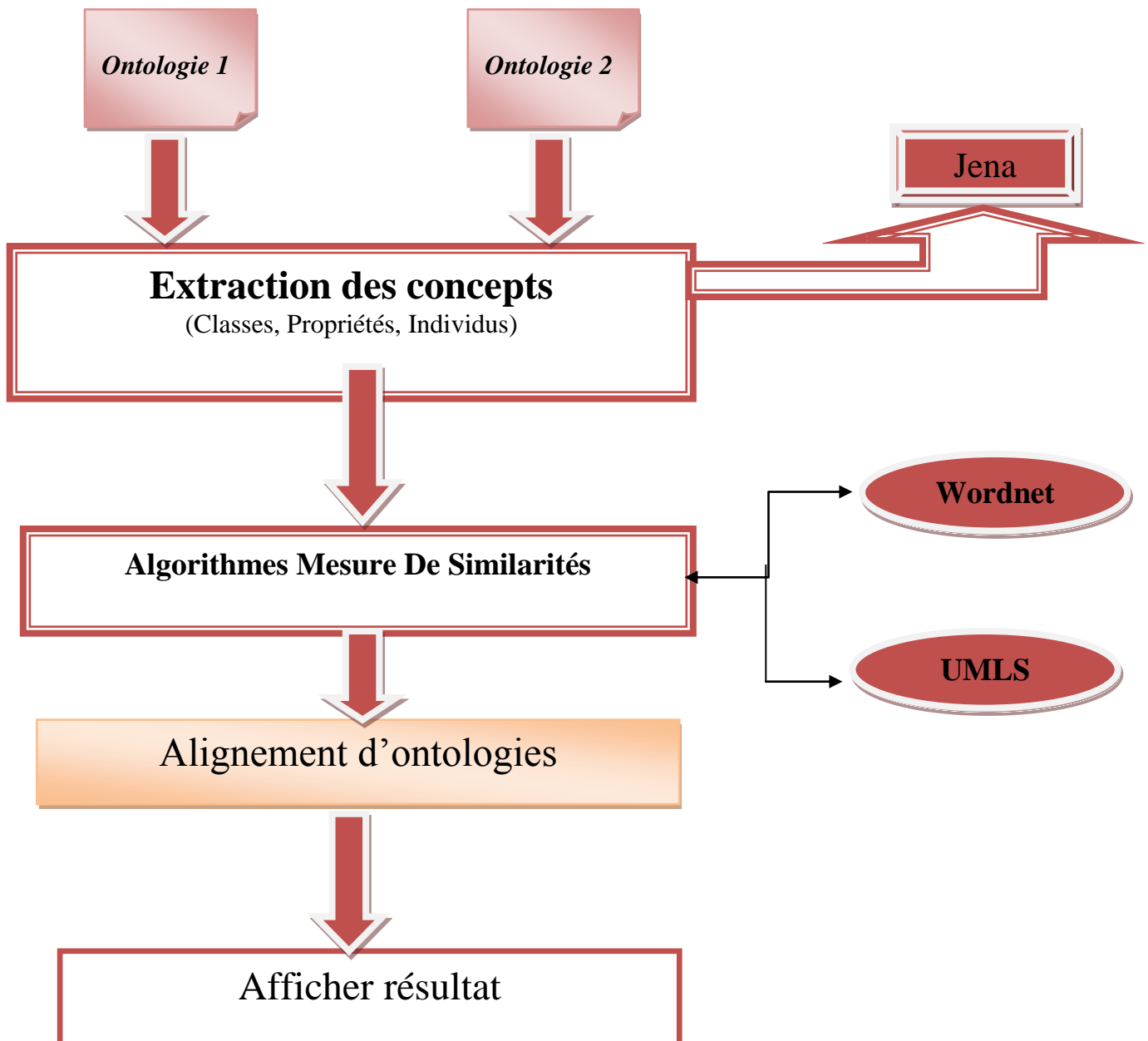
### III.2. Méthode d'alignement:

Pour l'alignement on a utilisé deux ontologies existante l'un sur la maladie asthme et l'autre sur la maladie autisme.

Notre application s'articule sur 7 grandes lignes :

- ✚ présentation du système d'alignement (schéma global, les entrées, les sorties, autres ressources : WordNet, UMLS).
- ✚ description de l'algorithme d'alignement (levenshtien) qui a pour entrées : deux ontologies O1, O2 sous forme owl.
- ✚ extractions des concepts de chacune et les stocke sous forme de vecteurs
- ✚ calculer la similarité terminologique entre les couples de concepts en utilisant l'algorithme de levenshtein .
- ✚ calculer la similarité lexicale en exploitant WordNet.
- ✚ calculer la similarité lexicale en exploitant UMLS.
- ✚ agrégation des mesures de similarité (combinaisons des mesures précédentes pour calculer la mesure de similarité globale).

En sortie, le résultat est un vecteur de quadruplets contenant concept 1 en provenance de l'ontologie O1 et concept 2 en provenance de l'ontologie O2, le type des concepts à aligner (classe, propriété, instance, ...), et la mesure de similarité (valeur [0,1]).



**Figure III.1 :** Méthode d'alignement

### III.2.1. L'UMLS et ontologie :

Unified Medical Language System, que l'on pourrait traduire par « Système d'unification de la langue médicale », est un produit terminologique extrêmement riche, construit de façon pragmatique, et que l'on peut appréhender de façon multiple. Nous donnons un aperçu de ce qu'est l'UMLS et mettons l'accent sur deux de ses aspects potentiellement antinomiques : sa relation aux ontologies et sa relation à la langue.

### III.2.2. Formats en entrée et en sortie :

Le type de données utilisé doit être précisé pour chaque méthode d'alignement.

Les ontologies à aligner peuvent être représentées avec des langages à balises. Les langages à balises sont XML, RDF(S), et OWL.

Les dictionnaires de synonymies ou de lexiques sont des informations supplémentaires pouvant parfois s'ajouter et qui sont nécessaires pour l'amélioration du rendu du processus d'alignement.

Le format et la structure du résultat de l'alignement sont précisés pour chaque méthode. Il faut préciser si l'alignement s'effectue entre les structures entières ou entre couples d'entités des deux ontologies. Le résultat pour la majorité des méthodes existantes est un fichier d'alignement (généralement en format XML), indiquant quelles sont les couples entités ontologiques qui correspondent. Toutes les méthodes d'alignement déterminent des correspondances entre les entités ontologiques en utilisant des mesures de similarité.

### III.2.3. Le problème de calcul des similarités :

Le calcul de la similarité sémantique entre les concepts et entre les phrases est un problème de longue durée dans le domaine du traitement du langage naturel.

Il existe plusieurs moyens pour évaluer la similarité entre deux entités. La technique la plus répandue consiste à définir une formulation mathématique pour évaluer le degré de similarité. Le champ d'analyse sémantique a un rôle crucial à jouer dans la recherche liée à l'analytique du texte. La similarité sémantique diffère lorsque le domaine d'opération diffère.

Une méthode pour calculer la similarité sémantique entre les mots et aussi entre les phrases peut être appliquée dans une variété de domaines. Chaque domaine nécessite des algorithmes différents et spécifiques bien que l'idée de base du calcul de la similarité sémantique reste la même. Pour déterminer la proximité entre les concepts en comparaison, nous avons besoin de mesure standard prédéfinie qui décrit aisément ces connexités de la signification. Les bases de données lexicales entrent dans l'image à ce stade de traitement.

Les bases de données lexicales ont des liens entre les mots qui peuvent être utilisés pour déterminer la similarité sémantique des mots.

Les valeurs de similarité calculées sont stockées dans une base de similarités (matrice). Elles sont calculées par des mesures normalisées, elles sont donc comprises dans l'intervalle de [0,1]. Ces valeurs de similarité sont ensuite agrégées par différentes stratégies de

combinaison pour atteindre une seule valeur de similarité finale entre deux entités, qui est aussi comprise entre 0 et 1. En examinant cette valeur de similarité finale par rapport à un seuil prédéfini, deux entités sont considérées comme similaires ou différentes.


Dans la section suivante nous allons présenter les méthodes d'alignement d'ontologies qui se base essentiellement sur les bases de données lexicales.


### III.3. La méthode utilisée pour le calcul de la distance :

Le système que nous traitons appartient à un domaine médical. La méthode d'alignement permet d'aligner deux ontologies décrites en format OWL. Elle produit un alignement représenté sous la forme d'un fichier texte.

En utilisant la base de données lexicale, la similarité est calculée en utilisant une hiérarchie prédéfinie, le sens et la relation avec d'autres concepts qui sont stockés dans une structure arborescente. Lors de comparaison de deux concepts, il prend en compte la distance de chemin entre les mots.

Les méthodes utilisées pour le calcul de la distance entre les concepts :

 **Methode de Levenshtein** : est une mesure avec un nombre minimal d'opérations d'insertions, de suppressions et de substitutions de caractères nécessaires pour la transformation d'une chaîne en une autre. La distance de levenshtein est une distance d'édition où tous les poids sont égaux à 1.

 **Wordnet** : c'est une ressource lexicale de langue anglaise, qui regroupe des termes (noms, verbes, adjectifs et adverbes) en ensembles de synonymes appelés synsets, elle regroupe tous les termes dénotant un concept donné. Les synsets sont reliés entre eux par des relations sémantiques: relation de généralisation /spécialisation, relation composant/composé. Les techniques basées sur les chaînes de caractères ne sont pas suffisantes quand les concepts sont sémantiquement proches et quand leurs noms sont différents, l'interrogation d'une ressource linguistique tel que WordNet peut indiquer que les concepts sont similaires. Pour le calcul de la similarité linguistique la fonction  $Syn(c)$  calcul l'ensemble des Synsets de WordNet du concept  $c$ ; soit  $S = Syn(c1) \cap Syn(c2)$  l'ensemble des sens communs

$$\lambda(S) = \frac{|Syn(c1) \cap Syn(c2)|}{\min(|Syn(c1)|, |Syn(c2)|)}$$

Soit  $\min(|Syn(c1)|, |Syn(c2)|)$  le minimum entre les cardinalité des deux ensembles  $Syn(c1)$  et  $Syn(c2)$  alors la similarité entre deux concepts  $c1$  et  $c2$  sera définie Comme suit :

$$\text{Sim}_{\text{ling}}(c1, c2) = \frac{\lambda(S)}{\min(|\text{Syn}(c1)|, |\text{Syn}(c2)|)}$$

Cette mesure retourne 1.0 si au moins c1 est le seul synonyme de c2 ou c2 est le seul synonyme de c1.

### III.3.1 Description de la méthode levenshtein pour calculer la similarité :

**Distance** (string1, string2)

**Début**

len1 = len(string1)

len2 = len(string2)

**Si** len1 = 0 et len2 = 0 **alors** return 0.0

**FinSi**

**Si** len1 == 0 ou len2 == 0:

return 1.0

**FinSi**

best = 0

i = 0

**Tant que** i < len1 et len1 - i > best **Faire**

j = 0

**Tant que** len2 - j > best **Faire**

k = i;

**Tant que** j < len2 and string1[k] != string2[j] **Faire**

j = j + 1

**Si** j != len2 **alors**

j = j + 1

k = k + 1

```

    Tant que j < len2 et k < len1 et string1[k] = string2[j]:
        j = j+1
        k = k+1
        best = max (best, k-i)
    FinSi
    i = i+1
    return 1.0 - ((2 * best) / (len1 + len2))
Fin
```

#### III.4. Conclusion :

Dans ce chapitre, nous avons présenté l'algorithme de levenshtein, l'approche utilisée pour aligner les ontologies en entrées qui sont représentées en OWL, le résultat est la liste des correspondances entre les entités des deux ontologies à aligner. Cet algorithme fait l'extraction de tous les concepts pour chaque ontologie et le stock dans des vecteurs ensuite, il calcule la similarité entre ces entités en vérifiant à chaque fois les synonymes de ces éléments avec les trois méthodes. Pour le calcul la valeur de similarité en mettant (1- distance obtenu), le résultat va être comparé avec un seuil donné ; finalement seuls les éléments qui ont une valeur de similarité supérieure ou égale à ce seuil vont être affichés.

# *Chapitre IV*

## **Réalisation et expérimentations**



## **IV.1. Introduction :**

Dans ce chapitre, nous présentons l'implémentation et l'évaluation de l'algorithme d'alignement d'ontologies sur deux ontologies existante .L'implémentation de l'algorithme a été réalisée en langage Java en utilisant l'API Jena, Les évaluations ont été faites sur une base de tests, qui se compose des petites ontologies.

## **IV.2. Environnement de réalisation :**

Avant de commencer l'implémentation de notre méthode, nous allons tout d'abord présenté les outils utilisés :

### **IV.2.1. présentation des outils de programmation :**

#### **✓ Langage de programmation java :**

Le langage Java est un langage généraliste de programmation synthétisant les principaux langages existants lors de sa création en 1995 par Sun Microsystems. Il permet une programmation orientée-objet (à l'instar de SmallTalk et, dans une moindre mesure, C++), modulaire (langage ADA) et reprend une syntaxe très proche de celle du langage C. Outre son orientation objet, le langage Java a l'avantage d'être modulaire (on peut écrire des portions de code génériques, c.-à-d. utilisables par plusieurs applications), rigoureux (la plupart des erreurs se produisent à la compilation et non à l'exécution) et portable (un même programme compilé peut s'exécuter sur différents environnements). En contrepartie, les applications Java ont le défaut d'être plus lentes à l'exécution que des applications programmées en éclipse par exemple.

Eclipse IDE est principalement écrit en Java (à l'aide de la bibliothèque graphique SWT, d'IBM), et ce langage, grâce à des bibliothèques spécifiques, est également utilisé pour écrire des extensions.

#### **✓ Apache Jena :**

Apache Jena est une Framework open-source du Web Sémantique pour Java. Il offre un API qui permet l'extraction de, ou bien l'écriture dans un graphe RDF. Les graphes sont représentés sous forme d'un « model » abstrait. Un modèle peut provenir des données à partir de fichiers, des bases de données, des URL ou une combinaison de ceux-ci. Un modèle peut également être interrogé par SPARQL.

Jena est similaire à Sesame (un Framework open-source pour l'interrogation et l'analyse des données RDF), mais, contrairement à Sesame, Jena fournit un support pour OWL (Web Ontology Language). Ce Framework a diverses raisonneurs internes et le raisonneur Pellet (un raisonneur Open-source Java OWL-DL) peut être mis en place dans Jena.

Jena supporte la sérialisation des graphes RDF en :

- Une base de données relationnelle
- RDF/XML
- Turtle
- ✓ **JWI (MIT Java Wordnet Interface) :**

JWI est une bibliothèque Java permettant une interface avec Wordnet. JWI prend en charge l'accès aux versions Wordnet parmi d'autres extensions Wordnet associées.

### **IV.2.2. présentation de l'interface du système :**

Maintenant on fera des tests sur l'application et nous allons les effectuer dans les prises d'écrans qui nous montrent les différentes étapes nécessaires à la réalisation de l'alignement d'ontologie et avoir les résultats dans une table qui permet à l'utilisateur de bien exploiter les similarités entre les concepts.

La figure suivante est l'interface du système de notre application. Elle contient deux boutons pour charger l'ontologie «ontologie 1» et «ontologie 2» qui permettent à l'utilisateur de choisir les fichiers OWL des deux ontologies à aligner voire (figure IV.2), et cela à travers un explorateur de fichiers.

Après qu'on charge les fichiers des ontologies à aligner, on appuie sur le bouton méthode pour choisir l'algorithme d'alignement après et choisi le seuil de similarité après lancer l'algorithme.

Après que l'algorithme finisse son travail, un table dans la même interface sera créée, dans lequel le résultat des correspondances est annoncé dans trios format (Excel, Text, Xml).

Après faire une agrégation pour les trois méthodes pour obtenir une bon similarité.

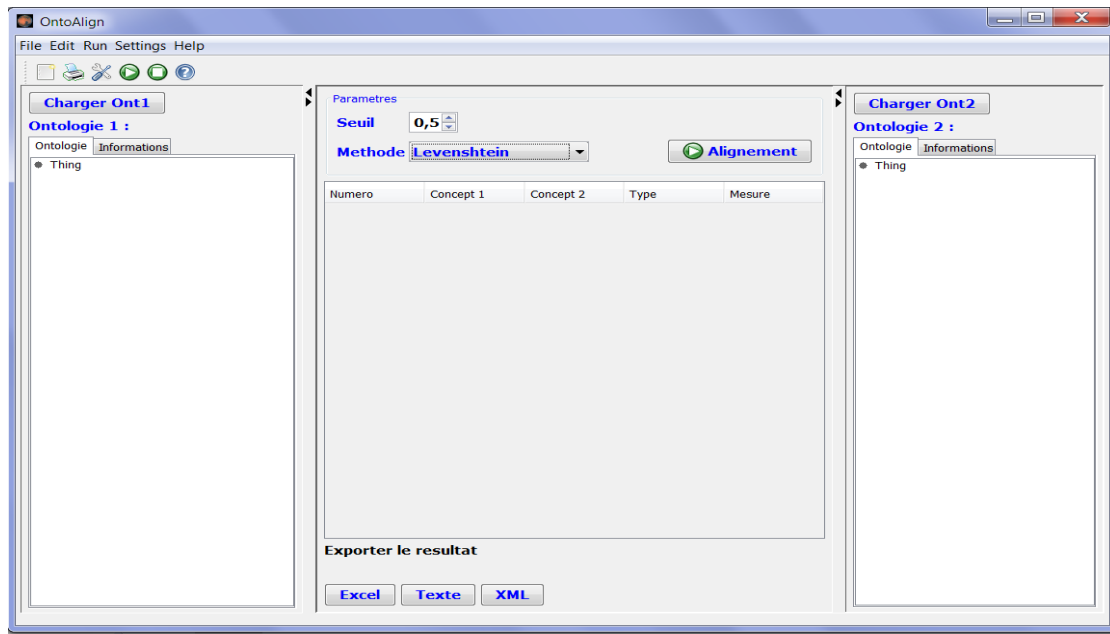


Figure IV.1 : L'interface principale de l'application

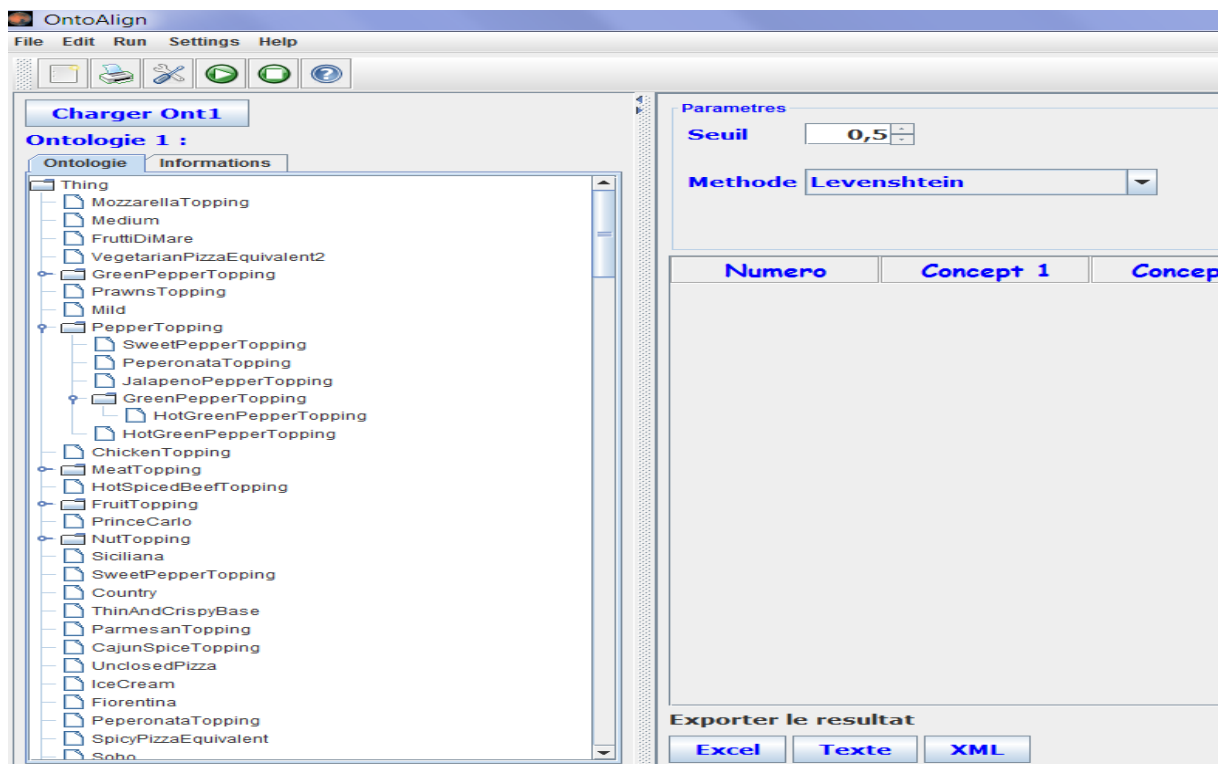


Figure IV.2 : Chargement d'une ontologie

### IV.3. Les ontologies de tests :

Pour appliquer l'alignement on a utilisé deux ontologies existante dans le domaine biomédicale l'un sur la maladie asthme contient 289 concepts et l'autre sur la maladie Autisme contient 791 concepts.

Pour tester l'algorithme implémenté, on a utilisé une paire d'ontologies, Ces ontologies se composent d'un nombre d'entités.

### IV.4. Résultats expérimentaux:

La figure suivant représente un exemple de la liste des correspondances trouvées par l'algorithme proposé seuil= 0.5 :

Numero	Concept 1	Concept 2	Type	Mesure
30276	RealItalianPizza	VegetarianPi...	CLASS	0.6875
30277	RealItalianPizza	RealItalianPi...	CLASS	1.0
30278	RealItalianPizza	VegetarianPi...	CLASS	0.6875
30279	RealItalianPizza	RealItalianPi...	CLASS	1.0
30280	RealItalianPizza	VegetarianPi...	CLASS	0.6875
30281	RealItalianPizza	RealItalianPi...	CLASS	1.0
30282	RealItalianPizza	VegetarianPi...	CLASS	0.6875
30283	RealItalianPizza	RealItalianPi...	CLASS	1.0
30284	RealItalianPizza	VegetarianPi...	CLASS	0.6875
30285	RealItalianPizza	RealItalianPi...	CLASS	1.0
30286	RealItalianPizza	VegetarianPi...	CLASS	0.6875
30287	RealItalianPizza	RealItalianPi...	CLASS	1.0
30288	RealItalianPizza	VegetarianPi...	CLASS	0.6875
30289	RealItalianPizza	RealItalianPi...	CLASS	1.0
30290	RealItalianPizza	VegetarianPi...	CLASS	0.6875
30291	RealItalianPizza	RealItalianPi...	CLASS	1.0
30292	RealItalianPizza	VegetarianPi...	CLASS	0.6875
30293	RealItalianPizza	RealItalianPi...	CLASS	1.0

Figure IV. 3: Résultats des correspondances obtenus par l'algorithme proposé

### IV.5. Conclusion :

Dans ce chapitre, nous avons présenté les outils qui nous avons utilisés pour l'implémentation de notre méthode d'alignement, qui sont Le vocabulaire « Wordnet » et « UMLS » qui nous a permis de trouver tout les synonymes d'un mot ou d'un terme en anglais, On a fait un petit aperçu sur notre interface. Pour les ontologies de teste on a utilisé deux ontologies existantes et énormes pour appliquer l'alignement.

# **Conclusion Générale**

## **Conclusion Générale**

Les travaux qui ont été menés dans le cadre de ce mémoire se situent dans l'intersection de deux domaines à savoir : l'ingénierie des connaissances et du Web sémantique. L'objectif est de proposer des méthodes d'alignement d'ontologies. Ces méthodes permettent s'assurer une interopérabilité sémantique des connaissances contenues dans les ontologies à aligner.

Elles exploitent des mesures de similarité et les structures internes des ontologies à aligner. Elles reposent sur une technique d'agrégation et de composition de différentes mesures de similarité.

L'algorithme implémenté en Java pour évaluer la qualité des correspondances entre deux ontologies .Dans les tests effectués, l'algorithme s'est montré être performant, surtout dans le cas de la dernière paire d'ontologies de test qui dans ce test nous donne le résultat correct.

Finalement, bien qu'on constate qu'il y a plusieurs travaux et de recherches qui traitent ce sujet, le problème d'alignement d'ontologies est loin d'être résolu , il reste beaucoup de travaux à faire dans ce domaine.

### **Les Perspectives :**

Plusieurs améliorations sont possibles dans notre méthode d'alignement. Ces améliorations incluent : le calcul plus riche et plus complet de la similarité terminologique, l'alignement des ontologies plus complexes en prenant en compte les instances par l'utilisation des différentes techniques d'apprentissage exploitants les différents types d'informations : les mots, leur position, leur format, leur fréquence, etc.

## ***BIBLIOGRAPHIE***

[1] : Fouzia BOUDRIES; Mapping d'ontologies dans un environnement distribué; Mémoire de Magistère En Informatique Option Réseaux et Systèmes Distribués, Bejaïa 2008-2009.

[2] : Khadim Drame. Contribution à la construction d'ontologies et à la recherche d'information : application au domaine médical. Autres [stat.ML]. Université de Bordeaux, Français 2014.

[3] : Moharned TOUZANI« Alignement des ontologies OWL-Lite », Faculté des arts et des sciences, Avril, 2005.

[4] :Hacine Gherbi Ahcine ; Mémoire Présenté pour l'obtention du Diplôme de Magister ; Construction d'une Ontologie pour le WEB Sémantique ; École doctorale d'informatique.

[5] :hadjui Fatima Zohra ; Mémoire de fin d'études pour l'obtention du diplôme de Master en Informatique ; Gestion des Ontologies Médicales ; Septembre 2012.

[6] : Yassine namir, Achraf ragui. « Alignement d'ontologies du web sémantique ».Master en Informatique université Sidi Mohamed Ben Abdallah 2016.

[7] : Rahamane Chahrazed , DAOUADJI Mohammed « Les ontologies pour modéliser les processus de soins en établissement de santé » MASTER en GENIE BIOMEDICAL Spécialité : Informatique Biomédicale, Université Abou Bakr Belkaïd de Tlemcen, 26 mai 2016.

[8] : Kolli Manel, « Intégration d'ontologies dans le cadre du web sémantique : une détection des relations sémantiques basée sur le RÀPC », Doctorat en Sciences en Informatique , Université Mentouri Constantine.

[9] : Siham Amrouch-berrouk ; THESE Présentée en vue de l'obtention du diplôme de DOCTORAT en Informatique ; « Conception d'outils pour la fusion automatique d'ontologies » ; Faculté des Sciences de l'Ingénieur ; Année : 2016/2017.

[10] : Siham Amrouch ; Mémoire En vue de l'obtention Du diplôme de Magister

Spécialité Informatique option Intelligence Artificielle ; « fission d'ontologies de domaine pour la gestion de connaissances application aux turbines a vapeur » ; Année 2008/2009.

[11] : Chaour Moundir ; Kermani Mohamed Hachem ; Projet de fin d'études pour l'obtention du diplôme de Master Recherche en Informatique ; « Développement d'une ontologie floue et classification d'instances : Application au domaine des analyses médicales » ; Université Constantine 2 ; Session Juin 2014.

[12] : Sami Zghal, « Contributions à l'alignement d'ontologies OWL par agrégation de Similarités ».Thèse Doctorat en Informatique université d'ARTOIS 2010.

[13] : MinaZiani ; Danielle Boulanger ; Guilaine Talens : « Système d'aide à l'alignement d'ontologies métier », Marseille ,mai 2010.

[14]: Hecht, Thomas and Buche, Patrice and DibiaBarthélemy, Juliette and Ibanescu, Liliana and Trojahn, Cassia. « Open Archive TOULOUSE Archive Ouverte (OATAO) ». 28 January 2014 - 31 January 2014 (Rennes, France).

[15]: MEKLID, Abdessalam , Approche agent Pour L'alignement d'ontologies :équivalence sémantique, mémoire de master, université de Biskra ,2011.

[16] :Khat Abderrahmane; Thèse Pour l'obtention du DOCTORAT LMD ; « Développement d'un Environnement pour l'Alignement des Ontologies Une Approche à Base d'Instances » ; Soutenanue le 11 Janvier 2017 .

[17] :Sami Zghal,Sadok Ben Yahia, Engelbert Mephu Nguifo, Yahya Slimani. « S ODA : Une approche structurelle pour l'alignement d'ontologies OWL-DL ». CRIL CNRS FRE 2499, Université d'Artois, IUT de Lens, Département des Sciences de l'Informatique, Faculté de Sciences de Tunis.

[18] : Jérôme Euzenat, Pavel Shvaiko; while at Département of Engineering and Computer Science (DISI) ; « Ontology Matching » ; 2001.

[19]: Jorge Martinez-gil and F. Aldana-montesL; Université of M´alaga ; « An Overview of Current Ontology Meta-Matching Solutions » ; 2010.



[20] :Soumaya kasri1, Fouzia Benchikha ; Université de Skikda, département d'informatique, Algérie ; « Un Algorithme de Partitionnement d'Ontologies Orienté Alignement » .

[21] : Warith-Eddine DJEDDI. « Alignement sémantique des ontologies de grande Taille ». Thèse Doctorat en Informatique Université Badji Mokhtar, Annaba 2013.

[22] : Jérôme Euzenat ; Towards a principled approach to semantic interoperability ; 655 avenue de l'Europe, 38330 Montbonnot Saint-Martin (France).

[23] : Saida Gharbi ; Université Badji Mokhtar-Annaba ; « Appariement d'ontologies hétérogènes » Année : 2011.

[24] : Bach, T.L.: Construction d'un Web Sémantique Multi-Points de Vue. These de doctorat Informatique, Ecole des Mines de Paris a Sophia Antipolis, (2006).

[25] : Bouquet et al. 2003 Bouquet P., Giunchiglia F., Van Harmelen F., Serafini L. and Stuckenschmidt H, "C-OWL – contextualizing ontologies".

[26]: Aissa Fellah, Mimoun Malki, Ahmed ZAHAF, « Alignement des ontologies: utilisation de WordNet eT une nouvelle mesure structurelle », 2008.

[27]: Soumaya Kasri, « Etude des techniques d'alignement des ontologies: proposition d'un algorithme de passage à l'échelle ». Thèse Magister en Informatique université 20 Août 1955 Skikda.2010.

[28] : Sami Zghal ,Karim Kamoun, Sadok Ben Yahia ,Engelbert Mephu Nguifo,Yahya Slimani ; « EDOLA : Une nouvelle method d'alignement d'ontologies OWL-Lite » ; CRIL CNRS FRE 2499, Université d'Artois, IUT de Lens Rue de l'Université - S.P. 16, 62307 Lens Cedex, France ; Faculté de Sciences de Tunis ; 21 décembre 2010.

[29] : Zohra Bellahsene, « Rôle et techniques de l'alignement d'ontologies : un survol de l'état de l'art » ; INFORSID 2017.

[30]: Klein, M. (2001). Combining and relating ontologies: an analysis of problems and solutions. In Proceedings of the IJCAI-Workshop on Ontology and Information Sharing, pages 5362, Seattle, US.

[31]: H.Benhmidi ; « Les ontologies » ; [dspace.univ-tlemcen.dz/bitstream/112/1062/5.pdf](http://dspace.univ-tlemcen.dz/bitstream/112/1062/5.pdf).