



HAL
open science

Etat de l'art sur l'alignement des ontologies

Abdoulaye Diallo, Mouhamadou Thiam

► **To cite this version:**

Abdoulaye Diallo, Mouhamadou Thiam. Etat de l'art sur l'alignement des ontologies. 2021. hal-03330388

HAL Id: hal-03330388

<https://inria.hal.science/hal-03330388>

Preprint submitted on 31 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Etat de l'art sur l'alignement des ontologies

Abdoulaye Diallo¹ et Mouhamadou THIAM²

¹ Université de THIES, Thiès, Sénégal
abdoulaye.diallo4@univ-thies.sn

² Université de THIES, Thiès, Sénégal
mthiam@univ-thies.sn

Résumé : Le web permet d'accéder à différents types de données et à des services se présentant sous forme de documents variés grâce à internet. Cependant l'interface d'accès à ces documents est écrite dans un langage qui est destiné à des humains et de plus en plus aux machines. Le web sémantique est une extension du web qui améliore la coopération entre les humains et les ordinateurs. Cependant dans le cadre de cette nouvelle représentation de la connaissance apparaissent les ontologies qui permettent de conceptualisation des connaissances d'un domaine. Cette représentation n'est pas toujours uniforme pour un domaine donné. Les ontologies sont des nouvelles formes d'organisation et de contrôle intelligent de l'information. Elles présentent un savoir préalable requis pour un traitement systématique de l'information à des fins de navigation, d'exploitation, etc. Toutefois, les ontologies sont confrontées de façon continue aux problèmes d'hétérogénéité syntaxique, terminologique, conceptuelle et sémiotique ou pragmatique. Ainsi l'alignement s'impose comme une solution, afin de régler ce problème d'hétérogénéité pour une meilleure utilisation des données. Dans ce papier nous présentons l'état de l'art l'alignement des ontologies et des propositions d'amélioration sont proposées.

Mots clés : ontologie, hétérogénéité, alignement d'ontologies, OAEL.

1 Introduction

Dans cette section nous présentons le contexte et la problématique de l'alignement d'ontologie. Ensuite nous donnons les objectifs du papier.

1.1 Contexte et problématique

L'évolution exponentielle du développement de nouvelles technologies internet a créé un besoin imminent de recherche sur le partage et l'intégration des ressources du web qui sont de nature dispersée dans un environnement partagé : c'est ce qui a engendré la création du web sémantique initiée par Tim Berners-Lee [1]. Ainsi les ontologies ont été considérées comme indispensables car permettant de représenter de façon sémantique les données ou documents qui se trouvent dans le web.

Dans la littérature il existe différents langages qui permettent de représenter les ontologies (RDF(S), OWL, etc.) [2]. Avec ces langages différentes méthodes

d'alignement d'ontologies ont été développées avec pour chacune d'elles un format précis de données en entrée. Partant aussi du fait que plusieurs ontologies peuvent être développées dans un même domaine, il existe différents niveaux d'hétérogénéité. Ceci nécessite de développer des outils permettant de réconcilier ces ontologies : on parle d'alignement des ontologies. Ces différents niveaux d'hétérogénéité dans le web sémantique sont les niveaux syntaxique, terminologique, conceptuel et sémiotique ou pragmatique [3]. Ces différents niveaux sont présentés ci-dessous.

Niveau d'hétérogénéité :

Le niveau syntaxique : il s'agit de toutes les formes d'hétérogénéité relatives au choix du format de représentation. En effet, il existe différentes façons de représenter les ontologies (OWL, RDF(S), KIF, etc.) et chaque langage est basé sur une syntaxe différente.

Le niveau terminologique : l'hétérogénéité, à ce niveau, intervient dans le fait d'affecter des noms aux entités (classes, propriétés, relations, etc.) qui constituent une ontologie. Nommer une entité revient à lui associer un objet linguistique à partir d'un langage public. Des exemples de ce type d'hétérogénéité :

- Différents noms sont utilisés pour désigner une même entité (synonymie).
- Même nom utilisé pour désigner deux entités distinctes (polysémie).
- Mots provenant de différentes langues (Français, Anglais, Italien, etc.) utilisés pour désigner une même entité.
- Variations syntaxiques du même mot (différentes prononciations, abréviations, utilisation des préfixes et des suffixes, etc.).

Le niveau conceptuel : les divergences à ce niveau peuvent être résumées en trois aspects :

- La couverture : la différence entre deux ontologies peut être au niveau de la portée de la couverture du domaine décrit. Elles peuvent couvrir des parties différentes (du monde réel ou d'un domaine) ou alors des parties qui se chevauchent, par exemple : une ontologie sur le sport couvre le sport de la course automobile qu'une autre ignorait complètement.
- La granularité : deux ontologies peuvent décrire les mêmes entités avec des niveaux de détail différents, par exemple : une ontologie concernée par la comptabilité va considérer le concept générique du document alors qu'une ontologie décrivant le domaine des bibliothèques va distinguer entre les différents types de documents : romans, nouvelles, biographies, manuscrits, etc.
- La perspective : deux ontologies peuvent décrire un domaine de deux points de vue différents.

Le niveau sémiotique ou pragmatique : ce type d'hétérogénéité intervient lorsqu'il y a une différence d'interprétation de la même ontologie par différentes personnes ou différentes communautés. Ces différences d'interprétations sont souvent liées au choix du formalisme de représentation des connaissances.

L'alignement des ontologies fait une réconciliation du point de vue sémantique les opinions de plusieurs experts [4]. Chacune d'entre elles présentent une limite liée à la méthode utilisée pour le calcul de la similarité ou de la dissimilarité.

1.2 Objectifs du papier

Il existe différentes méthodes utilisées dans le domaine du web sémantique pour faire de la réconciliation des ontologies. Les terminologies les plus utilisées pour désigner ce processus sont : Correspondances ou Mappings, Appariement ou Matching, méthodes de comparaison ou Matchers, Alignement d'ontologies, Fusion d'ontologies, et Intégration d'ontologies [5].

Dans la suite, nous définissons ces concepts.

Correspondances ou Mappings : Les mappings correspondent à des indications de similarité selon une mesure donnée entre les éléments de deux représentations (Ontologies, schémas de bases de données, etc.).

Appariement ou Matching : Le Matching d'ontologies est le processus de définition d'un ensemble de matchers permettant de spécifier des correspondances entre termes. Un matcher est une fonction utilisée pour calculer la distance entre deux fonctions.

Les méthodes de comparaison ou matchers : Un matcher est une fonction utilisée pour calculer la distance entre deux entités. Les matchers peuvent être combinées dans le processus de matching.

Alignement d'ontologies : L'alignement d'ontologies est le processus d'établissement de liens de correspondances entre deux ontologies. Il est appliqué si les ontologies concernées sont homogènes entre elles et ceci tout en les gardant séparées. Cette catégorie de mapping d'ontologies est faite habituellement quand les ontologies sources appartiennent à des domaines complémentaires.

Fusion d'ontologies : La fusion d'ontologies est le processus de création d'une seule ontologie rassemblant les connaissances de deux ou plusieurs ontologies existantes et différentes qui décrivent le même sujet ou appartiennent au même domaine d'application. L'ontologie générée inclut les informations de toutes les ontologies sources.

Intégration d'ontologies : L'intégration d'ontologies est un processus de construction d'une nouvelle ontologie qui n'est pas forcément destinée à remplacer les autres (ces dernières peuvent continuer à être utilisées par ailleurs, à être mises à jour, à évoluer, etc.). Ces différentes ontologies peuvent être connexes.

Dans ce papier il s'agira de faire une présentation des différentes approches déjà proposées et de montrer leurs limites respectives. Des perspectives de propositions d'une nouvelle approche seront dégagées.

2 Présentation des approches existantes

Dans le cadre de la réconciliation des ontologies différentes approches ont été proposées. Une bonne partie de ces approches ont été développées dans des systèmes et présentées au niveau de l'OAEI [6] à partir de 2004.

2.1 Présentation de l'évolution des approches avant OAEI

De 2001 à nos jours des approches ont été proposées pour régler le problème de l'hétérogénéité. Après la présentation des premières solutions, des améliorations ont

été proposées mais aussi des nouvelles approches qui exploitent au mieux les méthodes de calcul de la similarité et les langages de représentations des ontologies suivant leur niveau d'expressivité ont vu le jour.

L'évolution du nombre d'approches proposées par année est résumée dans la **figure 1**.

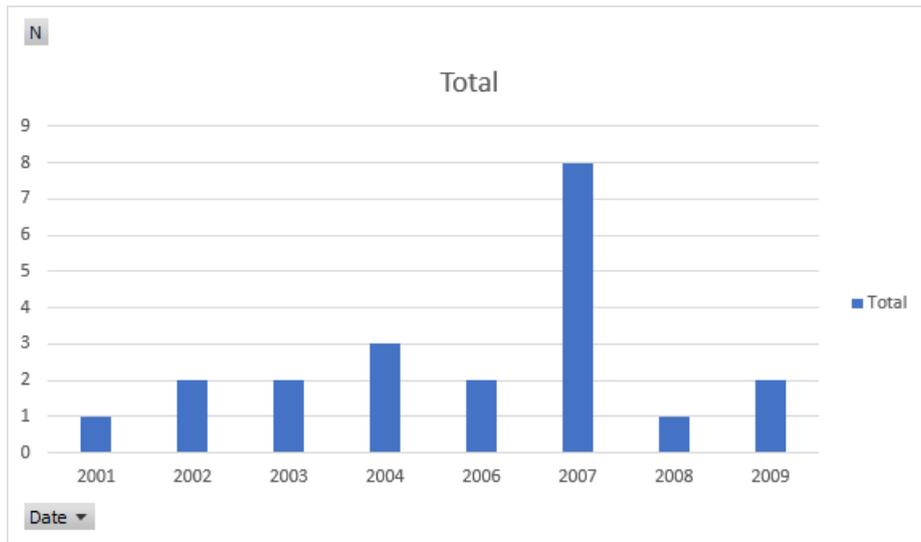


Figure 1. Evolution du nombre d'approches par année

Les approches de la figure 1 sont données dans le tableau 1 suivant les dates d'apparition.

Les années de publication de ces approches sont présentées dans le **tableau 1**.

APPROCHES	Date
Anchor-Prompt [7]	2001
GLUE [8] COMA [9]	2002
IF-MAP [10] S-MATCH [11]	2003
QOM [12] NOM [13] OLA [14]	2004
ASCO1[15] ASCO2[16]	2006
EDOLA [17] AROMA [18] SODA [19] ASMOV [20] FALCON-OA [21] COMA++ [22] ASCO3[23] TAXOMAP [24]	2007
OLA2[25]	2008
ROMIE [26] OACAS [27]	2009

Tableau 1 : Date publication des approches

Ces différentes approches proposées peuvent être comparées suivant plusieurs axes : la date et l'auteur, l'automatisme et le type (base ou améliorée).

Cette comparaison est présentée dans le **Tableau 2**.

Année	Type	Automatique	Semi-automatique
2001	Base		
	Améliorée		Anchor-Prompt (version améliorée de SMART [28], Noy et Musen)
2002	Base		
	Améliorée		GLUE (Amélioration de LSD [29], Doan et al.) COMA (combinaison de plusieurs matchers, Do et Rahm)
2003	Base	IF-Map (Kalfoglou et al.) S-Match (Giunchiglia et al.)	
	Améliorée		
2004	Base	NOM (Ehrig et al.) OLA (Euzenat et al., 2004a, Euzenat et al., 2004b)	
	Améliorée	QOM (amélioration de NOM, Ehrig et al.)	
2005	Base	ASCO1(Bach)	
	Améliorée	ASCO2(amélioration de ASCO1, Bach)	
2006	Base	ASCO1	
	Améliorée	ASCO2 (amélioration de ASCO1)	
2007	Base	AROMA (Jérôme David et al)	FALCON-OA (Hu et al.) ASMOV (Hu et al.) Taxomap (Kasri)
	Améliorée	EDOLA(amélioration de OLA, Zghal et al) SODA (Amélioration de EDOLA, Zghal et al.) ASCO 3(amélioration de ASCO2 , Bach Thanh Le, Rose-Dieng-Kuntz) OLA2(amélioration de OLA)	COMA++ (amélioration de COMA, Do and Rahm)
2008	Base		
	Améliorée	OLA2(amélioration de OLA, Djoufak et al)	
2009	Base		

	Améliorée	OACAS (amélioration de SODA, Zghal et al.)	ROMIE (Amélioration de OMIE, Elbyed et al)
--	-----------	--	--

Tableau 2: Comparaison suivant la typologie et l'automatisme

Une comparaison de ces approches suivant les axes comme les données d'entrée, les méthodes de calcul de la similarité, les données en sortie et la complexité est proposée dans le tableau en **annexe 1**.

2.2 Les approches présentées à l'OAEI

Avec le grand nombre de publications dans le domaine de la réconciliation ontologique du web sémantique est née une conférence annuelle qui permet aux différents chercheurs de présenter les approches/systèmes qui permettent de faire de la réconciliation des ontologies: l'OEIL. Depuis 2004 plus de quatre vingt(80) approches ou systèmes ont été publiées. **La figure 2** montre le nombre d'années d'apparition des différents systèmes.

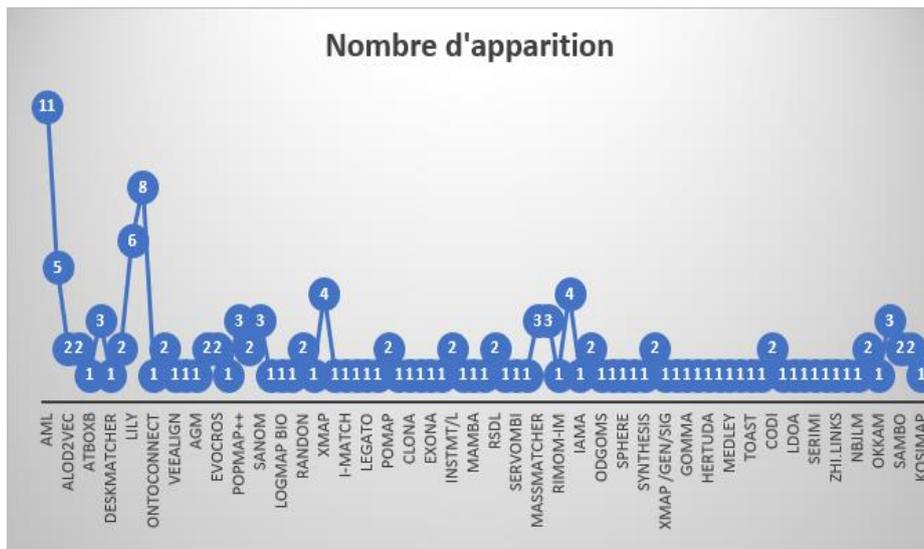


Figure 2 Nombre apparition

Parmi ces différentes approches/systèmes présentés au niveau de l'OAEI nous constatons que l'approche la mieux représentée est apparue onze fois, AML [30]. Nous nous intéressons à ce dernier. Cependant certaines approches n'ont été présentées qu'une fois. Les approches présentées plusieurs fois ont été continuellement améliorées. En guise d'illustration nous allons présenter dans le **tableau 4** les différentes améliorations subies par AML.

L'évolution de ce système est résumée dans le **tableau 4**

Année	Particularité
2009	<p>Agreement Maker 1 est la première version Elle intègre des stratégies d'appariement automatique efficaces avec une interface utilisateur polyvalente et un module pour évaluer les appariements. Agreement Maker 1 comprend une large gamme d'algorithmes de correspondance automatique appelés matchers, une architecture extensible et modulaire, une interface utilisateur polyvalente, un ensemble de stratégies d'évaluation et diverses fonctionnalités manuelles (comparaison visuelle) et semi-automatiques (boucle de rétroaction des utilisateurs).</p>
2010	<p>Le nouveau système Agreement Maker prend en charge : (1) les besoins des utilisateurs, tels que exprimés par les experts du domaine ; (2) une large gamme de formats d'entrée (ontologie) et de sortie (choisi par l'utilisateur) ; (3) un large choix de méthodes d'appariement en fonction de différentes granularités de l'ensemble de composants appariés (local vs global), des différentes caractéristiques considérées dans la comparaison (conceptuelle et structurelle), du niveau d'intervention qu'elles exigeait des utilisateurs (manuel vs automatique), sur l'utilisation (autonome ou composé) et sur les types de composants à considérer (schéma uniquement ou schéma et instances) ; (4) une amélioration des performances, c'est-à-dire de la précision (précision, rappel, F-mesure) et de l'efficacité (temps d'exécution) des méthodes automatiques ; (5) une architecture extensible pour intégrer facilement de nouvelles méthodes et ajuster leurs performances ; (6) la capacité d'évaluer, de comparer et de combiner différentes stratégies et résultats correspondants ; (7) une interface utilisateur complète qui prend en charge des techniques de visualisation avancées et un panneau de commande qui pilote toutes les méthodes d'appariement et les stratégies d'évaluation ; (8) une boucle de rétroaction qui accepte les suggestions et les corrections des utilisateurs et extrapole de nouvelles cartographies.</p>
2011	<p>Par rapport aux années précédentes, plusieurs techniques ont été proposées afin d'améliorer les algorithmes d'appariement. Le système Agreement Maker peut être exécuté avec différentes configurations qui optimisent la précision et la couverture du système en fonction des ontologies spécifiques à aligner. Changer la composition de la pile de matcher (par exemple, un matcher d'instance est utilisé uniquement lorsque des instances sont disponibles) a un impact important sur les performances du système. L'approche de cette année est présentée comme suit : les ontologies à aligner sont profilées à l'aide de plusieurs métriques proposées dans la littérature (richesse relationnelle, richesse d'héritage,</p>

	<p>couverture WorldNet, etc.). Les profils basés sur des métriques sont utilisés pour classer automatiquement la tâche correspondante dans une classe de configuration avec des paramètres spécifiques. La classification est basée sur un cadre d'apprentissage automatique supervisé formé avec un sous-ensemble de l'ensemble de données OEAI pour lequel un alignement de référence est disponible. Elle fournit un cadre d'apprentissage flexible ou on peut utiliser plusieurs combinaisons de matchers.</p>
2013	<p>La version de cette année contient six étapes: chargement de l'ontologie, mise en correspondance et profilage de base, mise en correspondance des connaissances de base (facultatif), mise en correspondance et sélection d'extensions, mise en correspondance des propriétés (conditionnelle) et réparation (optionnel).</p>
2014	<p>L'AML est principalement basée sur des techniques d'appariement lexical, en mettant l'accent sur l'utilisation de ressources externes comme connaissances de base. Il met également l'accent sur la cohérence de l'alignement, avec un module de réparation d'alignement amélioré.</p> <p>Le système comprend maintenant neuf étapes différentes : chargement et profilage d'ontologies, traduction, correspondance de base, correspondance de connaissances de base, correspondance de mots et de chaînes, correspondance structurelle, correspondance de propriétés, sélection et réparation. Les principales différences par rapport au flux de travail de l'année dernière sont l'introduction des étapes de traduction et de mise en correspondance structurelle.</p>
2015	<p>La version de cette année s'est principalement concentrée sur la piste Interactive Matching, en raison de son expansion pour inclure les ensembles de données Anatomy et Large Biomedical Ontologies. La gestion des commentaires des utilisateurs sur les tâches à grande échelle est un défi critique dans la correspondance d'ontologies, et était un aspect dans lequel AML avait encore une marge d'amélioration.</p> <p>Toujours les neuf étapes de l'année dernière sont conservées.</p>
2016	<p>Elle est basée principalement sur des techniques d'appariement lexical, en mettant l'accent sur l'utilisation de ressources comme connaissances de base.</p> <p>La version de cette année s'est principalement concentrée sur la mise en correspondance d'instances, un aspect de la mise en correspondance d'ontologies qui manquait à son portefeuille.</p>
2017	<p>Agreement Maker Light (AML) est un système de correspondance d'ontologies inspiré de Agreement Maker mais plus axé sur l'efficacité, afin de résoudre les problèmes d'appariement à grande échelle. Alors qu'il se concentrait à l'origine principalement sur le domaine biomédical, il a depuis été étendu pour traiter un large éventail de problèmes de correspondance d'ontologies et d'instances.</p>

	<p>L'AML s'appuie fortement sur des techniques d'appariement lexical, en mettant l'accent sur l'utilisation des connaissances de base, mais comprend également des composants structurels.</p> <p>La version de cette année a subi quelques modifications majeures dont on peut citer :</p> <p>Les ontologies d'entrée sont maintenant traduites en Anglais. Ceci augmente la probabilité de trouver des correspondances directes entre les entités des ontologies en utilisant Word Net.</p> <p>Des modifications ont été apportées sur l'analyseur d'ontologies pour le régler le problème de l'omission typique dans les ensembles de données correspondant aux instances, telles que les propriétés non déclarées. Ce problème faisait que AML donnait un résultat d'analyse erronée car par défaut, l'API OWL interprétait les propriétés non déclarées comme des propriétés d'annotation.</p>
2018	Des modifications majeures par rapport aux anciennes versions n'étaient pas présentées.
2019	Cette année, la version d'AML s'est principalement concentrée sur l'élargissement de l'arsenal d'algorithmes d'appariement complexes d'AML afin d'améliorer ses performances dans la nouvelle piste d'appariement complexe. La version de correspondance complexe d'AML, baptisée AMLC, reste distincte de la soumission AML principale, car nous n'avons pas encore été en mesure d'intégrer le code complexe dans la base de code principale.
2020	En 2020 AML est apparue sous le nom AMLC qui est une nouvelle version d'AML développée pour s'attaquer à la mise en correspondance d'ontologies complexes. Pour le moment, il reste séparé de la base de code AML principale. La particularité de la version de cette année est l'implémentation d'algorithmes de filtrage d'ontologies de pattern mining dans AMLC, basés sur des règles d'association et inspirés des travaux de Zhou et al...

Tableau 4 : Evolution de AML

3 Principes et limites de approches hors OA EI

Les approches utilisent chacune un principe pour la recherche de correspondances entre deux ontologies quel que soit le format d'entrée. Dans cette partie nous présentons un résumé des approches les plus saillantes les plus intéressantes ainsi que leurs limites respectives.

Anchor-Prompt

Définition

Il est une extension de PROMPT [31] connu aussi sous le nom SMART.

Principe

Anchor-PROMPT construit un graphe étiqueté orienté représentant l'ontologie à partir de la hiérarchie des concepts (appelés classes dans l'algorithme) et de la hiérarchie des relations (appelées slots dans l'algorithme), où les nœuds dans le graphe sont des concepts et les arcs dénotent des relations entre les concepts (les étiquettes des arcs sont les noms des relations).

Une liste initiale des paires d'ancres (des paires de concepts similaires) définies par les utilisateurs ou automatiquement identifiées par la mise en correspondance lexicologique sert d'entrée à l'algorithme. Anchor-PROMPT analyse alors les chemins dans les sous-graphes limités par les ancres et détermine quels concepts apparaissent fréquemment en positions similaires sur les chemins similaires. En s'appuyant sur ces fréquences, l'algorithme décide si ces concepts sont sémantiquement similaires ou non.

Limites

L'algorithme ne cherche que des correspondances entre les concepts pas de correspondances entre les relations. Les résultats sont limités si les ontologies ont des structures différentes (par exemple l'une est profonde avec beaucoup de concepts au milieu, et l'autre est peu profonde). L'algorithme rencontre des problèmes si une hiérarchie a seulement quelques niveaux et si la plupart des relations sont associées aux concepts situés au-dessus de la hiérarchie. Anchor-Prompt requiert un minimum de couples de concepts similaires déterminés par l'utilisateur tout au début.

GLUEDéfinition

Elle est une extension de LSD¹.

Principe

Elle combine les techniques d'apprentissage pour des ontologies décrites dans n'importe quel formalisme Techniques d'apprentissage et d'analyse statistique. Glue est un système d'alignement qui comprend trois modules pour trouver le meilleur appariement des ontologies : l'estimateur de distribution (qui applique les techniques d'apprentissage pour calculer la probabilité avec l'aide d'un apprenant), l'estimateur de similarité (qui produit une matrice de similarité en calculant la similarité pour chaque couple) et le relaxeur de libellés (qui exploite les contraintes et les connaissances heuristiques pour trouver le meilleur appariement).

Limites

GLUE se fonde principalement sur les instances des ontologies, qui ne sont pas toujours abondamment disponibles pour plusieurs ontologies. Cette approche ne se base que sur l'apprentissage machine et l'analyse d'instances mais n'exploite pas les autres aspects sur la structure de l'ontologie. Le résultat de l'alignement dépend de la qualité des apprenants. Enfin GLUE ne tire pas aussi profit des éléments structurels de l'ontologie comme les relations entre les entités ce qui explique son manque d'adaptation dans le cadre des ontologies OWL.

COMADéfinition

Elle est une bibliothèque contenant six apparieurs.

Principe

¹ Learning Source Descriptions

La plupart des appareilleurs implémentent les techniques d'analyse syntaxique. COMA permet de mettre en correspondance des schémas (des bases des données, de XML) automatiquement ou bien manuellement. Le système fournit une bibliothèque d'algorithmes de mise en correspondance de base (appelés matchers) et quelques mécanismes pour combiner des résultats de ces algorithmes de base afin d'obtenir une valeur de similarité finale de deux schémas. La bibliothèque des matchers se compose de 6 matchers simples qui emploient des techniques linguistiques.

Limites

L'agrégation des similarités calculées avec les différents matchers nécessite un bon choix des méthodes d'agrégation. L'intervention de l'utilisateur aussi après chaque itération est aussi un inconvénient.

S-Match

Définition

S-Match est une approche qui permet de faire de la recherche de correspondance sémantique entre deux schémas.

Principe

Il prend en entrée deux graphes des concepts (schémas), et produit en sortie des relations entre les concepts comme l'équivalence.

S-Match est un système de correspondance basé sur des schémas. Il prend en entrée deux structures de type graphique : des schémas XML ou des ontologies et produit à la fin des correspondances comme l'équivalence ou la subsomption entre les nœuds des graphiques qui correspondent sémantiquement. La méthode S-Match fait une analyse du sens des concepts qui est codifié dans les éléments et les structures des schémas d'entrée.

En particulier, les étiquettes aux nœuds, écrites en langage naturel, sont traduites en formules propositionnelles qui codifient explicitement la signification voulue de l'étiquette. Cela permet une traduction du problème d'appariement en un problème d'insatisfiabilité propositionnelle, qui peut ensuite être résolu efficacement à l'aide de décideurs de satisfiabilité propositionnelle. S-Match est un système hybride qui contient une dizaine de matchers.

Limites

La méthode S-Match ne s'intéresse principalement qu'à la sémantique des données d'entrée. Elle ne s'intéresse pas aux informations terminologiques de l'ontologie.

QOM/NOM

Définition

L'approche NOM et QOM sont deux systèmes qui permettent de faire de l'alignement d'ontologies. L'approche QOM² une amélioration de l'approche NOM³. Cette dernière a montré qu'elle est effective, mais inefficace.

Principe

QOM comme son prédécesseur exploite les 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

² Quick Ontology Matching

³ Naive Ontology Matching

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.

Limites

Les ontologies en entrée sont en format RDF qui est peu expressif. L'intervention de l'utilisateur constitue une limite. Aussi cette approche ne se limite qu'aux méthodes de calcul de la similarité terminologique et structurelle. L'efficacité de la méthode est sacrifiée au détriment de l'efficacité de l'alignement.

OLA

Définition

La méthode OLA fait partie des premières méthodes utilisant une agrégation de similarités en calculant une similarité locale et une similarité globale.

Principe

Son modèle de calcul assigne pour chaque catégorie de nœuds une fonction d'agrégation. Cette fonction exploite toutes les informations descriptives du couple d'entités à aligner en les englobant dans deux niveaux de calcul, un niveau local, et un autre global. Le calcul de la similarité globale se fait par catégorie (classe, objet, propriété, relation, etc.). Un poids est associé à chaque nœud du OL-Graph. La similarité entre les couples d'entités d'une même catégorie de chacune des ontologies représentées sous forme d'un OL-Graph est calculée. La similarité de leurs descripteurs locaux et des entités voisines est aussi exploitée. A la fin du calcul de la similarité locale des différents couples d'entités on calcule la similarité globale en faisant une agrégation des similarités locales calculées.

Limites

La méthode OLA a été testée sur des ontologies de petite taille.

Le résultat de l'alignement est impacté par : Le choix d'un poids pour les différentes catégories d'entités au début de l'alignement et le choix de WordNet pour la similarité terminologique. En plus, le choix d'un seuil par défaut pour le calcul de la similarité a un impact sur les résultats car certaines informations ne seront pas prises en compte.

ASCO 1

Définition

Elle fait partie des premières approches qui font de la réconciliation des ontologies au format RDF(S).

Principe

D'abord cet algorithme divise les voisines en trois groupes : les super-classes directes (les super-classes directes, qui sont les classes parent directes ayant un lien de subsomption vers la classe en question), les sous-classes directes (qui sont les descendants ayant le lien de subsomption de la classe en question vers ces classes), les classes sœurs (qui ont la même classe mère que celle de la classe en question). Il raisonne selon l'hypothèse suivante : deux classes sont similaires si leurs voisines sont similaires. Si les voisines de deux classes sont similaires, ces deux classes sont aussi considérées comme similaires.

Limites

ASCO1 n'exploite pas la similarité des instances pour déduire la similarité des classes ou des relations correspondantes à ces instances. Il a aussi une complexité quadratique ($O(n^2)$) du fait du nombre élevé de calculs à effectuer. Le calcul de la similarité finale se compose des calculs sur les noms, les étiquettes, les commentaires, les voisins et les entités dans les chemins vers les racines. Parmi ces calculs, le calcul de la similarité des commentaires, basé sur la technique TF/IDF, est le plus lourd. L'inconvénient de ASCO1 repose aussi sur le fait qu'il prend en entrée des ontologies au format RDF(S) qui est un langage peu expressif.

ASCO 2

Définition

La méthode ASCO2 est une amélioration de la méthode ASCO1 et règle le problème du format d'entrée des ontologies à aligner.

Principe

Elle prend en entrée des ontologies au format OWL Lite ou DL. Elle effectue des calculs de similarités entre les entités (classes, propriétés) des ontologies données en entrée. À partir des valeurs de similarité calculées, les correspondances entre les ontologies seront déduites si cette valeur dépasse un seuil fixé. Le calcul de la similarité avec ASCO2 se fait en faisant une exploitation maximale des descriptions en OWL.

Les définitions des entités sont considérées comme des ensembles des triplets RDF, dont les prédicats sont les primitives prédéfinies de OWL et de RDF(S). La similarité de deux entités correspond donc à la similarité des deux ensembles de triplets décrivant les entités en question. Cette dernière est calculée en agrégeant des valeurs de similarité partielle calculées à partir de la similarité des ensembles des objets des triplets ayant le même prédicat. L'agrégation est effectuée selon une méthode proposée appelée la somme pondérée avec les poids variables. Cela permet d'avoir une agrégation plus flexible, efficace et précise. Le calcul de la similarité de deux entités dépend donc des similarités des objets des triplets qui sont (dans la plupart des cas) aussi des entités. Le calcul est donc effectué d'une manière récursive. ASCO 2 effectue alors le calcul du point fixe pour mettre à jour les valeurs de similarité après chaque itération. Une matrice de similarité est utilisée pour stocker les valeurs de similarité entre les entités. Ces valeurs sont mises à jour jusqu'à ce qu'il n'y ait plus des changements des valeurs dans la matrice ou jusqu'à ce qu'un certain nombre d'itération soit atteint.

Limites

ASCO 2 définit tout au début un seuil ce qui fait que tant que ce seuil n'est pas dépassé des correspondances ne seront pas déduites. La valeur de similarité de deux classes peut être propagée vers les paires des autres classes (via la primitive `rdfs:subClassOf`) ou vers des paires des relations (via `rdfs:domain`, `rdfs:range`), mais la valeur de similarité de deux relations ne peut être propagée que vers des paires d'autres relations (via la primitive `rdfs:subPropertyOf`) mais vers des paires des classes. Ce qui entraînera la proposition de ASCO3 .

ASCO 3

Définition

C'est une amélioration de la méthode ASCO2.

Principe

ASCO3 prend en entrée des ontologies au format OWL. L'algorithme ASCO3 s'intéresse à la structure globale de l'ontologie en faisant une analyse des rapports sémantiques pour trouver des correspondances entre des ontologies. ASCO3 représente les ontologies en OWL par des réseaux sémantiques et puis cherche le sous-graphe commun le plus large de ces réseaux en prenant en compte la sémantique des arcs (introduction de la notion de saut). Les correspondances seront déduites du sous-graphe trouvé. L'algorithme raisonne selon le fait que deux ontologies sont similaires si leurs réseaux sémantiques sont aussi similaires

Limites

Cet algorithme ne cherche qu'une seule entité correspondante dans la deuxième ontologie pour chaque entité de la première ontologie. Alors si pour une entité de la 1ère ontologie on a plusieurs entités considérées similaires de la 2ème ontologie, l'algorithme ne peut trouver qu'une seule entité. Cet algorithme ne recherche des correspondances que pour deux entités qui ont la même nature.

La complexité de cet algorithme correspond à la complexité de l'algorithme de recherche de la clique maximale dans le graphe d'association. Le temps d'exécution de l'algorithme augmente exponentiellement avec la taille du graphe d'association. Donc si les ontologies à aligner se composent de beaucoup d'entités, le temps d'exécution de l'algorithme sera très important.

EDOLA

Définition

Elle est une approche qui fait de l'agrégation de similarité. C'est une amélioration de la méthode OLA.

Principe

EDOLA est une approche qui prend en entrée des ontologies OWL-Lite. L'extension du diamètre des nœuds des couples à aligner représente une innovation au niveau de la méthode EDOLA. En effet, ces diamètres permettent l'exploration étendue du voisinage des deux nœuds dans l'objectif d'explorer toutes les informations qu'elles incorporent. Les diamètres sont pris en compte pour le calcul de la similarité globale, ainsi que la propagation de la similarité à travers le voisinage. Le diamètre d'un nœud dans le O-Graph est par défaut la profondeur de ce nœud c'est-à-dire le nombre de nœuds non encore visités dans le graphe à partir du nœud considéré. La méthode EDOLA fait partie des méthodes d'alignement qui font de l'agrégation de similarité c'est-à-dire elle calcule les similarités locales et agrège ces similarités pour trouver la similarité globale afin de déduire des correspondances. Elle permet de prendre en charge aussi le problème de la circularité dans la définition des entités.

Limites

La méthode EDOLA utilise la technique de propagation de similarité à travers le voisinage des entités à aligner en fonction de leurs diamètres respectifs qui a comme inconvénient l'agrandissement de l'espace dans lequel cet agrandissement est fait.

AROMA

Définition

Elle est une méthode qui augmente les options sur les formats d'entrée.

Principe

AROMA est une approche qui a la particularité de prendre plusieurs formats de données en entrée : soit des hiérarchies textuelles soit des ontologies au format RDF/OWL. L'autre particularité de AROMA est qu'elle permet de détecter aussi les relations d'implications entre les entités issues de deux hiérarchies en utilisant les données textuelles contenues dans l'extension (instances) et dans les annotations (nom, commentaire, etc.) dans son processus d'alignement.

La méthode AROMA repose principalement sur la description extensionnelle des hiérarchies. Le processus d'alignement se déroule en trois phases : la phase de prétraitement des hiérarchies qui permet de les redéfinir sur un ensemble de termes communs. La phase de découverte des règles d'association entre les entités issues de deux hiérarchies. Enfin la phase de post-traitement qui permet d'obtenir un alignement consistant et minimal (sans redondance).

Limites

Le modèle proposé par AROMA n'est efficace que lorsque les alignements se limitent à l'équivalence. Le modèle ne prend pas en compte le fait qu'une double implication est identique à une équivalence. Exemple : $x \Rightarrow y$ et $x \Leftarrow y$ est sémantiquement identique à l'équivalence $x \Leftrightarrow y$. Ce manque de déduction qu'une double implication est une équivalence représente une des limites phares de la méthode AROMA.

SODA

Définition

La méthode SODA est une extension de la méthode EDOLA.

Principe

Elle prend en entrée des ontologies au format OWL-DL. Elle est une méthode structurelle pour l'alignement d'ontologies. La méthode SODA opère dans son processus d'alignement sur un voisinage moins étendu mais elle prend en charge la richesse du langage OWL-DL et exploite plusieurs mesures de similarité pour le calcul de la similarité linguistique.

La méthode SODA est une approche structurelle pour l'alignement d'ontologies OWL-DL. Elle opère en trois étapes successives et exploite la structure des ontologies à travers leurs graphes O-Graph. La première étape permet de calculer la similarité linguistique à travers des mesures de similarité plus adaptées aux descripteurs des constituants des ontologies à apparier. La seconde étape détermine la similarité structurelle en exploitant la structure des deux graphes O-Graph. La troisième étape déduit la similarité sémantique, en prenant en considération les deux types de similarité déjà calculées. La méthode SODA fait aussi partie des méthodes qui font de l'agrégation de similarités.

Limites

La méthode SODA n'est efficace que pour des ontologies de petites et moyennes tailles elle n'est pas adaptée aux ontologies de grande taille. Une autre limite de SODA est qu'elle n'exploite que la structure de l'ontologie.

ASMOV

Définition

Elle est un outil semi-automatique qui permet à un expert du domaine de faire de la réconciliation ontologique.

Principe

L'algorithme ASMOV calcule de manière itérative la similitude entre les concepts d'une paire d'ontologies en analysant quatre caractéristiques : description textuelle (id, label et commentaire), structure externe (parents et enfants), structure interne (restrictions de propriété pour les classes ; types, domaines, et les plages pour les propriétés) et la similitude individuelle. Les mesures obtenues en comparant ces quatre caractéristiques sont combinées en une seule valeur de confiance en utilisant une somme pondérée. Dans le cas de l'ASMOV, les poids initiaux ont été choisis arbitrairement et ont été optimisés en fonction des résultats du test de référence. Lors d'une phase de pré-traitement automatisé, ASMOV contient un mécanisme qui ajuste automatiquement les poids en fonction des informations contenues dans les ontologies. Par exemple, lors de l'analyse des informations textuelles dans la phase de pré-traitement, si ASMOV ne peut pas trouver de mots significatifs, il diminue le poids de similarité textuelle en fonction de règles prédéterminées.

Dans la phase de pré-traitement, les ontologies sont chargées en mémoire à l'aide de Jena⁴.

Chaque classe et propriété est encapsulée et étiquetée avec la signification de son identifiant et de ses étiquettes. La signification de ces textes est récupérée à l'aide du Metathesaurus UMLS⁵ pour le test d'anatomie et de WordNet[32] pour les autres tests dont les tests de référence. Grâce à un paramètre de configuration, un utilisateur peut forcer le système ASMOV à utiliser l'un, aucun ou les deux systèmes lexicaux. Au cours de la phase de pré-traitement, une analyse rapide des ontologies cartographiées est effectuée. Cette analyse consiste à vérifier la présence de propriétés et de mots significatifs dans la description textuelle des classes et des propriétés ; les poids sont ajustés en fonction du résultat de cette analyse.

Étant donné que le calcul de similarité globale d'ASMOV (ou valeur de confiance) est basé sur une somme pondérée et que les poids sont automatiquement ajustés en fonction de la structure des ontologies alignées, tous les tests ont été exécutés dans les mêmes conditions, la configuration par défaut du système

Limites

Il existe une configuration par défaut du système ASMOV qui justifie les résultats obtenus. Des règles statiques sont aussi présentes dans ASMOV ce qui a impacté sur les résultats obtenus.

FALCON-OA

Définition

Falcon-AO est un outil automatique qui fait de l'alignement d'ontologies.

Principe

Elle prend en entrée des ontologies OWL-DL.

Il existe deux matchers au sein de l'outil FALCON-AO. Le premier se base sur la correspondance linguistique pour les ontologies et l'autre est basé sur la correspondance des graphes des ontologies. Ce second matcher prend en entrée les alignements générés par le premier.

Limites

⁴ <https://jena.apache.org/documentation/ontology/>

⁵ Unified Medical Language System (UMLS) <http://umlsks.nlm.nih.gov/>

Les tests effectués pour FALCON-OA ont été effectués sur des ontologies modifiées dans le cadre du test. Une des limites aussi de Falcon-AO est le fait que les résultats du deuxième matcher sont dépendants des données entrées qui proviennent du premier matcher.

OACAS

Définition

Elle est une approche qui fait de l'agrégation de similarités. C'est une extension de la méthode EDOLA et OLA.

Principe

Elle prend en entrée des ontologies OWL-DL.

Les deux ontologies seront représentées sous forme de O-Graph. OACAS se base sur le O-Graph pour et produit l'alignement en trois étapes. D'abord au niveau de la première étape elle calcule la similarité composée en considérant tous les descripteurs des entités des ontologies. Ensuite elle détermine la similarité du voisinage en différents niveaux. Enfin elle fait une agrégation de la similarité composée et une de la similarité du voisinage.

Limites

La méthode OACAS nécessite une pondération des descripteurs pendant son exécution (nom, commentaire, étiquette). Cette pondération a aussi un impact sur les résultats obtenus. En l'absence de noms, de commentaires et d'étiquettes, la méthode OACAS fournit des résultats moins satisfaisants.

ROMIE

Définition

Elle est une extension du système OMIE⁶ où l'on exploite les instances(ressources) pour améliorer le processus de mapping.

Principe

ROMIE est une extension du système OMIE C'est une approche qui regroupe les méthodes d'alignement syntaxiques, linguistiques, structurelles et sémantiques pour offrir un alignement semi-automatique.

Cette approche se base principalement sur les informations réelles décrites par les ontologies en entrée. Le système ROMIE est capable d'identifier des liens de similarité entre les concepts des ontologies en analysant leurs instances associées.

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

⁶ Ontology Mapping within and Interactive and Extensible environment

alignements les moins pertinents. Ces méthodes sont basées sur des relations structurelles et sémantiques entre les concepts. Enfin, l'ensemble des hypothèses d'alignements fournies par l'étape précédente est ordonné selon les valeurs de similarité.

Limites

Pour le bon fonctionnement de ROMIE il faudra une étape de filtrage et de validation pour les candidats au mapping afin d'assurer le bon fonctionnement de ce processus.

4 Conclusion et perspectives

Dans ce papier nous avons présenté les différentes approches qui permettent de faire de l'alignement des ontologies. Ces approches présentées utilisent chacune une méthode spécifique pour trouver des correspondances. Chaque méthode dispose d'un format d'entrée, ce qui entraîne une multitude de solutions pour un même problème. Les différentes méthodes ne sont comparables que si elles ont des éléments de comparaison en commun comme le format d'entrée, le format de sortie, etc. Depuis l'avènement l'OAEI offre un cadre de comparaison des différentes méthodes. Ces différentes méthodes sont comparées avec les mêmes ontologies en entrée. Ce qui permet d'établir un tableau comparatif standard de la plupart des méthodes et systèmes d'alignement d'ontologies.

Ces différentes méthodes se distinguent de par leur type (base ou améliorée), leur caractère automatique ou non. Plusieurs méthodes de calcul de la similarité apparaissent aussi dans la littérature. Nous avons des techniques terminologiques comme la comparaison des affixes, des tokens et des trigrammes. D'autres techniques structurelles sont présentes comme la distance d'édition et la comparaison des statistiques structurelles. Nous avons aussi des méthodes extensionnelles comme la comparaison des ensembles d'instances. Enfin nous avons des méthodes sémantiques utilisant des ressources externes et la logique propositionnelle, etc. Ces techniques peuvent être utilisées directement ou être combinées. La complexité du calcul est un aspect important dans la comparaison des approches même si elle est occultée dans certaines publications. La complexité de ces approches varie généralement entre $O(n \log 2n)$ et $O(n^2)$. Un compromis est souvent nécessaire entre l'expressivité des langages de représentation et la complexité des approches.

Dans le cadre de nos travaux futurs nous comptons proposer ou améliorer les approches existantes en :

- Essayant de combiner plusieurs approches en une seule tout en gardant leur efficacité et en réduisant leur complexité ;
- Faisant varier les méthodes de calcul de la similarité pour augmenter leur efficacité ;
- Exploitant autant que possible la structure de l'ontologie pour détecter encore plus de correspondances et améliorer les résultats de ces approches.

Références

- [1] Tim Berners-Lee semantic web SA Journal of management December 2004.
- [2] Hanène GHORBEL --- Afef BAHRI --- Rafik BOUAZIZ : Les langages de description des ontologies : RDF & OWL Conference : GEI'08Tunisia 2008.
- [3] Euzenat, J. and al., "Heterogeneity in the semantic web, deliverable 2.2", Knowledge Web, December 2007.
- [4] Touzani M., « Alignement des ontologies OWL-Lite », Master's thesis, University of Montreal, 2005 p.17.
- [5] Abdeltif Elbyed. ROMIE, une approche d'alignement d'ontologies à base d'instances. Autre [cs.OH].Institut National des Télécommunications, 2009. Français. NNT : 2009TELE0014. tel-00541874.p51-52.
- [6] <http://oaei.ontologymatching.org/>
- [7] Noy N. F., Musen M. A., « Anchor-PROMPT : Using Non-Local Context for Semantic Matching », Proceedings of the Workshop on Ontologies and Information Sharing at the Seventeenth International Joint Conference on Artificial Intelligence, Seattle, WA, August, 2001.
- [8] Doan, A., Madhavan, J., Dhamankar, R. et al. Learning to match ontologies on the Semantic Web. VLDB 12, 303–319 (2003). <https://doi.org/10.1007/s00778-003-0104-2>
- [9] Do, H.H. et Rahm, E. Coma – a system for flexible combination of schema matching approaches. Dans Proc. VLDB, pages 610–621, 2002
- [10] Kalfoglou Y., Schorlemmer M., « IF-Map : an ontology mapping method based on information flow theory », Journal of data semantics, vol. 1, n° 1, p. 98-127, 2003.
- [11] Fausto Giunchiglia, Pavel Shvaiko and Mikalai Yatskevich S-MATCH : AN ALGORITHM AND AN IMPLEMENTATION OF SEMANTIC MATCHING February 2004
- [12] Ehrig M., Staab S., « QOM : Quick Ontology Mapping », Proceedings of The 3rd ISWC, GI Jahrestagung (1), Hiroshima, Japon, p. 356-361, November, 2004a
- [13] Ehrig, M., Sure, Y.: Ontology mapping - an integrated approach. In Bussler, C., Davis, J., Fensel, D., Studer, R., eds.: Proceedings of the 1st ESWS. Volume 3053 of Lecture Notes in Computer Science., Heraklion, Greece, Springer Verlag (2004) 76–91
- [14] Touzani M., « Alignement des ontologies OWL-Lite », Master's thesis, University of Montreal, 2005.
- [15] Bach T. L., Construction d'un web sémantique multi-points de vue, PhD thesis, École des Mines de Nice à Sophia Antipolis, 2006 p.50-73.
- [16] Bach T. L., Construction d'un web sémantique multi-points de vue, PhD thesis, École des Mines de Nice à Sophia Antipolis, 2006 p.77-90.
- [17] Sami Zghal, Karim Kamoun, Sadok Ben, Yahia Engelbert , Mephu Nguifo et Yahya Slimani : EDOLA :Une nouvelle méthode d'alignement d'ontologies OWL Lite.
- [18] David, J. (2007). AROMA : une méthode pour la découverte d'alignements orientés entre ontologies à partir de règles d'association. Ph. D. thesis, Université de Nantes.
- [19] Zghal, S., Ben Yahia, S., Mephu Nguifo, E., Slimani, Y. : SODA : Une approche structurelle pour l'alignement d'ontologies OWL-DL. In: Proceedings of the first French Conference on Ontology (JFO 2007), Sousse, Tunisia (October 2007).
- [20] Yves R. Jean-Mary, Mansur R. Kabuka ASMOV Results for OAEI 2007

- [21] Ningsheng Jian, Wei Hu, Gong Cheng, Yuzhong Qu Falcon-AO: Aligning Ontologies with Falcon Conference: Integrating Ontologies '05, Proceedings of the K-CAP 2005 Workshop on Integrating Ontologies, Banff, Canada, October 2, 2005[22] Schema and ontology matching with COMA++ SIGMOD '05: Proceedings of the 2005 ACM SIGMOD international conference on Management of data June 2005 Pages 906–908 <https://doi.org/10.1145/1066157.1066283>.
- [23] Bach T. L., Construction d'un web sémantique multi-points de vue, PhD thesis, École des Mines de Nice à Sophia Antipolis, 2006 p.91-113.
- [24] Fayçal Hamdi, Harfa Zargayouna, Brigitte Safar, and Chantal Reynaud1 TaxoMap in the OAEI 2008 alignment contest
- [25] Jean François Djoufak Kengue , Jérôme Euzenat, and Petko Valtchev OLA in the OAEI 2007 Evaluation Contest
- [26] Abdeltif Elbyed. ROMIE, une approche d'alignement d'ontologies à base d'instances. Autre [cs.OH]. Institut National des Télécommunications, 2009. Français. NNT : 2009TELE0014. tel-00541874
- [27] Sami Zghal Contributions à l'alignement d'ontologies OWL par agrégation de similarités, Phd thesis Université d'Artois CRIL-CNRS ,Lens ,2010.
- [28] Noy, N. F., & Musen, M. A. (1999). Smart : Automated support for ontology merging and alignment. Technical Report SMI-1999-0813, Stanford Medical Informatics.
- [29] Doan, A.H., Domingos, P. et Halevy, A. Learning source descriptions for data integration. In: Proc WebDB Workshop, pp. 81–92, 2000.
- [30] Isabel F. Cruz, Flavio Palandri Antonelli, Cosmin Stroe, Ulas C. Keles, and Angela Maduko: Using AgreementMaker to Align Ontologies for OAEI 2009: Overview, Results, and Outlook.
- [31] PROMPT : Algorithm and Tool for Automated Ontology Merging and Alignment Natasha Noy Mark Alan Musen Conference : Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence, July 30 - August 3, 2000, Austin, Texas, USA.

Annexes

Approches	Données d'entrées	Méthode de calcul de la similarité		Données de sorties		Complexité
		Techniques Utilisées	Mesure de la Similarité	Format	Relation	
COMA	SR, XML	Terminologique /structurelle	Comparaison des affixes, tokens, Trigrammes, synonymes, types, edit Distance, soundex comparaison des statistiques structurelles	RDF/XML	↔	-
NOM	Ontologies	Terminologique /structurelle	Distance d'édition Somme pondérée	RDF/XML	↔	$O(n^2 \cdot \log_2(n))$
QOM	RDFS/OWL	Terminologique /structurelle /méthodes extensionnelles/méthodes combinées	Distance d'édition Somme pondérée, Comparaison des ensembles d'instances concepts / relations	RDF/XML	↔	$O(n \cdot \log(n))$

Anchor-Prompt	RDF(s)/OWL	Terminologique /structurelle	String Comparaison des classes/relations comparaison des positions	RDF/XML	↔	$O(n^2 \cdot \log_2(n))$
GLUE	Taxonomies	Terminologique Structurelle /extensionnelle	Tokenisation string Comparaison des tokens	RDF/XML	↔	$O(n^2)$
ASCO1	RDF(s)	Simples (Terminologiques structurelles) Combinées	Jaro-Winkler (id,labels,ynset),TF,IDF WordNet Synonymie TALN ,chemin, voisine	RDF/XML	↔	-
ASCO2	OWL-DL	Simples (Terminologique structurelle) Combinées (Somme pondérée avec pondération variée)	Jaro-Winkler (id,labels,ynset), TF, IDF, TALN (normalisation des espaces) Structure topologique voisine, ensemble	RDF/XML	↔	-
TAXOMAP	OWL/RDF/	Terminologique /structurelle	Techniques basées sur la mesure de	RDF/XML	↔	-

	Taxonomie		similarité de Lin (SimLinLike)			
FALCON-OA	RDF/OWL	Terminologique /structurelle	Edition distance,TF, IDF	RDF/XML	↔	-
ROMIE	Ensemble de ressources locales et d'ontologies	Terminologique /structurelle/ Méthodes Combinées	Somme pondérée Word Net, distance (Hamming, subString, N-gramme, Levenshtein Comparaison des noeuds voisins	RDF/XML		-
IF-MAP	Ontologies	Structurelle	Isomorphisme	RDF/XML	↔	-
S-MATCH	HT, RDF(S), OWL	Terminologique /structurelle /sémantique	Tokenisation Lemmatisation wordnet logique propositionnelle	RDF/XML	↔ ⇒	-
OLA	RDFS/OWL	Terminologique /structurelle /Sémantique	Hamming entre les synsets,système d'équations interdépendantes	RDF/XML	↔	-

EDOLA	OWL-Lite	Terminologique /structurelle	distance de Levenshtein distance d'édition API WordNet	RDF/X ML	↔	-
SODA	OWL-DL	Simple (Terminologique/st ructurelle /méthodes /sémantique) Et Combinées	Jaro- Winkler pour les noms et labels, Monge- Elkan pour les commentair es Comparaiso n des noeuds voisins somme pondérée	RDF/X ML	↔	-
ASCO3	OWL DL/Lite	Simple (Terminologique structurelle) et Combinées (Somme pondérée)	Jaro- Winkler TF/IDF TALN (Normalisati on des espaces) Graphe isomorphis me structure topologique	RDF/X ML	↔	-
AROMA	OWL- Lite/DL	Terminologique /structurelle	Jaro- Winkler TAL, TF/IDF	RDF/X ML	↔ ⇒	-

ASMOV	OWL	Terminologique /structurelle Interne conceptuelle et extensionnelle	UMLS Meta thesarus Word net	RDF/X ML	-	-
COMA ++	OWL-Lite	Terminologique /structurelle	Word Net + matchers	RDF/X ML	↔	-
OACAS	OWL-DL	Terminologique/ Structurelle	Jaro Winkler distance de Levenshtein	RDF/X ML	↔	-

Tableau 3 : Comparaison technique