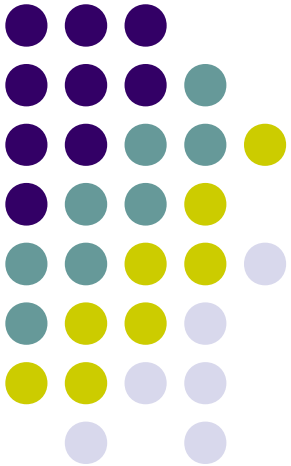
Ontology Matching and Evaluation

Cássia Trojahn dos Santos
EXMO Group - INRIA Rhône Alpes & LIG
Grenoble, France
cassia.trojahn@inrialpes.fr



INRIA

- Institut National de Recherche en Informatique et en Automatique

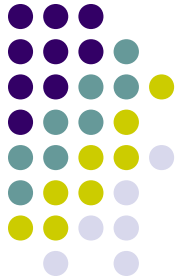
30 groups of research
~210 researchers
~210 engineers and administratives
~240 docs and post docs (45% étrangers)

- EXMO Group: Computer mediated exchange of structured knowledge
  - Leader: Jérôme Euzenat
  - http://www.inrialpes.fr/exmo/

Outline

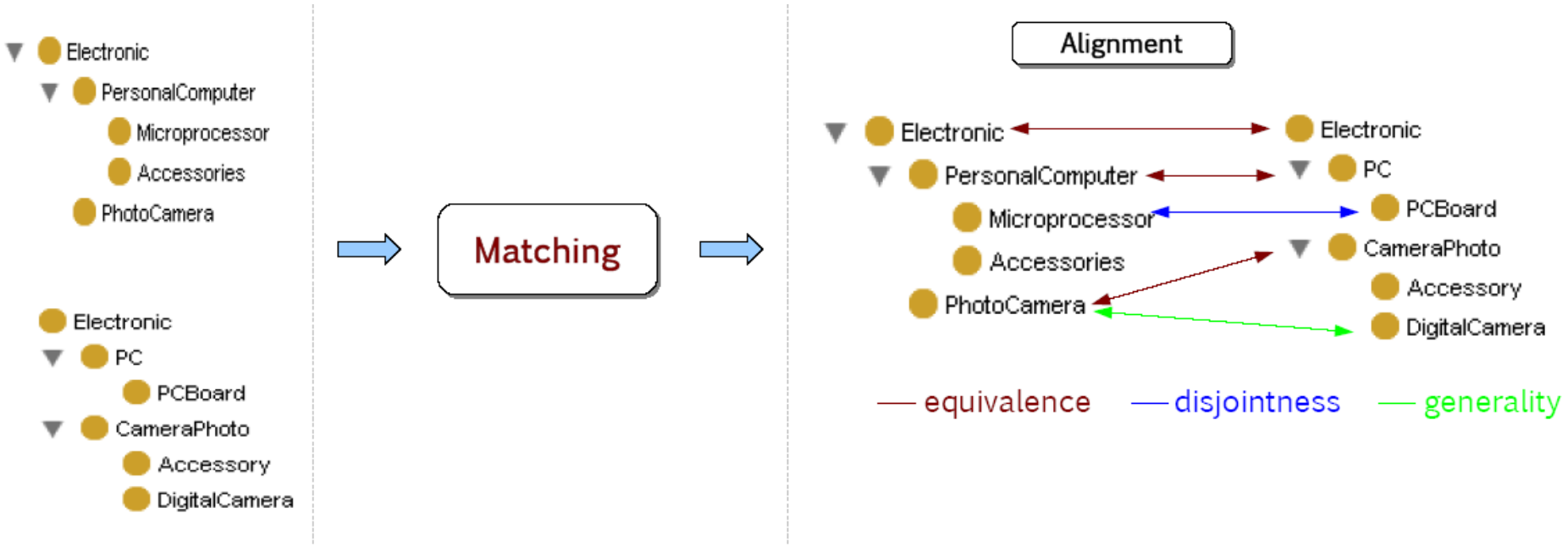
- What is ontology matching
- Matching approaches and strategies
- Ontology Alignment Evaluation Initiative (OAEI)
- SEALS project
- Alignment API
- 10 challenges for ontology matching

Context



- Ontology
  - a formal, explicit specification of a shared conceptualization (Gruber, 1993)
  - key role in knowledge based systems
- Problem
  - semantic heterogeneity between ontology-based systems
- Solution
  - ontology matching

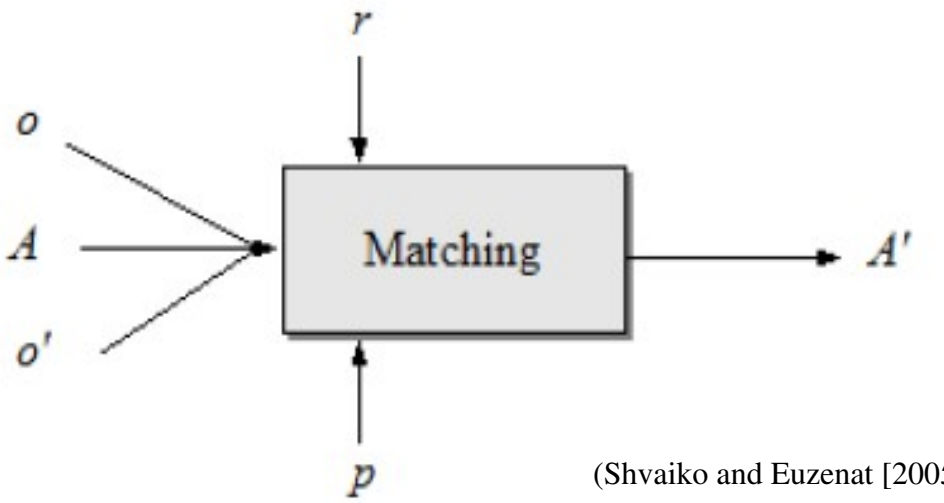
Ontology matching



Traditional application domains:
- Schema integration
- Data warehouse
- Mediator generator

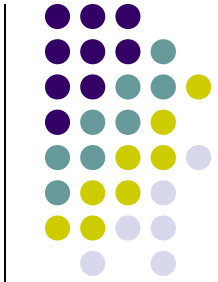
Emergent
- P2P databases
- Agent communication
- Web services integration

Matching process



(Shvaiko and Euzenat [2005])

Correspondence

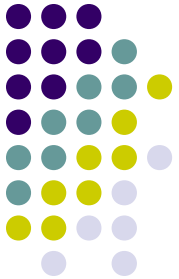


Definition (Correspondence)

Given two ontologies *o* and *o'* , a **correspondence** between *o* and *o'* is a 5-tuple <id, e, e',e, n> where

- ***id*** is na identifier of the correspondence
- ***e*** and ***e'*** are entities of o and o'
- ***r*** is a relation (equivalence, more general, disjointness)
- ***n*** is a confidence measure (usually in the [0,1] range)

Alignment

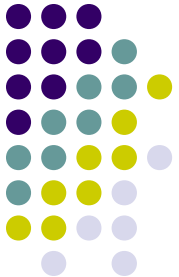


Definition (Alignment)

Given two ontologies *o* and *o'* , an **alignment** between *o* and *o'* is

- a set of correspondences on *o* and *o'*
- includes some metadata (multiplicity: 1-1, 1-*: method, data, etc)

Alignment



```
<owl:Class rdf:ID="MastersThesis">
   <rdfs:subClassOf rdf:resource="#Academic" />
</owl:Class>

<owl:Class rdf:ID="PhdThesis">
   <rdfs:subClassOf rdf:resource="#Academic" />
</owl:Class>

<owl:Class rdf:ID="Misc">
   <rdfs:subClassOf rdf:resource="#Reference" />
</owl:Class>
```

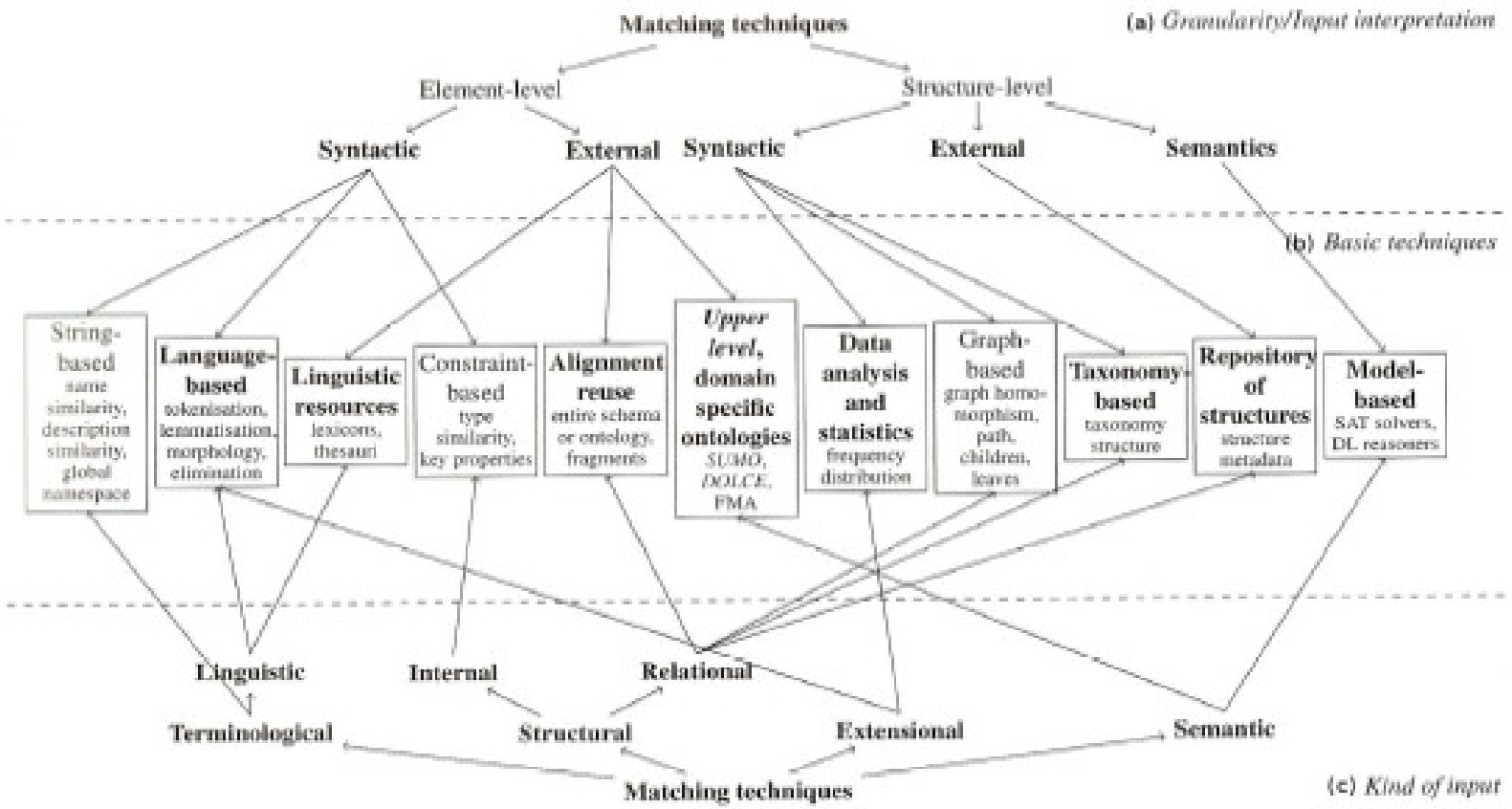
```
<owl:Class rdf:ID="MémoireDeMastère">
   <rdfs:subClassOf rdf:resource="#Mémoire"/>
</owl:Class>

<owl:Class rdf:ID="MémoireDeDoctorat">
   <rdfs:subClassOf rdf:resource="#Mémoire"/>
</owl:Class>

<owl:Class rdf:ID="Divers">
   <rdfs:subClassOf rdf:resource="#Référence"/>
</owl:Class>
```

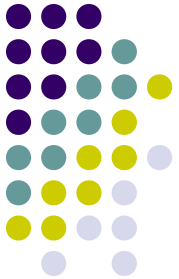
```
<map>
    <Cell>
      <entity1 rdf:resource="http://oaei.ontologymatching.org/2009/benchmarks/206/onto.rdf#MémoireDeMastère"/>
      <entity2 rdf:resource="http://oaei.ontologymatching.org/2009/benchmarks/206_pt/onto.rdf#TeseDeMestrado"/>
      <measure rdf:datatype="http://www.w3.org/2001/XMLSchema#float">1.0</measure>
      <relation>=</relation>
    </Cell>
  </map>
```

Matching approaches



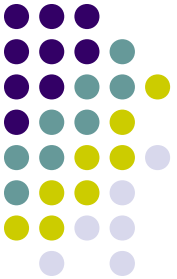
(Euzenat and Shvaiko [2007])

Matching approaches



- Representative categories
  - Syntactic/lexical (string-based similarity)
  - Semantic (WordNet based)
  - Structural (positions of the terms in the ontology hierarchy)

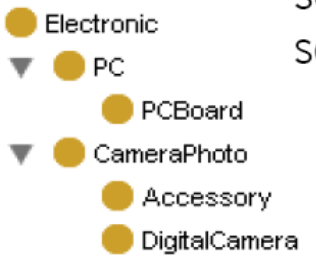
Syntactic and semantic



- Syntactic/lexical matcher (string-based similarity)
  - e = "**photo-camera**" and e' = "**camera-photo**"
    - **exactMatch**, with strength = matches/max length = 2 / 2 = 1
  - e = "**science**" and e' = "**computer-science**"
    - **broadMatch**, with strength = matches/max length = 1 / 2 = 0.5
- Semantic matcher(WordNet based)
  - e = "**personal-computer**" and e' = "**pc**"
    - **exactMatch**, with strength = 1, direct synonymous WordNet

Structural matcher

- Structural matcher
  - Based on taxonomy overlap and semantic cotopy (Maedche and Staab, 2002)



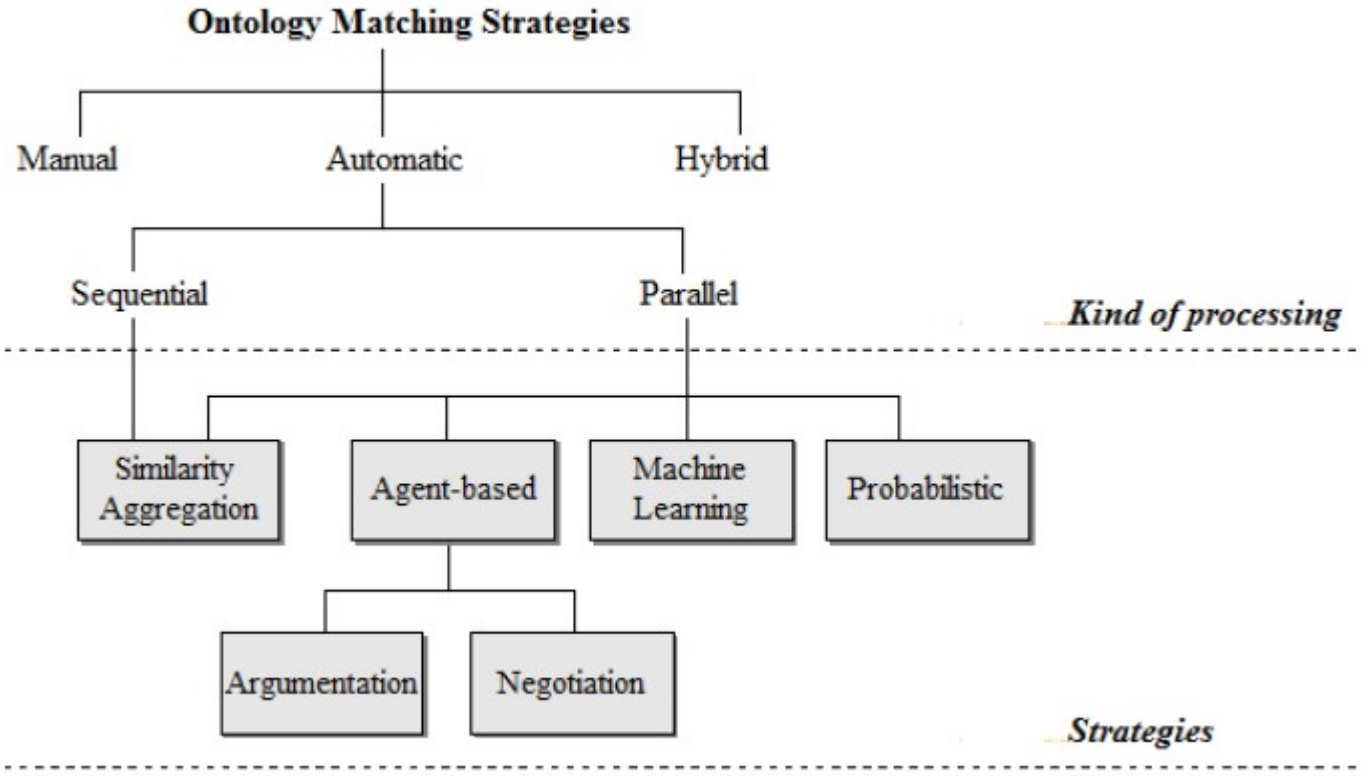
SC(PersonalComputer,O_s) = {Electronic, PersonalComputer}

SC(PC,O_t) = {Electronic, PC}

$$TO_n(e_s, e_t, O_s, O_t) = \frac{|\ Electronic, PersonalComputer \cap Electronic, PC\ |}{|\ Electronic, PersonalComputer \cup Electronic, PC\ |}$$

= 2 /2 = exactMatch

Matching strategies



(Trojahn, 2008)

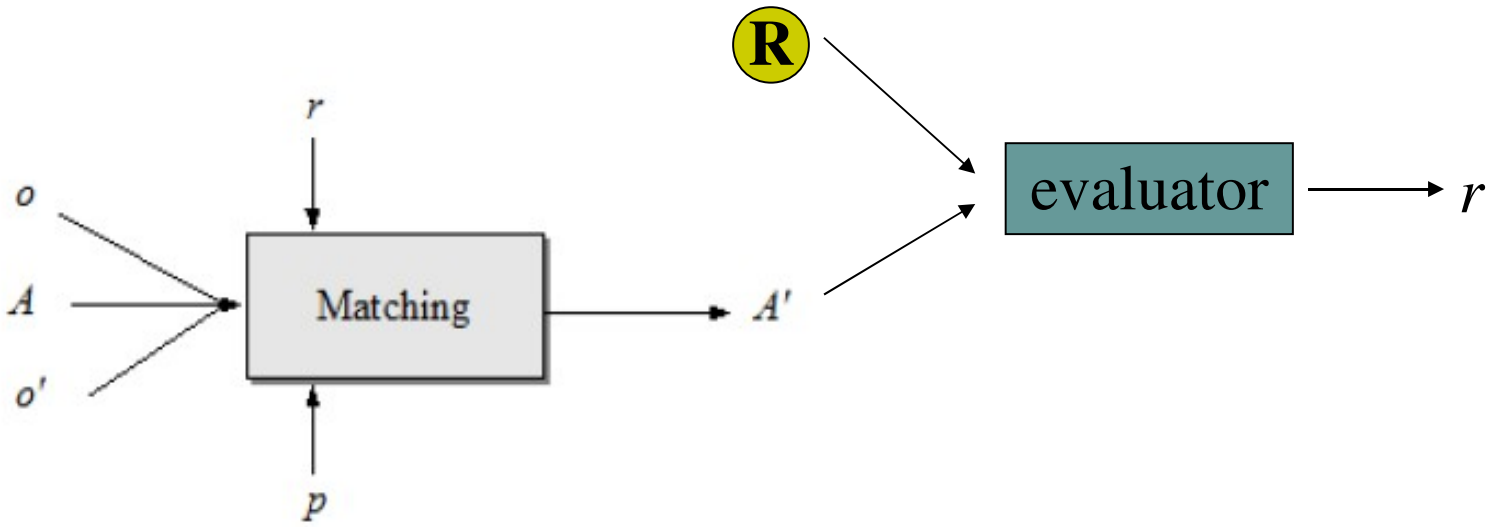
Goal of evaluation

- Improve the performance of systems
- Comparison of systems
- Various sets of tests and criterion
- Created the **Ontology Alignment Evaluation Initiative** (OAEI)

OAEI

- Organization of a yearly evaluation event
- Different domains of test data (complexity, size, ...)
- Participants submit their alignments in a standard format
- These are compared with available reference alignments
- Deviation is measured by classical measures, such as precision and recall
- Results are published on the web site and OM Workshop
- http://oaei.ontologymatching.org/

OAEI



OAEI metrics

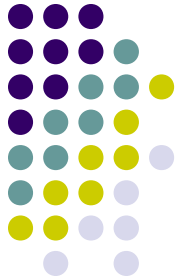
Definition (Precision and Recall)

Given a reference alignment **R**

- **precision** of some aligment **A** is given by $P(A,R) = \frac{|R \cap A|}{|A|}$
- **recall** is given by $R(A,R) = \frac{|R \cap A|}{|R|}.$

When no reference alignment is provided: consensus, task-oriented,..

Extended precision and recall



- **Problems with classical P and R**
  - Do not make difference between a nearly good alignment and a bad one
  - P and R do not recognise two equivalent alignments
  - **If they are not the same exact correspondence they score zero**
  - **How to know if an alignment is closer to the expected?**
- **Solution**
  - **Measuring the "proximity" of alignments: generalizing precision and recall**

Extended precision and recall

A_2 – reasonable

<o1:Car,o2:Thing,=,1.0>
<o1:hasSpeed,o2:hasProperty,=,1.0>
<o1:MotorKA1,o2:MarcsPorsche,=,1.0>
<o1:250kmh,o2:fast,=,1.0>

A_3 – wrong

<o1:Object,o2:Thing,=,1.0>
<o1:Owner,o2:Volkswagen,=,1.0>
<o1:Boat,o2:Porsche,=,1.0>
<o1:hasOwner,o2:hasMotor,=,1.0>
<o1:Marc,o2:fast,=,1.0>

R

<o1:Object,o2:Thing,=,1.0>
<o1:Car,o2:Automobile,=,1.0>
<o1:Speed,o2:Characteristic,=,1.0>
<o1:250kmh,o2:fast,=,1.0>
<o1:PorscheKA123,o2:MarcsPorsche,=,1.0>

Relaxed P and R [Ehrig and Euzenat 2005]

- Generalises classical P and R by using a **proximity** function ω instead of |A ∩ R|
- Three concrete extensions proposed.

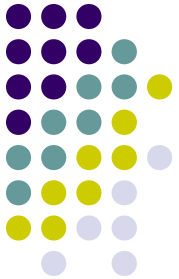
Definition (Relaxed precision and recall)

Given a reference alignment R and an overlap function ω between alignments

$$P_\omega(A,R) = \frac{\omega(R \cap A)}{|A|}$$

$$R_\omega(A,R) = \frac{\omega(R \cap A)}{|R|}$$

Relaxed P and R [Ehrig and Euzenat 2005]



$$\omega(A,R) = \sum_{<a,r>\in M(A,R)} \sigma(a,r)$$

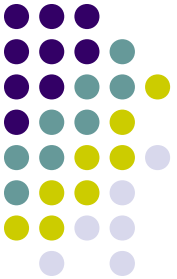
Symmetric

$\sigma_{pair}(<e_a,e'_a>,<e_r,e'_r>)$ distance 0 at class, 0.5 subclasse, 1 others

$\sigma_{rel}(r_a,r_r)$ 1 correct relations in both r_a and r_r, 0.5 if found equivalence but correct is subsumption

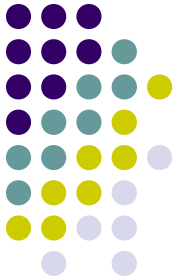
$\sigma_{conf}(n_a,n_r)$ complement of the difference

Relaxed P and R [Ehrig and Euzenat 2005]



ω	(R,R)		(R,A_1)		(R,A_2)		(R,A_3)	
	P	R	P	R	P	R	P	R
standard	1.0	1.0	0.2	0.2	0.25	0.2	0.2	0.2
symmetric	1.0	1.0	0.4	0.4	0.375	0.3	0.2	0.2
edit	1.0	1.0	0.44	0.44	0.35	0.28	0.2	0.2
oriented	1.0	1.0	0.5	0.5	0.375	0.4	0.2	0.2

Semantic P and R [Euzenat 2007]



- Previous solution: syntatic
- P and R semantically grounded
  - correspondences that are consequences of the evaluated alignment: recalled
  - correspondences that are consequences of the reference alignment: correct

Definition (Semantic precision and recall)

Given a reference alignment R, the precision of some alignment A is given by

$$P_{sem}(A,R) = \frac{|A \cap Cn(R)|}{|A|} \text{ and recall is given by } R_{sem}(A,R) = \frac{|Cn(A) \cap R|}{|R|}$$

OAEI 2009 dataset

test	formalism	relations	confidence	modalities	language
benchmarks	OWL	=	[0 1]	open	EN
anatomy	OWL	=	[0 1]	blind	EN
conference	OWL-DL	=, <=	[0 1]	blind+open	EN
fishery	OWL	=	1	expert	EN+FR+ES
directory	OWL	=	1	blind+open	EN
library	SKOS +OWL	exact-,narrow-, broadMatch	1	blind	EN+DU+FR
benchmarksubs	OWL	=,<,>	[0 1]	open	EN
ars	RDF	=	[0 1]	open	EN
tap	RDF	=	[0 1]	open	EN
iimb	RDF	=	[0 1]	open	EN
vlcr	SKOS +OWL	exact-, closeMatch	[0 1]	blind expert	DU+EN

OAEI 2009 participants

Software	confidence	benchmark	anatomy	conference	directory	library	bench-subs	ars	iimb	vlcr	total
aflood		√	√	√	√				√		5
AgrMaker	√	√	√	√							3
AMExt				√							1
AROMA	√	√	√	√							3
ASMOV	√	√	√	√	√		√	√	√		7
DSSim	√	√	√	√	√			√	√	√	7
FBEM								√	√		2
GeRoMe	√	√									1
GG2WW										√	1
HMatch									√	√	2
kosimap	√	√	√	√	√						4
Lily	√	√	√		√						3
MapPSO		√									1
RiMOM	√	√	√				√	√	√		5
SOBOM	√	√	√		√						3
TaxoMap	√	√	√		√	√	√				5
Total=16		12	10	7	7	1	3	5	6	2	53

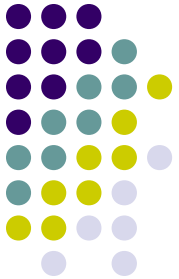
SEALS Project

Semantic Evaluation at Large Scale

http://www.seals-project.eu/

- Scalability and new metrics
- New open platform for semantic technology evaluation
- Automated test infrastructure
- Organize integrated evaluation campaigns
- Online evaluation
- Participants run the evaluation for themselves

Alignment API



- http://alignapi.gforge.inria.fr/

- Reads two OWL/RDF ontologies
- Computes the alignment between these ontologies
- Displays the result (OWL, SWRL, XSTL)
- Evaluate the alignment (precision, recall, f-measure, extended precision and recall)
- Display the evaluation results (graphs, tables)

Alignment API

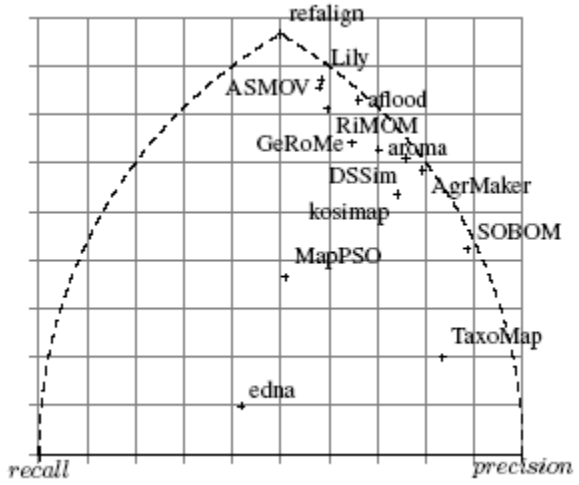
Alignment format

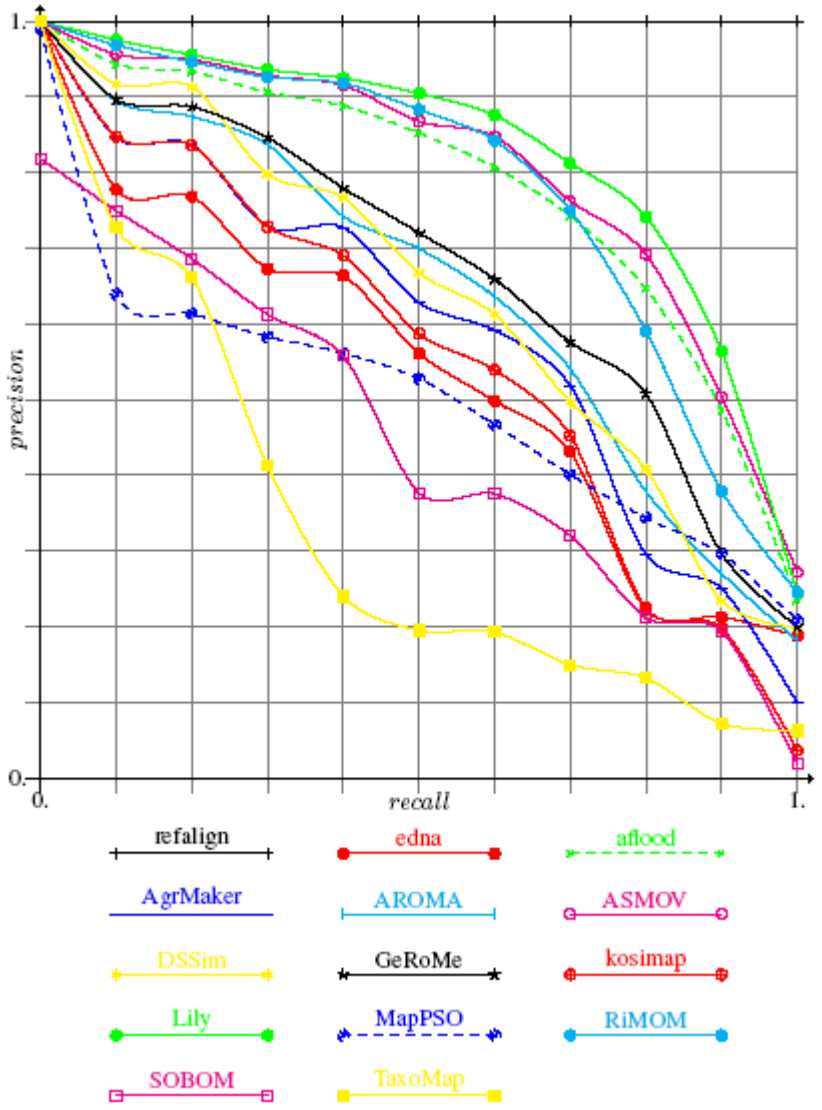
```
- <Alignment>
    <xml>yes</xml>
    <level>0</level>
    <type>11</type>
    <onto1>http://oaei.ontologymatching.org/2009/benchmarks/101/onto.rdf</onto1>
    <onto2>http://oaei.ontologymatching.org/2009/benchmarks/101/onto.rdf</onto2>
    <uri1>http://oaei.ontologymatching.org/2009/benchmarks/101/onto.rdf</uri1>
    <uri2>http://oaei.ontologymatching.org/2009/benchmarks/101/onto.rdf</uri2>
  - <map>
    - <Cell>
        <entity1 rdf:resource="http://oaei.ontologymatching.org/2009/benchmarks/101/onto.rdf#type" />
        <entity2 rdf:resource="http://oaei.ontologymatching.org/2009/benchmarks/101/onto.rdf#type" />
        <measure rdf:datatype="http://www.w3.org/2001/XMLSchema#float">1.0</measure>
        <relation>=</relation>
      </Cell>
    </map>

      ..
  </Alignment>
</rdf:RDF>
```

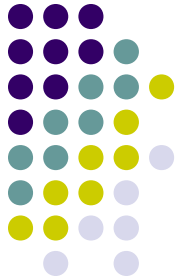
Alignment API

Examples evaluation output





10 Challenges [Shvaiko and Euzenat, 20008]



1. large-scale evaluation

2. performance of ontology-matching techniques

3. discovering missing background knowledge

4. uncertainty in ontology matching

5. matcher selection and self-configuration

6. user involvement

7. explanation of matching results

8. social and collaborative ontology matching

9. alignment management: infrastructure and support

10. reasoning with alignments

Acknowledges

- Thanks to Jérôme Euzenat for sharing some material.
- Thanks to Renata Vieira for the opportunity.
- Thank you for the attention.