

# Semantic Interactive Ontology Matching: Synergistic Combination of Techniques to Improve the Set of Candidate Correspondences

Jomar da Silva, Fernanda Baião, **Kate Revoredo**, Jérôme Euzenat

Graduate Program in Informatics, UNIRIO, Brazil

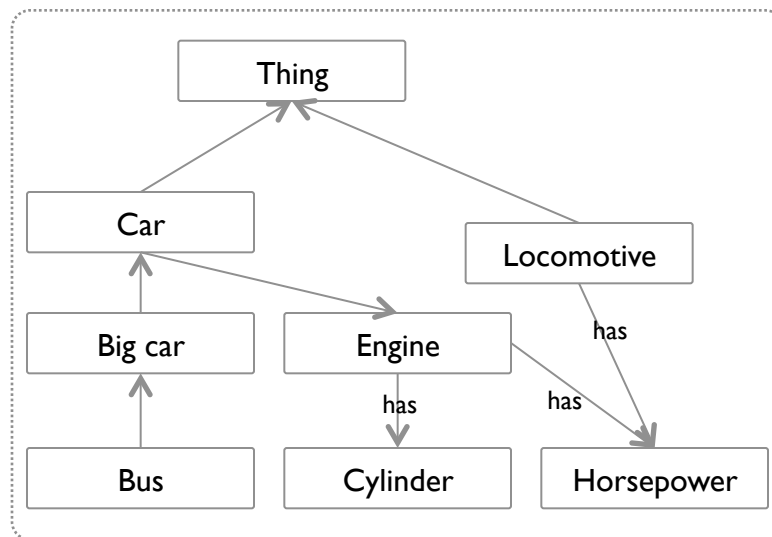
Laboratoire d'informatique de Grenoble - INRIA, GrenobleRhône-Alpes, France



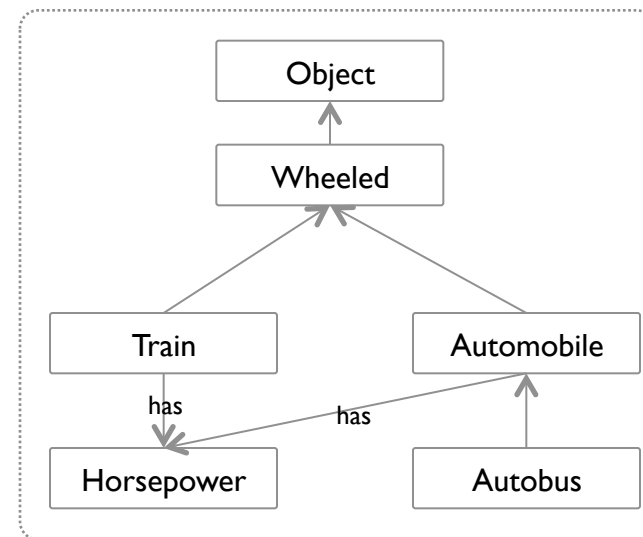
# Introduction

## Data Integration and systems interoperability problems

- Several ontology artifacts for the same universe of discourse.
- Differences in several perspectives: terminological, structural, semantic, ...



Ontology  $O_1$   
 $O_1 = \langle C, R, P, I, A \rangle$



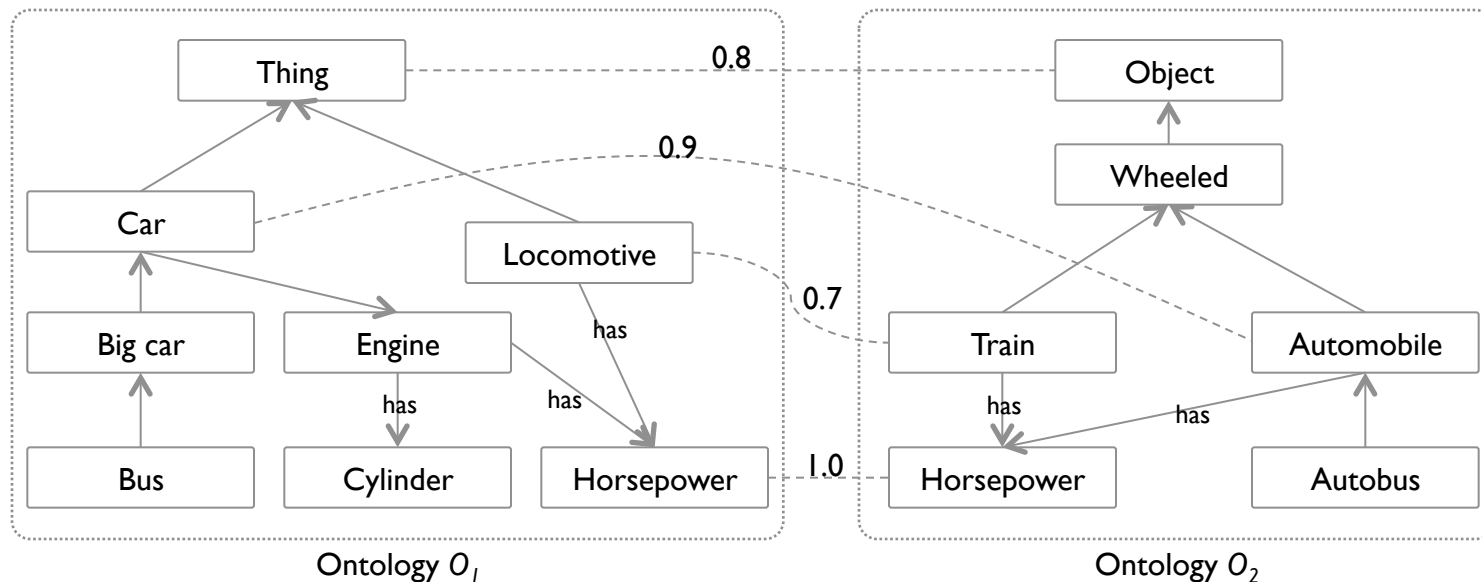
Ontology  $O_2$   
 $O_2 = \langle C', R', P', I', A' \rangle$

[Source: [www.webontology.org](http://www.webontology.org)]

# Motivation

Given 2 ontologies,  $O_1$  and  $O_2$ , Ontology Matching Process searches for correspondences between their entities

►  $\langle e, e', r, n \rangle$



The set of correspondences defines an alignment  $\bar{A}$

$$\bar{A} = \{ \langle O_1.\text{Thing}, O_2.\text{Object}, \equiv, 0.8 \rangle, \langle O_1.\text{Car}, O_2.\text{Automobile}, \equiv, 0.9 \rangle, \langle O_1.\text{Locomotive}, O_2.\text{Train}, \equiv, 0.7 \rangle, \langle O_1.\text{Horsepower}, O_2.\text{Horsepower}, \equiv, 1.0 \rangle \}$$

# Introduction

---

The Ontology Matching Process can be done

- Manually
- Semi-automatically
- Automatically

# Introduction

---

## The Ontology Matching Process can be done

- Manually
- Semi-automatically
  - Interactive strategy considering knowledge from domain experts
  - Improves the quality of the alignment in 30% when compared to automatic approaches.
- Automatically

# Introduction

---

In Interactive Ontology Matching, experts can provide feedback to a correspondence indicating whether or not it belongs to the alignment.

## **Issue: which correspondences should be shown to the expert?**

- Bad selection may cause an alignment
  - Imprecise
  - Incomplete
  - With slow convergence

# Problem

---

How to define an adequate set of correspondences to receive feedback from the expert?

# Goal

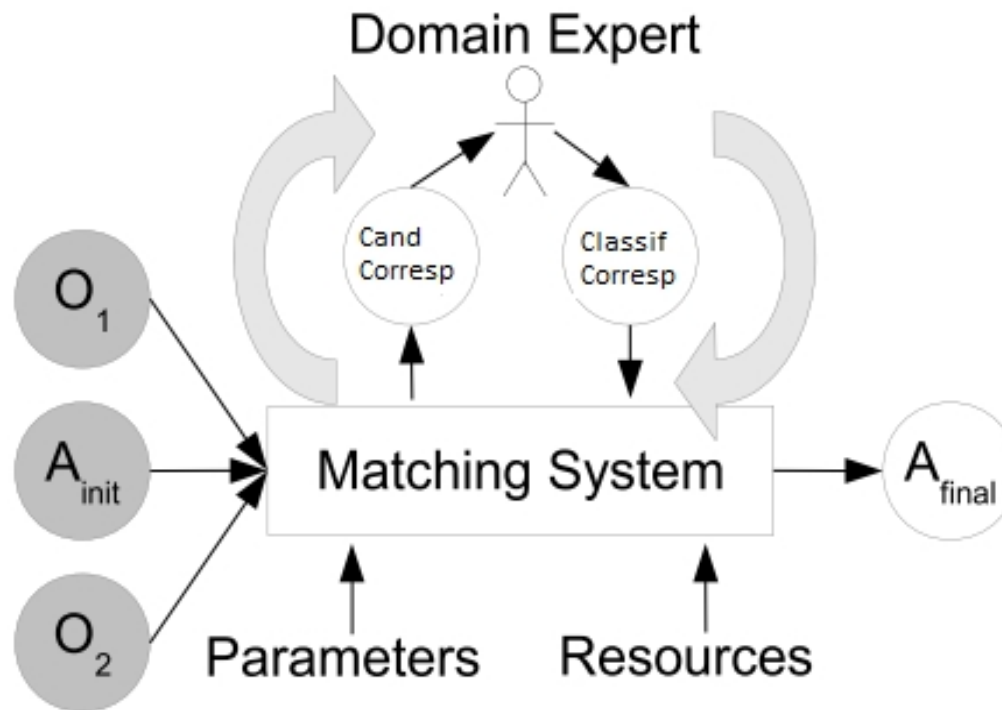
---

## The approach must guarantee

- fast convergence
  - discover correspondences with high chance of not being in the alignment and do not ask for the expert feedback
- quality of the alignment
  - good precision and recall



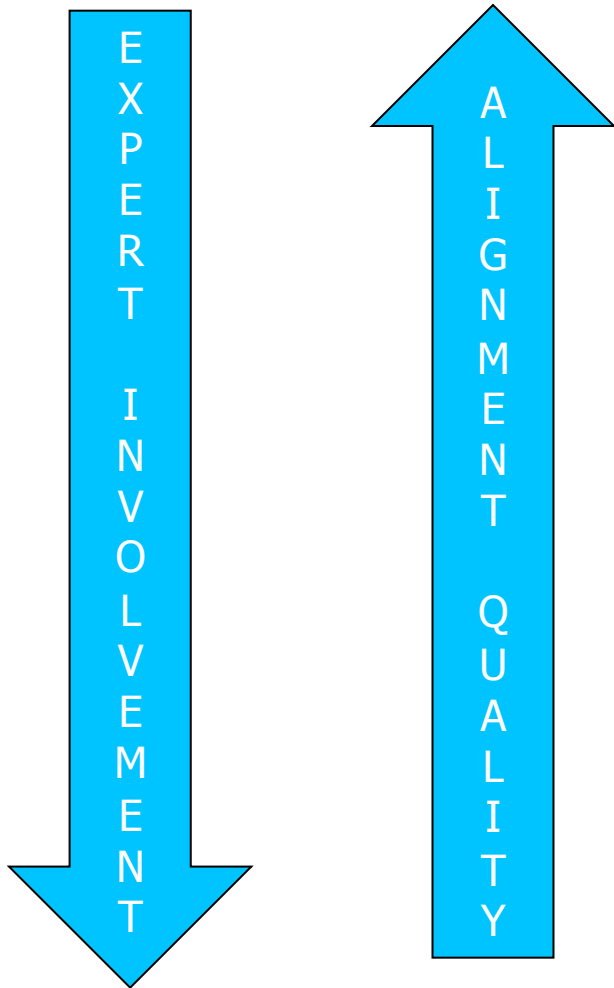
# Interactive Ontology Matching



- Candidate correspondence
  - possible correspondence selected to receive expert feedback.
- Classified correspondence
  - correspondence that received feedback from the expert

# Interactive Ontology Matching

---

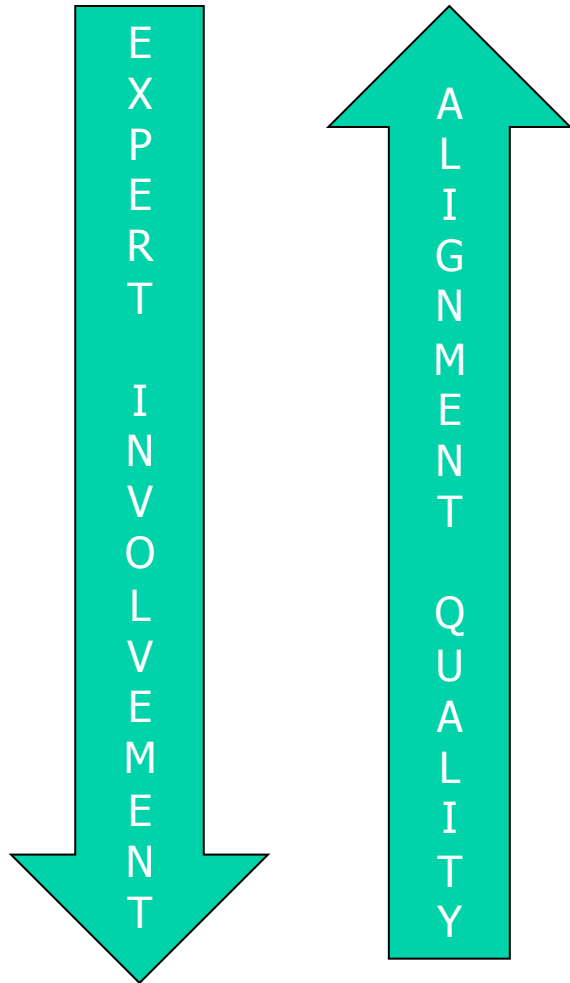


The approach should improve a quality measure...

... without increasing expert involvement in an accentuated way

# Interactive Ontology Matching Proposal

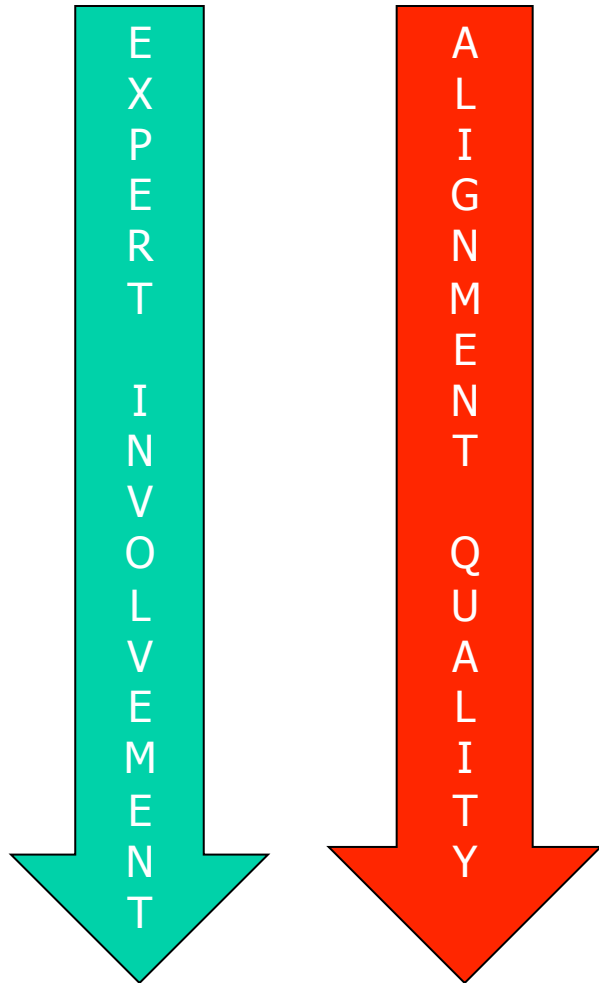
---



Combine two techniques to improve the alignment quality, without increasing expert involvement

# Interactive Ontology Matching Proposal

---

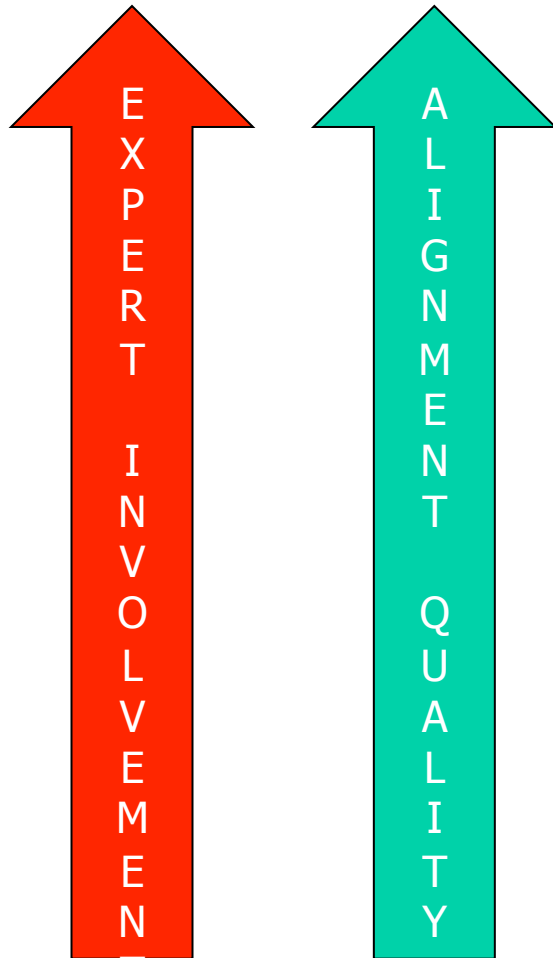


Combine two techniques to improve the alignment quality, without increasing expert involvement

- Semantic technique

# Interactive Ontology Matching Proposal

---

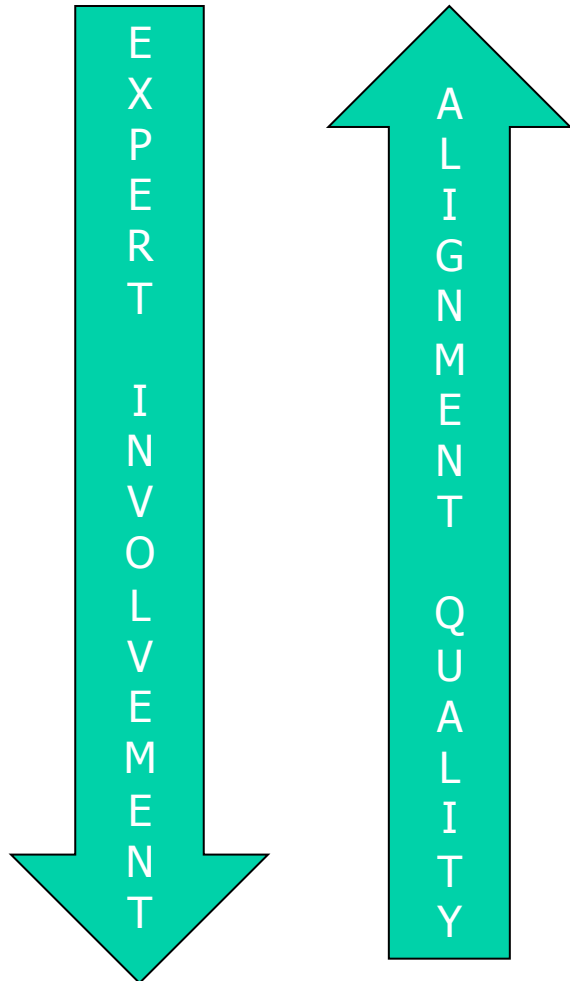


Combine two techniques to  
improve the alignment quality,  
without increasing expert  
involvement

- Structural technique

# Interactive Ontology Matching Proposal

---



Combine two techniques to improve the alignment quality, without increasing expert involvement

- Semantic technique
- +
- Structural technique

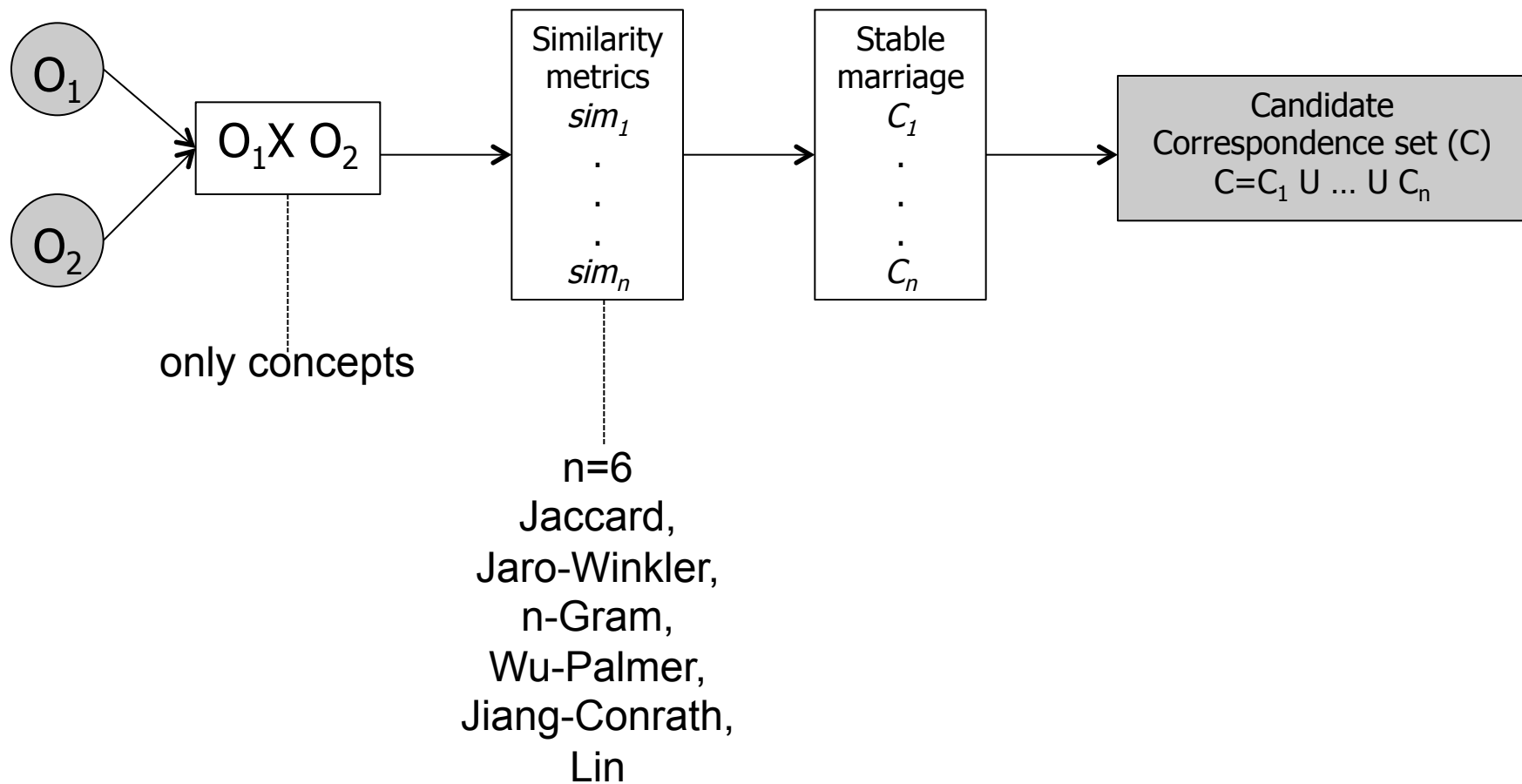
# Interactive Ontology Matching Proposal

---

ALINSyn algorithm, which has 3 main steps

- Generate initial candidate correspondence set
- Apply semantic technique
- Apply structural technique

# Generate initial candidate correspondence set





# Apply Semantic Technique

---

Candidate correspondences with **semantically different entity names** are removed from C and stored in a set T, called temporarily suspended correspondence set.

- Entity name = entity head noun
- Head noun = a noun that all other terms depend on it
- Semantically different = entities head nouns not in the same synset of the Wordnet

# Apply Semantic Technique Example

ID	O1.e	O2.e	C	After semantic step
4	Author	Regular_author	Candidate correspondence	Candidate correspondence
14	Chairman	Chair	Candidate correspondence	Temporarily suspended correspondence
18	Co_author	Regular_author	Candidate correspondence	Temporarily suspended correspondence
23	Paper	Paper	Candidate correspondence	Candidate correspondence
31	Paper_Abstract	Abstract	Candidate correspondence	Candidate correspondence
36	Person	Person	Candidate correspondence	Candidate correspondence
45	Subject_Area	Abstract	Candidate correspondence	Temporarily suspended correspondence
48	Subject_Area	Program_Committee	Candidate correspondence	Temporarily suspended correspondence

Same synset of the Wordnet

Different synsets of the Wordnet

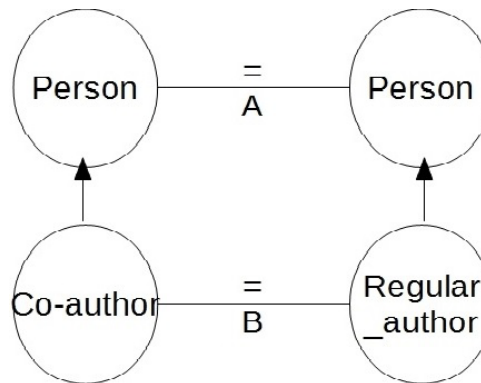
$C = \{4, 23, 31, 36\}$   
 $T = \{14, 18, 45, 48\}$

# Apply Structural Technique

---

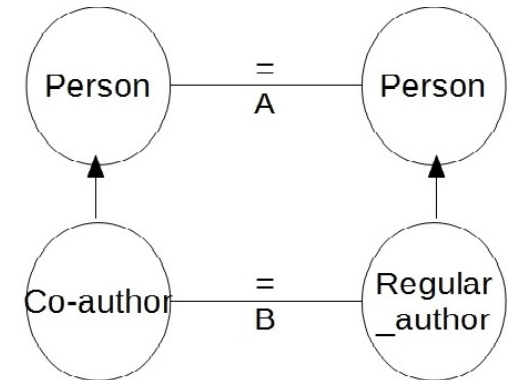
Correspondences in set T can be recovered based on expert feedback and structural properties.

- Structural property = entities of a correspondence in T that are sub-classes of the entities of a correspondence classified as true by the expert



# Apply Structural Technique Example

ID	O1.e	O2.e	C	After semantic step	Feedback expert	After structural step
4	Author	Regular_author	Candidate correspondence	Temporarily suspended correspondence		Candidate correspondence
14	Chairman	Chair	Candidate correspondence	Candidate correspondence		Temporarily suspended correspondence
18	Co_author	Regular_author	Candidate correspondence	Temporarily suspended correspondence		Candidate correspondence
23	Paper	Paper	Candidate correspondence	Candidate correspondence	TRUE	TRUE
31	Paper_Abstract	Abstract	Candidate correspondence	Temporarily suspended correspondence		Candidate correspondence
36	Person	Person	Candidate correspondence	Candidate correspondence	TRUE	TRUE
45	Subject_Area	Abstract	Candidate correspondence	Temporarily suspended correspondence		Temporarily suspended correspondence
48	Subject_Area	Program_Committee	Candidate correspondence	Temporarily suspended correspondence		Temporarily suspended correspondence



# ALINSyn

---

---

## Algorithm 4 ALINSyn

---

**Input:** Two ontologies to be aligned

**Output:** Alignment between the two ontologies

- 1: Run candidate correspondence generation (Algorithm 1)
  - 2: Run semantic step (Algorithm 3)
  - 3: **for** Each candidate correspondence **do**
  - 4:     Receive feedback (the candidate correspondence is transformed to classified correspondence)
  - 5:     Run structural Step (Algorithm 5 )
  - 6: **end for**
-

# Evaluation

---

RQ: How does each technique behave, wrt the expert involvement and the alignment quality?

Compare ALYNSyn with:

- ALINBasic: neither of the 2 techniques.
- ALINSem: semantic technique only.
- ALINStr: structural technique only.

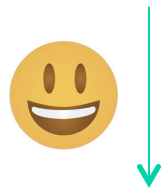
# Evaluation

---

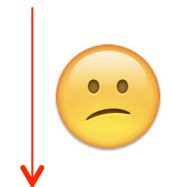
	NI	%	Precision	F-measure	Recall
ALINBasic	619	0,492%	0,92	0,79	0,7
ALINSem	152	0,121%	0,9	0,69	0,57
ALINStr	3539	2,812%	0,93	0,84	0,78
ALINSyn	219	0,174%	0,91	0,75	0,65

# Evaluation

---



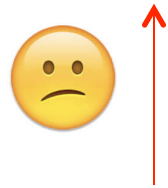
	NI	%	Precision	F-measure	Recall
ALINBasic	619	0,492%	0,92	0,79	0,7
ALINSem	152	0,121%	0,9	0,69	0,57
ALINStr	3539	2,812%	0,93	0,84	0,78
ALINSyn	219	0,174%	0,91	0,75	0,65



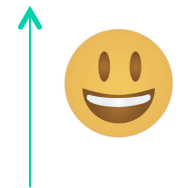


# Evaluation

---

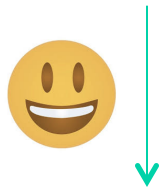


	NI	%	Precision	F-measure	Recall
ALINBasic	<b>619</b>	0,492%	0,92	0,79	<b>0,7</b>
ALINSem	152	0,121%	0,9	0,69	0,57
ALINStr	<b>3539</b>	2,812%	0,93	0,84	<b>0,78</b>
ALINSyn	219	0,174%	0,91	0,75	0,65

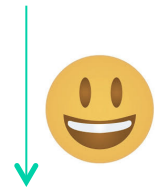


# Evaluation

---



	NI	%	Precision	F-measure	Recall
ALINBasic	<b>619</b>	0,492%	0,92	0,79	<b>0,7</b>
ALINSem	152	0,121%	0,9	0,69	0,57
ALINStr	3539	2,812%	0,93	0,84	0,78
ALINSyn	<b>219</b>	0,174%	0,91	0,75	<b>0,65</b>



# Evaluation

---

RQ: How does ALINSyn compare to the state of the art proposals, wrt the expert involvement and the alignment quality?

- OAEI Conference dataset
- 100% correctness from expert
- Compare with the interactive systems from OAEI 2016

# Evaluation

---

	Number of questions	NI	%	Precision	F-measure	Recall
AML	270	271	0,275	0,912	0,799	0,711
ALINSyn	483	<b>219</b>	0,174	<b>0,915</b>	<b>0,754</b>	<b>0,652</b>
LogMap	142	142	0,113	0,886	0,723	0,61
Xmap	4	4	0,003	0,837	0,681	0,574

# Conclusion

---

## ALINSyn combines

- A semantic approach to reduce the expert involvement
- A structural approach to improve the alignment quality

ALINSyn showed its potential compared to state of the art systems using OAEI 2016 results

# Future Work

---

## Other semantic techniques

- Entity name disambiguation
- Enrich entity meaning with contextual information

## Other structural techniques

- Anti-patterns

## Other entity types

- attributes, relationships

## Deal with expert error

# Semantic Interactive Ontology Matching: Synergistic Combination of Techniques to Improve the Set of Candidate Correspondences

Jomar da Silva, Fernanda Baião, **Kate Revoredo**, Jérôme Euzenat

Graduate Program in Informatics, UNIRIO, Brazil

Laboratoire d'informatique de Grenoble - INRIA, GrenobleRhône-Alpes, France

