Alignement de ressources sémantiques à partir de règles

Valentina Ceausu*, Sylvie Després*

*CRIP 5, Unuversité Paris V
45 Rue des Saints Pères
75006 Paris
ceausu@math-info.univ-paris5.fr
sd@math-info.univ-paris5.fr

Résumé. Ce papier présente une approche automatique pour aligner des ressources sémantiques. L'alignement se traduit par la mise en correspondance des entités (termes, concepts, rôles) appartenant à des ressources d'un même domaine qui peuvent avoir des niveaux de formalisation différents. Les entités correspondantes sont de même nature et un coefficient caractérise leur degré de ressemblance.
L'approche proposée est fondée sur des règles d'appariement entre les entités des deux ressources. Dans une première phase, ces règles d'appariement sont identifiées empiriquement. Des algorithmes combinant les différentes règles identifiées sont ensuite définis afin d'établir des correspondances entre les entités des ressources considérées.
Ce papier présente un ensemble de règles d'appariement exploitant des éléments situés à différents niveaux conceptuels. Cet ensemble constitue un cadre pour l'alignement automatique des ressources sémantiques. Les résultats d'une première expérimentation qui a porté sur l'alignement de deux ressources du domaine de l'accidentologie sont également présentés.

1 Introduction

Avec l'émergence du web sémantique, l'exploitation des ressources sémantiques pour annoter des documents, personnaliser des services ou décrire des ressources disponibles sur le web est devenue essentielle. Créées par des communautés distinctes, existant déjà en grand nombre, les ressources sémantiques se différencient par des niveaux différents de formalisation et de conceptualisation engendrant un certain degré d'hétérogénéité.
Ainsi, l'utilisation conjointe des éléments (documents, services web) décrits par des ressources distinctes est soumise à leur mise en correspondance. Un nouveau problème émerge, qui concerne la mise en correspondance des ressources sémantiques.
Parmi les techniques proposées pour apporter des solutions à ce problème, l'alignement sera considéré dans ce papier. L'alignement établit des appariements entre les entités appartenant à deux ressources distinctes. Au-delà d'une certaine complexité, taille, ou nombre de ressources il est impossible d'établir manuellement ces appariements. La nécessite des méthodes automatiques (ou semi automatiques) pour l'alignement des ressources sémantiques est évidente.

Dans le cadre de ce travail, une ressource sémantique représente un modèle de connaissance spécifique à un domaine. Elle est constituée de concepts, qui modélisent les objets spécifiques au domaine et de rôles, qui correspondent à des relations entre ces objets. Les concepts et les rôles peuvent être structurés dans une hiérarchie.
L'alignement est réalisé automatiquement et met en correspondance des entités ayant la même nature. Ainsi, soient S_s et S_c deux ressources qui décrivent le même domaine. L'alignement associe à chaque entité de e_s de S_s une entité e_c appartenant à S_c, si et seulement si les deux entités ont la même nature et modélisent des objets ou des relations qui sont similaires, voir identiques.
Un ensemble de règles d'appariement est utilisé pour établir ces correspondances. Les règles sont déduites empiriquement et interviennent à plusieurs niveaux : formel, conceptuel, etc.. Des algorithmes sont ensuite implémentés qui combinent ces règles afin de mettre en oeuvre l'alignement.

Le papier est structuré en trois parties : la première définit l'alignement et introduit l'ensemble des règles d'appariement utilisées. La deuxième présente l'alignement de deux ressources du domaine de l'accidentologie et met en évidence les problèmes spécifiques à cette étude de cas. Des travaux connexes sont présentés dans la troisième partie.

2 Alignement de ressources sémantiques

Nous avons proposé une approche pour aligner deux ressources d'un même domaine qui est fondée sur le cadre proposé par Ehrig et Sure (2004). Ce cadre définit des règles empiriques pour estimer le degré de ressemblance entre les entités (concepts ou rôles) des deux ressources. Il a été choisi car : les règles définies prennent en compte des éléments situés à différents niveaux conceptuels ; il n'existe pas de contrainte concernant la modélisation et la formalisation des ressources ; il est possible d'enrichir le cadre initial en ajoutant de nouvelles règles d'appariement.
Par la suite on introduit l'alignement des ressources, tel qu'il est défini pour ce travail, le cadre proposé par Ehrig et Sure (2004) ainsi que les règles d'appariement ajoutées pour l'enrichir.

2.1 Définition

L'alignement des ressources sémantiques peut être défini formellement comme suit : soient S_s (source) et S_c (cible) deux ressources sémantiques du même domaine ; E_{Ss}, E_{S_c} les ensembles d'entités (concepts et rôles), appartenant à la ressource S_s, respectivement S_c.
L'alignement s'exprime par une fonction :

$$Aligner : E_{S_s} \rightarrow E_{S_c}, \text{ où } Aligner(e_j^s) = e_i^c \quad (1)$$

si et seulement si les entités e_j^s et e_i^c ont la même nature et modélisent des objets (si les deux entités sont des concepts) ou des relations (si les deux entités sont des rôles) similaires, voir identiques.
Le degré de similarité est exprimé par un coefficient dont les valeurs sont comprises entre 0 et

1. La valeur 1 signifie que les concepts (respectivement les rôles) mis en correspondance modélisent la même notion et la valeur 0 signifie qu'il n'existe pas de similarités entre les notions représentées par e_j^s, respectivemment e_i^c.
Le processus d'alignement est orienté d'une ressource source S_s vers une ressource cible S_c. Le résultat de l'alignement est constitué d'un ensemble de triplets :

$$(e_j^s, e_i^c, coefficient) \tag{2}$$

Dans le cadre de ce travail, l'alignement associe une seule entité de S_c à une entité de e_j^s. Ceci représente un choix de modélisation, déterminé par l'utilisation ultérieure des résultats issus de cet alignement.

2.2 Règles pour l'alignement des ressources sémantiques

Des règles empiriques pour supporter l'alignement des ressources sémantiques ont été identifiées par Ehrig et Sure (2004). Les règles définies estiment la similarité entre deux entités en prenant en compte des éléments qui se trouvent à différents niveaux conceptuels. Ces niveaux sont repartis sur une échelle, présentée dans la fig. 1.

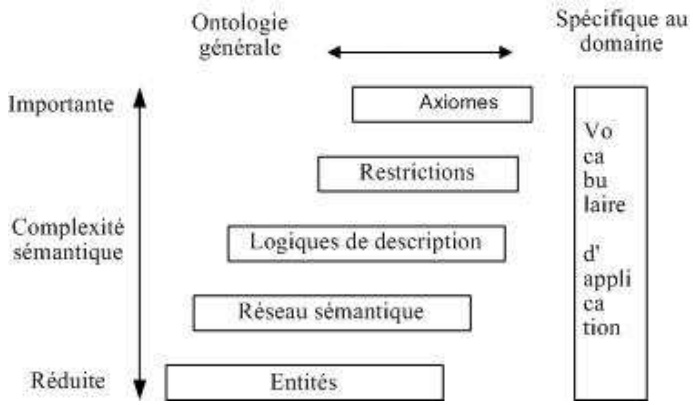

FIG. 1 – *Niveaux conceptuels, d'après Ehrig et Sure (2004)*

Les éléments situés sur chaque niveau ainsi que les règles définies à partir de ces éléments sont présentés *infra*.

Au niveau des entités Les entités constituent le premier niveau de l'échelle. Elles représentent les concepts et les rôles modélisés dans la ressource. Les entités s'identifient à l'aide des termes (étiquettes) qui sont attribués par les humains lors de la construction de la ressource. L'approche d'alignement étant proposée pour des ressources décrivant le même domaine, une première règle peut être définie. Elle s'énonce :

R_1 : *si deux entités ont des étiquettes identiques, alors les entités seront considérées identiques.*

Pour comparer les étiquettes des entités, des mesures lexicales introduites dans Cohen et al. (2003) peuvent être utilisées.

Le niveau du réseau sémantique Le deuxième niveau conceptuel concerne les entités organisées sous forme de réseau sémantique, tel qu'il a été introduit dans Quillan (1967).
Un réseau sémantique modélise les concepts du domaine et les relations entre ces concepts. Les relations sont modélisées par des rôles, chaque rôle étant une application ayant un domaine et un co-domaine. Le domaine de définition et le co-domaine représentent les concepts liés par le rôle dans l'ontologie considérée. Les règles définies à ce niveau s'énoncent :
R_2 : *si les rôles des deux concepts sont identiques, alors les concepts seront considérés identiques*.
R_3 : *si deux rôles ont les domaine de définition et le co-domaine représentés par des concepts, respectivement identiques, alors les deux rôles seront considérés identiques*.

Le niveau des logiques de description Le troisième niveau concerne les ressources exprimées dans le formalisme des logiques de description, voir Baader et al. (2003). Dans ce cas, les concepts sont structurés dans une hiérarchie, dont la racine est un concept générique, le *topConcept*. Les liens hiérarchiques correspondent à des relations de généralisation/spécialisation entre concepts. Par conséquent, tout concept, excepté la racine, a des concepts pères (ascendants) et des concepts fils (descendants). Un concept de l'ontologie est plus spécifique que ses concepts ascendants et plus générique que ses concepts descendants. Les règles qui peuvent être définies à ce niveau sont :

R_4 : *si les concepts pères des deux concepts sont identiques, alors les concepts seront considérés identiques*.
R_5 : *si les concepts fils des deux concepts sont identiques, alors les concepts seront considérés identiques*.
Certaines ressources modélisent les relations être concepts par des rôles qui sont également structurés hiérarchiquement. Deux règles peuvent être déduites, comme suit :
R_6 : *si deux rôles ont des super rôles identiques, alors les deux rôles sont identiques*.
R_7 : *si deux rôles ont des sous rôles identiques, alors les deux rôles sont identiques.*

Le niveau des restrictions Ce niveau peut être considéré dans le cas des ressources exprimées en *OWL*, voir Dean et Schreiber (2004) et qui utilisent les primitives du langage afin d'ajouter des restrictions. En *OWL*, ces primitives, telles que *OWL :sameIndividualAs* ou *OWL :sameClassAs* affirment, explicitement, que deux entités sont identiques. Des règles peuvent être définies en prenant en compte ces primitives.

Le niveau des axiomes Si des relations entre des entités ont été représentées sous la forme d'axiomes, ces relations peuvent être utilisées pour estimer le degré de ressemblance entre les entités. Néanmoins, dans la pratique, de telles modélisations sont presque inexistantes.

Nous avons proposé une approche d'alignement qui enrichit l'ensemble de règles définies par ce cadre. Deux règles ont été ajoutées, qui seront appelées transversales, car elles font ap-

pel à des éléments qui se trouvent à deux niveaux conceptuels. Ces règles s'énoncent comme suit :
R_8 : *tout concept de la ressource S_s qui n'a pas été assigné par la règle R_1 à un concept de la ressource S_c sera assigné au concept auquel son concept père a été assigné par la règle R_1.*
Si les rôles de la ressource sont organisés dans une taxinomie, on déduit :

R_9 : *tout rôle de la ressource S_s qui n'a pas été assigné par la règle R_1 à un rôle de la ressource S_c sera assigné au rôle auquel son rôle père a été assigné par la règle R_1.*
Ces règles prennent en compte à la fois les étiquettes des entités, car elles exploitent les résultats fournis par la règle R_1 et la structuration hiérarchique des entités. Elles sont utilisées si la règle R_1 ne réussit pas à associer à un concept (rôle) de S_s un concept (rôle) similaire, voire identique, appartenant à S_c. Dans ce cas, les règles transversales essayent d'assigner un concept (rôle) de S_s à un concept (rôle) de S_c qui le généralise, en utilisant le concept (rôle) auquel son concept père (rôle père) a été assigné.

3 Aligner des ressources sémantiques de l'accidentologie

L'approche proposée a été employée pour aligner une ontologie et une ressource termino-ontologique (RTO) de l'accidentologie. L'ontologie de l'accidentologie modélise des connaissances expertes. La RTO a été construite à partir d'un corpus constitué de procès verbaux (PV) d'accidents de la route. Les deux ressources modélisent des connaissances propres à des communautés distinctes et ont été construites par des approches différentes.
Ce paragraphe présente les ressources utilisées, la manière dont l'approche proposée a été mise en oeuvre, les résultats obtenus et l'évaluation de ces résultats.

3.1 Ressources sémantiques de l'accidentologie

L'ontologie de l'accidentologie, voir Desprès (2002) modélise les connaissances expertes du domaine. Elle a été construite *ex nihilo* (*from scratch* en anglais), et est fondée sur des entretiens réalisés avec des chercheurs en sécurité routière et en utilisant comme principale ressource textuelle les scénarios d'accidents.
L'éditeur *Protégé*, Noy et al. (2000) a été utilisé pour construire cette ontologie et *OWL* est le langage de représentation choisi. Les connaissances sont modélisées selon un point de vue systémique, les principaux concepts du domaine (*l'Humain*, le *Véhicule* et l'*Environnement*) étant mis en évidence ainsi que les relations qui les lient. Les concepts sont dénommés par des termes du domaine, (*conducteur*, *piéton*). Des attributs sont définit pour chaque concept (le concept *Humain* a l'attribut *âge*), qui sont implémentés à l'aide du type *DataTypeProperty*. Les relations entre concepts sont modélisées par des rôles à partir des verbes du domaine. Les rôles sont implémentés à l'aide du type *ObjectProperty* et sont organisés hiérarchiquement.

La ressource termino-ontologique, voir Ceausu et Desprès (2005), a été construite à partir de procès verbaux (PV) d'accidents de la route rédigés par les gendarmes ou les policiers.
Elle décrit les accidents de la route en mettant en évidence les particularités du vocabulaire employé par les forces d'ordre pour décrire les accidents de la route. L'éditeur *Terminae*,

Aussenac-Gilles et al. (2002) qui offre des facilités pour la construction des ressources sémantiques à partir de textes et leur gestion a été choisi pour structurer les connaissances. Un modèle du domaine a été élaboré utilisant deux types d'entités : les concepts et les rôles. Les concepts sont structurés dans une hiérarchie. Un concept est dénommé par un terme du domaine, et il ne possède pas d'attributs. Les concepts sont liés par des rôles créés à partir de verbes du domaine. La RTO est représentée en *OWL*.

Le tab. 1 met en évidence les différences entre les deux ressources d'un point de vue conceptuel. Les différences d'un point de vue technique sont présentées dans le tab. 2.

	Ontologie	RTO
Connaissances	Expertes	Profanes
Communauté	Experts	Forces de l'ordre
Construction	*ex nihilo*	à partir de textes
Domaine	accidentologie	accidentologie

TAB. 1 – *Ontologie vs RTO : niveau conceptuel*

	Ontologie	RTO
Editeur	Protégé2000 3.1	Terminae
Modélisation	Concept,Attribut (DataProperty)	Concept
	Rôles (ObjectProperty)	Rôles
Structuration	Hiérarchie de rôles et de concepts	Hiérarchie de concepts
Langage	OWL	OWL

TAB. 2 – *Ontologie vs RTO : niveau technique*

Le nombre de concepts et de rôles de la RTO construite à partir des PV est plus important que celui de l'ontologie. L'explication tient au fait que la RTO est construite à partir de textes rédigés en langage courant par des communautés de personne différentes dont l'objectif n'est pas la rigueur dans la présentation mais la description de l'accident dans lequel ils sont impliqués. Tandis que les chercheurs utilisent un langage de spécialité et se contraignent à la concision pour écrire leurs textes.

4 Sélection des règles pour aligner les ressources

Pour aligner les deux ressources, des règles d'appariement introduites dans le paragraphe 2.2 sont utilisées. L'alignement réalisé a comme *ressource source* la RTO construite à partir de PV et comme *ressource cible* l'ontologie de l'accidentologie. Chaque concept (respectivement rôle) appartenant à la RTO est assigné à un concept (respectivement rôle) de l'ontologie. Orienté de cette manière, l'alignement permet de mettre en évidence la manière dont les connaissances expertes modélisées dans l'ontologie de l'accidentologie sont exprimées dans le

langage commun employé par les forces d'ordre.
Le choix des règles utilisées est guidé par les particularités des deux ressources et par le sens de l'alignement.
Les deux ressources modélisent des entités (concept et rôles) du domaine, par conséquent il est possible d'utiliser les règles définies au premier niveau. Ces règles prennent en compte les étiquettes des entités pour estimer leur degré de ressemblance.
Les deux ressources adoptent le formalisme des logiques de description et sont représentées en *OWL*. Cependant, les règles définies au niveaux 2 (réseau sémantique) ne sont pas utilisées, car utilisées conjointement elles engendrent des calculs circulaires.
Les rôles modélisés par la RTO ne sont pas structurés hiérarchiquement, par conséquent les règles définies au niveau de la logique de description ne sont pas appliquées.
Les règles des deux derniers niveaux sont ignorées, car les ressources ne font pas appel à des directives du langage *OWL* tels que *OWL :sameClassAs* et elles ne contient pas d'axiomes.
Le nombre d'entités modélisées par la RTO étant plus important que le nombre d'entités de l'ontologie, les règles transversales (R_8 et R_9) sont choisies pour mettre en oeuvre l'alignement.
L'ensemble des règles utilisé pour aligner les deux ressources est constitué de R_1, R_8 et R_9.Ces règles sont utilisées par deux algorithmes développés pour aligner les concepts, respectivement les rôles des deux ressources, qui sont présentés dans le paragraphe suivant.

Aligner les concepts des ressources L'alignement des concepts est fondé sur la règle R_1 et R_8. L'application de la première (R_1) identifie pour chaque concept de la RTO un concept de l'ontologie ayant la même étiquette. Un coefficient égal à 1 sera assigné à chaque couple de concepts.
La règle R_8 s'applique successivement en prenant en compte les assignations ainsi obtenues. Chaque application de cette règle entraîne une diminution de la valeur du coefficient qui caractérise le degré de similarité des entités. Un concept modélisé dans la RTO sera assigné à un concept de l'ontologie d'autant plus général que le nombre d'applications de la règle R_8 est important.
L'alignement est terminé si tout concept de la RTO est assigné à un concept de l'ontologie. Un extrait des résultats obtenus en alignant les concepts est présenté dans le tab. 3.

Aligner les rôles des ressources Les rôles sont structurés hiérarchiquement dans l'ontologie de l'accidentologie, mais ils se trouvent au même niveau dans la RTO. Par conséquent, les règles d'alignement fondées sur la hiérarchie ne peuvent pas être appliquées.
Pour pallier cet inconvénient, la règle R_1 a été adaptée. Ainsi, elle est implémentée en utilisant comme mesure de similarité lexicale la mesure de *Monge-Elkan*, voir par exemple Ceausu et Desprès (2006). Cette mesure fait des comparaisons récursives au niveau des sous-chaînes et fournit la valeur maximale si la chaîne s_1 est une sous-chaîne de la chaîne s_2. La règle devient : les rôles dénommés par des étiquettes dont la similarité (calculée par le coefficient *Monge-Elkan*) est supérieure à une valeur seuil donnée seront considérés similaires (voir identiques). Les valeurs calculées par *Monge-Elkan* représentent les coefficients caractérisant le degré de ressemblance des rôles.
Le tab. 4 montre un extrait des résultats obtenus en alignant les rôles des deux ressources.

Concept RTO	Concept ontologie	Coefficient
train	transport en commun	0.50
motocyclette	deux roues	0.50
vélomoteur	deux roues	0.50
gauche	direction	0.50
automobile	véhicule léger	0.50
ruisseau	environnement	0.50
cyclomotoriste	conducteur	0.50
chauffeur	conducteur	0.50
autocar	transport en commun	0.50
motocycliste	conducteur	0.50

TAB. 3 – *Alignement des concepts*

Rôles RTO	Rôles ontologie	Coefficient
accrocher avec	accrocher	1.0
parler avec	discuter avec	0.76
percuter par	percuter	1.0
traverser sans	traverser	1.0
virer à	tourner à	0.83
virer vers	diriger vers	0.89
aller de	venir de	0.79

TAB. 4 – *Alignement des Rôles*

4.1 Evaluation des résultats

Les résultats de l'alignement de chaque type d'entité sont évalués indépendamment. Cette évaluation distincte est nécessaire car les concepts et les rôles sont structurés différemment et les algorithmes utilisés pour les aligner sont distincts. Les résultats de l'alignement ne peuvent pas être évalués globalement car ils sont issus de deux approches différentes.
L'évaluation réalisée concerne seulement l'étude de cas présentée, car la qualité des résultats de l'alignement est influencée par la complexité des ressources alignées. Le scénario d'évaluation proposé compare, pour chaque type d'entité, les résultats fournis par l'alignement des ressources avec des résultats obtenus en établissant, manuellement, des correspondances entre des entités.
L'évaluation fait appel à des mesures proposées dans le domaine de la recherche de l'information, qui reposent sur les notions classiques de $Rappel$ et de $Precision$.
Ces mesures ont été adaptées au nouveau contexte de l'alignement des ressources sémantiques. Ainsi, on peut définir le $Rappel$ comme suit :

$$Rappel = \frac{NoCorrects}{NoRef} \tag{3}$$

où $NoCorrects$ représente le nombre d'alignements corrects et $NoRef$ représente le nombre d'alignements de référence. La $Precision$ est définie par :

$$Precision = \frac{NoCorrects}{NoPro} \tag{4}$$

où $NoCorrects$ représente le nombre d'alignements corrects et $NoRef$ représente le nombre d'alignements proposés.
L'évaluation des résultats peut aussi être exprimée en terme de $Bruit$ ou $Silence$, comme suit :

$$Silence = 1 - Rappel;\ Bruit = 1 - Precision \tag{5}$$

F-measure est une mesure d'efficacité globale qui combine *Precision* et *Rappel* en une mesure unique donnée par :

$$F = \frac{2 * Precision * Rappel}{Precision + Rappel} \tag{6}$$

On suppose qu'un alignement de référence a été établit manuellement, et que c'est par rapport à cet alignement que seront évalués les résultats obtenus. La comparaison par rapport à l'alignement de référence sera effectuée en comparant les couples de concepts, respectivement rôles, engendrés par l'alignement. La valeur du coefficient qui exprime le degré de ressemblance est ignorée lors de l'évaluation. Ainsi, un alignement $(e_s, e_c, coefficient)$ sera considéré correct si, dans le set de référence il existe un couple (e_s, e_c), quelque soit la valeur du coefficient $coefficient$. Les valeurs obtenues sont présentées dans le tab. 5. Dans le même tableau on

	Prescision	Rappel	Silence	Bruit	F-Measure
Concepts	0.79	0.67	0.33	0.21	0.72
Rôles	0.71	0.52	0.29	0.48	0.60

TAB. 5 – *Evaluation des résultats*

peut observer que l'algorithme utilisé pour aligner les concepts est plus performant. Cela s'explique par la structuration hiérarchique des concepts dans les deux ressources, qui fait possible l'utilisation de la règle transversale R_8. Par conséquent, l'alignement des concepts est réalisé par un algorithme combinant deux règles d'appariement.
Malgré l'adaptation de la règle R_1 (qui utilise une mesure lexicale particulière pour apparier des rôles), l'algorithme proposé pour aligner les rôles est moins performant. D'un point de vue pratique, un nombre important de rôles de la ressource source (la RTO construite à partir de PV) sont assignés au rôle générique de la ressource cible (l'ontologie de l'accidentologie).

5 Travaux connexes

Un recueil des méthodes et des outils permettant la mise en correspondance de sémantiques est réalisé dans : Kalfoglou et Schorlemmer (2003) et Euzenat (2004). Des comparaisons entre les outils sont présentées dans Do et al. (2002), Rahm et Bernstein (2001). Parmi les outils proposés, on retrouve :

Anchor Prompt , Fridman Noy et Musen (2001) est un outil permettant l'alignement et l'intégration des ontologies. Il reçoit en entrée deux ontologies et une liste de paires de termes. Les paires de termes sont fournies par l'utilisateur ou identifiées en utilisant des métriques lexicales. Un processus semi-automatique permet d'identifier des concepts qui sont similaires en utilisant ces paires de termes, les structures des ontologies et les choix de l'utilisateur.

Chimaera , McGuinness et al. (2000) est un outil qui permet d'aligner des ontologies de grande taille. Un algorithme engendre des paires de concepts similaires en comparant : les termes utilisés pour designer les concepts, les définitions des concepts, et, selon le cas, les acronymes ou les expansions de ces noms. Chimaera est aussi capable d'identifier des concepts qui sont corrélés par d'autres types de relations, telles que la subsomption, ou des termes qui sont disjoints.

Cupid, Madhavan et al. (2000) est un système qui implémente un algorithme d'alignement fondé sur les similarités lexicales et structurelles. Des coefficients de similarité sont calculés et trois étapes sont exécutées pour générer des paires de concepts similaires. Une première étape calcule des similarités au niveau lexical en utilisant les noms des entités, des mesures de similarité lexicale et en faisant appel à un thesaurus ; la deuxième étape estime la similarité d'un point de vue structurel, en considérant les contextes d'apparition des concepts dans les ontologies. La dernière phase engendre les paires des concepts similaires, en choisissant, parmi les paires générées, celles ayant un coefficient de similarité supérieur à une valeur seuil donnée.

Asco est un système développé à *INRIA Sophia Antipolis*, qui peut identifier des correspondances entre : deux concepts appartenant à deux ontologies distinctes ; deux relations modélisées dans deux ontologies distinctes ; un concept et une relation appartennant à deux ontologies distinctes. *Asco* implémente un algorithme qui met en correspondance les entités en exploitant le maximum d'éléments disponibles : les noms des entités ; les structures des ontologies ; les structures des entités,des concepts ou des rôles. Cet algorithme est implémenté en *Java* et est fondé sur le moteur de recherche sémantique *CORESE*, décrit dans Corby et Faron (2002).

Des méthodes ont été proposées qui estiment la similarité en prenant en compte les instances des concepts, voir par exemple les systèmes *Glue*, Doan et al. (2002) et *FCA Merge*, Stumme et Maedche (2002) ou les axiomes présentes dans une ontologie, voir Furst (2002).

L'approche que nous avons proposée pour aligner deux ressources sémantiques est fondée sur les travaux de Ehrig et Sure (2004). Ce choix a été guidé par les particularités des ressources à aligner qui sont constituées de concepts et de rôles. Elles ne contiennent pas d'axiomes ou

des individus, par conséquent les méthodes faisant appel à ces éléments ne peuvent pas être appliquées.
L'alignement doit mettre en évidence des similarités entre les concepts et les rôles des ressources. Des outils tels que *Anchor Prompt*, *Chimaera* ou *Cupid* s'avèrent inappropriés, car ils établissent des correspondances seulement entre les concepts des ressources.
Asco est un outil récent qui n'a pas pu être considéré dans le cadre de ce travail.
L'approche proposée par Ehrig et Sure (2004) à l'avantage de mettre en correspondance les concepts et les rôles appartenant à deux ressources. Nous avons adaptée cette approche en enrichissant l'ensemble de règles d'appariement utilisées.

6 Conclusion

Ce papier présente une approche pour aligner deux ressources sémantiques. L'alignement est fondé sur des règles d'appariement entre les entités des deux ressources et se traduit par des correspondances entre ces entités.
Une première étude a été réalisée qui emploie cette approche pour aligner une ontologie et une ressource termino-ontologique de l'accidentologie. L'évaluation des résultats issus de cette expérimentation montre qu'elle est sensible à la manière dont les entités sont modélisées. En perspective, la méthode d'alignement des ressources peut être améliorée en enrichissant l'ensemble de règles utilisées. De nouveaux algorithmes faisant appel à ces règles peuvent être proposés. Une méthode d'évaluation prenant en compte à la fois les couples d'entités obtenus ainsi que les coefficient caractérisant leur degré de similarité est également envisageable.

Références

Aussenac-Gilles, N., B. Biébow, et S. Szulman (2002). Terminae. In *Workshop on Evaluation of Ontology Engineering Environments, Knowledge Engineering and Knowledge Management : Methods, Models and Tools, 13th International Conference, EKAWŠ2002*, Siguenza, France, Octobre 2002.

Baader, F., D. Calvanese, D. McGuinness, D. Nardi, et P.-F. Patel-Schneider (Eds.) (2003). *The Description Logic Handbook : Theory, Implementation, and Applications*. Cambridge University Press.

Ceausu, V. et S. Desprès (2005). Towards a text mining driven approach for terminology construction. In *7th International conference on Terminology and Knowledge Engineering*.

Ceausu, V. et S. Desprès (2006). Reconnaissance automatique de concepts à partir d'une ontologie. In *Revue RNTI (Revue des Nouvelles Technologies de l'Information), numéro spécial EGC'2005*.

Cohen, W., , P. Ravikumar, et S. Fienberg (2003). A comparison of string distance metrics for name-matching tasks. In *IJCAI,Workshop on Information Integration on the Web*.

Corby, O. et C. Faron (2002). Corese : A corporate semantic web engine. In *Proceedings of WWW International Workshop on Real World RDF and Semantic Web Applications*, USA.

Dean, M. et G. Schreiber (2004). Owl web ontology language reference. Technical report, W3C. W3C Proposed Recommendation.

Desprès, S. (2002). *Contribution à la conception de méthodes et d'outils pour la gestion des connaissances*. Habilitation à diriger des recherches, Université René Descartes.

Do, H., S. Melnik, et E. Rahm (2002). Comparison of schema matching evaluations. In *Proceedings of the 2nd International Workshop on Web Databases*.

Doan, A., J. Madhavan, P. Domingos, et A. Halevy (2002). Learning to map between ontologies on the semantic web. In *Proceedings of the 11th International WWW Conference*.

Ehrig, M. et Y. Sure (2004). Ontology mapping - an integrated approach. In *Proceedings of the First European Semantic Web Symposium*, Heraklion, Greece, pp. 76–91.

Euzenat, J. (2004). State of the art on current alignment techniques. Technical report, IST Knowledge Web Network of Excellence no FP6-507482, D2.2.3.

Fridman Noy, N. et M. A. Musen (2001). Anchor-prompt : Using non-local context for semantic matching. In *IJCAI 2001 workshop on ontology and information sharing*.

Furst, F. (2002). *Contribution à l'ingénierie des ontologies : une méthode et un outil d'opérationnalisation*. Ph. D. thesis, Ecole Polytechnique de l'Université de Nantes.

Kalfoglou, Y. et M. Schorlemmer (2003). Ontology mapping : the state of the art. *The Knowledge Engineering Review 18(1)*, 1–31.

Madhavan, J., P. Bernstein, et E. Rahm (2000). Generic schema matching using cupid. In *Proceedings of 27th International Conference on Very Large Data Bases*, Roma, Italy.

McGuinness, D. L., R. Fikes, J. Rice, et S. Wilder (2000). An environment for merging and testing large ontologies. In *Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning*, San Francisco, USA, pp. 483–493.

Noy, N., R. W. Fergerson, et M. A. Musen (2000). The knowledge model of protégé-2000 : Combining interoperability and flexibility. In *In Proceedings of the International Conference on Knowledge Engineering and Knowledge Management*.

Quillan, M. R. (1967). Word concepts : A theory and simulation of some basic capabilities. *Behavioral Science* (12).

Rahm, E. et P. Bernstein (2001). A survey of approaches to automatic schema matching. *VLDB Journal : Very Large Data Bases 10*(4), 334–350.

Stumme, G. et A. Maedche (2002). Fca-merge : Bottom-up merging of ontologies. In *Proceedings of 7th International Joint Conference on Artificial Intelligence*, USA, pp. 225–234.

Summary

This paper presents an automatic approach to align semantic resources. The alignment establishes associations between entities (concepts, properties) of two different resources of the same field, which could have different levels of formalisation. Corresponding entities have the same nature and a trust assessment is assigned to each association.
This approach is based on marching rules. First, a set of matching rules is identified. Then, algorithms are combining those rules to automatically align two resources.
This paper presents the set of matching rules and a first experimentation carried our in order to align two semantic resources of accidentology.