

Un jeu de données d'évaluation de correspondances complexes entre ontologies

Elodie Thiéblin Ollivier Haemmerlé Nathalie Hernandez Cássia Trojahn
28èmes Journées Francophones d'Ingénierie des Connaissances

IRIT & Université de Toulouse 2, Toulouse, France
firstname.lastname@irit.fr

Un jeu de données d'évaluation de
correspondances complexes entre
ontologies

- Le LOD
- Hétérogénéité entre ontologies

Un jeu de données d'évaluation de correspondances complexes entre ontologies

- Alignements : correspondances simples et complexes
- Patrons de correspondance



Alignement d'ontologies

Alignement [Euzenat and Shvaiko, 2013]

Ensemble de correspondances entre deux ontologies : une ontologie source o_1 et une ontologie cible o_2

Correspondance

$\langle e_{o_1}, e_{o_2}, r \rangle, \quad r \in \{\equiv, \leq, \geq\}$

e_{o_1}, e_{o_2} : membres : entités et constructeurs

Entités classes, relations, propriétés, instances, valeurs

Constructeurs opérateurs logiques (OR, AND, etc.) et fonctions de transformation (concaténation de chaînes de caractères, sommes, etc.)



Cardinalité d'une correspondance

Cardinalité d'une correspondance

La **cardinalité** d'une correspondance dépend du nombre d'entités et de constructeurs présents dans chaque membre de la correspondance

Cardinalité d'une correspondance

Cardinalité d'une correspondance

La **cardinalité** d'une correspondance dépend du nombre d'entités et de constructeurs présents dans chaque membre de la correspondance

Correspondance simple (1:1)

Correspondance entre 2 entités atomiques

Cardinalité (1:1) : pas de constructeur

$$\forall x, o1:Paper(x) \equiv o2:Article(x)$$



Correspondances complexes

Correspondance complexe (1:n), (m:1), (m:n)

Correspondance dont un des membres a au moins deux entités ou au moins un constructeur

Cardinalité (1:n), (m:1), (m:n)

$$\forall x, o1:Accepted_Paper(x) \equiv \exists y, o2:hasDecision(x,y) \wedge o2:Acceptance(y)$$



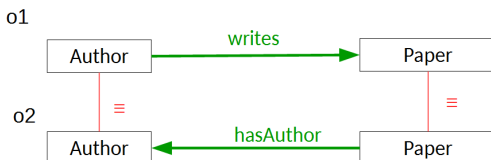
Correspondances complexes

Correspondance complexe (1:n), (m:1), (m:n)

Correspondance dont un des membres a au moins deux entités ou au moins un constructeur

Cardinalité (1:n), (m:1), (m:n)

$$\forall x,y, o1:writes^-(x,y) \equiv o2:hasAuthor(x,y)$$



Patrons de correspondance [Scharffe, 2009]

Class By Attribute Type (CAT)

Restriction de type sur les objets d'une relation (propriété sur les objets)

Correspondances de la forme $\forall x, A(x) \equiv \exists y, b(x, y) \wedge C(y)$.

$$\forall x, o1:Accepted_Paper(x) \equiv \exists y, o2:hasDecision(x, y) \wedge o2:Acceptance(y)$$



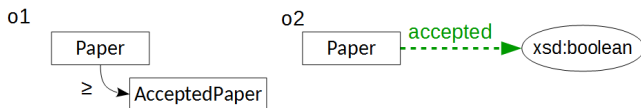
Patrons de correspondance [Scharffe, 2009]

Class By Attribute Value (CAV)

Restriction de valeur sur les objets d'un attribut (relation ou propriété)

Correspondances de la forme $\forall x, A(x) \equiv \exists y, b(x, v)$.

$$\forall x, o1:Accepted_Paper(x) \equiv \exists y, o2:accepted(x, true)$$



D'autres patrons existent (CAO, PC, IP...)

Alignement complexe d'ontologie et patrons de correspondances

- L'alignement d'ontologies permet leur **interopérabilité**
- Les correspondances complexes permettent une **expressivité** plus fine
- Les patrons de correspondance permettent de les caractériser et les **analyser**

Un jeu de données d'évaluation correspondances complexes entre ontologies

- Approches d'alignement complexe
- Les jeux d'évaluation existants

Approches d'alignement complexe

Nom	Type de sortie	Techniques	Entrées
[Ritze et al., 2009]	CAV, CAT, CAE, etc.	Structure, Syntaxe, Patrons	Alignement simple
[Ritze et al., 2010]	CAV, CAT, CAE, etc.	Structure, Linguistique, Patrons	Alignement simple (opt.)
[Parundekar et al., 2010]	AV= Union ($A_i V_i$)	Patrons, Statistiques	Alignement d'instances
[Parundekar et al., 2012]	AV= $A\{v_1, v_2, \dots\}$	Patrons, Statistiques	Alignement d'instances
[Walshe, 2014]	CAV	Patrons, Statistiques	Alignement d'instances
[Nunes et al., 2011]	fonctions de transformation	Statistiques	Alignement d'instances
[Qin et al., 2007]	Chemins	Structure, Statistiques	Instances communes

Approches d'alignement complexe

Nom	Type de sortie	Techniques	Entrées
[Ritze et al., 2009]	CAV, CAT, CAE, etc.	Structure, Syntaxe, Patrons	Alignement simple
[Ritze et al., 2010]	CAV, CAT, CAE, etc.	Structure, Linguistique, Patrons	Alignement simple (opt.)
[Parundekar et al., 2010]	AV= Union (AiVi)	Patrons, Statistiques	Alignement d'instances
[Parundekar et al., 2012]	AV= A{v1,v2,...}	Patrons, Statistiques	Alignement d'instances
[Walshe, 2014]	CAV	Patrons, Statistiques	Alignement d'instances
[Nunes et al., 2011]	fonctions de transformation	Statistiques	Alignement d'instances
[Qin et al., 2007]	Chemins	Structure, Statistiques	Instances communes

Évaluation des approches d'alignement complexe

Nom	Jeu de données	Évaluation
[Ritze et al., 2009]	OAEI Conference ¹	Manuelle (précision)
[Ritze et al., 2010]		Manuelle (précision)
[Parundekar et al., 2010]	DBpedia, Geonames, Geospecies, LinkedGeoData, GenID, MGI	Aucune
[Parundekar et al., 2012]		Manuelle (précision) sur un sous-ensemble
[Walshe, 2014]	DBpedia, Yago	Automatique sur jeu de données (CAV) créé automatiquement
[Nunes et al., 2011]		Aucune
[Qin et al., 2007]	UMD, CMU	Manuelle (précision)

¹<http://oaei.ontologymatching.org/>

Approches d'alignement complexe et leur évaluation

- De plus en plus d'approches d'alignement complexe
- Pas de jeu de données de référence
- Pas d'évaluation automatique générique

Un jeu de données d'évaluation
correspondances complexes entre
ontologies.

- Méthodologie
- Le jeu de données
- Utilisation du jeu de données

Méthodologie de construction du jeu de données

5 étapes pour trouver un alignement de correspondances (1:n) :

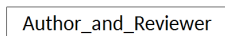
1. Correspondances simples o1-o2
2. Recherche pour chaque entité de o1, une équivalence dans o2
Si pas d'équivalence, chercher une subsomption
3. Vérification manuelle de la cohérence de chaque correspondance
4. Vérification de la cohérence globale en fusionnant les 2 ontologies
5. Filtrer les correspondances pour éviter la redondance

Cohérence des correspondances

Aucun membre d'une correspondance ne doit renvoyer l'ensemble vide

$$\forall x, o1:Author_and_Reviewer(x) \equiv o2:Author(x) \wedge o2:Reviewer(x)$$

o1



o2

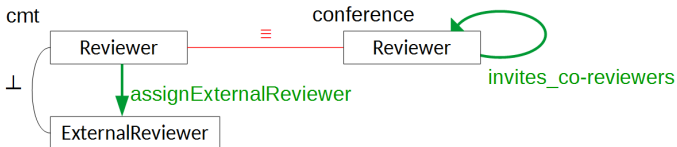


Cohérence des correspondances

Aucun membre d'une correspondance ne doit renvoyer l'ensemble vide

Deux correspondances ne doivent pas se contredire

1. $\forall x, \text{cmt:Reviewer}(x) \equiv \text{conf:Reviewer}(x)$
2. $\forall x, y, \text{cmt:assignExternalReviewer}(x, y) \equiv \text{conf:invites_co-reviewer}(x, y)$
3. (1) $\Rightarrow \forall x, \text{conf:ExternalReviewer}(x) \perp \text{conf:Reviewer}(x)$
4. (2) $\Rightarrow \forall x, \text{conf:ExternalReviewer}(x) \equiv \text{conf:Reviewer}(x)$



Filtrage des correspondances pour alignement minimal

Choix pour éviter la redondance

1. Préférer les équivalences aux subsumptions
2. Préférer les correspondances simples aux complexes
 - ✓ $\forall x, o1:Reviewer(x) \equiv o2:Reviewer(x)$
 - ✗ $\forall x, o1:Reviewer(x) \equiv \exists y, o2:reviews(x, y)$
3. Écrire les constructions le plus simplement possible
 - ✓ $\forall x, o1:Reviewer(x) \equiv \exists y, o2:reviews(x, y)$
 - ✗ $\forall x, o1:Reviewer(x) \equiv \exists y, o2:hasReviewer^-(x, y)$

Les ontologies du jeu de données

Jeu de données Conférence

- Utilisé dans la campagne de l'OAEI
- 16 ontologies
- Alignements simples de référence entre 7 ontologies

	cmt	conference	edas
Nombre de classes	30	60	104
Nombre de relations	49	46	30
Nombre de propriétés	10	18	20

Le jeu de données

Paire	Classes	Relations	Propriétés
cmt-conference	13 simples, 4 CAT, 1 union, 2 CAE (composés)	2 dom, 2 range, 2 chaines, 4 dom-range, 2 unions	1 simple, 3 transfo
cmt-edas	9 simples, 2 CAT, 1 CAE, 1 union	5 simples, 1 dom, 2 dom-range, 4 unions	2 dom, 3 transfo

Correspondances de **différents patrons** (atomiques ou composés).

Pas tous les patrons (e.g. CAV)

Apport alignements complexes

Entités de *cmt* mises en équivalence (\equiv) avec *conference* :

- simples : 15.7%
- simples + complexes : 30.3%

Évaluation de [Ritze et al., 2009]

CMT - CONFERENCE Précision 0 - Rappel 0

✗ $\forall x, \text{cmt:Reviewer}(x) \equiv \exists y, \text{conf:contributes}(x, y) \wedge \text{conf:Reviewed_contribution}(y)$

CMT - EDAS Précision 0.75 (3/4) - Rappel 0.27 (3/11)

✓ $\forall x, \text{edas:RejectedPaper}(x) \equiv \exists y, \text{cmt:hasDecision}(x, y) \wedge \text{cmt:Rejection}(y)$

✓ $\forall x, \text{cmt:ConferenceMember}(x) \equiv \exists y, \text{edas:hasMember}^-(x, y) \wedge \text{edas:Person}(x)$
 $\forall x, \text{cmt:ConferenceMember}(x) \equiv \exists y, \text{edas:isMemberOf}(x, y)$

Correspondances CAT² dans le jeu de données non-détectées

²Class By Attribute Type

Évaluation de [Ritze et al., 2010]

Ajout de contraintes linguistiques

CMT - CONFERENCE Précision 0 - Rappel 0

Aucune

CMT - EDAS Précision 1 (2/2) - Rappel 0.18 (2/11)

✓ $\forall x, edas:AcceptedPaper(x) \equiv \exists y, cmt:hasDecision(x, y) \wedge cmt:Acceptance(y)$

Diminution du rappel mais augmentation de la précision

Conclusion et perspectives

Premiers pas vers l'évaluation d'approches d'alignement complexe

- Alignements cohérents et exhaustifs pour les correspondances (1:n)
- Première évaluation d'approches d'alignement complexe

Conclusion et perspectives

Premiers pas vers l'évaluation d'approches d'alignement complexe

- Alignements cohérents et exhaustifs pour les correspondances (1:n)
- Première évaluation d'approches d'alignement complexe

Encore du chemin à parcourir

- Extension du jeu de données (instances, correspondances (m:n))
- Ajout d'autres paires d'ontologies
- Formalisation des correspondances (e.g. EDOAL, etc.)
- Étude de la confiance des correspondances
- Méthode d'évaluation automatique des correspondances complexes
- Approches d'alignement complexe

MILLE MERCI DU FOND
DU COEUR D'AVOIR PRIS
LE TEMPS SANS REGARDER
VOS MAÎLS DE M'ECOUTER
DIDADIDADOOM

QUELLE HEURE EST-IL ?
C'EST MA PRESENTATION,
MERCİ DE VOTRE ATTENTION.

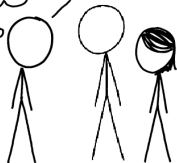
ET VOILA, C'EST FINI.
★★★★☆ BON PUBLIC

J'ESPERE QU'ILS ONT
COMPRIŞ CE QU'EST UN
ALIGNEMENT COMPLEXE

LE GENERAL CHAIRA DU
RECEVOIR MON POT DE VIN...
J'ESPERE QUE ÇA ALLAIT.

MERCİ, QUESTIONS ?

DAMMIT.



JE NE SAIS JAMAIS QUOI DIRE
APRES UNE PRESENTATION

Merci, questions ?

elodie.thieblin@irit.fr

-  Euzenat, J. and Shvaiko, P. (2013).
Ontology Matching.
Springer Berlin Heidelberg.
-  Nunes, B. P., Mera, A., Casanova, M. A., Breitman, K. K., and Leme, L. A. P. (2011).
Complex Matching of RDF Datatype Properties.
In 6th ISWC workshop on ontology matching (OM).
-  Parundekar, R., Knoblock, C. A., and Ambite, J. L. (2010).
Linking and building ontologies of linked data.
In International Semantic Web Conference, pages 598–614. Springer.



Parundekar, R., Knoblock, C. A., and Ambite, J. L. (2012).

Discovering concept coverings in ontologies of linked data sources.


In *International Semantic Web Conference*, pages 427–443. Springer.



Qin, H., Dou, D., and LePendu, P. (2007).

Discovering executable semantic mappings between ontologies.

In *On the Move to Meaningful Internet Systems*, pages 832–849. Springer.

 Ritze, D., Meilicke, C., Šváb Zamazal, O., and Stuckenschmidt, H. (2009).

A pattern-based ontology matching approach for detecting complex correspondences.

In *4th ISWC workshop on ontology matching (OM)*, pages 25–36.

 Ritze, D., Völker, J., Meilicke, C., and Šváb Zamazal, O. (2010).

Linguistic analysis for complex ontology matching.

In *5th ISWC workshop on ontology matching (OM)*, pages 1–12.

 Scharffe, F. (2009).

Correspondence Patterns Representation.

PhD thesis, Faculty of Mathematics, Computer Science and University of Innsbruck.



Walshe, B. (2014).

Detecting Restriction Class Correspondences in Linked Open Data.

PhD thesis, Trinity College, Dublin.