

# Évaluation automatique d'alignements complexes : une approche basée sur les instances

---

**Elodie Thiéblin**   Ollivier Haemmerlé   Cassia Trojahn  
Institut de Recherche en Informatique de Toulouse  
IC2022 : 29 juin 2022

Article traduit : *Thiéblin, E., Haemmerlé, O., & Trojahn, C. (2021). Automatic evaluation of complex alignments: An instance-based approach. Semantic Web, 12(5), 767-787.*

## Contexte

---

# Ontologies hétérogènes

Contexte

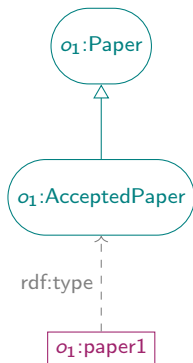
Proposition

Jeu de données

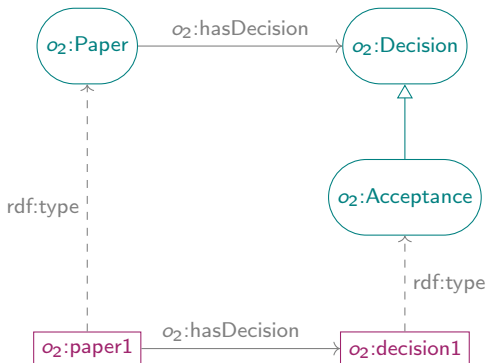
Évaluation

Conclusion

Base décrite avec  $o_1$



Base décrite avec  $o_2$



# Correspondances simples et complexes

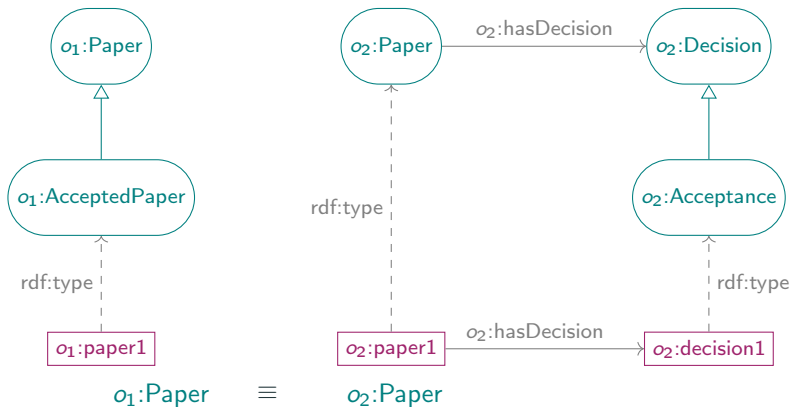
Contexte

Proposition

Jeu de données

Évaluation

Conclusion



Correspondance simple: 1 URI à 1 URI (s:s)

# Correspondances simples et complexes

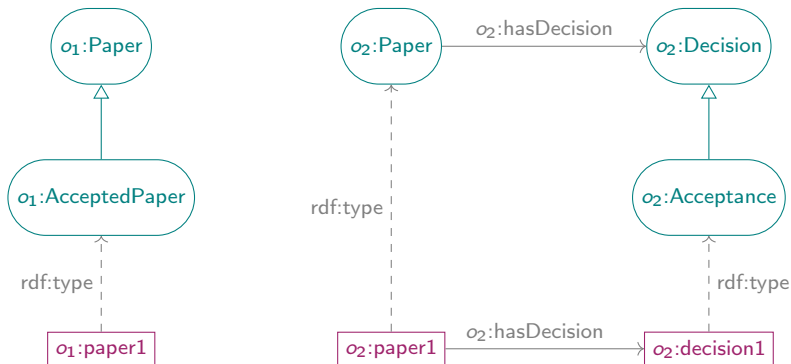
Contexte

Proposition

Jeu de données

Évaluation

Conclusion



$o_1:\text{AcceptedPaper} \quad \equiv \quad o_2:\text{Paper} \sqcap \exists o_2:\text{hasDecision}.o_2:\text{Acceptance}$

**Correspondance complexe:** constructions logiques ou transformations (s:c)

# Correspondances simples et complexes

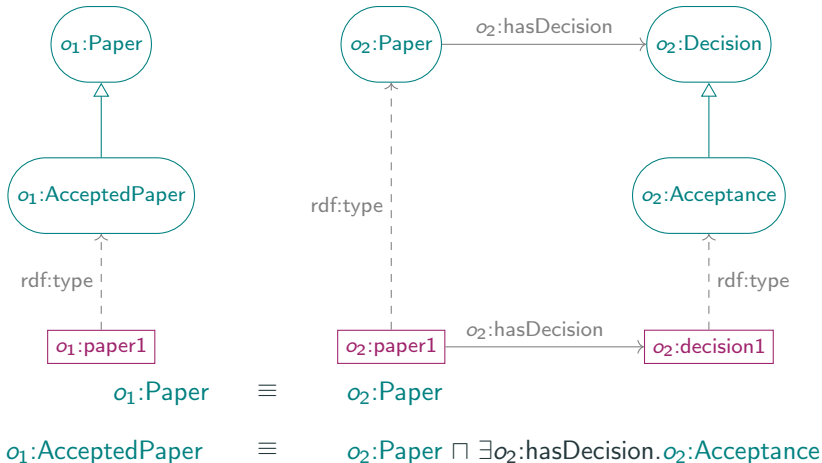
Contexte

Proposition

Jeu de données

Évaluation

Conclusion



**Alignement:** ensemble de correspondances

- Des outils permettent de trouver des alignements automatiquement [Faria et al., 2018, Walshe et al., 2016, Ritze et al., 2010], ...
- L'OAEI <sup>1</sup> évalue les outils d'alignement
- Les méthodes d'évaluation automatique des alignements sont concentrées sur les **simples**

---

<sup>1</sup>Ontology Alignment Evaluation Initiative <http://oaei.ontologymatching.org>

- **Intrinsèque** Propriétés de l'alignement  
[Solimando et al., 2014, Meilicke and Stuckenschmidt, 2008]
- **Extrinsèque** Comparaison avec une référence (e.g., alignement de référence)

Difficulté pour les alignements complexes: **comparaison** et **référence**



# Proposition

---

**Problème:** Création d'un alignement de référence **exhaustif** (sans restriction d'expressivité)

# Couverture : Question de Compétence pour Alignement

Contexte

Proposition

Jeu de données

Évaluation

Conclusion

**Problème:** Création d'un alignement de référence **exhaustif** (sans restriction d'expressivité)

**CQA:** Comme les questions de compétence pour la **conception d'ontologie**, mais à laquelle **deux ontologies ou plus** peuvent répondre

**Problème:** Création d'un alignement de référence **exhaustif** (sans restriction d'expressivité)

**CQA:** Comme les questions de compétence pour la **conception d'ontologie**, mais à laquelle **deux ontologies ou plus** peuvent répondre

Peut-être une **question en langage naturel** ou des **requêtes SPARQL**.

- “Quels sont les papiers acceptés ?”
- `SELECT ?x WHERE {?x a o1:AcceptedPaper.}`
- `SELECT ?x WHERE {?x o2:hasDecision ?y. ?y a o2:Acceptance.}`

**Problème:** Création d'un alignement de référence **exhaustif** (sans restriction d'expressivité)

**CQA:** Comme les questions de compétence pour la **conception d'ontologie**, mais à laquelle **deux ontologies ou plus** peuvent répondre

Peut-être une **question en langage naturel** ou des **requêtes SPARQL**.

- “Quels sont les papiers acceptés ?”
- `SELECT ?x WHERE {?x a o1:AcceptedPaper.}`
- `SELECT ?x WHERE {?x o2:hasDecision ?y. ?y a o2:Acceptance.}`

**Choix :** Utiliser un ensemble de CQA comme référence

# Comparaison de correspondances

Contexte

**Proposition**

Jeu de données

Évaluation

Conclusion

**Problème:** Comparer des correspondances complexes

**Problème:** Comparer des correspondances complexes

- **Syntaxique**  
 $\langle o_1:Author, \exists o_2:writes.\top, \equiv \rangle \langle o_1:Author, \exists o_2:isWrittenBy^-. \top, \equiv \rangle$
- **Sémantique** expressivité sur *SROIQ*, pas de gestion des transformations
- **Instances** demande des bases peuplées et instances de qualité

**Problème:** Comparer des correspondances complexes

- **Syntaxique**  
 $\langle \alpha_1:Author, \exists \alpha_2:writes.\top, \equiv \rangle \langle \alpha_1:Author, \exists \alpha_2:isWrittenBy^-\top, \equiv \rangle$
- **Sémantique** expressivité sur *SROIQ*, pas de gestion des transformations
- **Instances** demande des bases peuplées et instances de qualité

**Choix :** Comparaison d'ensembles d'instances



# Comparaison d'instances

Contexte

Proposition

Jeu de données

Évaluation

Conclusion

2 ensembles d'instances  $I_{ref}$  et  $I_{ev}$ .

$$\text{query Fmeasure}(I_{ref}, I_{ev}) = 2 \times \frac{QR \times QP}{QR + QP}$$

$$QP = \frac{|I_{ev} \cap I_{ref}|}{|I_{ev}|}$$

$$QR = \frac{|I_{ev} \cap I_{ref}|}{|I_{ref}|}$$

$$\text{equiv}(I_{ref}, I_{ev}) = \begin{cases} 1 & \text{si } I_{ev} \equiv I_{ref} \\ 0 & \text{sinon} \end{cases}$$

$$\text{non disjoint}(I_{ref}, I_{ev}) = \begin{cases} 1 & \text{si } I_{ev} \not\cap I_{ref} \\ 1 & \text{si } I_{ev} \stackrel{\emptyset}{=} I_{ref} \\ 0 & \text{sinon} \end{cases}$$

$$\text{intersection}(I_{ref}, I_{ev}) = \begin{cases} 1 & \text{si } I_{ev} \not\cap I_{ref} \\ 0 & \text{sinon} \end{cases}$$

*cqa<sub>T</sub>*

```
SELECT ?x WHERE{ ?x  
o2:hasDecision ?y. ?y  
a o2:Acceptance}
```

*cqa<sub>S</sub>*

```
SELECT ?x WHERE{ ?x a  
o1:AcceptedPaper}
```

A

Référence: paire de requêtes SPARQL équivalentes représentant une CQA

*cqaT*

```
SELECT ?x WHERE{ ?x  
o2:hasDecision ?y. ?y  
a o2:Acceptance}
```

*cqaS*

```
SELECT ?x WHERE{ ?x a  
o1:AcceptedPaper}
```

A

*réécr.*

*q1T*

```
SELECT ?x WHERE{ ?x  
a o2:Paper}
```

*q2T*

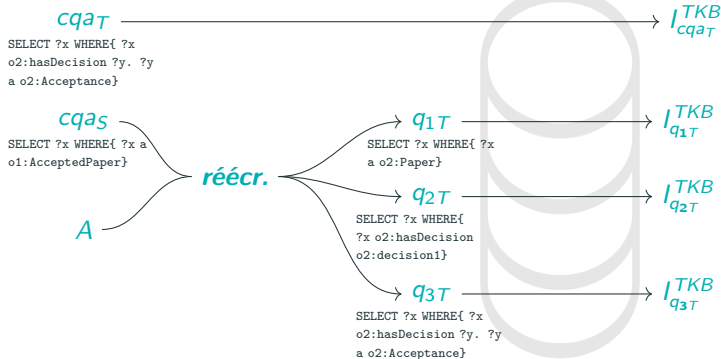
```
SELECT ?x WHERE{  
?x o2:hasDecision  
o2:decision1}
```

*q3T*

```
SELECT ?x WHERE{ ?x  
o2:hasDecision ?y. ?y  
a o2:Acceptance}
```

Réécrire la CQA source avec l'alignement évalué

## Target KB



Récupérer les réponses de la requête réécrite et de la CQA cible

# Couverture de CQA

Contexte

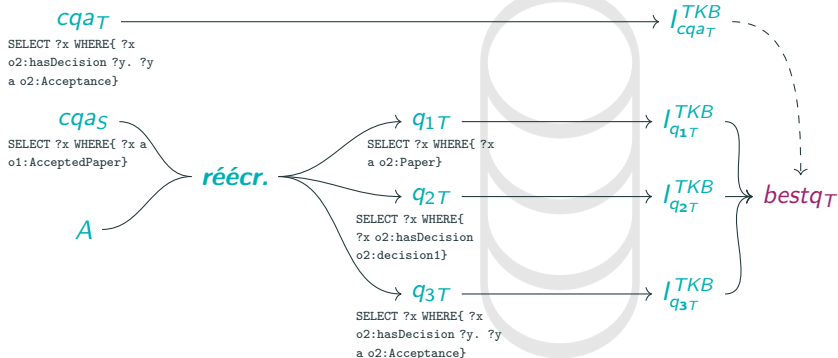
Proposition

Jeu de données

Évaluation

Conclusion

Target KB



$$bestq_T = \underset{q_T \in \text{rewrite}(cqas, A)}{\operatorname{argmax}} \quad queryFmeasure(I_{cqa_T}^{TKB}, I_{q_T}^{TKB})$$

Sélectionner la requête avec le meilleur score *query Fmeasure*

# Couverture de CQA

Contexte

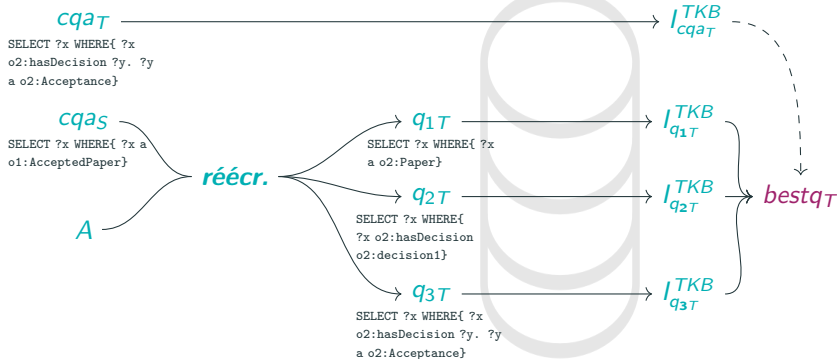
Proposition

Jeu de données

Évaluation

Conclusion

Target KB



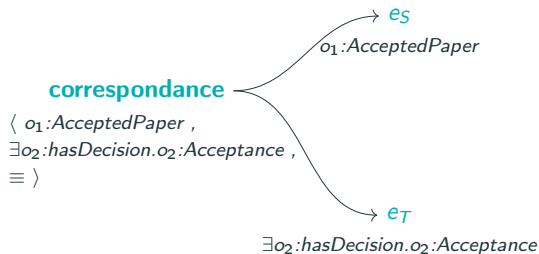
$$coverage(A, cqa_{pairs}, SKB, TKB) = \text{average}_{(cqa_S, cqa_T) \in cqa_{pairs}} bestq_T$$

$$bestq_T = \operatorname{argmax}_{q_T \in \text{rewrite}(cqa_S, A, SKB)} queryFmeasure(I_{cqa_T}^{TKB}, I_{q_T}^{TKB})$$

## correspondance

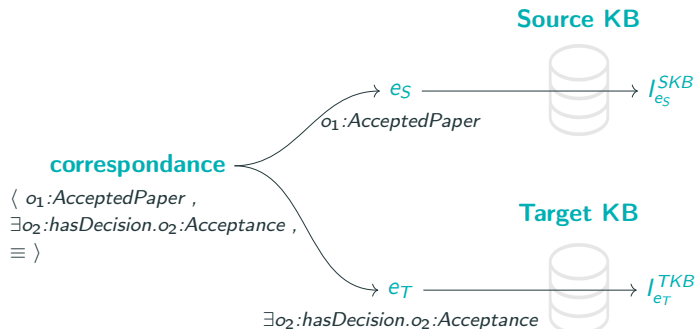
$\langle o_1:AcceptedPaper ,$   
 $\exists o_2:hasDecision.o_2:Acceptance ,$   
 $\equiv \rangle$

Évaluation pour chaque correspondance

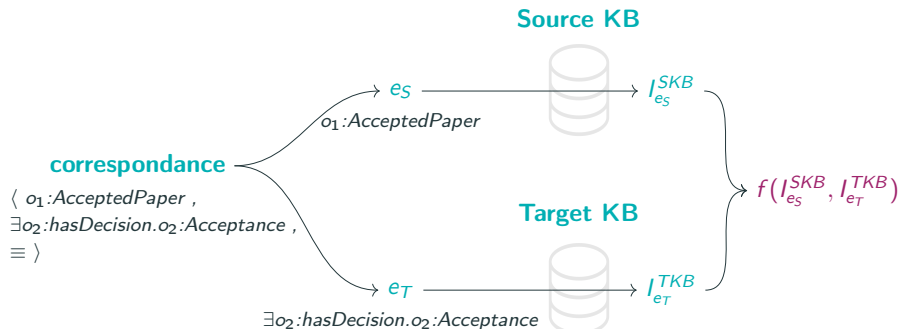


Membres sources et cibles

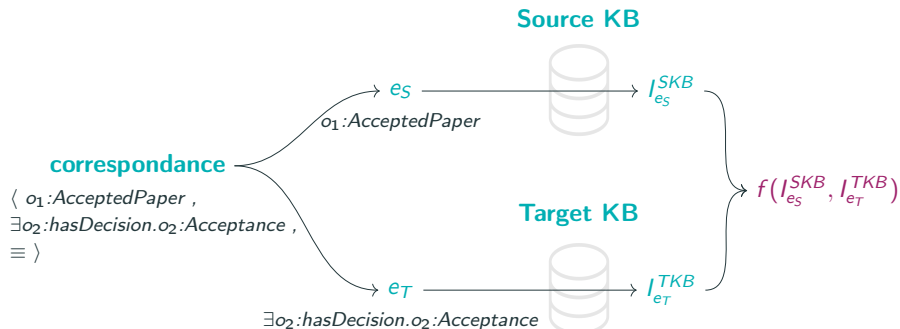




Récupérer les instances décrites par les membres



Comparer les instances et y appliquer un score



$$precision(A, SKB, TKB, f) = \text{average}_{\langle e_S, e_T \rangle \in A} f(I_{e_S}^{SKB}, I_{e_T}^{TKB})$$

# Avantages, Inconvénients et Moyenne Harmonique

Contexte

Proposition

Jeu de données

Évaluation

Conclusion

	Avantages	Inconvénients
Couverture de CQA	Estime la couverture Comparaison d'instances de la BC cible	Bruit introduit par la réécriture Dépend de la qualité des instances Ne prend pas en compte la relation de la correspondance
Précision Intrinsèque	Équilibre la couverture de CQA	Nécessite que les 2 BC aient les mêmes instances Dépend de la qualité des instances Ne prend pas en compte la relation de la correspondance

**Moyenne Harmonique** de la couverture de CQA et de la précision intrinsèque

$$HMean(A, cqa_{pairs}, SKB, TKB, f) = 2 \times \frac{coverage(A, cqa_{pairs}, SKB, TKB) \times precision(A, SKB, TKB, f)}{coverage(A, cqa_{pairs}, SKB, TKB) + precision(A, SKB, TKB, f)}$$

# Jeu de données

---

Jeu de données de l'OAEI sur l'organisation de conférences[Šváb Zamazal et al., 2005] très utilisé et décliné [Zamazal and Svátek, 2017]

Peuplement de 5 ontologies (cmt, conference, confOf, edas, ekaw)

Peuplement basé sur 152 CQA: peuplement équivalent pour les ontologies couvrant la CQA

100 CQA ont été gardées pour l'évaluation

# Évaluation

---

- 2 alignements complexes de référence [Thiéblin et al., 2018]  
**Query rew.** et **Onto.merg.**
- 1 alignement simple de référence **ra1** [Šváb Zamazal et al., 2005]
- 2 alignements issus d'aligneurs complexes  
[Ritze et al., 2010, Faria et al., 2018]



# Évaluation d'alignements existants

Contexte

Proposition

Jeu de données

Évaluation

Conclusion

	Query rew.	Onto. merg.	ra1	Faria 2018	Ritze 2010
C. CQA	<b>0.69</b>	0.63	0.42	0.41	0.48
P. équi.	0.42	0.43	0.56	0.65	0.75
P. inter.	0.90	0.86	<b>0.92</b>	0.71	0.75
P. non-dis.	0.94	0.91	<b>0.96</b>	0.71	0.75
MH inter.	<b>0.78</b>	0.73	0.58	0.52	0.59
MH non-dis.	<b>0.80</b>	0.74	0.58	0.52	0.59

# Conclusion

---

- Un système d'évaluation **automatique** d'alignements complexes
- Évaluation basée sur les instances
- Score de **Couverture** basé sur les **CQA**
- Précision Intrinsèque équilibre la Couverture
- **Jeu de données** adapté au système d'évaluation
- Création d'une tâche dédiée dans l'OAEI

- Réflexion sur les systèmes de réécriture de requête
- Exploration d'approches plus sémantiques

**Merci pour votre attention !**

[elodie@thieblin.fr](mailto:elodie@thieblin.fr)



Faria, D., Pesquita, C., Balasubramani, B. S., Tervo, T., Carriço, D., Garrilha, R., Couto, F. M., and Cruz, I. F. (2018).

**Results of AML participation in OAEI 2018.**

In Shvaiko, P., Euzenat, J., Jiménez-Ruiz, E., Cheatham, M., and Hassanzadeh, O., editors, *Proceedings of the 13th International Workshop on Ontology Matching co-located with the 17th International Semantic Web Conference, OM@ISWC 2018, Monterey, CA, USA, October 8, 2018*, volume 2288 of *CEUR Workshop Proceedings*, pages 125–131. CEUR-WS.org.



Meilicke, C. and Stuckenschmidt, H. (2008).

**Incoherence as a basis for measuring the quality of ontology mappings.**

In *Proceedings of the 3rd International Conference on Ontology Matching-Volume 431*, pages 1–12. CEUR-WS. org.



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

**Linguistic analysis for complex ontology matching.**

In Shvaiko, P., Euzenat, J., Giunchiglia, F., Stuckenschmidt, H., Mao, M., and Cruz, I. F., editors, *Proceedings of the 5th International Workshop on Ontology Matching (OM-2010), Shanghai, China, November 7, 2010*, volume 689 of *CEUR Workshop Proceedings*. CEUR-WS.org.



Solimando, A., Jiménez-Ruiz, E., and Guerrini, G. (2014).

**Detecting and correcting conservativity principle violations in ontology-to-ontology mappings.**

In Mika, P., Tudorache, T., Bernstein, A., Welty, C., Knoblock, C. A., Vrandečić, D., Groth, P. T., Noy, N. F., Janowicz, K., and Goble, C. A., editors, *The Semantic Web - ISWC 2014 - 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part II*, volume 8797 of *Lecture Notes in Computer Science*, pages 1–16. Springer.



Thiéblin, E., Cheatham, M., Trojahn, C., Zamazal, O., and Zhou, L. (2018).

**The first version of the OAEI complex alignment benchmark.**

In van Erp, M., Atre, M., López, V., Srinivas, K., and Fortuna, C., editors, *Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks co-located with 17th International Semantic Web Conference (ISWC 2018), Monterey, USA, October 8th - to - 12th, 2018*, volume 2180 of *CEUR Workshop Proceedings*. CEUR-WS.org.



Walshe, B., Brennan, R., and O'Sullivan, D. (2016).

**Bayes-recce: A bayesian model for detecting restriction class correspondences in linked open data knowledge bases.**

*Int. J. Semant. Web Inf. Syst.*, 12(2):25–52.





Zamazal, O. and Svátek, V. (2017).

**The Ten-Year OntoFarm and its Fertilization within the Onto-Sphere.**

*Web Semantics: Science, Services and Agents on the World Wide Web*,  
43:46–53.



Šváb Zamazal, O., Svátek, V., Berka, P., Rak, D., and Tomášek, P.  
(2005).

**Ontofarm: Towards an experimental collection of parallel ontologies.**

*Poster Track of ISWC, 2005.*