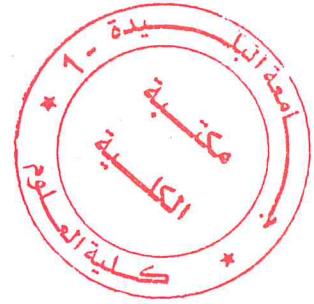


République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Saad Dahlab Blida

N° D'ordre :.....



Faculté des sciences

Département d'informatique

Mémoire Présenté par :

Arabji Mohamed Yacine

En vue d'obtenir le diplôme de master

Domaine : Mathématique et informatique

Filière : Informatique

Spécialité : Informatique

Option : Ingénierie de logiciel

***Thème : Alignement des ontologies basé sur
les fonctions de croyances***

Soutenu le : 02/11/2017, devant :

Mme. Chikhi. I
Mme Cherfa
Mme. Farch
Mme. Zahra

Président
Examineur
Promotrice
Encadreur

Promotion

2016 / 2017



Remerciements

Mes premiers remerciements vont à Dieu tout puissant qui m'a permis de mener à terme non seulement ce mémoire mais mes études en Master.

Je tiens ensuite à exprimer ma profonde gratitude à Mme Fareh et Mme Zahra pour le temps qu'ils ont investi pour suivre et analyser les résultats de ce travail ainsi que pour leurs conseils qu'ils n'ont cessé de nous prodiguer.

Mes remerciements vont aussi à tous les membres Du jury qui ont accepté d'examiner mon mémoire.

Enfin, je ne saurai oublier de remercier tous mes amis et ma famille qui, de près ou de loin, m'ont soutenu pour la réalisation du présent mémoire.

ملخص

إن الأنطولوجيات هي في قلب الشبكة الدلالية، فهي تمثل معرفة الويب، فإن هذا التمثيل يجمع معا مجموعة من المفاهيم والعلاقات التي تصف مجالا معيناً. إن محاذاة الأنطولوجيات هو حل لحل التغايرية في البيانات، سواء على المستوى النحوي أو الدلالي. والهدف من هذا العمل هو تصميم وتحقيق نظام يسمح تلقائياً بإنشاء علاقات المراسلات بين مفهومي اثنين من الأنطولوجيا ممثلة في لغة أونتولوجي ويب (أول)، وذلك باستخدام تقنيات المصطلحية والهيكلية بين المفاهيم، وهذه مختلفة يتم الجمع بين التقنيات باستخدام وظائف الاعتقاد لنموذج مشكلة المحاذاة. يتم إجراء تجربة على المعايير (أونتولوجي محاذاة مبادرة التقييم) للتحقق من صحة نمذجة لدينا.

الكلمات الدالة: علم الأنطولوجيا، تدابير التشابه، وظيفة الاعتقاد، الجمع،

Résumé

Les ontologies sont au cœur du web sémantique, elles permettent de représenter les connaissances du web, cette représentation regroupe un ensemble de concepts et de relations décrivant un domaine particulier. L'alignement des ontologies est une solution pour résoudre l'hétérogénéité des données, que ce soit au niveau syntaxique ou sémantique.

L'objectif de ce travail est de concevoir et réaliser un système permettant de générer automatiquement des relations de correspondances entre les concepts de deux ontologies représentées par le langage OWL (Ontology Web Language), en utilisant des mesures terminologiques et structurelles entre concepts, ces différentes mesures sont combinées en utilisant les fonctions de croyances pour modéliser le problème d'alignement. Une expérimentation sur les benchmarks de l'OAEI (Ontology Alignment Evaluation Initiative) est réalisée pour valider notre modélisation.

Mots clé : Ontologies, , mesure de similarité , Fonction de croyance, Combinaison, Alignement

Abstract:

Ontologies are at the heart of the semantic web, they represent the knowledge of the web, this representation groups together a set of concepts and relationships describing a particular domain. The alignment of ontologies is a solution to resolve the heterogeneity of the data, whether at the syntactic or semantic level.

The aim of this work is to design and realize a system allowing to automatically generate relations of correspondences between the concepts of two ontologies represented by the Ontology Web Language (OWL), using terminological and structural techniques between concepts, these different techniques are combined using the belief functions to model the

alignment problem. An experiment on the benchmarks of the OAEI (Ontology Alignment Evaluation Initiative) is carried out to validate our modeling.

Keywords : Ontology, similarity measures , belief function , combination ,Mapping

Table des matières

Liste des figures

Liste des Tableau

INTRODUCTION GÉNÉRALE

1. Contexte du travail.....	1
2. Problématique	1
3. Objectif	2
4. organisation.....	2

PARTIE 1 Etat de l'Art

Chapitre1 : Les ontologies

1. Introduction.....	6
2. Définition d'ontologie.....	6
3. Les constituants d'une ontologie	7
3.1. Les classes/ les concepts.....	7
3.2. Les relations.....	7
3.3. Les axiomes	7
3.4. Les instances.....	8
4. Rôle des ontologies	8
5. Les types d'ontologie.....	8
6. Langages d'ontologie.....	10
6.1. KIF.....	10
6.2. RDF/RDF Schéma.....	10
6.3. DAML + OIL.....	11
6.4. OWL	11
7. Domaine d'application des ontologies.....	12
7.1. La classification automatique	12
7.2. L'intégration de données	12
8. Conclusion	13

Chapitre2 : Alignement d'ontologie

1. Introduction.....	14
2. Définition de l'alignement	14
2.1. L'Alignement.....	15

3. Dimension de l'alignement	15
4. La Similarité.....	17
4.1. Définition de la similarité	17
4.2. Définition de dissimilarité :	17
5. Les Techniques de mesures de similarité.....	17
5.1. Les méthodes terminologiques :	18
5.2. Méthodes basées sur des chaînes de caractères :	18
5.3. Méthodes basées sur la langue :	18
5.4. Les méthodes extensionnelles :	18
5.5. Les méthodes hybrides :	18
6. Systèmes d'alignement	19
6.1. PROMPT :	19
6.2. OLA (OWL Lite Alignment) :	19
6.3. AROMA (Association Rule Ontology Matching Approach)	19
7. Synthèse et analyse	20
8. Conclusion	21

Chapitre3 : Fonction de croyance et les techniques existante

1. Introduction.....	23
2. La fonction de croyance et la fonction de plausibilité :	23
2.1. La fonction de croyance $Bel(A)$:	24
2.2. La fonction de plausibilité $Pl(A)$	24
3. Fusion des données	24
3.1. Combinaison conjonctive :	25
3.2. Combinaison disjonctive :	25
4. Méthodes existantes qui utilisent la théorie de Dempster-Shafer	26
4.1. Paolo Besana's ontology matching Framework	26
4.2. Wang et al.'s approach	26
4.3. DSSim-ontology :	26
5. Comparaison entre les techniques.....	28
6. Analyse	28
7. Conclusion	28

PARTIE 2 Solution proposée

Chapitre 4 : la Conception

1. Introduction.....	31
2. Architecture détaillée du système	31
2.1. Calcul des mesures	32
2.1.1. Mesure de similarité de Wu & Palmer (structurelle) :.....	32
2.1.2. Mesure de Jaccard (Syntaxique) :.....	33
2.1.3. Mesure de Jaro (Syntaxique) :.....	33
2.2. Combinaison des mesures :	34
3. Modélisation de l'alignement selon la théorie de la fonction de croyances :	35
3.1. L'ontologie Cmt	35
3.2. L'ontologie OpenConf.....	36
3.3. Cadre de discernement :	36
3.4. Fonction de la masse :	36
3.5. Combinaison conjonctive normalisée (Shafer & Dempster) :.....	39
3.6. Combinaison conjonctive	41
3.7. Combinaison disjonctive :	43
4. Analyse et synthèse.....	44
5. Conclusion	45

Chapitre 5 : Implémentation

1. Introduction.....	47
2. Outils de développement de notre système.....	47
2.1. NetBeans et JDK :	47
2.2. OWL2Prefuse	47
2.3. JENA	47
2.4. WordNet 3.0	47
2.5. JAWS.....	47
3. Les Benchmarks.....	48
4. La mise en œuvre du système	49
4.1. L'authentification	49
4.2. Interfaces de l'administrateur	50
4.2.1. L'interface globale de l'administrateur	50
4.2.2. L'interface « Applications »	51
4.2.3. L'interface d'alignement :	57
5. Conclusion	59

Chapitre 6 : test et évaluation

1. Introduction.....	61
----------------------	----

2.	L'évaluation du système	61
3.	Qualité d'alignement d'ontologies.....	62
4.	Test.....	63
5.	Validation.....	64
6.	Conclusion	64

Conclusion générale

Références bibliographiques

Liste des figures

Figure 1 : Processus d'alignements	16
Figure 2 Combinaison des Information.....	25
Figure 3 :processus de la technique DSSim-ontology	27
Figure 4 : Architecture du système d'alignement.....	31
Figure 5 Taxonomie de Cmt.....	35
Figure 6 Taxonomie d'OpenConf	36
Figure 7 : Interface « Authentification	49
Figure 8 : Interface « Administrateur »	50
Figure 9: Interface « Applications ».....	51
Figure 10 : Interface "ajouter une ontologie"	52
Figure 11 : Interface "supprimer une Ontologie "	53
Figure 12 : Interface «Ajouter une Base d'ontologie"	53
Figure 13 : Interface "Supprimer une Base"	54
Figure 14 : Interface Voir Taxonomie.....	55
Figure 15: Interface Voir Taxonomie.....	56
Figure 16 : Interface d'application.....	57
Figure 17 : Interface d'application.....	58
Figure 18 : Interface de combinaison des mesures	59

Liste des Tableaux

Tableau 1 Compraison entre les techniques	28
Tableau 2 : valeur des mesures de similarité.....	37
Tableau 3 : Fonction de masse	38
Tableau 4 : Fonction de masse de chaque mesure et sa fonction de l'ignorance.....	39
Tableau 5 : combinaison entre Jaro et Jaccard	40
Tableau 6 : résultats de la combinaison conjonctive normalisée	40
Tableau 7 : résultats de la combinaison conjonctive normalisé	40
Tableau 8 : Résultats de la combinaison conjonctive	42
Tableau 9 : Résultats de la combinaison disjonctive.....	44
Tableau 11 : La comparaison des correspondances trouvé manuellement et celle trouvé avec notre système d'alignement de deux ontologies	64

INTRODUCTION GÉNÉRALE

1. Contexte du travail

Les ontologies ont été reconnues comme une composante essentielle pour le partage des connaissances et la réalisation de cette vision. En définissant les concepts associés à des domaines particuliers, elles permettent à la fois de décrire le contenu des sources à intégrer et d'explicitier le vocabulaire utilisable dans des requêtes par des utilisateurs. Toutefois, il est peu probable qu'une ontologie globale couvrant l'ensemble des systèmes distribués puisse être développée.

Dans la pratique, les ontologies de différents systèmes sont développées indépendamment les unes des autres par des communautés différentes. Plusieurs ontologies couvrant un même domaine ou des domaines connexes sont développés indépendamment les unes des autres. Elles sont très souvent source d'hétérogénéité. Cette hétérogénéité posera des problèmes pour l'échange, le traitement, l'intégration, et la recherche d'information, c'est la problématique d'interopérabilité. Pour cette raison, un certain nombre de techniques ont été proposées dont l'une d'elles l'alignement d'ontologies.

La mise en correspondance vise à apporter des solutions à ce problème par l'introduction des systèmes qui, en prenant deux ontologies hétérogènes en entrée, découvrent de manière automatique les liens sémantiques entre leurs éléments.

2. Problématique

Les techniques actuelles du mapping d'ontologie s'appuient en général sur des mesures calculant la similarité de couples de concepts issus des deux ontologies. Ces mesures sont la plupart fondées sur les caractéristiques lexicales des labels des concepts et /ou des caractéristique structurelles des ontologies.

Le besoin le plus important par rapport aux méthodes de mise en correspondance est la combinaison efficace de ces méthodes, dont le but d'avoir une seule valeur qui représente le mapping entre les deux concepts.

La théorie de Dempster-Shafer est une théorie mathématique utilisant les fonctions de croyance et le raisonnement plausible. Le but de cette théorie est de permettre de combiner des preuves distinctes pour calculer la probabilité d'un évènement.

Une des principales difficultés consiste à modéliser le problème de mapping d'ontologies en utilisant les fonctions de croyances.

3. Objectif

Le but de ce projet est de réaliser un système d'alignement d'ontologies, étant données deux ontologies représentées par le langage OWL (Ontology Web langage), décrivant un domaine donnée, en se basant sur les fonctions de croyances pour la combinaison des différentes techniques de similarité, afin de trouver des mapping plus pertinent, plus fiable, et en prenant en considération la sémantique des informations.

L'ensemble des alignements obtenus est modélisé dans le cadre de la théorie des fonctions de croyance en utilisant les règles de combinaison utilisées pour combiner les résultats d'alignement.

4. Organisation

Afin d'atteindre le but de notre travail, le mémoire sera présenter en deux parties :

La première partie : Contient trois chapitres, elle présente le contexte du travail, elle a pour but de présenter les ontologies, et de montrer l'intérêt de l'alignement.

Chapitre 1- Les Ontologies : Ce chapitre propose une généralité sur les ontologies. Nous présentons les ontologies avec leurs caractéristiques, leurs constructions et leurs méthodes d'application, ainsi que les domaines d'application de ces ontologies. Nous allons décrire quelques langages de représentation des ontologies utilisées dans le Web sémantique.

Chapitre 2- alignement d'ontologie : Ce chapitre a pour objectif de dresser un état de l'art sur les techniques de mapping d'ontologies. Tout au long de ce chapitre, nous éclaircissons la notion du mapping en présentant des différentes définitions des concepts liées au contexte du mapping.

Chapitre 3- Fonction de la croyance et les techniques existantes : Ce chapitre a pour objectif de présenter les fonctions de la croyance et éclaircir ses notions de base comme la fonction de masse et fonction de l'ignorance et aussi les types de combinaison de fonction de masse, nous présentons ainsi les méthodes de mapping existantes qui utilisent les fonction de la croyance.

La deuxième partie :

Après les généralités présentées dans la première partie. Cette partie est consacrée pour la conception et le développement du système. Cette partie met l'accent sur le processus de développement de notre système d'alignement qui se présente comme une suite de phases enchaînées dans un déroulement linéaire, basé sur une succession de phases.

Chapitre 4 : La conception : Ce chapitre est une explication sur les étapes de construction les fonctions de masse à partir des mesures de similarité et ensuite faire les combinaison jusqu'à ou on trouve une résultats

Chapitre 5 : Implémentation : Ce chapitre est le résultat de la conception pour réaliser le système de mapping d'ontologies, il s'agit de transformer les éléments décrits lors de la conception en éléments du langage

Chapitre 6 : Test et Evaluation : cette phase est basée sur des expérimentations effectuées sur le système et des tests sur deux ontologies concrètes permettent de vérifier les résultats de de l'implémentation en testant la construction du système.

Partie 1 :

État de l'art

Chapitre1 :

Les Ontologies

1. Introduction

Le développement et l'exploitation des connaissances en informatique ont tellement évolué, qu'un des principaux objectifs de la recherche actuelle est de ne plus considérer l'ordinateur comme une boîte noire, dans laquelle, sont stockées des informations, mais plutôt comme une machine intelligente avec laquelle on peut dialoguer et créer une coopération. Le système doit alors, avoir accès non seulement aux termes utilisés par l'être humain mais aussi à la sémantique qui leur est associée, afin qu'une communication efficace soit possible. la connaissance visée par les ontologies est un thème de recherche dans divers axes tels que l'ingénierie des connaissances, la recherche d'information, les systèmes d'information coopératifs, l'intégration intelligente d'information, la gestion des connaissances etc... elles offrent une connaissance partagée sur un domaine qui peut être échangée entre des personnes et des systèmes hétérogènes.

2. Définition d'ontologie

Ontologie est une branche de la métaphysique qui s'intéresse à l'existence, à l'être en tant qu'être et aux catégories fondamentales de l'existant. En effet, ce terme est construit à partir des racines grecques « ontos » 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.

Dans le cadre de l'intelligence artificielle, Neeches et ses collègues [Neeches et al, 1993] furent les premiers à proposer une définition à savoir : « Une ontologie définit les termes et les relations de base du vocabulaire d'un domaine ainsi que les règles qui indiquent comment combiner les termes et les relations de façon à pouvoir étendre le vocabulaire »

En 1993, Gruber [Gruber T. 1993] propose la définition suivante : « Spécification explicite d'une conceptualisation ». Cette définition a été reprise et améliorée par Borst [Borst W. N. 1997] comme « Une ontologie est définie comme étant une spécification formelle d'une conceptualisation partagée ».

Ces deux dernières définitions sont regroupées dans celle de Studer comme « une ontologie est une spécification formelle et explicite d'une conceptualisation partagée ».

- **Formelle:** l'ontologie doit être lisible par une machine, ce qui exclut le langage naturel.

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

3. Les constituants d'une ontologie

Les ontologies définissent, le sens des termes et les relations entre eux, elles leur fournissent un vocabulaire commun. Bien que les ontologies peuvent être très différentes surtout au niveau du traitement de leurs composantes de base telles que les choses, les relations..., elles caractérisent un même univers.

3.1. Les classes/ les concepts

Également appelés classes dans certains travaux. Un concept peut désigner un objet, une idée ou bien une notion abstraite. Dans une ontologie, on désigne un concept par un label (terme, nom). Ce concept est défini par ses différentes relations avec les autres concepts de l'ontologie, ses attributs et les contraintes qui lui sont associées. Un concept peut représenter plusieurs objets.

3.2. Les relations

L'interaction entre les notions d'un domaine définit les relations, On trouve des relations binaires, tertiaire. Certaines relations binaires entre des objets sont considérées comme des rôles comme le définit Borgida [A. Borgida, 1996]. Les fonctions sont des cas particuliers de relations dans lesquelles un élément de la relation est défini à partir des autres éléments.

3.3. Les axiomes

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

3.4. Les instances

Les instances représentent les éléments des concepts et des relations dans un domaine donné. [Blas et al.1998]

4. Rôle des ontologies

Nées des besoins de représentation des connaissances, les ontologies sont à l'heure actuelle au cœur des travaux menés dans le Web sémantique. Visant à établir des représentations à travers lesquelles les machines peuvent manipuler la sémantique des informations, la construction des ontologies demande 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 (IA) : scientifiques comme étude des connaissances humaines et plus largement de l'esprit humain, ce qui rattache l'IA aux sciences humaines, et techniques comme création d'artefacts possédant certaines propriétés et capacités en vue d'un certain usage.

Le champ d'application des ontologies ne cesse de s'élargir et couvre les systèmes conseillers (systèmes d'aide à la décision, systèmes d'enseignement assisté par ordinateur – *e-learning* –, etc.), les systèmes de résolution de problèmes et les systèmes de gestion de connaissances (par exemple dans le domaine du biomédical). 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.

5. Les types d'ontologie

➤ *Ontologie de représentation de connaissances*

Modélise les représentations primitives utilisées pour la Formalisation des connaissances sous un paradigme donné. Par exemple, une ontologie sur le formalisme des Topic Maps comportera les concepts : Topic, Type de Topic, Association, Occurrence, Type Occurrence,

➤ ***Ontologie de haut niveau***

Elle exprime des conceptualisations valables dans différents domaines. Elle décrit des concepts très généraux comme l'espace, le temps, la matière, les objets, les événements, les actions, etc...

Ces concepts ne dépendent pas d'un problème ou d'un domaine particulier, et doivent être, du moins en théorie, consensuels à de grandes communautés d'utilisateurs. Ce type d'ontologies est fondé sur la théorie de la dépendance. Son sujet est l'étude des catégories des choses qui existent dans le monde. Comme les concepts de haute abstraction tels que les entités, les événements, les états, les processus, les actions, le temps, l'espace, les relations, les propriétés...

➤ ***Ontologie Générique (Generic ontology)***

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, événements, temps, espace, causalité, comportement, fonction, etc.

➤ ***Ontologie du domaine (Domain ontology)***

Cette ontologie exprime des conceptualisations spécifiques à un domaine, elle est pour plusieurs applications de ce domaine. Elle fournit les concepts et les relations permettant de couvrir les vocabulaires, activités et théories de ces domaines. Selon Mzoguchi, l'ontologie du domaine caractérise la connaissance du domaine ou la tâche est réalisée. Par exemple, dans le contexte du e-Learning, le domaine peut être celui de formation.

➤ ***Ontologie de Tâches (Task ontology)***

L'ontologie de tâches fournit un vocabulaire systématisé des termes employés pour résoudre des problèmes liés aux tâches qui peuvent être ou non du même domaine. Elle fournit un ensemble de termes au moyen desquelles nous pouvons décrire généralement comment résoudre un type de problèmes.

Elle inclut des noms, des verbes et des adjectifs génériques dans les descriptions de tâches.

➤ *Ontologie d'application (Application ontology)*

C'est l'ontologie la plus spécifique, elle contient des concepts dépendants d'un domaine et d'une tâche particuliers, elle est spécifique et non réutilisable. Ces concepts correspondent souvent aux rôles joués par les entités du domaine lors de l'exécution d'une certaine activité. Il s'agit donc ici de mettre en relation les concepts liés à une tâche particulière de manière à en décrire l'exécution

6. Langages d'ontologie

Des langages sont développés pour :

- décrire les ressources et services présents sur le Web,
- exprimer des modèles abstraits (ontologies) de ces descriptions,
- échanger / partager ces descriptions sur le Web

Il existe plusieurs langages de représentation des ontologies [ELB, 2009], les plus connus et les plus utilisés sont :

6.1. *KIF*

KIF [M.Genesereth et al ; 1992] est un langage basé sur les prédicats du premier ordre avec des extensions pour représenter des définitions et des méta-connaissances, la logique du premier ordre étant un langage de bas niveau pour l'expression d'ontologies.

Une extension du langage KIF, ONTOLINGUA, est utilisée dans le serveur d'édition d'ontologies, ONTOLINGUA du même nom.

6.2. *RDF/RDF Schéma*

Le W3C [World Wide Web Consortium] a adopté le langage RDF (Ressource Description Framework) comme un des formalismes standards de représentation de connaissances sur le Web. Utilisant la syntaxe XML (Extended Markup Language) qui constitue déjà un standard, le RDF permet de décrire des ressources Web en termes de ressources, propriétés et valeurs.

Une ressource peut être une page Web (identifiée par son URI, United Ressource Identifier) ou une partie de page (identifiée par une balise).

Les propriétés couvrent les notions d'attributs, relations ou aspects et servent à décrire une caractéristique d'une ressource en précisant sa valeur.

Les valeurs peuvent être des ressources ou des littéraux. RDF dispose d'une sémantique formelle analogue à celle des graphes conceptuels, c'est-à-dire identique à celle d'un fragment de la logique du premier ordre.

Un schéma de base incluant les primitives sémantiques généralement utilisées, a ainsi été ajouté au RDF et constitue ce qu'on appelle le RDF SCHEMA.

6.3. DAML + OIL

Dans l'optique d'une utilisation d'ontologies sur le Web, le langage RDF-S [W3C] a été enrichi par l'apport du langage OIL (Ontology Interchange Language) qui permet d'exprimer une sémantique à travers le modèle des frames tout en utilisant la syntaxe de RDF-S.

OIL offre de nouvelles primitives permettant de définir des classes à l'aide de mécanismes ensemblistes issus des logiques de description (intersection de classes, union de classes, complémentaire d'une classe). Il permet également d'affiner les propriétés de RDF-S en contraignant la cardinalité ou en restreignant la portée.

Le langage OIL a été fusionné avec le langage DAML pour former le DAML+OIL.

DAML (Darpa Agent Markup Language) est conçu pour permettre l'expression d'ontologies dans une extension du langage RDF. Il offre les primitives usuelles d'une représentation à base de frames et utilise la syntaxe RDF.

L'intégration de OIL rend possible les inférences compatibles avec les logiques de description, essentiellement les calculs de liens de subsumption.

6.4. OWL

La combinaison de RDF/RDF-S et de DAML+OIL a permis l'émergence d'OWL (Web Ontology Language), un langage standard de représentation de connaissances pour le Web développé par le groupe de travail sur le Web Sémantique du W3C, OWL peut être utilisé pour représenter explicitement les sens des termes des vocabulaires et les relations entre ces termes.

Les activités de spécification de langages d'ontologies pour le Web sont le fruit du consortium W3C (World Wide Web) qui émet des recommandations : RDF(S), OWL.

OWL s'inscrit dans les langages du Web permettant de définir des Métadonnées sur les informations disponibles sur Internet. Basé sur RDF, OWL étend le pouvoir de description du langage de modélisation. Ainsi, si RDF-S (la partie sémantique de RDF) s'est attaché à définir des vocabulaires, le W3C s'est rapidement aperçu de la faiblesse du pouvoir d'expression de ce langage de modélisation qui ne permettait pas de pouvoir aisément mettre en relation des vocabulaires, des ontologies. Pour pallier cette faiblesse, OWL a été standardisé qui étend le pouvoir de RDF-S tout en lui restant compatible pour cela, OWL a des moyens plus puissants pour exprimer la signification et la sémantique que XML, RDF, et RDF-S.

OWL permet donc de définir des ontologies qui sont l'association de définitions de termes d'un domaine particulier et des relations entre ces termes.

7. Domaine d'application des ontologies

Parmi les domaines d'application d'ontologie on peut citer

7.1. La classification automatique

La mise en œuvre de mécanismes de raisonnement sur des bases de connaissances peut tirer parti de la sémantique disponible dans une ontologie pour classer de manière automatique des instances en fonction de leur propriété.

7.2. L'intégration de données

L'intégration de données est une problématique majeure en informatique. Les ontologies sont des outils particulièrement intéressants pour réconcilier les représentations hétérogènes des données. Comme elles proposent une sémantique en plus des données (des données aux connaissances) il existe des mécanismes qui proposent d'identifier les choses identiques.

CHAPITRE2 :

L'alignement des ontologies

1. Introduction

Les progrès de l'information et de la communication technologies a mis à disposition une grande perte d'informations. Le problème de la gestion Hétérogénéité entre les diverses ressources d'information est en augmentant. Par exemple, la plupart des recherches de bases de données posse la question d'hétérogénéité sémantique, Qui est d'une manipulation les variations de sens ou d'ambiguïté dans interprétation d'entité, En conséquence beaucoup solutions ont été proposées pour faciliter le traitement à cette situation, et en particulier Une ontologie fournit généralement un vocabulaire qui décrit un domaine d'intérêt et une spécification du sens des termes utilisés dans le vocabulaire. En fonction la précision de la présente description, la notion d'ontologie englobe plusieurs données et modèles conceptuelles, y compris, des ensembles de termes, classifications, thesaurus Les schémas de base de données, ou des théories totalement axiomatisées. Lorsque plusieurs ontologies concurrentes sont utilisées dans différentes applications, le plus souvent ces applications ne peuvent pas inter-opéré immédiatement.

2. Définition de l'alignement

L'alignement d'ontologies est le processus de mise en correspondance sémantique des entités qui les composent [Euzenat, J. Valtchev, 2004] , le processus est exécuté selon une stratégie ou une combinaison de techniques de calcul de mesures de similarité et utilise un ensemble de paramètres (ex : paramètres de pondération, seuils ...) et un ensemble de ressources externes (ex : thesaurus, lexique...). Au final, nous obtenons un ensemble de liens sémantiques reliant les entités qui composent les ontologies.

Ces derniers comprennent des relations d'équivalence, de généralisation/spécialisation, de chevauchement ou encore d'incompatibilité. De nombreux travaux ont été développés dans le domaine de l'alignement d'ontologies et portent sur les techniques de recherche de similarité et sur les outils ou sur les Framework qui les intègrent.

2.1. L'Alignement

Soit deux ontologies O1 et O2, l'alignement A entre O1 et O2 est :

- Un ensemble de correspondances entre O1 et O2,
- Associées à une multiplicité : 1-1, 1-*, etc.

Peut être considéré comme une fonction f,

$$A' = f(O, O', A, p, r)$$

- Un pair d'ontologie : “ O et O' ”
- un alignement en entrée “A” (optionnel)
- un ensemble de paramètres “p”
- un ensemble de ressources externes “r”

3. Dimension de l'alignement

L'alignement regroupe trois dimensions : *l'input, le processus d'alignement et l'output*

➤ L'input :

Est constitué essentiellement des structures destinées à être alignées et qui peuvent être, comme énoncé précédemment, des schémas XML, des schémas relationnels, des ontologies (décrites en OWL, RDFS,...etc.) ou des instances d'une base de données ou d'une ontologie.

➤ Le processus d'alignement :

Peut être considéré comme une fonction f, qui à partir d'une paire d'ontologies O et O', un alignement en entrée A (optionnel), un ensemble de paramètres p (ex : paramètres de pondération, seuils ...) et un ensemble de ressources externes r (ex : thesaurus, lexique...), détermine un alignement A' entre ces deux ontologies.

$$A' = f(O, O', A, p, r)$$

Ceci peut être représenté schématiquement de la manière suivante :

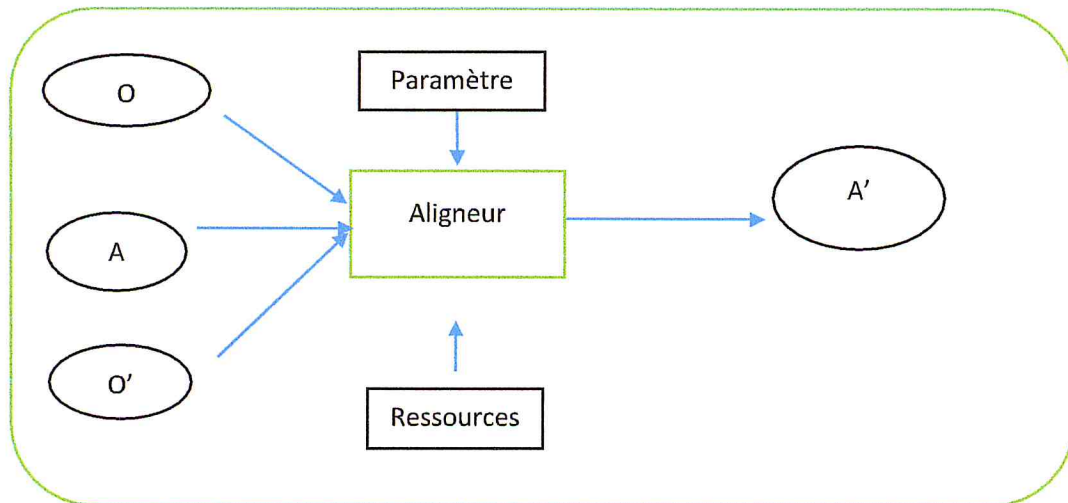


Figure 1 : Processus d'alignements

Le processus d'alignement est exécuté selon une stratégie ou une combinaison de techniques d'alignement de base.

➤ ***L'output :***

Est un ensemble d'alignements reliant les entités qui composent les deux ontologies. Un alignement est décrit comme un ensemble de cinq éléments :

$$\langle Id, e, e', r, n \rangle$$

- Id : identifiant unique d'un mapping
- e : une entité à aligner, appartenant à O (classe, propriété, contrainte, instance)
- e' : une entité à aligner, appartenant à O'
- r : la relation qui relie e a e'
- n : la mesure de confiance de la relation r, généralement une valeur réelle comprise dans l'intervalle [0,1]. Plus le n est proche du 1, plus la relation est considérée comme étant forte

➤ **La multiplicité**

Contraintes sur les relations entre les entités des deux ontologies

Si on considère, d'une part, les valeurs suivantes :

- 1 : une et une seule relation
- ? de 0 à 1 relation
- + : de 1 à plusieurs relations
- * : de 0 à plusieurs relations

D'autre part, les deux orientations possibles d'un alignement entre deux ontologies (O_O' et O'_O),

La multiplicité peut prendre les valeurs suivantes :

1:1, 1:+, +:1, 1:*, *:1, +:?, ?:*, *:?, +:*, *:+, +:+, *:.*.

➤ **le format de l'output :**

Il existe plusieurs possibilités de représenter l'output, dont : OWL, XML, etc.

4. La Similarité

4.1. Définition de la similarité

La similarité $S : O \times O \rightarrow R$ [Shvaiko et Euzenat, 2005] est une fonction à deux arguments vers un nombre réel satisfaisant les conditions suivantes :

- La positivité : $\forall a, b \in O, S(a, b) \geq 0$
- La maximalité : $\forall a, b \in O, S(a, a) \geq S(a, b)$.
- La symétrie : $\forall a, b \in O, S(a, b) = S(b, a)$

4.2. Définition de dissimilarité :

La dissimilarité $dS : O \times O \rightarrow R$ est une fonction à deux arguments vers un nombre réel satisfaisant les conditions suivantes :

- La positivité : $\forall a, b \in O, S(a, b) \geq 0$
- La minamilité : $\forall a, b \in O, S(a, a) \leq S(a, b)$.
- La symétrie : $\forall a, b \in O, S(a, b) = S(b, a)$

5. Les Techniques de mesures de similarité

Il existe plusieurs méthodes de calcul de la similarité entre les entités de plusieurs ontologies.

5.1. Les méthodes terminologiques :

Elles sont employées pour calculer la valeur de similitude des entités textuelles, telles que des noms, des métadonnées sur les noms, des étiquettes, des commentaires,...

Le problème de ces comparaisons est l'existence des synonymes et des homonymes, il n'est pas possible d'assumer que deux entités sont similaires si elles portent le même nom, ou qu'elles sont différentes si elles possèdent des noms différents

Ces méthodes peuvent encore être divisées en deux sous-catégories : l'une contient des méthodes qui comparent des termes en se basant sur les caractères contenus dans ces termes et l'autre utilise certaines connaissances linguistiques

5.2. Méthodes basées sur des chaînes de caractères :

Ces méthodes à base de chaînes profitent de la structure de la chaîne (en tant que séquence de caractères).

Avant de comparer les chaînes caractères, qui ont un sens dans le langage naturel, il faut passer par des étapes qui sont :

5.3. Méthodes basées sur la langue :

Les informations exploitées peuvent être celles intrinsèques (des propriétés linguistiques internes des termes telles que des propriétés morphologiques ou syntaxiques) ou celles extrinsèques en effectuant la correspondance à travers les relations lexicales (par exemple, synonymie, hyponymie, etc.) en faisant appel à des ressources auxiliaires telles que les dictionnaires de synonymes et d'hyponymes, mais aussi à des thésaurus et des ressources sémantiques ainsi qu'à des dictionnaires spécifiques aux domaines étudiés

5.4. Les méthodes extensionnelles :

Elles déduisent la similarité entre deux entités qui sont notamment des concepts ou des classes en analysant leurs extensions (leurs ensembles d'instances). Chaque instance peut être représentée par un vecteur de noms et/ou de valeurs. Des calculs de similarités entre vecteurs permettent de comparer les instances.

5.5. Les méthodes hybrides :

Elles combinent plusieurs mesures lorsqu'une seule est insuffisante.

Ces méthodes sont intégrées dans des outils permettant la mise en correspondance d'ontologies. Il existe des outils qui combinent plusieurs méthodes de similarité et des Framework implémentant plusieurs mesures et permettant ainsi de suggérer à l'expert plusieurs mapping.

6. Systèmes d'alignement

Différents systèmes ont été développés dans le but d'aligner plusieurs ontologies.

6.1. *PROMPT* :

C'est un système interactif constituant une aide pour la comparaison, l'alignement, la fusion et l'évolution de plusieurs formalismes de représentation des connaissances

[Noy et al. 2001]. Son module d'alignement appelé Anchor-Prompt permet de rapprocher des ontologies de la façon suivante : d'abord, des 'matchers' linguistiques permettent de déterminer un ensemble initial de concepts similaires. Ensuite, à partir de cette liste, un algorithme analyse les chemins dans les sous graphes délimités par ces concepts et détermine quelles classes apparaissent fréquemment dans les mêmes positions sur des chemins

6.2. *OLA (OWL Lite Alignment)* :

C'est un système implémentant un algorithme d'alignement des ontologies décrites en OWL. OLA mesure la similarité entre deux entités à partir des calculs de similarité entre leurs caractéristiques (leurs types : classe, relation ou instance, leurs liens avec d'autres entités : sous-classes, domaine,...). La valeur de similarité finale est la somme pondérée des valeurs de similarité de chaque caractéristique. Les poids sont associés suivant le type d'entité à comparer et ses caractéristiques. L'algorithme applique un calcul du point fixe, avec des itérations pour améliorer la similarité de deux entités. Quand il n'y a plus d'améliorations possibles, des alignements entre deux ontologies sont proposés [Euzenat et al. ,2004].

6.3. *AROMA (Association Rule Ontology Matching Approach)*

C'est une approche d'alignement pour des ontologies représentées en OWL. [David ; 2007] Elle permet de découvrir des liens sémantiques de type « subsomption » ou « équivalence » entre deux entités (classes ou propriétés). Le processus d'alignement se déroule en trois étapes : la première procède à l'acquisition des termes contenus dans les descriptions et instances des entités à partir d'outils de Traitement Automatique du Langage (TAL).

Ensuite, pour chaque entité, ainsi qu'à ses ancêtres est associé un ensemble de termes dits représentatifs. La deuxième étape permet de créer des relations de subsomptions entre les entités à partir de règles d'association construites sur le principe qu'une entité X est plus spécifique ou équivalente à une entité Y si le vocabulaire utilisé dans les descriptions et les instances de X a tendance à être inclus dans celui de Y. Des 'matchers' terminologiques sont utilisés pour comparer les différentes descriptions. Enfin, la dernière étape vise à analyser les règles d'associations trouvées afin de :

- déduire des relations d'équivalence ;
- trouver des incohérences (cycles) et de les éliminer
- supprimer les relations redondantes ;
- sélectionner le meilleur alignement pour chaque entité.

7. Synthèse et analyse

L'alignement des ontologies représente un grand intérêt pour plusieurs domaines d'applications qui manipulent des connaissances hétérogènes.

Les techniques d'alignement présentées sont les outils dont dispose les concepteurs de systèmes d'alignement pour trouver les solutions aux problèmes de l'hétérogénéité des ontologies. La conception et le développement de ces systèmes nécessite un traitement plus global, en effet, ces systèmes doivent pouvoir :

- *Sélectionner et combiner* les techniques de base de l'alignement afin d'améliorer la qualité de l'alignement résultant.
- *Apprendre* à partir des données les meilleurs paramètres
- Certaines méthodes permettent de découvrir des similarités qui peuvent être ignorées par d'autres, et réciproquement. De plus les méthodes les plus complexes reposent fréquemment sur l'exploitation d'information donnée par des méthodes plus simples.

Le besoin le plus important par rapport aux méthodes de mise en correspondance est la combinaison efficace de plusieurs méthodes complémentaires, cette combinaison se base sur des propositions des formules pour la combinaison de différentes techniques de similarité utilisées, selon des tests effectués avec l'expert du domaine.

8. Conclusion

Ce chapitre a eu pour objectif de dresser un état de l'art sur les techniques d'alignement d'ontologies. Tout au long de ce chapitre, nous avons essayé d'éclaircir la notion du mapping en présentant des différentes définitions des concepts liées au contexte du mapping. Nous avons présenté les classifications des techniques d'alignements selon plusieurs vues, et pour finir une que les techniques de base utilisées.

Chapitre 3 : fonction de croyance et les Techniques existantes

1. Introduction

La théorie Dempster [Dempster ; 1966] complétée par les propositions de Shafer [G.Shafer ; 1976] a donné naissance à un modèle mathématique de croyance connue sur le nom de la théorie de l'évidence. Cette théorie est un modèle d'inférence statistique qui généralise l'inférence Bayésienne et qui connaît des applications diverse tels que :

- ✓ Années 80 : IA, modélisation des incertitudes dans les systèmes experts
- ✓ Années 90 : fusion d'informations (télédétection, identification de cibles, imagerie, médicale, ...)

Elle permet de représenter explicitement, à partir d'outils mathématiques, l'incertitude liée aux connaissances. Par cette méthode on modélise le degré de croyance,

2. La fonction de croyance et la fonction de plausibilité :

La fonction de croyance [Th.Denœux ; 2011] est une fonction qui, pour une information ou proposition émise ou déduite d'une analyse, permet d'attribuer des degrés de croyance ou de plausibilité comprise entre 0 et 1.

Soit Ω un ensemble d'hypothèses $\{H_1, H_2, \dots, H_n\}$ appelé cadre de discernement. L'ensemble des parties A de Ω est noté 2^Ω .

Si on définit la fonction $m : 2^\Omega \rightarrow [0, 1]$ une distribution d'évidence sur les parties Ω (ou la partition de Ω), alors $A \rightarrow m(A)$ représente la distribution d'évidence sur la partie A .

- $m(A)$ est la confiance portée strictement dans A sans que celle-ci puisse être répartie sur les hypothèses qui la composent.
- A est un élément focal si $m(A) \neq 0$

Dans le cas où pour $m : 2^\Omega \rightarrow [0, 1]$ les deux conditions suivantes sont remplies :

$$m(\emptyset) = 0$$

Et

$$\sum_{A_i \subset \Omega} m(A_i) = 1$$

Il est appelé fonction de masse sur Ω .

2.1. La fonction de croyance $Bel(A)$:

C'est par conséquent la somme des masses des propositions incluses dans A. Elle est définie par la formule suivante

$$Bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B) \quad \forall A \subseteq \Omega$$

2.2. La fonction de plausibilité $Pl(A)$

C'est la somme de des masses des propositions dont l'intersection avec A n'est pas nulle, elle est donnée par la formule :

$$Pl(A) = \sum_{B \cup A \neq \emptyset} m(B) \quad \forall A \subseteq \Omega$$

3. Fusion des données

La théorie de la fonction de croyance est un outil utile pour la fusion de données imparfaites. En effet, pour un problème donné et pour le même cadre de discernement, il est possible Pour obtenir une fonction de masse synthétisant les connaissances à partir de sources séparées et indépendantes de L'utilisation d'une règle de combinaison. Il existe trois modes de combinaison :

Combinaison conjonctive, combinaison conjonctive normalisée et combinaison disjonctive

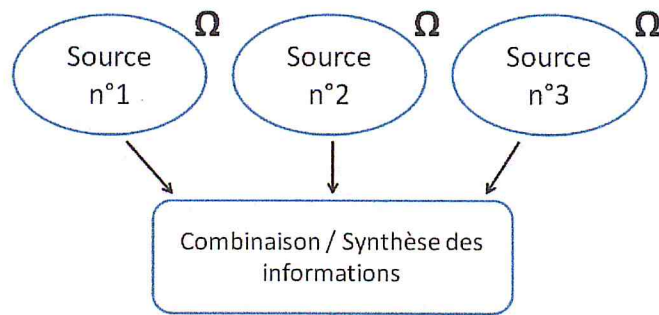


Figure 2 Combinaison des Information

3.1. Combinaison conjonctive :

$$m_1 \overset{\textcircled{\cap}}{2} (A) = \sum_{B \cap C = A} m_1(B) \times m_2(C), \quad \forall A \subseteq \Omega$$

Forme normalisé :

$$m_1 \oplus_2 = \begin{cases} \frac{m_1 \overset{\textcircled{\cap}}{2}(A)}{1 - k}, & A \neq \emptyset \\ 0, & A = \emptyset \end{cases}$$

Avec $K = m_1 \overset{\textcircled{\cap}}{2}(\Phi)$ le degré de conflit

3.2. Combinaison disjonctive :

$$m_1 \overset{\textcircled{\cup}}{2} (A) = \sum_{B \cup C = A} m_1(B) \times m_2(C); \forall A \subseteq \Omega$$

4. Méthodes existantes qui utilisent la théorie de Dempster-Shafer

4.1. *Paolo Besana's ontology matching Framework*

[Besana, P. 2006]. Basana a proposé un Framework basé sur la théorie Dempster-Shafer pour l'alignement des ontologies. A cet effet, il a suggéré de combiner les résultats des différents matchers dans pour obtenir de meilleurs résultats.

- Le cadre du discernement représente le produit cartésien $e \times O_{target}$ où e est Une entité de l'ontologie source O_{source} et O_{target} représentent toutes les entités de la Ontologie cible. Chaque hypothèse du cadre est le couple $\langle e, ei \rangle$ tel que e est une entité de l'ontologie source et ei est une entité de l'ontologie cible.
- La fonction de masse est déduite à partir de la mesure de similarité obtenue lors de l'application du la Méthode. Selon le matcher, le résultat obtenu est entre $[0, 1]$.
- La combinaison des fonctions de masse générées par les matchers est effectuée grâce à l'application de la règle de combinaison de Dempster

4.2. *Wang et al.'s approach*

[Wang et al., 2007] Les auteurs ont proposé d'améliorer les résultats d'alignement en combinant les résultats de trois différents matchers (deux matchers name-based et un matcher). Le but De cette méthode est de détecter des correspondances simples où chaque entité de la première ontologie O_1 est alignée sur une entité de la deuxième ontologie O_2 .

Ils spécifient le cadre du discernement comme un ensemble de paires d'entités. Pour chaque entité e de la première ontologie O_1 . Ses mappages avec toutes les entités de la deuxième ontologie O_2 sont formés de sorte que chaque paire soit formée à partir d'une entité :

$\Theta = e \times O_2$ pour chaque paire d'entités, nous aurons trois mesures de similarité normalisées considérées comme des fonctions de masses. Afin d'obtenir un résultat d'alignement unifié pour une paire d'entités, les fonctions de masse sont combinées à l'aide de la règle de combinaison de Dempster.

4.3. *DSSim-ontology :*

C'est un système d'alignements d'ontologie qui combine les valeurs de similarité fournies Par des algorithmes de similarité syntaxique et sémantique puis raffiner l'exactitude des résultats

Au moyen d'une fonction de croyances.

Pour faire un alignement des ontologies, DSSim [Nagy,M et al ; 2007] fonctionne comme suit :

1. Prenez un concept (ou une propriété) de l'ontologie 1 et considérez comme la une requête de fragment qui serait normalement posé par un utilisateur.
2. Prenez des concepts et des propriétés syntaxiques semblables et ses synonymes au graphe de requête de Ontologie 2 et construire un graphe qui contient à la fois les concepts (propriétés) et ses synonymes.
3. Différents algorithmes de similarité sont utilisés pour évaluer les valeurs de similarité quantitative (converties en fonction de masse de croyance) entre les nœuds de la requête et l'ontologie qui est considéré comme une évaluation incertaine et objective. Après L'information produite par les différents algorithmes est combinée en utilisant la règle de Dempster de combinaison.
4. nous évaluons la similarité sémantique entre la requête de fragment d'ontologie 1 et le graphe de fragment du l'ontologie 2 et on choisit celles dans lesquelles nous calculons la fonction de croyance la plus élevée.
5. Les concepts sélectionnés sont ajoutés dans l'alignement.

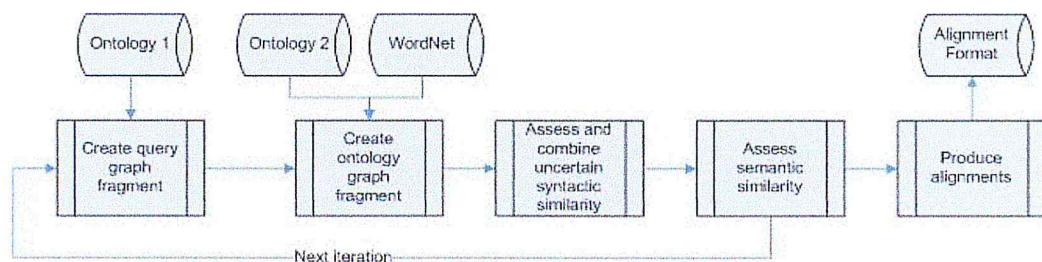


Figure 3 : processus de la technique DSSim-ontology

5. Comparaison entre les techniques

Dans le tableau suivant on a fait la comparaison entre les différentes techniques définies précédemment selon des critères qui ont été bien choisis :

Méthode	Besana	Wang and al	DSSim
Auteur	Paolo Besana	Wang and al	Nagy and al.
Contexte	Domaine général	Domaine général	Domaine particulier
Alignement	Simple	Simple	Simple
Mesure de similarité	Différentes mesures	Name-based matchers Structured-based matchers	Différentes mesures
Théorie de l'incertitude	Théorie de croyances	Théorie de croyances	Théorie de croyances
Relation	équivalence	équivalence	équivalence Subsumption
Manipuler l'incertitude	Combinaison des résultats	Combinaison des résultats	Combinaison des résultats
La gestion des conflits	Non	Non	Oui

Tableau 1 : Comparaison entre les techniques

6. Analyse

La plupart des méthodes existantes étudiées, qui utilisent les fonctions de croyance pour résoudre le problème de combinaison entre différentes mesures utilisées pour chaque méthode, ont plusieurs critères en commun ce qui est remarquable à partir de la table de comparaison, et elles utilisent les mêmes démarches de transformation des valeurs de mesures en fonction de masse pour réaliser la combinaison. Cependant elles utilisent des types de mesures différents pour le calcul de la similarité entre les concepts des ontologies.

7. Conclusion

Ce chapitre a eu pour objectif de dresser un état de l'art sur les fonctions de croyances. Tout au long de ce chapitre, nous avons essayé d'éclaircir la notion de la fonction de masse et les types de combinaison. Nous avons présenté les classifications des techniques existantes.

Partie 2 :

Solution proposée

Chapitre 4 :

Conception

1. Introduction

La conception est une étape très importante du processus de modélisation d'un système, ce chapitre porte sur la conception de notre système qui consiste à faire un alignement entre deux ontologies basé sur les fonctions des croyances.

2. Architecture détaillée du système

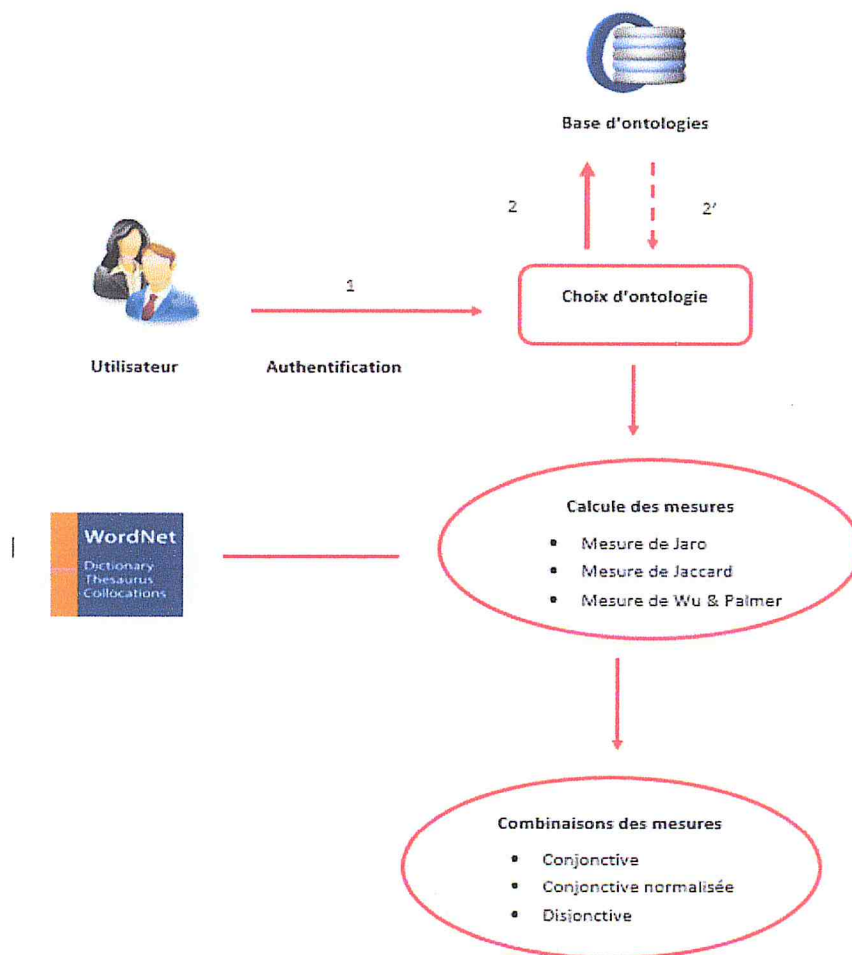


Figure 4 : Architecture du système d'alignement

Numéro	Explication
1	l'utilisateur lance le processus d'alignement après l'authentification
2	L'utilisateur choisit deux ontologies à partir de la base d'ontologies, soit existante ou bien chargée dans la base d'ontologies par l'utilisateur.
2'	Chargement des deux ontologies à aligner.
3	Calculer les mesures de similarité
4	La combinaison des mesures

Définition des éléments de l'architecture :

- **Utilisateur** : C'est la personne qui utilise notre système pour l'alignement des ontologies.
- **WordNet** : C'est une base de données lexicale développée par des linguistes. Son but est de répertorier, classifier et mettre en relation de diverses manières le contenu sémantique et lexical de la langue anglaise, nous utiliserons cette base pour chercher la synonymie des concepts.
- **Base d'ontologies** : c'est la base qui contient les ontologies qu'un utilisateur va aligner

2.1. Calcul des mesures

Dans cette section nous expliquons les mesures utilisées, nous avons choisi 2 types de mesures :

- Syntaxique (La mesure de Jaro et la mesure de Jaccard)
- Structurelle (La mesure de Wu & Palmer)

2.1.1. Mesure de similarité de Wu & Palmer (structurelle) :

La longueur de chemin entre deux concepts dans une hiérarchie est une mesure intuitive pour calculer la similarité. C'est une mesure utile et facile à implémenter en même temps. Une mesure qui est fonction de la longueur de chemin est proposée par [Wu & Palmer, 1994] qui s'appuie sur la longueur du chemin entre deux concepts d'une même hiérarchie. Cette mesure est définie comme suit :

$$sim_{wu \& palmer}(C1, C2) = \frac{2 * depth(C)}{depth(C1) + depth(C2)}$$

Où 'C' est le subsumant commun le plus spécifique, depth(C) est la longueur du Chemin entre C et la racine de la hiérarchie, depth(Ci) est le nombre d'arcs entre Ci et la racine en passant par C. Cette mesure est facile à implémenter mais elle ne tient pas compte des descriptions de concepts.

Exemple de calcul entre plusieurs concepts :

	Reviewer	Person	Bid
Reviewer	1.0	0.842	0.2667
Administrator	0.7619	0.8889	0.2857

2.1.2. Mesure de Jaccard (Syntaxique) :

Pour comparer les ontologies à partir des instances associées aux concepts : L'approche que nous avons utilisée dans notre travail s'inscrit dans la catégorie des approches à base d'instances lorsque les deux ontologies à aligner partagent beaucoup d'instances communes (même domaine d'application). Elle utilise des mesures telles que la métrique de JACCARD pour calculer les similarités entre les entités à aligner.[Slimani et al , 2007]

$$Sim_{JAC}(C1, C2) = \frac{|i_1 \cap i_2|}{|i_1 \cup i_2|}$$

Prenant l'exemple de calcul de mesure de Jaccard sur les termes des concepts suivants :

	Reviewer	Person	Bid
Reviewer	1.0	0.4	0.1
Administrator	0.23	0.35	0.23

2.1.3. Mesure de Jaro (Syntaxique) :

Elle mesure le nombre et l'ordre des caractères communs entre deux chaînes de caractères. Nous définissons la distance de similarité de Jaro entre deux mots w1, w2 comme la distance Jaro entre les deux chaînes de caractères des deux mots.[Arnaud et al ,2012]

$$dj = (\frac{m}{|S_1|} + \frac{m}{|S_2|} + \frac{m - t}{m})$$

La distance de Jaro entre chaînes S1 et S2 est définie par

Ou

- $|S_i|$ est la longueur de la chaîne du caractère S_i
- m est la nombre de caractère correspondants
- t est le nombre de transpositions

Prenant l'exemple de calcul de mesure de Jaro sur les termes des concepts suivants :

	Reviewer	Person	Bid
Reviewer	1.0	0.431	0.486
Administrator	0.468	0.496	0.607

2.2. Combinaison des mesures :

Dans notre approche, nous avons utilisé la théorie des fonctions des croyances pour combiner les mesures de similarité.

En premier lieu, nous avons calculé les similarités (terminologique et structurelle) ensuite on prend les valeurs maximales de chaque mesures donc au aura 3 concepts (concept à la fin on fait les combinaisons de ces 3 valeurs selon les règles suivantes :

- Combinaison Conjonctive.
- Combinaison Conjonctive normalisée (règle de Shafer & Dempster) .
- Combinaison Disjonctive.

3. Modélisation de l'alignement selon la théorie de la fonction de croyances :

Utiliser une seule mesure de similarité ne permet pas d'obtenir un alignement parfait, vu que chaque mesure a ses propres caractéristiques. Afin d'améliorer le résultat de l'alignement, il est intéressant d'exploiter la complémentarité des différentes mesures.

Cependant, l'utilisation de plusieurs mesures fait apparaître un conflit entre les différents résultats produits par chacune de ces mesures. Ce conflit doit être modélisé et résolu.

Afin de résoudre le désaccord entre les différents résultats, nous modélisons les alignements dans le cadre de la théorie des fonctions de croyance et de gérer le conflit par combinaison des différents résultats.

Pour comprendre le principe de modélisation et de combinaison en utilisant les fonctions de croyance nous prenons un exemple les ontologies suivantes « cmt » et « OpenConf »

3.1. L'ontologie Cmt

La figure suivante représente la taxonomie des concepts de l'ontologie « Cmt »

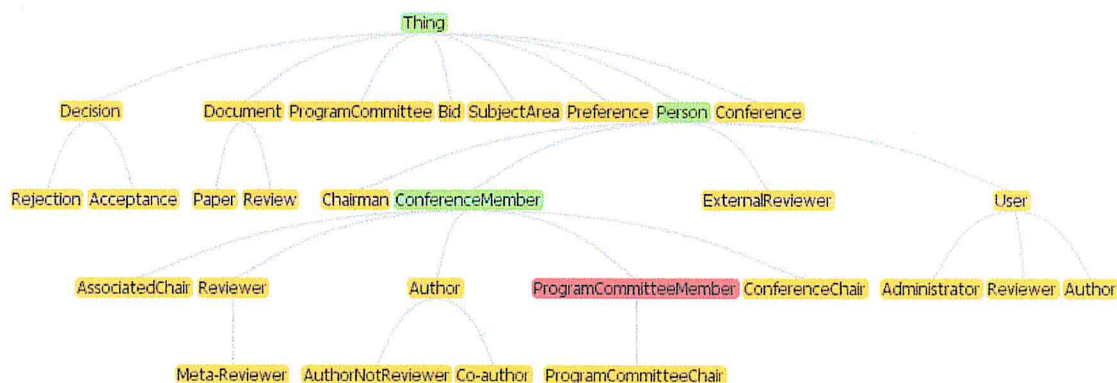


Figure 5 Taxonomie de Cmt

3.2. L'ontologie OpenConf

La figure suivante représente la taxonomie des concepts de l'ontologie « OpenConf »¹.

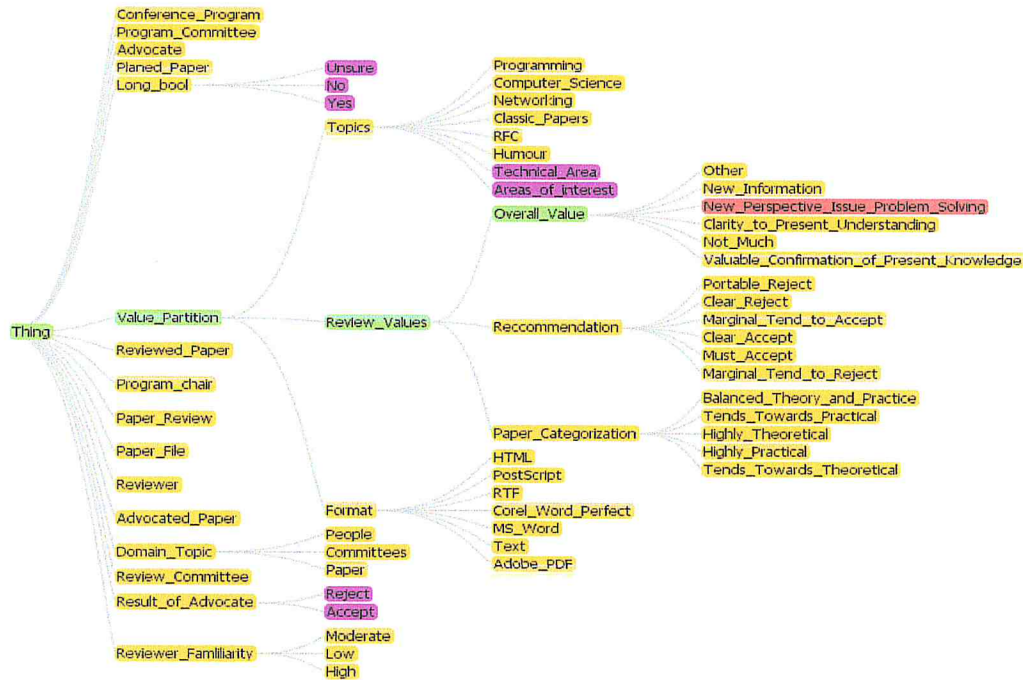


Figure 6 Taxonomie d'OpenConf

3.3. Cadre de discernement :

Notre cadre c'est un ensemble qui contient tous les réponses possibles à notre problème, Dans notre cas les éléments de O_2 , donc notre cadre de discernement contient tous les éléments de O_2 ,

$$\Omega = \{Organization, Person, Chair, Presentation, Rejected_Contribution, Reviewer, Conference, Poster, Conference_fees, Program_Committe.\}$$

3.4. Fonction de la masse :

Une fois on obtient toutes les valeurs de similarité, on prend pour chaque entité $e_i \in O_1$ son correspondant avec $e_j \in O_2$ ensuite pour chaque couple on lui donne sa valeur.

Pour dire que deux concepts sont alignés entre eux s'il existe une importante valeur de similarité selon les techniques utilisées. Cette valeur représente une distance entre ces concepts et dans notre modélisation selon les fonctions de croyance une mesure de similarité peut être considérée comme une fonction de masse.

Cette table suivante représente les valeurs de similarité de Jaro , de Jaccard et Wu & Palmer entre concepts de deux ontologies « Cmt » et « OpenConf » .

Tableau 2 : valeur des mesures de similarité

$e_1 \in O_1$	Mesure de similarité	$e_2 \in O_2$	Valeur de similarité
Administrator	Jaro	Clarity_to_Present_Understanding	0.658
	Jaccard	Format	0.583
	Wu et palmer	Advocate	0.8
Person	Jaro	Clarity_to_Present_Understanding	0.729
	Jaccard	Paper	0.375
	Wu et palmer	Advocate	0.941
Review	Jaro	Reviewer	0.917
	Jaccard	Reviewer	0.75
	Wu et palmer	Computer_Science	0.5
Reviewer	Jaro	Reviewer	1.0
	Jaccard	Reviewer	1.0
	Wu et palmer	Reviewer	1.0
Conference	Jaro	Conference_Program	0.852
	Jaccard	Conference_Program	0.555
	Wu et palmer	People	0.615
Chairman	Jaro	Valuable_Confirmation_of_Present_Knowledge	0.681
	Jaccard	Programming	0.461
	Wu et palmer	Advocate	0.8
Bid	Jaro	Reviewed_Paper	0.603
	Jaccard	High	0.166
	Wu et palmer	Programming	0.667
Co-author	Jaro	Contact_Author	0.82
	Jaccard	Author	0.66
	Wu et palmer	Author	0.909
User	Jaro	Other	0.633
	Jaccard	Other	0.2857
	Wu et palmer	Advocate	0.889
Author	Jaro	Author	1.0

	Jaccard	Author	1.0
	Wu et palmer	Author	1.0

Afin de construire la fonction de masse et garantir que la somme soit égale à 1, une masse sera allouée à l'ignorance totale $m(\Omega)$:

$$m(\Omega) = 1 - m(e_1, e_2)$$

Tableau 3 : Fonction de masse

$e_1 \in O_1$	Mesure de similarité	Fonction de masse
Administrator	Jaro	$m(\text{Clarity_to_Present_Understanding}) = 0.658$ $m(\Omega) = 0.342$
	Jaccard	$m(\text{Format}) = 0.583$; $m(\Omega) = 0.417$
	Wu et palmer	$m(\text{Advocate}) = 0.8$; $m(\Omega) = 0.2$
Person	Jaro	$m(\text{Clarity_to_Present_Understanding}) = 0.729$; $m(\Omega) = 0.271$
	Jaccard	$m(\text{Paper}) = 0.375$; $m(\Omega) = 0.625$
	Wu et palmer	$m(\text{Advocate}) = 0.941$; $m(\Omega) = 0.059$
Review	Jaro	$m(\text{Reviewer}) = 0.917$; $m(\Omega) = 0.083$
	Jaccard	$m(\text{Reviewer}) = 0.75$; $m(\Omega) = 0.25$
	Wu et palmer	$m(\text{Computer_Science}) = 0.5$; $m(\Omega) = 0.5$
Reviewer	Jaro	$m(\text{Reviewer}) = 1.0$
	Jaccard	$m(\text{Reviewer}) = 1.0$
	Wu et palmer	$m(\text{Reviewer}) = 1.0$
Conference	Jaro	$m(\text{Conference_Program}) = 0.852$; $m(\Omega) = 0.148$
	Jaccard	$m(\text{Conference_Program}) = 0.555$; $m(\Omega) = 0.455$
	Wu et palmer	$m(\text{People}) = 0.615$; $m(\Omega) = 0.475$
Chairman	Jaro	$m(\text{Valuable_Confirmation_of_Present_Knowledge}) = 0.681$; $m(\Omega) = 0.319$
	Jaccard	$m(\text{Programming}) = 0.461$; $m(\Omega) = 0.549$
	Wu et palmer	$m(\text{Advocate}) = 0.8$; $m(\Omega) = 0.2$
Bid	Jaro	$m(\text{Reviewed_Paper}) = 0.603$; $m(\Omega) = 0.397$

	Jaccard	$m(\text{High}) = 0.166$; $m(\Omega) = 0.844$
	Wu et palmer	$m(\text{Programming}) = 0.667$; $m(\Omega) = 0.333$
Co-author	Jaro	$m(\text{Contact_Author}) = 0.82$; $m(\Omega) = 0.18$
	Jaccard	$m(\text{Author}) = 0.66$; $m(\Omega) = 0.34$
	Wu et palmer	$m(\text{Author}) = 0.909$; $m(\Omega) = 0.091$
User	Jaro	$m(\text{Other}) = 0.633$; $m(\Omega) = 0.367$
	Jaccard	$m(\text{Other}) = 0.2875$; $m(\Omega) = 0.7125$
	Wu et palmer	$m(\text{Advocate}) = 0.889$; $m(\Omega) = 0.111$
Author	Jaro	$m(\text{Author}) = 1.0$
	Jaccard	$m(\text{Author}) = 1.0$
	Wu et palmer	$m(\text{Author}) = 1.0$

3.5. Combinaison conjonctive normalisée (Shafer & Dempster) :

Nous avons choisis les mesures suivantes :

- Mesure 1 = Jaro
- Mesure 2 = Jaccard
- Mesure 3 = Wu & palmer

Et leurs résultats de mesures de similarité suivante :

- $M_{\text{jaro}}(\text{Review}, \text{Reviewer}) = 0.915$
- $M_{\text{jaccard}}(\text{Review}, \text{Reviewer}) = 0.75$
- $M_{\text{wu\&palmer}}(\text{Review}, \text{Computer_Science}) = 0.5$

Premièrement on choisit un cadre de discernement qui est dans notre cas

$$\Omega = \{\text{Reviewer} ; \text{Computer_Science}\}$$

Tableau 4 : Fonction de masse de chaque mesure et sa fonction de l'ignorance

	Jaro (1)	Jaccard (2)	Wu&palmer (3)
Reviewer	0.917	0.75	0.0
Computer_Science	0.0	0.0	0.5
Ω	0.083	0.25	0.5

On combine les valeurs de similarité Jaro et Jaccard, puis le résultat sera combiné avec la 3^{ème} mesure qui est la mesure de Wu & Palmer

$$M_{12}(\text{Reviewer}) = \frac{(m_1(\text{Reviewer}) * m_2(\Omega)) + (m_2(\text{Reviewer}) * m_1(\Omega)) + m_1(\text{Reviewer}) * m_2(\text{Reviewer})}{1}$$

- $M_{12}(\text{Reviewer}) = 0.97925$
- $M_{12}(\text{Computer_Science}) = 0.0$
- $M_{12}(\Omega) = 0.02075$

Tableau 5 : combinaison entre Jaro et Jaccard

	M_{12}	Wu&palmer (3)
Reviewer	0.97925	0.0
Computer_Science	0.0	0.5
Ω	0.02075	0.5

$$M_{123}(\text{Reviewer}) = \frac{(m_{12}(\text{Reviewer}) * m_3(\Omega)) + (m_3(\text{Reviewer}) * m_{12}(\Omega)) + m_{12}(\text{Reviewer}) * m_2(\text{Reviewer})}{1 - (m_{12}(\text{Reviewer}) * m_3(\text{Reviewer}))}$$

- $M_{123}(\text{Reviewer}) = 0.95$
- $M_{123}(\text{Computer_Science}) = 0.04$
- $M_{123}(\Omega) = 0.01$

Tableau 6 : résultats de la combinaison conjonctive normalisée

M_{123}	Combinaisons conjonctive normalisé
Reviewer	0.95
Computer_Science	0.04
Ω	0.01

Dans ce tableau ci-dessous les résultats de combinaison conjonctive normalisé pour les deux ontologies

Tableau 7 : résultats de la combinaison conjonctive normalisé

$e_1 \in O_1$	Combinaison des masses
Administrator	$m(\text{Clarity_to_Present_Understanding}) = 0.231$

	$m(\text{Format}) = 0.168$
	$m(\text{Advocate}) = 0.481$
	$m(\Omega) = 0.12$
Person	$m(\text{Clarity_to_Present_Understanding}) = 0.133$
	$m(\text{Paper}) = 0.030$
	$m(\text{Advocate}) = 0.789$
	$m(\Omega) = 0.048$
Review	$m(\text{Reviewer}) = 0.95$
	$m(\text{Computer_Science}) = 0.04$
	$m(\Omega) = 0.01$
Reviewer	$m(\text{Reviewer}) = 1.0$
Conference	$m(\text{Conference_Program}) = 0.845$
	$m(\text{People}) = 0.095$
	$m(\Omega) = 0.06$
Chairman	$m(\text{Valuable_Confirmation_of_Present_Knowledge}) = 0.267$
	$m(\text{Programming}) = 0.1073$
	$m(\text{Advocate}) = 0.50$
	$m(\Omega) = 0.1257$
Bid	$m(\text{Reviewed_Paper}) = 0.327$
	$m(\text{High}) = 0.042$
	$m(\text{Programming}) = 0.424$
	$m(\Omega) = 0.207$
Co-author	$m(\text{Contact_Author}) = 0.60$
	$m(\text{Author}) = 0.40$
	$m(\Omega) = 0.0$
User	$m(\text{Other}) = 0.238$
	$m(\text{Advocate}) = 0.677$
	$m(\Omega) = 0.085$
Author	$m(\text{Author}) = 1.0$

3.6. Combinaison conjonctive

On prend le même exemple précédent

$\Omega = \{\text{Reviewer ; Computer_Science}\}$

	Jaro (1)	Jaccard (2)	Wu&palmer (3)
Reviewer	0.917	0.75	0.0
Computer_Science	0.0	0.0	0.5
Ω	0.083	0.25	0.5

On combine les mesures des 2 premières mesures, puis le résultat sera combiné avec la troisième mesure

$$M_{12}(\text{Reviewer}) = ((m_1(\text{Reviewer}) * m_2(\Omega)) + (m_2(\text{Reviewer}) * m_1(\Omega)) + m_1(\text{Reviewer}) * m_2(\text{Reviewer}))$$

- $M_{12}(\text{Reviewer}) = 0.9792$
- $M_{12}(\text{Computer_Science}) = 0.0$
- $M_{12}(\Omega) = 0.02075$

	M_{12}	Wu&palmer (3)
Reviewer	0.9792	0.0
Computer_Science	0.0	0.5
Ω	0.02075	0.5

$$M_{123}(\text{Reviewer}) = ((m_{12}(\text{Reviewer}) * m_3(\Omega)) + (m_3(\text{Reviewer}) * m_{12}(\Omega)) + m_{12}(\text{Reviewer}) * m_2(\text{Reviewer}))$$

$$M_{123}(\text{Reviewer}) = 0.49$$

$$M_{123}(\text{Computer_Science}) = 0.03$$

$$M_{123}(\Omega) = 0.48$$

	Combinaisons conjonctive normalisé
Reviewer	0.49
Computer_Science	0.03
Ω	0.48

Dans ce tableau ci-dessous les résultats de combinaison conjonctive pour les deux ontologies

Tableau 8 : Résultats de la combinaison conjonctive

$e_1 \in O_1$	Combinaison des masses
Administrator	$m(\text{Clarity_to_Present_Understanding}) = 0.0548$

	$m(\text{Format}) = 0.0398$
	$m(\text{Advocate}) = 0.114$
	$m(\Omega) = 0.791$
Person	$m(\text{Clarity_to_Present_Understanding}) = 0.0268$
	$m(\text{Paper}) = 0.0059$
	$m(\text{Advocate}) = 0.1593$
	$m(\Omega) = 0.808$
Review	$m(\text{Reviewer}) = 0.49$
	$m(\text{Computer_Science}) = 0.03$
	$m(\Omega) = 0.48$
Reviewer	$m(\text{Reviewer}) = 1.0$
Conference	$m(\text{Conference_Program}) = 0.359$
	$m(\text{People}) = 0.121$
	$m(\Omega) = 0.52$
Chairman	$m(\text{Valuable_Confirmation_of_Present_Knowledge}) = 0.073$
	$m(\text{Programming}) = 0.029$
	$m(\text{Advocate}) = 0.137$
	$m(\Omega) = 0.761$
Bid	$m(\text{Reviewed_Paper}) = 0.167$
	$m(\text{High}) = 0.022$
	$m(\text{Programming}) = 0.220$
	$m(\Omega) = 0.591$
Co-author	$m(\text{Contact_Author}) = 0.024$
	$m(\text{Author}) = 0.010$
	$m(\Omega) = 0.966$
User	$m(\text{Other}) = 0.08$
	$m(\text{Advocate}) = 0.45$
	$m(\Omega) = 0.47$
Author	$m(\text{Author}) = 1.0$

3.7. Combinaison disjonctive :

$$M(\text{Reviewer U Computer_Science}) = M_{\text{aro}}(\text{Reviewer}) * M_{\text{jaccard}}(\text{Reviewer}) *$$

$$M_{\text{wu\&palmer}}(\text{Computer_Science})$$

$$M(\text{Reviewer U Computer_Science}) = 0.343$$

Dans ce tableau ci-dessous les résultats de combinaison disjonctive pour les deux ontologies

Tableau 9 : Résultats de la combinaison disjonctive

$e_1 \in O_1$	Combinaison des masses
Administrator	$m(\text{Clarity_to_Present_Understanding} \cup \text{Format} \cup \text{Advocate}) = 0.307$
	$m(\Omega) = 0.673$
Person	$m(\text{Clarity_to_Present_Understanding} \cup \text{Paper} \cup \text{Advocate}) = 0.257$
	$m(\Omega) = 0.743$
Review	$m(\text{Reviewer} \cup \text{Computer_Science}) = 0.343$
	$m(\Omega) = 0.657$
Reviewer	$m(\text{Reviewer}) = 1.0$
Conference	$m(\text{Conference_Program} \cup \text{People}) = 0.291$
	$m(\Omega) = 0.719$
Chairman	$m(\text{Valuable_Confirmation_of_Present_Knowledge} \cup \text{Programming} \cup \text{Advocate}) = 0.25$
	$m(\Omega) = 0.75$
Bid	$m(\text{Reviewed_Paper} \cup \text{High} \cup \text{Programming}) = 0.06$
	$m(\Omega) = 0.94$
Co-author	$m(\text{Contact_Author} \cup \text{Author}) = 0.496$
	$m(\Omega) = 0.504$
User	$m(\text{Other} \cup \text{Advocate}) = 0.168$
	$m(\Omega) = 0.832$
Author	$m(\text{Author}) = 1.0$

4. Analyse et synthèse

Après avoir effectué plusieurs tests sur les résultats obtenus par notre système d'alignement, nous avons trouvé que la combinaison disjonctive donne les meilleures résultats selon les benchmark, de plus les concepts qui ont une valeur de masse supérieure ou égale à 0.8 donnent un mapping valide et les mappings qui ne vérifient pas cette condition sont des mappings non valides

5. Conclusion

Après avoir spécifié et analysé les besoins du système, nous avons présenté la mise en œuvre de son architecture ainsi que la conception globale et détaillée, afin de pouvoir construire notre système de façon objective et organisée. Nous allons entamer donc dans le chapitre suivant l'implémentation de notre système.

Chapitre 5 :

Implémentation

1. Introduction

Implémenter c'est « effectuer l'ensemble des opérations qui permettent de définir un projet et de le réaliser, de l'analyse du besoin à l'installation et la mise en service du système ou du produit »,

Dans ce chapitre nous allons décrire de façon visuelle l'implémentation de notre système, en effectuant des captures d'écran des différentes interfaces du système.

2. Outils de développement de notre système

2.1. *NetBeans et JDK :*

Il offre de nombreuses fonctionnalités pratiques. Ainsi qu'un environnement «Java développent Kit » JDK est requis pour les développements en Java de notre système.

2.2. *OWL2Prefuse*

OWL2Prefuse est un package Java qui crée des graphes et des arbres Prefuse à partir de fichiers OWL. Il prend soin de convertir la structure de données OWL en une structure de données Prefuse.

2.3. *JENA*

Jena est une API Java qui peut être utilisée pour créer et manipuler des graphes RDF et les ontologies exprimé en OWL. Jena possède des classes pour représenter des graphes, des ressources, des propriétés et des littéraux. Dans Jena, un graphe est appelé un modèle et est représenté par l'interface Model.

2.4. *WordNet 3.0*

WordNet est une base de données lexicale. Son but est de répertorier, classer et mettre en relation de diverses manières le contenu sémantique et lexical de la langue anglaise.

2.5. *JAWS*

JAWS (Java API for WordNet Searching) comme son nom l'indique c'est un API Java qui permet aux applications Java de récupérer des données à partir de la base de données WordNet

3. Les Benchmarks

Le nombre croissant de méthodes disponibles pour le mapping d'ontologie nécessite d'établir un consensus pour l'évaluation de ces méthodes. Depuis 2004, Les chercheurs en alignement des ontologies, et avec le nombre croissant des systèmes automatiques d'alignement, organisent des campagnes annuelles d'évaluation appelées OAEI (Ontology Alignment Evaluation Initiative).

OAEI est une campagne qui vise à comparer les systèmes de mapping des ontologies sur des cas de test définis précisément. Les cas de test peuvent utiliser des ontologies de nature différente.

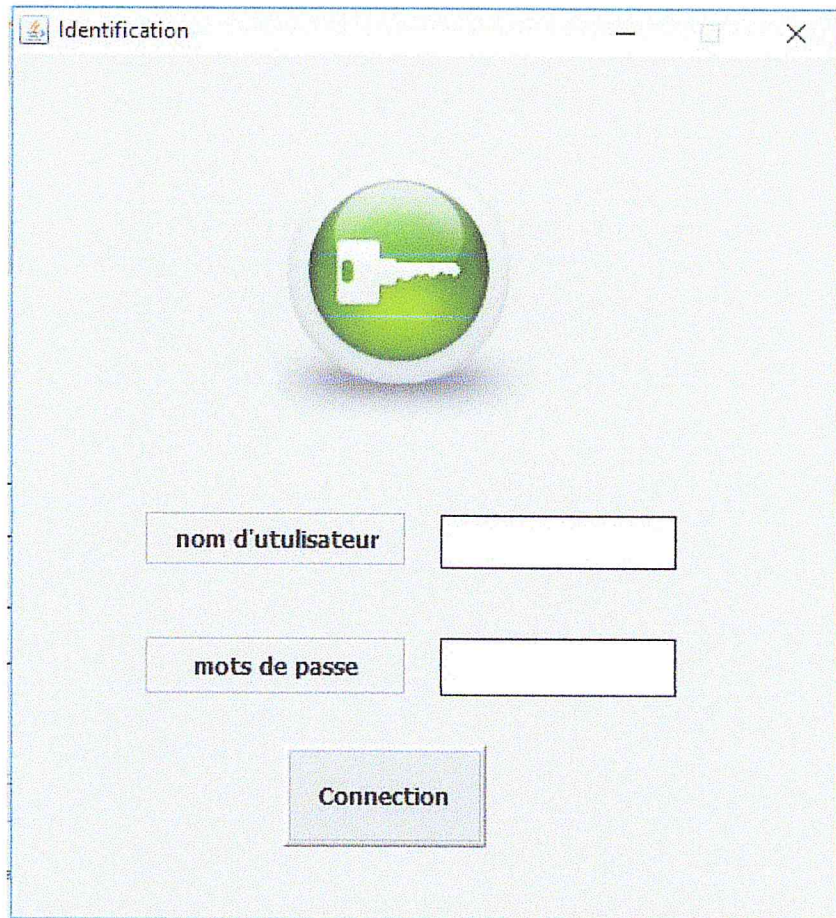
Les objectifs de l'initiative d'évaluation d'alignement d'ontologies sont :

- Mettre en évidence les forces et les faiblesses de ces systèmes.
- Comparer les performances des différentes techniques.
- Favoriser la communication entre les développeurs de ces systèmes.
- Développer l'évaluation des techniques d'alignement.
- D'une manière générale, faire progresser la recherche dans le domaine de l'alignement des ontologies.

Dans le cadre des expérimentations, les tests fournis dans la base benchmark mise à la disposition de la communauté d'OAEI sont utilisés. Le domaine de ce test est le domaine des références bibliographiques.

4. La mise en œuvre du système

4.1. L'authentification



The image shows a screenshot of a software window titled "Identification". The window has a light blue background. At the top center, there is a green circular icon with a white key inside. Below the icon, there are two input fields. The first field is labeled "nom d'utilisateur" and the second field is labeled "mots de passe". Below these fields, there is a button labeled "Connection". The window has standard Windows-style window controls (minimize, maximize, close) in the top right corner.

Figure 7 : Interface « Authentification »

C'est l'interface qui permet à un utilisateur de s'authentifier avec un nom d'utilisateur, et un mot de passe.

4.2. Interfaces de l'administrateur

4.2.1. L'interface globale de l'administrateur

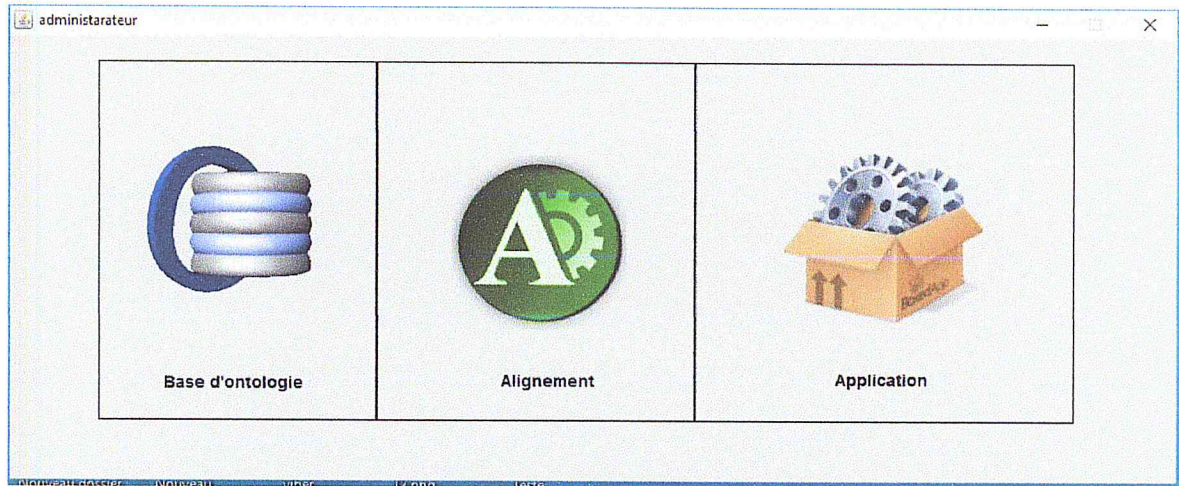


Figure 8 : Interface « Administrateur »

Administrateur a 4 boutons dans son interface :

- Base d'ontologie : elle permettra à l'administrateur de consulter les bases d'ontologie existante.
- Alignement : Il permettra à l'administrateur d'aligner deux ontologies à la fois.
- Applications : L'administrateur peut accéder aux applications proposées par notre système.



4.2.2. L'interface « Applications »

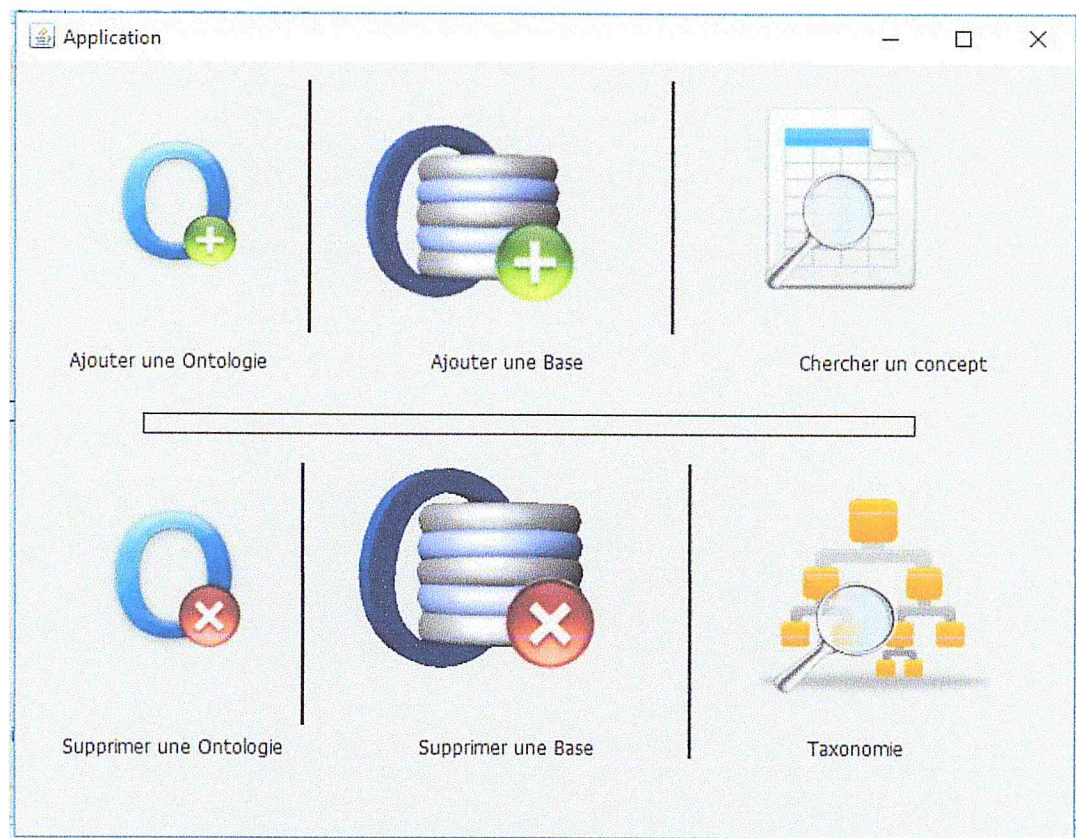


Figure 9: Interface « Applications »

L'administrateur a une interface Applications contenant différentes tâches à réaliser :

- Ajouter une ontologie à une base d'ontologie existante.

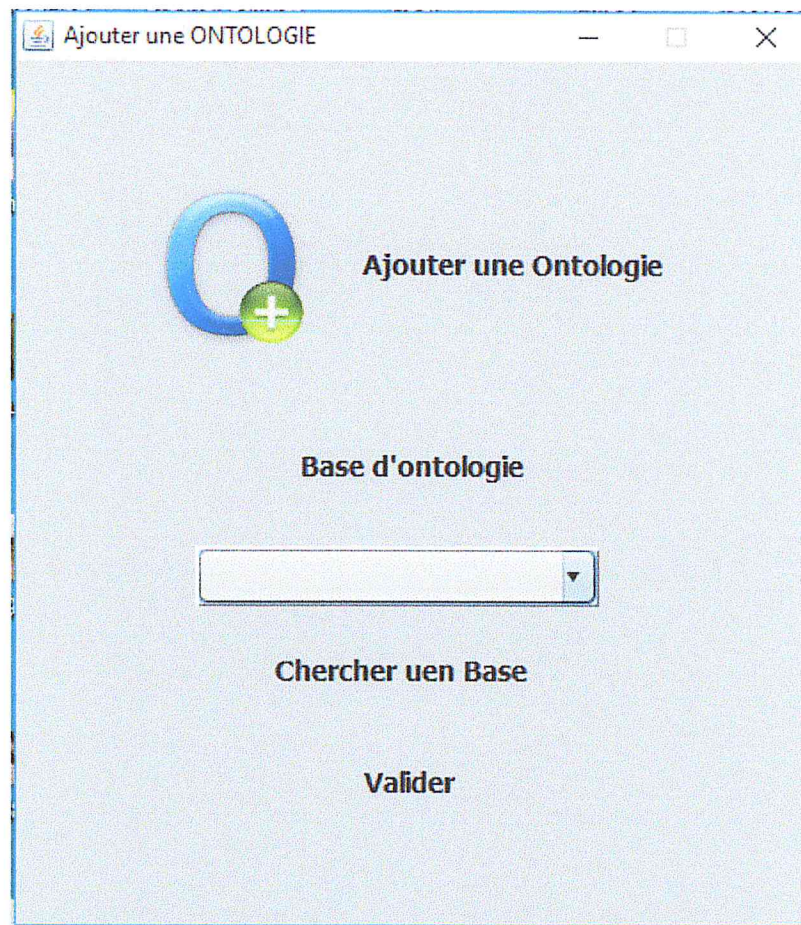


Figure 10 : Interface "ajouter une ontologie"

- Supprimer une ontologie existante.

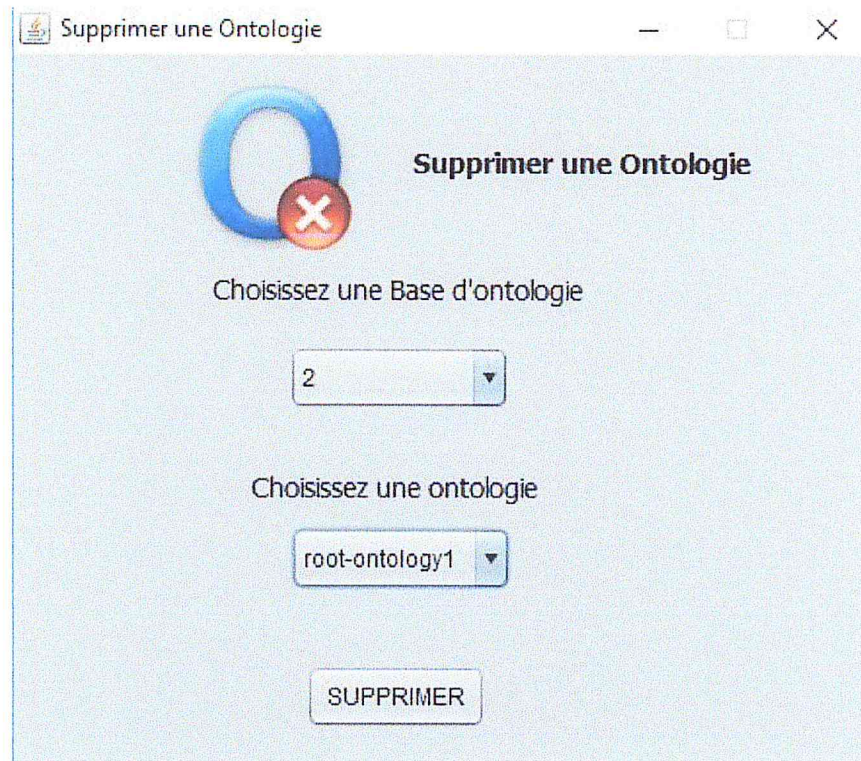


Figure 11 : Interface "supprimer une Ontologie "

- Ajouter une base d'ontologie.

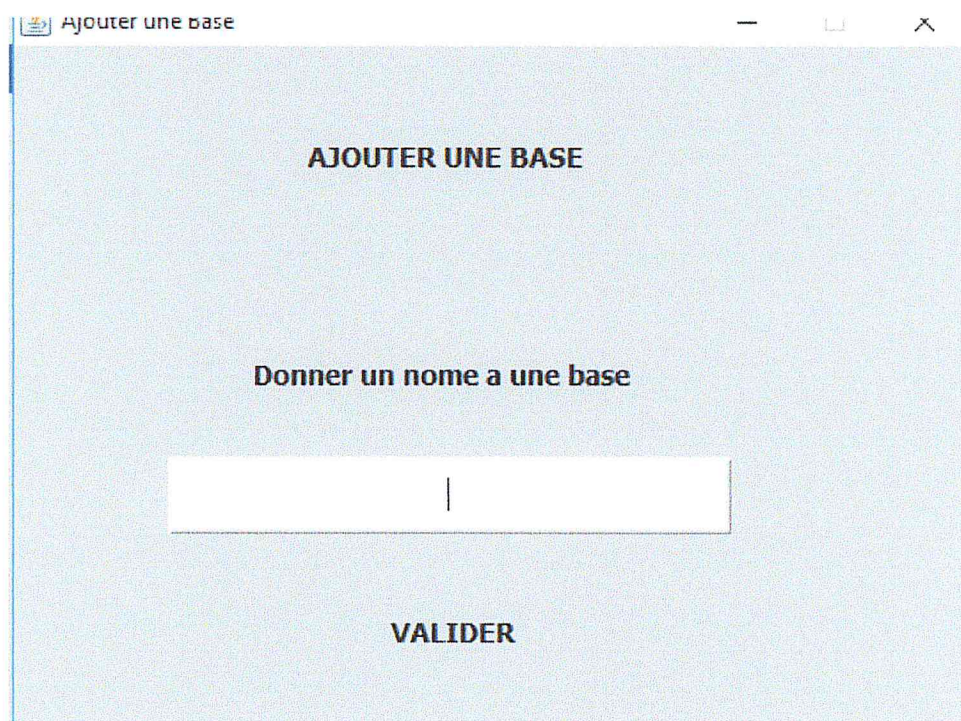


Figure 12 : Interface «Ajouter une Base d'ontologie"

- Supprimer une base d'ontologie.

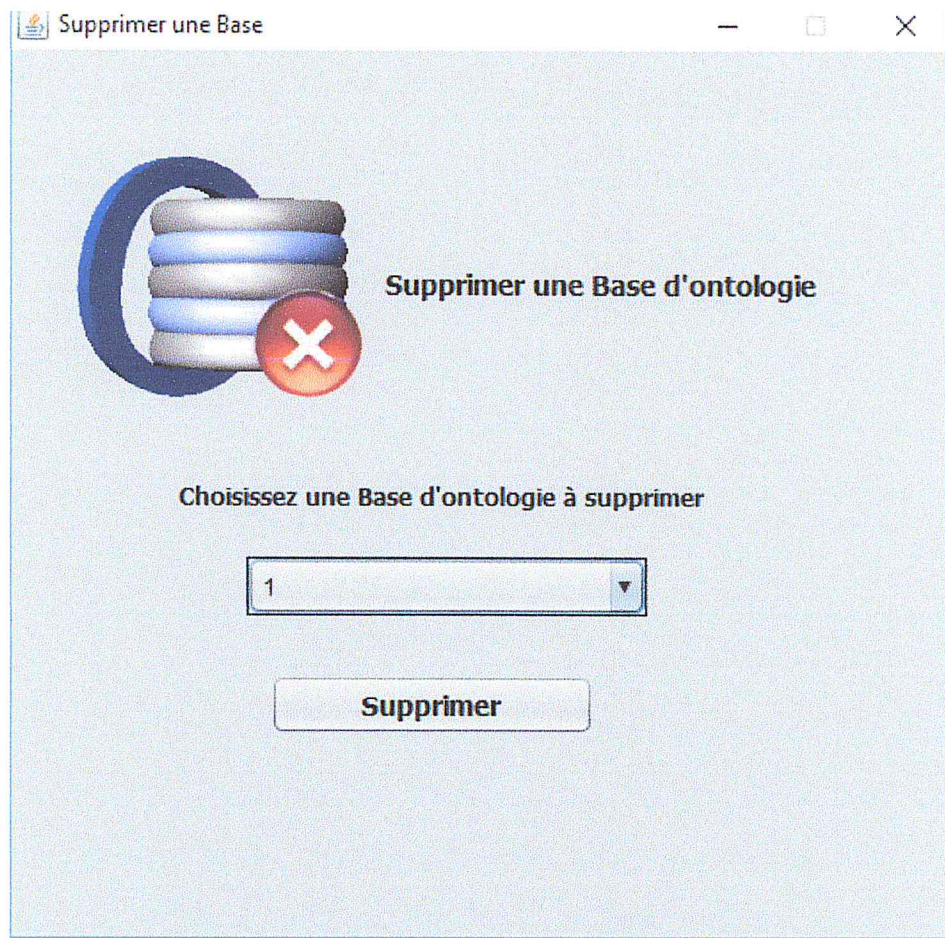


Figure 13 : Interface "Supprimer une Base"

➤ Voir Taxonomie :

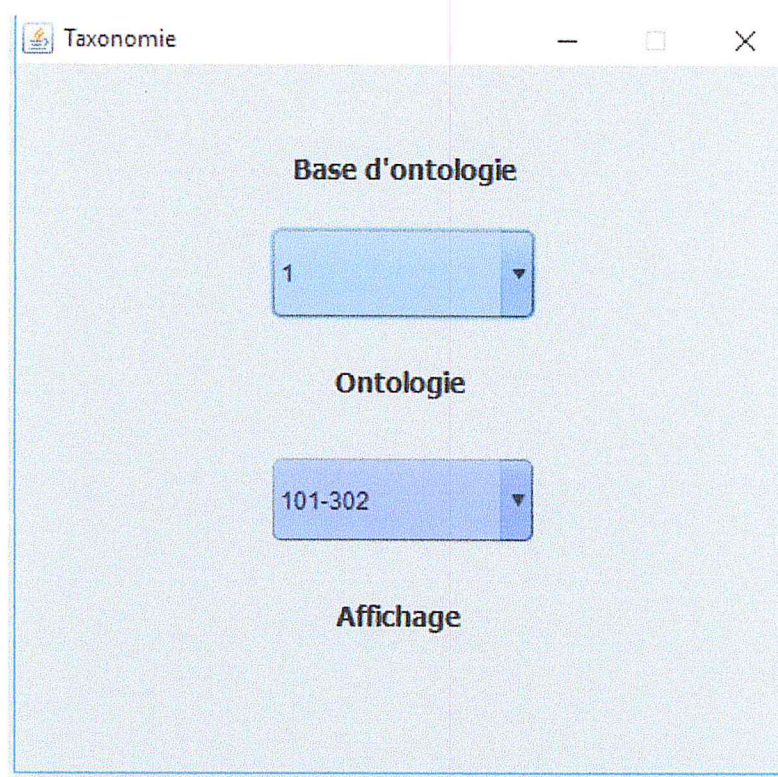


Figure 14 : Interface Voir Taxonomie

L'utilisateur peut voir une taxonomie d'une ontologie en suivant ces étapes :

- Choisir une base d'ontologie contenant l'ontologie voulue
- Choisir une Ontologie
- Cliquez sur le bouton Affichage

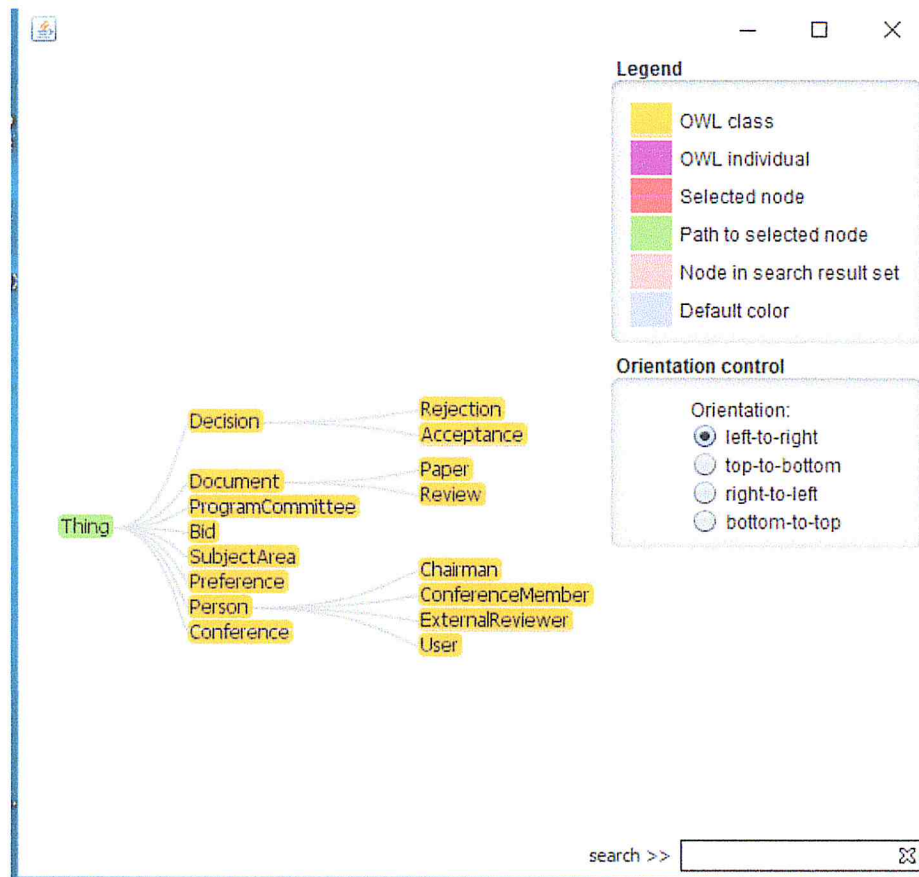


Figure 15: Interface Voir Taxonomie

4.2.3. L'interface d'alignement :

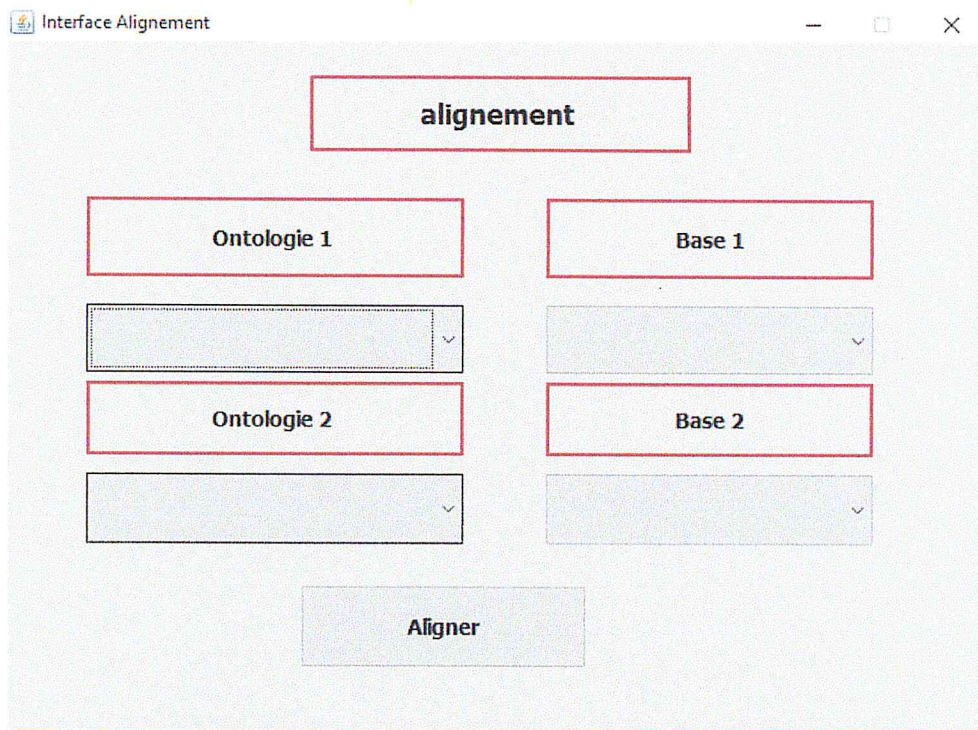


Figure 16 : Interface d'application

Afin qu'un utilisateur puisse aligner deux ontologies, il doit faire ce qui suit :

- Choisir la base où se situe l'ontologie 1 puis il choisit l'ontologie à sélectionner.
- Choisir la base où se situe l'ontologie 2 puis il choisit l'ontologie à sélectionner.
- Cliquer sur le bouton Aligner.

L'interface qui s'affichera est la suivante :

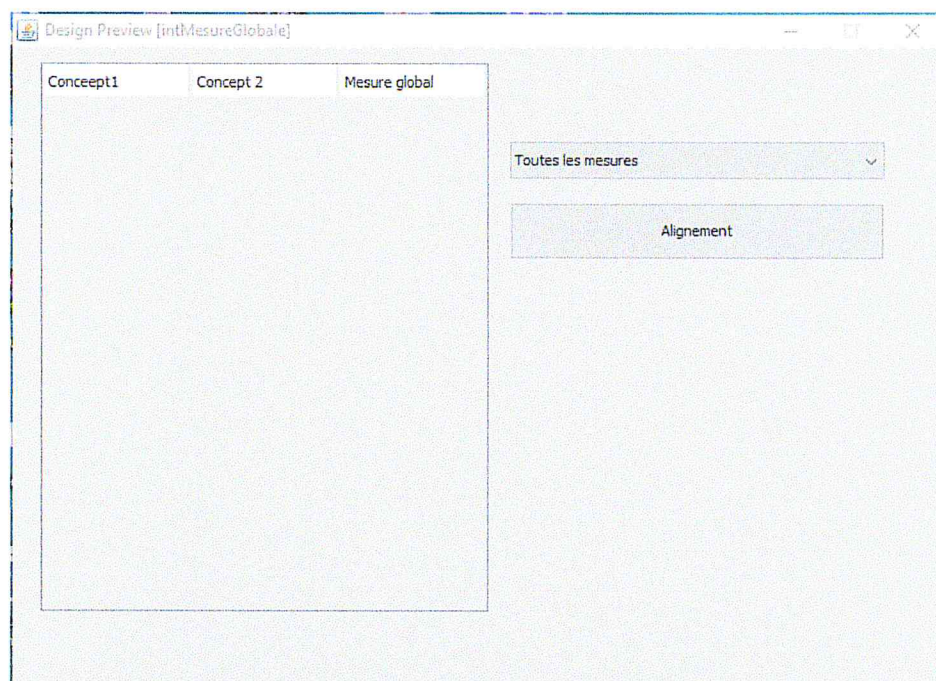


Figure 17 : Interface d'application

Le système lui affichera le résultat de l'alignement des deux ontologies choisies :

➤ L'utilisateur peut voir les détails des mesures de similarité en choisissant toutes les mesures ou bien une mesure spécifique dans la liste de type de mesure, puis il cliquera sur Afficher

concepte 1	concepte 2	mesure Jaro	mesure Jaccard	mesure wu et pa...	Combinaison sh...	Combinaison co...	Combinaison di...
MastersThesis	MastersThesis	1.0	--	--	1.0	1.0	--
MastersThesis	MastersThesis	--	1.0	--	1.0	1.0	--
MastersThesis	MastersThesis	--	--	1.0	1.0	1.0	--
MastersThesis	MastersThesis ...	--	--	--	--	--	1.0
PhdThesis	PhdThesis	1.0	--	--	1.0	1.0	--
PhdThesis	PhdThesis	--	1.0	--	1.0	1.0	--
PhdThesis	PhdThesis	--	--	1.0	1.0	1.0	--
PhdThesis	PhdThesis U Ph...	--	--	--	--	--	1.0
Unpublished	Unpublished	1.0	--	--	1.0	1.0	--
Unpublished	Unpublished	--	1.0	--	1.0	1.0	--
Unpublished	Unpublished	--	--	1.0	1.0	1.0	--
Unpublished	Unpublished U ...	--	--	--	--	--	1.0
List	List	1.0	--	--	1.0	1.0	--
List	List	--	1.0	--	1.0	1.0	--
List	List	--	--	1.0	1.0	1.0	--
List	List U List U List	--	--	--	--	--	1.0
PersonList	PersonList	1.0	--	--	1.0	1.0	--
PersonList	PersonList	--	1.0	--	1.0	1.0	--
PersonList	PersonList	--	--	1.0	1.0	1.0	--
PersonList	PersonList U Pe...	--	--	--	--	--	1.0
Chapter	Chapter	1.0	--	--	1.0	1.0	--
Chapter	Chapter	--	1.0	--	1.0	1.0	--
Chapter	Chapter	--	--	1.0	1.0	1.0	--
Chapter	Chapter U Chapt...	--	--	--	--	--	1.0
Book	Book	1.0	--	--	1.0	1.0	--
Book	Book	--	1.0	--	1.0	1.0	--
Book	Book	--	--	1.0	1.0	1.0	--
Book	Book U Book U ...	--	--	--	--	--	1.0
Misc	Misc	1.0	--	--	1.0	1.0	--
Misc	Misc	--	1.0	--	1.0	1.0	--
Misc	Misc	--	--	1.0	1.0	1.0	--

Figure 18 : Interface de combinaison des mesures

5. Conclusion

Nous avons vu dans ce chapitre les différents outils et technologies nécessaires pour la réalisation de notre système tels que l'environnement de développement, les langages de programmation, les bibliothèques open source utilisées...etc.

Nous avons présenté clairement et en détail toutes les interfaces nécessaires, en utilisant des captures d'écran.

Chapitre 6 :

Test et évaluation

1. Introduction

Dans cette dernière partie nous allons effectuer un test sur notre système implémenté, ainsi que sa validation pour pouvoir dire que nous avons atteint le but de notre projet. Afin d'évaluer l'alignement des ontologies s'il est pertinent, en utilisant les mesures de similarités sélectionnées tout le long de notre projet.

2. L'évaluation du système

La méthode que nous avons utilisée pour évaluer notre système c'est l'utilisation du corpus de test de la campagne d'évaluation des systèmes d'alignements OAEI 2016.[OAEI] Ce corpus offre la possibilité d'avoir des alignements prêts provenant de plusieurs outils d'alignement.

Dans cette évaluation nous avons utilisé les ontologies 101 et 304 fournis par l'OAEI, nous allons donc comparer les résultats donnés par la campagne et ceux effectués par notre système.

Source: <http://oaei.ontologymatching.org/tests/101/onto.rdf>

Target : <http://oaei.ontologymatching.org/tests/304/onto.rdf>

Correspondences

PersonList = PersonList	1.0	Unpublished = Unpublished	1.0
Chapter < InBook	1.0	InBook < InBook	1.0
Date = Date	1.0	PageRange = PageRange	1.0
Booklet = Booklet	1.0	LectureNotes = LectureNotes	1.0
MastersThesis = MastersThesis	1.0	TechReport = TechReport	1.0
lcn = lcn	1.0	Misc = Misc	1.0
Collection = Collection	1.0	Academic = Academic	1.0
MotionPicture = MotionPicture	1.0	Deliverable = Deliverable	1.0
Article = Article	1.0	Informal = Informal	1.0
Book = Book	1.0	School = School	1.0
PhdThesis = PhdThesis	1.0	Proceedings = Proceedings	1.0
InCollection = InCollection	1.0	Part = Part	1.0
Publisher = Publisher	1.0	Manual = Manual	1.0
InProceedings = InProceedings	1.0	Journal = Journal	1.0
Monograph = Monography	1.0		

3. Qualité d'alignement d'ontologies

Les mesures de Précision, Rappel et Fallout [Do H et al, 2002] ont été des métriques largement exploitées pour estimer la qualité des alignements obtenus. Le EON3 "Evaluation of Ontology-based Tools" [Euznat J, Shvaiko P et al, 2006] retient ces mesures pour l'évaluation de la qualité de l'alignement. L'objectif principal de ces mesures est l'automatisation du processus de comparaison des méthodes d'alignement ainsi que l'évaluation de la qualité des alignements produits. La première phase dans le processus d'évaluation de la qualité de l'alignement consiste à résoudre le problème manuellement. Le résultat obtenu manuellement est considéré comme l'alignement de référence.

La comparaison du résultat de l'alignement de référence avec celui de l'appariement obtenu par la méthode d'alignement produit trois ensembles : Nfound, Nexpected et Ncorrect. L'ensemble Nfound représente les paires alignées avec la méthode d'alignement. L'ensemble Nexpected désigne l'ensemble des couples appariés dans l'alignement de référence. L'ensemble Ncorrect est l'intersection des deux ensembles Nfound et Nexpected. Il représente l'ensemble des paires appartenant à la fois à l'alignement obtenu et l'alignement de référence.

La précision est le rapport du nombre de paires pertinentes trouvées, i.e., "Ncorrect", rapporté au nombre total de paires, i.e., "Nfound". Il renvoie ainsi, la partie des vraies correspondances parmi celles trouvées. Ainsi, la fonction précision est définie par :

$$Précision = \frac{|Ncorrect|}{|Nfound|}$$

Le rappel est le rapport du nombre de paires pertinentes trouvées, "Ncorrect", rapporté au nombre total de paires pertinentes, "Nexpected". Il spécifie ainsi, la part des vraies correspondances trouvées. La fonction rappel est définie par :

$$Rappel = \frac{|N_{correct}|}{|N_{expected}|}$$

4. Test

Le tableau suivant représente la comparaison des correspondances trouvées de façon manuelle et celle trouvée avec notre système d'alignement de deux ontologies 101 et 304.

Cet alignement représente un alignement de référence.

Le tableau suivant est juste une partie du résultat d'alignement avec notre système.

✓ : disponible

✗ : non disponible

Concept 1	Concept 2	Correspondances Manuelles	Mesure de similarité	Correspondances de notre système
MastersThesis	MastersThesis	✓	1.0	✓
PhdThesis	PhdThesis	✓	1.0	✓
Unpublished	Unpublished	✓	1.0	✓
List	List	✓	1.0	✓
PersonList	PersonList	✓	1.0	✓
Chapter	Chapter	✓	1.0	✓
Book	Book	✓	1.0	✓
Misc	Misc	✓	1.0	✓
Date	Date	✓	1.0	✓
Manual	Manual	✓	1.0	✓
Deliverable	Deliverable	✓	1.0	✓
Monograph	Monography	✓	0.87	✓
Monograph	Chapter	✗	0.87	✓
School	School	✓	1.0	✓
PageRange	PageRange	✓	1.0	✓
Person	Person	✓	1.0	✓
Proceedings	Proceedings	✓	1.0	✓
Collection	Collection	✓	1.0	✓
Institution	Institution	✓	1.0	✓
TechReport	TechReport	✓	1.0	✓
Journal	Journal	✓	1.0	✓
Journal	Chapter	✗	1.0	✓
Organization	Organization	✓	1.0	✓
Article	Article	✓	1.0	✓
InProceedings	InProceedings	✓	1.0	✓

Informal	Informal	✓	1.0	✓
Part	Part	✓	1.0	✓
InCollection	InCollection	✓	1.0	✓
Academic	Academic	✓	1.0	✓
Address	Address	✓	1.0	✓
Publisher	Publisher	✓	1.0	✓
LectureNotes	LectureNotes	✓	1.0	✓
MotionPicture	MotionPicture	✓	1.0	✓
InBook	InBook	✓	1.0	✓
Booklet	Booklet	✓	1.0	✓

Tableau 10 : La comparaison des correspondances trouvé manuellement et celle trouvé avec notre système d'alignement de deux ontologies

5. Validation

➤ Les mesures de rappel, précision, et Fallout :

On a : $\left\{ \begin{array}{l} |N_{found}| = 35 \\ |N_{expected}| = 33 \\ |N_{correct}| = 33 \end{array} \right.$

Donc : $\left\{ \begin{array}{l} \text{Précision} = \frac{|N_{correct}|}{|N_{found}|} = 0.94 \\ \text{Rappel} = \frac{|N_{expected}|}{|N_{correct}|} = 1.0 \end{array} \right.$

✓ Les résultats obtenus par l'application des opérations sont comparés aux résultats obtenus par le meilleur outil d'alignement selon l'OAEI. Nous avons montré que nos résultats seront raffinés.

6. Conclusion

Ce chapitre a été consacré aux tests et validation de notre système. En ce qui concerne les objectifs listés dans le chapitre d'introduction, l'évaluation et l'expérimentation de notre

système ont permis de répondre au but de notre système qui est trouvé un alignement entre deux ontologies

Conclusion générale

Ces dernières années, les ontologies ont suscité de nombreux travaux dans le domaine du web sémantique. Toutefois, en raison de la nature décentralisée du web sémantique, les ontologies sont très hétérogènes. L'alignement des ontologies est une solution pour résoudre l'hétérogénéité sémantique des données, en proposant des correspondances entre des entités sémantiques similaires de différentes ontologies.

L'hétérogénéité sémantique présente un défi majeur dans le processus d'alignement des ontologies. En effet, les ontologies existantes sont conçues indépendamment, par des concepteurs différents, ayant des objectifs applicatifs différents. Chacune peut donc avoir un point de vue différent sur le même concept.

Les travaux menés dans ce mémoire nous ont permis d'approfondir nos connaissances dans le domaine de l'ingénierie des connaissances (et plus particulièrement l'ingénierie Ontologique). Notre objectif est de proposer une architecture d'alignement d'ontologies pour deux ontologies en OWL qui décrivent le même domaine, en utilisant différentes techniques reposantes sur la définition de mesures de similarité et la combinaison de ses mesures selon la définition du fonction de croyance afin de trouver les relations entre les entités de ces ontologies.

Nous avons combiné des techniques de base de l'alignement tel que les techniques syntaxiques (mesure de Jaccard et Jaro), structurelles (mesure de Wu & Palmer), afin de rendre l'alignement plus pertinent. Notre but est d'appliquer la théorie du fonction du croyance et surtout la combinaison des fonction du masses avec les techniques de calcul des mesures de similarité pour avoir un bon résultat de l'alignement.

Références bibliographiques

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

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

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

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

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

[ELB, 2009]M Abdelatif ELBYED, ROMIE, une approche d’alignement d’ontologies à base d’instances, Thèse présentée pour l’obtention du grade de Docteur de l’INSTITUT NATIONAL DES TELECOMMUNICATIONS, (16 Octobre 2009)

[M.gensenth et al ; 1992] knowldge interchang format version 3.0 reference manual

[Noy et al , 2001]Noy, N. F., & Musen, M. A. (2000). Prompt: algorithm and tool for automated ontology merging and alignment. In Proceeding of Seventeenth National Conference on Artificial Intelligence AAAI.

[Euzenat, J. et al 2004]Similarity-based ontology alignment in OWLLite. In Proceedings of the 15th European conference on Artificial Intelligence (ECAI), Pages 333_337, Valence, Espagne.

[David ; 2007] Jérôme David AROMA : une methode pour la decouverte d’alignement orienté entre ontologies a partir de règles d’association PhD thesis, Universite de Nantes, 2007

[Dempster ; 1966] Arthur Dempster : upper and lower probailites inded by a multivaluet mapping

[G, shafer 1976] Shafer : a mathematical theory of evidance 1976

[Th.Denoeux ; 2011] Thierry Denoeux : introduction to belief function spring school BFTA 2011

[Besana, P. 2006]. A framework for combining ontology and schema matchers with dempster-shafer. In *Proceedings of the 1st International Workshop on Ontology Matching (OM-2006) collocated with the 5th International Semantic web Conference (ISWC-2006), Athens, Georgia, USA* (Vol. 225).

[Wang et al., 2007] Ying Wang, Weiru Liu, and David Bell Combining Uncertain Outputs from Multiple Ontology Matchers

[Nagy, M.,2007] Nagy, M., Vargas-Vera, M., & Motta, E. DSSim - Managing Uncertainty on the Semantic Web

[Shvaiko,Euzenat, 2005] Shvaiko P., Euzenat J., A Survey of Schema-based Matching Approaches , Journal on Data Semantics, (2005).

[Do H et al , 2002] Do H., Melnik S., Rahm E.2002, « Comparison of schema matching evaluations», Proceedings of the 2nd Int. Workshop on Web Databasess , German Informatics Society, Erfurt.

[Wu & Palmer, 1994] Z. Wu et M. Palmer.1994. Verb semantics and lexical selection. In Proceedings of the 32nd Annual Meeting of the Associations for Computational Linguistics, p 133-138.

[Slimani et al, 2007] Slimani T et al.2007.Une extension de mesure de similarité entre les concepts d'une ontologie. p 4-5.

[Arnaud M et al.2012] *Gestion du conflit dans l'appariement des ontologies.* p 4.

[OAEI]. OAEI. . <http://oaei.ontologymatching.org/>, consulte en.2017

[Euznat J, Shvaiko P et al, 2006] Euznat J, Shvaiko P et al, 2006, « Results of the Ontology Alignment Evaluation Initiative 2006 », in, Proceedings of the 1st ESWC 2006 international workshop on ontology matching, Athens, (GAU US)

