

Gestion du conflit dans l'appariement des ontologies

Amira Essaid, Boutheina Ben Yaghlane, Arnaud Martin

▶ **To cite this version:**

Amira Essaid, Boutheina Ben Yaghlane, Arnaud Martin. Gestion du conflit dans l'appariement des ontologies. Extraction et Gestion des Connaissances, Jan 2011, Brest, France. pp.50-60. hal-00657620

HAL Id: hal-00657620
https://hal.science/hal-00657620

Submitted on 7 Jan 2012

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.

Gestion du conflit dans l'appariement des ontologies

Amira Essaid*, Boutheina Ben Yaghlane**
Arnaud Martin***

*LARODEC, Université de Tunis/ISG Tunis
essaid_amira@yahoo.fr
**LARODEC, Université 7 Novembre à Carthage/IHEC Carthage
boutheina.yaghlane@ihec.rnu.tn
***UMR 6074 IRISA, Université de Rennes1 / IUT de Lannion
Arnaud.Martin@univ-rennes1.fr

Résumé. L'alignement des ontologies représente un grand intérêt dans le web sémantique. En effet, il permet de pallier le problème d'hétérogénéité sémantique et d'assurer une interopérabilité entre les systèmes utilisant des ontologies distribuées et hétérogènes. Cet alignement est réalisé grâce aux différentes approches fondées sur des mesures de similarité. L'appariement des ontologies fait naturellement apparaître du conflit entre les résultats fournis par ces mesures. Dans cet article, nous proposons d'appliquer la théorie des fonctions de croyance pour gérer ce conflit et ceci par la combinaison des résultats des différentes mesures. Plusieurs règles de combinaison peuvent être appliquées à cet effet.

1 Introduction

Le *web sémantique* introduit par Berners-Lee et al. (2001), tend à rendre le contenu des documents web accessible par les applications afin d'assurer un échange d'information et une interopérabilité entre les différents systèmes. Les ontologies proposées comme élément central du web sémantique, sont définies par Gruber (1993) comme "*une spécification explicite d'une conceptualisation*". En effet, selon Kengue et al. (2008), les ontologies permettent de décrire un domaine en définissant un ensemble de concepts et les relations qu'ils entretiennent avec d'autres concepts par spécialisation ou à travers des propriétés.

Les ontologies sont utilisées pour la représentation des connaissances. Plusieurs ontologies couvrant totalement ou partiellement un même domaine d'application peuvent être développées indépendamment les unes des autres. Les entités de ces ontologies peuvent être définies selon différents niveaux de granularité ou simplement décrites grâce à des différents langages de représentation. Afin de permettre l'interopérabilité et le partage des données, il est primordial que l'hétérogénéité entre les connaissances décrites dans les ontologies soit résolue et ceci par la création des liens sémantiques entre les entités, c'est le but de l'*alignement*.

Etant données deux ontologies, l'alignement (appariement ou mise en correspondance) consiste en la production d'un ensemble de correspondances entre les entités. Ces entités

peuvent être des concepts, des propriétés ou encore des instances. Cet ensemble de correspondances ou encore alignement peut être par la suite utilisé pour fusionner les ontologies, créer une troisième ontologie à partir des entités des deux ontologies en entrée et la liste des correspondances ou encore effectuer des tâches de raisonnement entre les ontologies appariées.

La découverte des alignements d'une manière manuelle est une tâche coûteuse en temps, inefficace et peut produire des erreurs. C'est pour cette raison que plusieurs méthodes d'appariement ont été proposées. Euzenat et Shvaiko (2007) donnent un état de l'art exhaustif des différentes techniques d'alignement. Ces méthodes sont fondées sur des mesures de similarité. Dans le but d'améliorer le résultat de l'alignement, plusieurs mesures peuvent être utilisées simultanément ce qui fait apparaitre inévitablement un conflit entre les différents résultats produits par chacune de ces mesures.

La théorie des fonctions de croyance introduite par Dempster (1967) et Shafer (1976) permet de combiner des informations provenant de plusieurs sources hétérogènes. Ces informations fournies peuvent être en conflit. Dans le cadre de cette théorie, la gestion du conflit peut se faire de deux manières différentes : soit nous nous intéressons à gérer le conflit avant la combinaison des informations des sources ce qui revient à le réduire et donc d'affaiblir ces informations selon le degré de fiabilité de la source, soit nous pouvons tenir compte du conflit lors de la combinaison et ceci par application des règles de combinaison qui permettent de supprimer le conflit en le redistribuant de manières différentes sur les informations disponibles.

Dans cet article, nous proposons de gérer le conflit dans l'alignement des ontologies par l'utilisation de la théorie des fonctions de croyance. Tout d'abord, nous considérons chacune des mesures de similarité comme une source d'évidence ayant un degré de confiance par rapport aux résultats d'alignement fournis. Ces résultats font naturellement apparaitre du conflit qu'on gérera lors de l'étape de combinaison de ces résultats et ceci par application des règles de combinaison adéquates.

La suite de cet article est organisée comme suit : dans la deuxième section, nous rappelons la définition de l'alignement des ontologies ainsi que les mesures de similarité que nous avons utilisées. La troisième section est dédiée à la présentation des concepts de base de la théorie des fonctions de croyance. Enfin, dans la quatrième section nous décrivons notre approche de gestion du conflit dans l'alignement des ontologies pour ainsi présenter dans la section suivante quelques directives des travaux de recherche envisageables dans le futur.

2 L'alignement des ontologies et les mesures de similarité

2.1 L'alignement des ontologies

Afin de faciliter l'échange entre les applications, plusieurs langages de représentation des connaissances ont été développées parmi lesquels le langage *OWL*[1]. Ce dernier fournit un grand nombre de constructeurs pour représenter les entités (*e.g.* "owl :class", "rdfs :subClassOf"...).

Selon Ehrig et Staab (2004), une ontologie est définie formellement par le tuple suivant :

$$O := (C,\ H_C,\ R_C,\ H_R,\ I,\ R_I,\ A)$$

1. http ://www.w3.org/TR/owl-ref/

Une ontologie est un ensemble de concepts *C* (des instances de *'owl :class'*) reliés par une hiérarchie de subsomptions H_C (relation binaire correspondant à *'rdfs :subClassOf'*). Les relations R_C (des instances de *'owl :objectProperty'*) relient les concepts. Ces relations sont organisées selon une hiérarchie de subsomption H_R (*'rdfs :subPropertyOf'*). Un individu *i* appartenant à l'ensemble des individus *I* est une instance de la classe *c* tel que $c \in C$. Cette instance *i* peut être reliée à une instance *j* par une relation *r* appartenant à R_I. Le triplet (*i, r, j*) est une instance de propriété. Une ontologie permet aussi d'inférer de nouvelles connaissances grâce à l'ensemble des axiomes noté par *A*.

Aujourd'hui, il existe un très grand nombre d'ontologies disponibles sur le web. L'utilisation d'une seule ontologie ne permet pas de soutenir les tâches demandées par le web sémantique. Les applications doivent alors utiliser plusieurs ontologies distribuées et hétérogènes. Ceci est réalisé par un appariement des ontologies qui selon Euzenat et Shvaiko (2007) consiste en une fonction *f* définie par :

$$A = f\ (O_1, O_2, A', p, r)$$

Cette fonction tend à partir de deux ontologies O_1 et O_2, d'un ensemble d'alignements *A'*, d'un paramètre *p* (seuil minimal de similarité) ainsi que des ressources externes *r* (*e.g.* thesaurus) à renvoyer un alignement *A* : en d'autres termes à chercher pour chaque entité *e* (concept, relation, ou encore instance) de l'ontologie O_1, les entités correspondantes se trouvant dans O_2 ayant des sémantiques similaires. La figure 1 montre l'appariement de deux ontologies relatives aux références bibliographiques. L'alignement doit en effet identifier la classe *Congress* comme correspondante à la classe *Conference* et que la classe *Address* doit être reliée à la classe *Directions*.

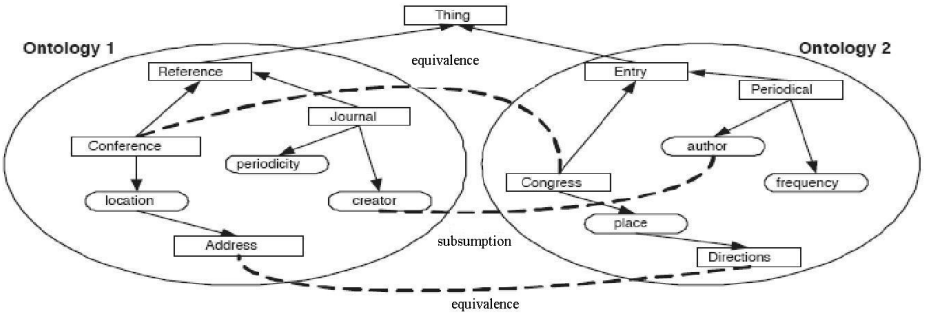


FIG. 1 – *Extrait d'appariement de deux ontologies bibliographiques.*

Dans cet article, nous nous intéressons à aligner uniquement les concepts représentés dans une ontologie en *OWL* par *"owl :class"*.

2.2 Les mesures de similarité

Rahm et Bernstein (2001) proposent la classification suivante des différentes techniques d'alignement :

1. **Les méthodes terminologiques** : Elles comparent les labels des différentes entités. Elles sont décomposées en :
   (a) *Les méthodes syntaxiques* : Elles effectuent la correspondance à travers les mesures de dissimilarité des chaînes.
   (b) *Les méthodes lexicales* : Elles effectuent la correspondance à travers les relations lexicales (synonymie, hyponymie ...).
2. **Les méthodes structurelles** : Elles se fondent sur la comparaison des structures des entités. On distingue :
   (a) *Les méthodes de comparaison de structures internes* : Elles se fondent sur la structure interne des entités (cardinalité des attributs, transitivité des propriétés, ...).
   (b) *Les méthodes de comparaison de structures externes* : La comparaison de similarité entre deux entités des deux ontologies dépend des relations de ces entités avec d'autres entités. Des types de relations existent : les relations taxonomiques, les relations méréologiques, ...
3. **Les méthodes sémantiques** : Elles sont fondées sur le modèle des entités utilisé pour valider et justifier les résultats d'alignement. Ce sont des méthodes déductives. Parmi ces techniques, nous citons la satisfiabilité propositionnelle qui tend à trouver les alignements en transformant deux ontologies ainsi que la requête d'alignement (le couple d'entités à apparier et la relation possible entre ces entités) en des formules propositionnelles et par la suite de vérifier sa validité.

Dans cet article, nous avons appliqué des méthodes terminologiques afin de chercher les correspondances entre les entités de deux ontologies. Nous avons utilisé trois mesures syntaxiques à savoir *la distance de Levenshtein*, *la distance Jaro* ainsi que *la distance de Hamming*. Quant à la mesure lexicale, nous avons eu accès au dictionnaire *WordNet 2.0* proposé par Miller (1995) pour identifier les correspondances. Ci dessous, nous rappelons brièvement les mesures syntaxiques utilisées.

– **Distance de Levenshtein** : Hall et Dowling (1980) définissent cette distance comme étant la mesure la similarité entre deux chaînes de caractères. Elle est égale au nombre d'opérations nécessaires pour transformer une chaîne en une autre. Les opérations consistent en suppression, insertion ou remplacement. Nous définissons la mesure de similarité entre deux mots w_1, w_2 comme :

$$sim_{ed}(w_1, w_2) = \frac{1}{1 + ed(w_1, w_2)} \quad (1)$$

Notons que $ed(w_1, w_2)$ est la distance de Levenshtein entre les deux mots w_1 et w_2.

– **Distance Jaro** : 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 w_1, w_2 comme la distance Jaro entre les deux chaînes de caractères des deux mots.

– **Distance de Hamming** : Elle mesure le nombre de positions au niveau desquelles les deux chaînes de caractères diffèrent.

3 La théorie des fonctions de croyance

3.1 Les concepts de base

La théorie des fonctions de croyance appelée aussi théorie de l'évidence est issue des travaux de Dempster (1967) et fut reprise par Shafer (1976). C'est un outil qui permet de modéliser aussi bien l'incertitude que l'imprécision. Il prend en compte des ambiguïtés et des conflits entre les sources. Nous présentons dans ce qui suit les concepts de base de cette théorie.

Pour un problème donné, la théorie des fonctions de croyance définit un *cadre de discernement* noté Θ comme étant l'ensemble des *N* hypothèses exhaustives et exclusives.

$$\Theta = \{H_1, H_2, \ldots H_N\}$$

Un ensemble 2^Θ, défini à partir de Θ contient les hypothèses singletons de Θ, toutes les disjonctions possibles de ces hypothèses ainsi que l'ensemble vide.

La théorie des fonctions de croyance est fondée sur la manipulation des fonctions de masse appelée encore *masse élémentaire de croyance*. La fonction de masse *m* définie sur 2^Θ et à valeurs dans $[0, 1]$, vérifie les propriétés suivantes :

$$m(\emptyset) = 0 \tag{2}$$

$$\sum_{A \subseteq \Theta} m(A) = 1 \tag{3}$$

Les éléments *A* ayant une masse de croyance non nulle sont appelés *les éléments focaux*.

D'autres mesures de croyance peuvent être calculées à partir des fonctions de masse, à savoir la crédibilité (*bel*) et la plausibilité (*pl*).

La fonction de crédibilité (*bel*) mesure la croyance minimale apportée à un sous ensemble de 2^Θ . Cette fonction est définie pour tout $A \in 2^\Theta$ et à valeurs dans $[0, 1]$ par :

$$bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B) \tag{4}$$

La fonction de plausibilité (*pl*) mesure la croyance maximale apportée à un sous ensemble de 2^Θ. Cette fonction est définie pour tout $A \in 2^\Theta$ et à valeurs dans $[0, 1]$ par :

$$pl(A) = \sum_{A \cap B \neq \emptyset} m(B) \tag{5}$$

Afin de bien illustrer les notions présentées ci dessus, nous proposons dans ce qui suit un exemple d'organisation des conférences. Afin de participer à une conférence, les auteurs soumettent leurs articles. Ces contributions seront par la suite évaluées par des relecteurs. Nous pouvons identifier trois types de travaux à savoir un poster (PS), un papier long (PL) ou encore un papier court (PC). Suite à l'évaluation, un papier long peut être accepté, par exemple, à ce qu'il soit réécrit comme un papier court. Le problème peut être modélisé par un cadre de discernement $\Theta = \{PS, PL, PC\}$. L'ensemble 2^Θ correspondant est :
$2^\Theta = \{\emptyset, PS, PL, PC, PS \cup PL, PS \cup PC, PL \cup PC, \Theta\}$.

Supposons qu'un des relecteurs a évalué un papier et a exprimé son degré de croyance par les masses suivantes :

$$m(PL) = 0.3,\ \ m(PL \cup PC) = 0.5,\ \ m(PS \cup PC \cup PL) = 0.2$$

Le relecteur a un degré de croyance de 0.3 que le papier doit être long, de 0.5 qu'il peut être accepté en tant que papier long ou réécrit en tant que papier court et un degré d'ignorance totale (le papier peut être un poster, un papier long ou encore un papier court) de 0.2. Nous présentons dans le tableau 1 les fonctions de masse fournies par le relecteur ainsi que les fonctions de crédibilité et de plausibilité.

	m	bel	pl
$\emptyset$	0	0	0
PL	0.3	0.3	1
PC	0	0	0.7
PL $\cup$ PC	0.5	0.8	1
PS	0	0	0.2
PL $\cup$ PS	0	0.3	1
PC $\cup$ PS	0	0	0.7
Θ	0.2	1	1

TAB. 1 – *Les fonctions de masse, de crédibilité et de plausibilité relatives à l'évaluation d'un article.*

3.2 Les règles de combinaison

La fusion des données imparfaites (incertaines, imprécises et incomplètes) est une solution pour obtenir une information plus pertinente et plus fiable. La théorie des fonctions de croyance est un outil intéressant de fusion de données. En effet, pour un problème donné et pour un même cadre de discernement, il est possible d'obtenir une fonction de masse synthétisant les différentes connaissances issues des différentes sources d'information *distinctes* et *indépendantes* et ceci par application d'une règle de combinaison.

Il existe plusieurs modes de combinaison développés dans le cadre de la théorie des fonctions de croyance. Nous présentons dans ce qui suit la combinaison conjonctive, la combinaison disjonctive ainsi que la combinaison mixte. Afin d'illustrer les différentes règles, nous supposons que nous avons deux sources S_1 et S_2 distinctes et indépendantes. Ces sources fournissent des masses m_1 et m_2 pour un même cadre de discernement Θ.

3.2.1 Combinaison conjonctive

Elle fut introduite par Dempster (1967) et reprise par Shafer (1976), elle se fonde sur la combinaison des fonctions de masse en considérant les intersections des éléments de 2^{Θ}.

Règle orthogonale normalisée de Dempster-Shafer Etant données deux fonctions de masse m_1 et m_2, cette règle de combinaison est présentée comme suit :

$$m_{1\oplus 2}(A) = \begin{cases} \dfrac{\sum\limits_{B\cap C=A} m_1(B)\times m_2(C)}{1-\sum\limits_{B\cap C=\emptyset} m_1(B)\times m_2(C)} & \forall A\subseteq\Theta,\ A\neq\emptyset \\ 0 & si\ A=\emptyset \end{cases} \tag{6}$$

Notons que m_1(B) et m_2(C) sont les fonctions de masse associées aux deux sources S_1 et S_2 respectivement et que $\sum\limits_{B\cap C=\emptyset} m_1(B)\times m_2(C)$ est une mesure de conflit entre les sources. Cette règle est normalisée par $1-\sum\limits_{B\cap C=\emptyset} m_1(B)\times m_2(C)$. Cette normalisation masque le conflit et n'est intéressante que sous l'hypothèse du monde fermé (toutes les hypothèses possibles du problème appartiennent au cadre de discernement Θ) et aussi lorsqu'il n'y a pas de conflit entre les sources. En effet, la fonction de masse obtenue suite à la combinaison renforce la croyance sur les décisions pour lesquelles les sources sont concordantes et l'atténue en cas de conflit.

Règle de Smets Smets propose la forme non normalisée de la règle précédemment décrite. Cette règle est utilisée sous l'hypothèse du monde ouvert où une masse non nulle est affectée à l'ensemble vide. La règle conjonctive de combinaison est définie alors comme suit :

$$m_{1\bigcirc\!\!\!\!\cap 2}(A) = \sum_{B\cap C=A} m_1(B)\times m_2(C) \tag{7}$$

Cette règle est principalement utilisée lorsque les deux sources sont fiables.

Règle de Yager Yager (1987) propose une règle où la mesure de conflit est affectée au cadre de discernement Θ. Pour deux fonctions de masse m_1 et m_2, la règle de Yager est définie comme suit :

$$\begin{cases} m_Y(A) = m_{1\bigcirc\!\!\!\!\cap 2}(A) & \forall A\in 2^\Theta,\ A\neq\Theta\ et\ A\neq\emptyset \\ m_Y(\Theta) = m_{1\bigcirc\!\!\!\!\cap 2}(\Theta) + m_{1\bigcirc\!\!\!\!\cap 2}(\emptyset) & \\ m_Y(\emptyset) = 0 & \end{cases} \tag{8}$$

3.2.2 Combinaison disjonctive

La règle de combinaison disjonctive a été proposée par Smets (1993) pour combiner des fonctions de masse dont l'une au moins est fiable, sans pour autant quantifier la fiabilité ni identifier laquelle des deux sources est fiable. Cette règle est définie pour $\forall A\in 2^\Theta$ par l'équation suivante :

$$m_{1\bigcirc\!\!\!\!\cup 2}(A) = \sum_{B\cup C=A} m_1(B)\times m_2(C) \tag{9}$$

3.2.3 Combinaison mixte

Dubois et Prade (1988) ont proposé un compromis afin de tenir compte des avantages des deux modes de combinaison à savoir la combinaison conjonctive et la combinaison disjonctive. Cette combinaison est donnée pour tout $A \in 2^{\Theta}$ par la formule suivante :

$$\begin{cases} m_{DP}(X) = m_{1\textcircled{\cap}2}(X) + \sum_{A\cap X=\emptyset,\ A\cup X=B} m_1(X)m_2(A) \quad \forall A \subseteq \Theta,\ A \neq \emptyset \\ m_{DP}(\emptyset) = 0 \end{cases} \tag{10}$$

Ce mode de combinaison suppose que le conflit provient de la non fiabilité des sources.

4 Une approche de gestion du conflit dans l'alignement des ontologies

Dans le but d'apparier les ontologies, plusieurs mesures de similarité peuvent être utilisées afin de tenir compte des aspects terminologiques des différents concepts de deux ontologies à aligner. L'utilisation de ces mesures peut faire apparaitre du conflit. Dans le cadre de la théorie des fonctions de croyance, nous proposons de gérer le conflit une fois que nous avons appliqué ces mesures. Pour ce faire, nous devons modéliser le problème d'appariement dans le cadre de cette théorie. En effet, les mesures de similarité utilisées seront considérées comme les sources d'évidence fournissant une information quant aux classes à apparier. Les résultats de ces mesures seront interprétés comme des fonctions de masse fournies par une source.

Afin d'illustrer notre modélisation, nous nous fondons sur deux ontologies *cmt* et *conference* relatives à l'organisation des conférences [2]. *cmt* est une ontologie source pour laquelle nous cherchons pour chacune de ses classes ses correspondantes au niveau de l'ontologie cible *conference*. Le tableau 2 met en évidence quelques classes de l'ontologie *cmt* (Paper, PaperAbstract, Document, Preference) ainsi que leurs correspondantes de l'ontologie *conference* suite à l'application des mesures terminologiques.

	Distance Levenshtein	Distance Jaro	Distance Hamming	Méthode Lexicale
Paper	Paper	Paper	Paper	Paper
PaperAbstract	Abstract	Paper	Paper	Abstract
Document	Conference_document	Committee	Co-chair	Paper
Preference	Conference	Review_preference	Conference	Review_preference

TAB. 2 – *Appariement des ontologies cmt et conference.*

La distance Levenshtein, la distance jaro, la distance Hamming ainsi que la méthode lexicale sont représentées comme les sources d'évidence ($S_{lev}, S_{jaro}, S_{ham}, S_{lex}$). Les quatre sources considèrent que la classe *Paper* de l'ontologie *cmt* doit être appariée à la classe *Paper* de l'ontologie *conference*. S_{lev} et S_{lex} relient la classe PaperAbstract de l'ontologie source à la classe Abstract de l'ontologie cible tantdis que les deux autres sources S_{jaro} et S_{ham} relient

2. http ://nb.vse.cz/svabo/oaei2010/

PaperAbstract à *Paper*. Quant à la classe *Document*, le tableau 2 montre bien que les quatre sources ont des points de vue différents.

Nous définissons le cadre de discernement comme étant l'ensemble de tous les couples formés à partir des classes des deux ontologies. $\Theta = \{(c_i, e_1), (c_i, e_2), \ldots (c_i, e_n)\}$ où c_i et e_j représentent respectivement une classe de l'ontologie source et son correspondant de l'ontologie cible fournie par la mesure de similarité. Les classes pour lesquelles les quatre sources donnent un même résultat seront exclues de cette modélisation comme le cas de *Paper*. Par exemple, le Θ correspondant au tableau 2 est représenté comme suit :

$\Theta = \{(PaperAbstract, Abstract), (PaperAbstract, Paper),$
$(Document, Conference_document), (Document, Committee),$
$(Document, Co - chair), (Document, Paper),$
$(Preference, Conference), (Preference, Review_preference)$

Le tableau 3 illustre les résultats de ces différentes mesures. Dans le cadre de la théorie des fonctions de croyance, ces résultats sont interprétés comme des fonctions de masse.

	Distance Levenshtein	Distance Jaro	Distance Hamming	Méthode Lexicale
Paper	1.000	1.000	1.000	1.000
PaperAbstract	0.615	0.795	0.385	0.762
Document	0.421	0.638	0.125	0.910
Preference	0.700	0.670	0.700	0.740

TAB. 3 – *Résultats d'appariement des ontologies cmt et conference.*

Par exemple pour la classe *Paper*, les quatre sources sont en accord et attribuent une masse égale à 1.000 à la classe *Paper* de l'ontologie cible comme suit :

$m^{S_{lev}}$(Paper, Paper) = 1.000, $m^{S_{jaro}}$(Paper, Paper) = 1.000,
$m^{S_{ham}}$(Paper, Paper) = 1.000, $m^{S_{lex}}$(Paper, Paper) = 1.000.

Quant à la classe *PaperAbstract* et malgré le choix d'une même classe *Paper* à apparier à *PaperAbstract*, les deux sources S_{jaro} et S_{ham} n'attribuent pas une même masse. En effet, nous avons :

$m^{S_{lev}}$(PaperAbstract, Abstract) = 0.615, $m^{S_{jaro}}$(PaperAbstract, Paper) = 0.795,
$m^{S_{ham}}$(PaperAbstract, Paper) = 0.385, $m^{S_{lex}}$(PaperAbstract, Abstract) = 0.762.

Etant donné le conflit entre les différentes mesures, on gérera ce conflit lors de la combinaison des résultats des différentes sources et ceci par application des règles de combinaison précédemment présentées.

5 Discussion

Dans cet article, nous nous sommes limités à apparier les ontologies tout en cherchant à relier les concepts qui sont équivalents en appliquant des techniques terminologiques. Nous

pouvons chercher dans le futur à apparier les individus ou encore les propriétés de deux ontologies tout en utilisant des techniques structurelles ou sémantiques. Le fait d'appliquer uniquement les techniques terminologiques a permis de détecter un conflit. L'utilisation d'autres méthodes d'alignement permettra de tenir compte des différents aspects des entités tel que la relation d'une entité avec les entités qui lui sont directement reliées. Dans ce cas, la gestion du conflit sera intéressante.

Dans cet article, nous nous sommes contentés de proposer une approche où la gestion du conflit est effectuée lors la combinaison des informations issues des différentes mesures de similarité. Cette approche peut être étendue en s'intéressant à gérer le conflit avant même de combiner en estimant la fiabilité des mesures de similarité tout en affaiblissant les informations des mesures et donc de réduire le conflit.

6 Conclusion

Dans cet article, nous avons proposé une approche de gestion du conflit dans l'appariement des ontologies. Nous nous sommes limités à chercher les correspondances entre les classes de deux ontologies en utilisant quatre mesures de similarité terminologiques. Une fois que nous avons identifié pour chaque classe de l'ontologie source son correspondant de l'ontologie cible, une gestion du conflit est effectuée par application de la théorie des fonctions de croyance. En effet, les mesures de similarité sont assimilées à des sources d'évidence et les résultats fournis par ces mesures sont traités comme des fonctions de masse. La gestion du conflit est effectuée lors de la combinaison des fonctions de masse des différentes sources par application des règles de combinaison.

Comme perspective, nous envisageons à appliquer différentes règles de combinaison et de comparer les résultats fournis par ces règles. Des tests expérimentaux seront effectués sur différents couples d'ontologies fournis par l'OAEI [3].

Références

Berners-Lee, T., J. Hendler, et O. Lassila (2001). The semantic web. *Scientific American 284*, 34–43.

Dempster, A. (1967). Upper and Lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics 38*, 325–339.

Dubois, D. et H. Prade (1988). Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence 4*, 244–264.

Ehrig, M. et S. Staab (2004). Qom - quick ontology mapping. In *Proceedings of the International Semantic Web Conference (ISWC04), Springer-Verlag LNCS 3298*, pp. 289–303.

Euzenat, J. et P. Shvaiko (2007). Ontology Matching. Springer-Verlag.

Gruber (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition 5*, 199–220.

3. Ontology Alignment Evaluation Initiative : http ://oaei.ontologymatching.org/2010/

Hall, P. A. V. et G. R. Dowling (1980). Approximate string matching. *ACM Computing Surveys 12*(4), 381–402.

Kengue, J. F. D., J. Euzenat, et P. Valtchev (2008). Alignement d'ontologies dirigé par la structure. In Y. A. Ameur (Ed.), *CAL*, Volume RNTI-L-2 of *Revue des Nouvelles Technologies de l'Information*, pp. 155. Cépaduès-Éditions.

Miller, G. A. (1995). Wordnet : A lexical database for english. *Communications of the ACM 38*, 39–41.

Rahm, E. et P. A. Bernstein (2001). A survey of approaches to automatic schema matching. *The International Journal on Very Large Data Bases 10*, 334–350.

Shafer, G. (1976). A Mathematical Theory of Evidence. *Princeton University Press*.

Smets, P. (1993). Belief functions : The disjunctive rule of combination and the generalized bayesian theorem. *International Journal of Approximate Reasoning 9*(1), 1–35.

Yager, R. R. (1987). On the dempster-shafer framework and new combination rules. *Informations Sciences 41*, 93–137.

Summary

Mapping ontologies is a crucial step to facilitate semantic interoperability between systems. Different matchers can be used in order to find the correspondences between two ontologies. These matchers can be contradictory, thus leading to a conflict. In order to manage this conflict, we propose an approach by using the Dempster-Shafer theory. Every matcher is considered as a source of evidence and the match results as the mass functions. Managing conflict is dealt when combining between the different results using for that purpose different rules of combination.