# *Semantic Measures and Ontology Matching: an Overview and some Examples*


**Jorge Gracia del Río**

Ontology Engineering Group (OEG)
Departamento de Inteligencia Artificial
Facultad de Informática
Universidad Politécnica de Madrid (UPM), Spain

**2nd December 2009**

# *Outline*

- Introduction
- Semantic Measures
- Ontology Matching
- References

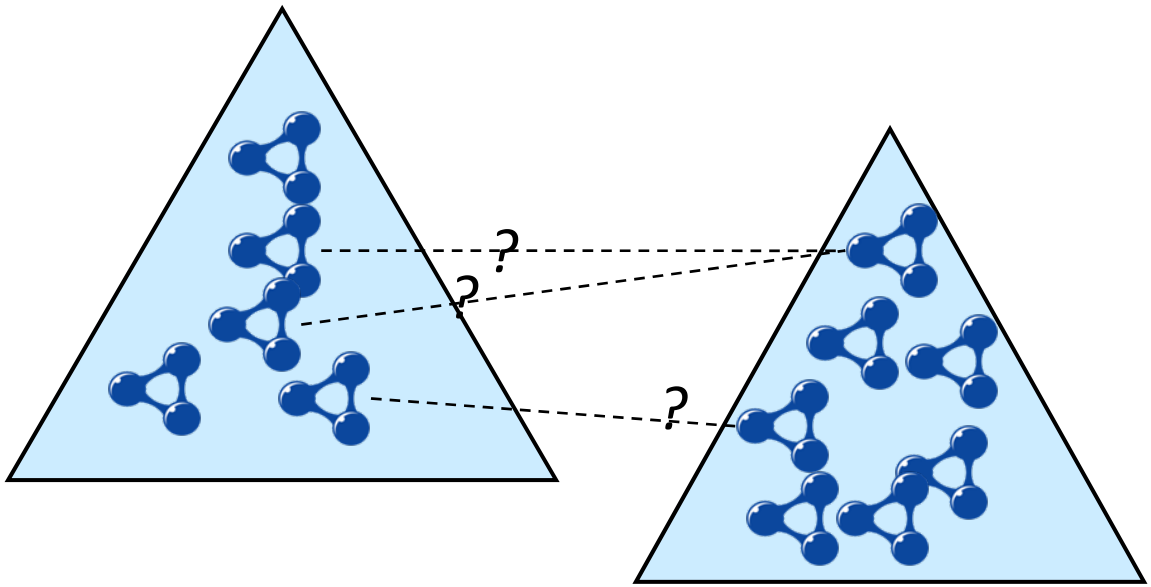
# Introduction

# *Introduction*

- The heterogeneity problem [Euzenat & Shvaiko, 2007]:
  - Resources being expressed in different ways must being reconciled before being used
  - Mismatch between different formalized knowledge may occur when:
    - Different languages are used
    - Different terminologies are used
    - Different modelling is used

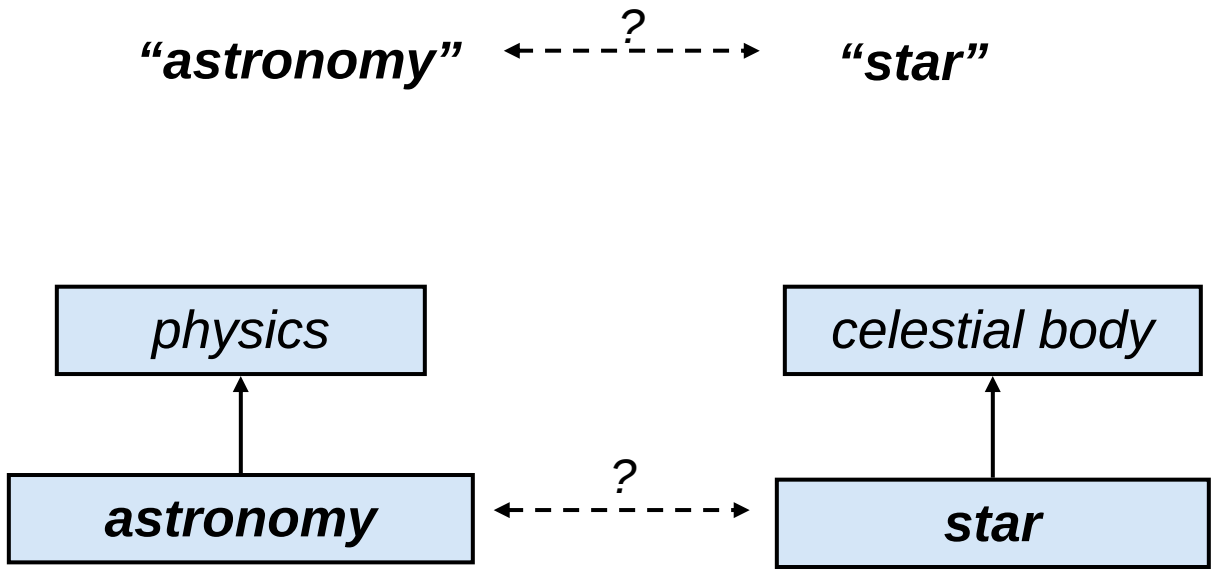
# *Introduction*

- Ontology Matching (OM) is the task of discovering correspondences between terms from different ontologies.



# Introduction

- Semantic Measures
  - Evaluate numerically how semantically related are two entities (words, ontology terms, etc.):



# *Introduction*

- Semantic Measures. Applications:
  - Ontology matching
  - Word sense disambiguation
  - Information retrieval
  - Annotation
  - Automatic indexing
  - Analysis of structure of texts
  - Spelling correction

# Semantic Measures

# *Semantic Measures*

- Semantic similarity
  - Hierarchy-based relationships
  - *doctor* is highly similar to *nurse*
  - *doctor* is not similar to *hospital*
- Semantic relatedness
  - Considers similarity + any other relationship
  - *doctor* is highly related to *hospital*
- Semantic distance
  - Inverse of relatedness

# Semantic Measures

- Semantic relationships

***Exercise: Put in the right places...***

| Name | Example |
|---|---|
| Identity | |
| Hypernymy | |
| Hyponymy | |
| Synonymy | |
| Compatibility | |
| Taxonomical connection | |
| Disjointness | |
| Antonymy | |
| Holonymy | |
| Meronymy | |
| Positive association | |

# Semantic Measures

- Semantic relationships

*Semantic relatedness measure*

| Symbol | Name | Example |
|---|---|---|
| $R_i$ | Identity | (person, person) |
| $R_H$ | Hypernymy | (vehicle, car) |
| $R_h$ | Hyponymy | (hospital, building) |
| $R_{syn}$ | Synonymy | (doctor, physician) |
| $R_{com}$ | Compatibility | (resort, private accommodation) |
| $R_{conn}$ | Taxonomical connection | (hospital, monastery) |
| $R_{dis}$ | Disjointness | (liquid, solid) |
| $R_{ant}$ | Antonymy | (hot, cold) |
| $R_{hol}$ | Holonymy | (keyboard, key) |
| $R_{mer}$ | Meronymy | (finger, hand) |
| $R_{assoc}$ | Positive association | (penguin, Antarctica) |

*Semantic similarity measure*

# *Semantic Measures*

Some definitions [Gracia 09]

- Considering the previous semantic relationships as mathematical binary relations, and given $E$ a set of entities with an associated semantics, we define:
- Semantic relatedness relation: set of ordered pairs

  $R \subseteq E \times E$ such that

$$\mathbf{R = R_i \cup R_H \cup R_h \cup R_{syn} \cup R_{conn} \cup R_{com} \cup R_{ant} \cup R_{dis} \cup R_{hol} \cup R_{mer} \cup R_{assoc}}$$

- Semantic similarity relation: set of ordered pairs

  $R_S \subseteq E \times E$ such that

$$\mathbf{R_S = R_i \cup R_H \cup R_h \cup R_{syn} \cup R_{conn} \cup R_{com}}$$

# Semantic Measures

- Semantic relatedness measure. Given $R$, $E$

$$rel: E \times E \rightarrow \mathbb{R} \quad \text{such that:}$$

| | |
|---|---|
| $\forall x, y \in E, rel(x,y) \geq 0$ | (positiveness) |
| $\forall x, y, z \in E, rel(x,x) \geq rel(y,z)$ | (maximality) |
| $\forall x, y \in E, rel(x,y) = rel(y,x)$ | (symmetry) |
| $\forall x, y \in E, (x, y) \notin R, rel(x,y) = 0$ | |

- Semantic similarity measure. Given $R_S$, $E$

$$sim: E \times E \rightarrow \mathbb{R} \quad \text{such that:}$$

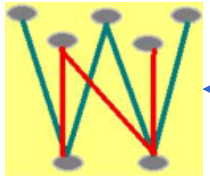
| | |
|---|---|
| $\forall x, y \in E, sim(x,y) \geq 0$ | (positiveness) |
| $\forall x, y, z \in E, sim(x,x) \geq sim(y,z)$ | (maximality) |
| $\forall x, y \in E, sim(x,y) = sim(y,x)$ | (symmetry) |
| $\forall x, y \in E, (x, y) \notin R_S, sim(x,y) = 0$ | |

# Semantic Measures

- Use of external resources

*Few named entities: “The Rolling Stones”*
*Few very specialized terms: “exocitosys”*

*Many relations remain hidden:*
*“aspirin” + “stomach disease”*




*Lexical resources (WordNet, corpus, …)*

*Wikipedia*

*The Web*

*increasing **coverage***

# *Semantic Measures*

- Two examples of semantic measures, designed for its use on the Semantic Web [Gracia 09, Gracia & Mena 08, Trillo et al. 07]:
  - Context and inference-based semantic similarity measure
  - Web-based semantic relatedness measure

# Semantic Measures: Similarity

- Semantic similarity between two classes
  - 1st) Extract Ontological Contexts
  - 2nd) Enrich them by applying inference
  - 3rd) Compute the following:

**$sim(t_1, t_2)$ = Linear Combination of**

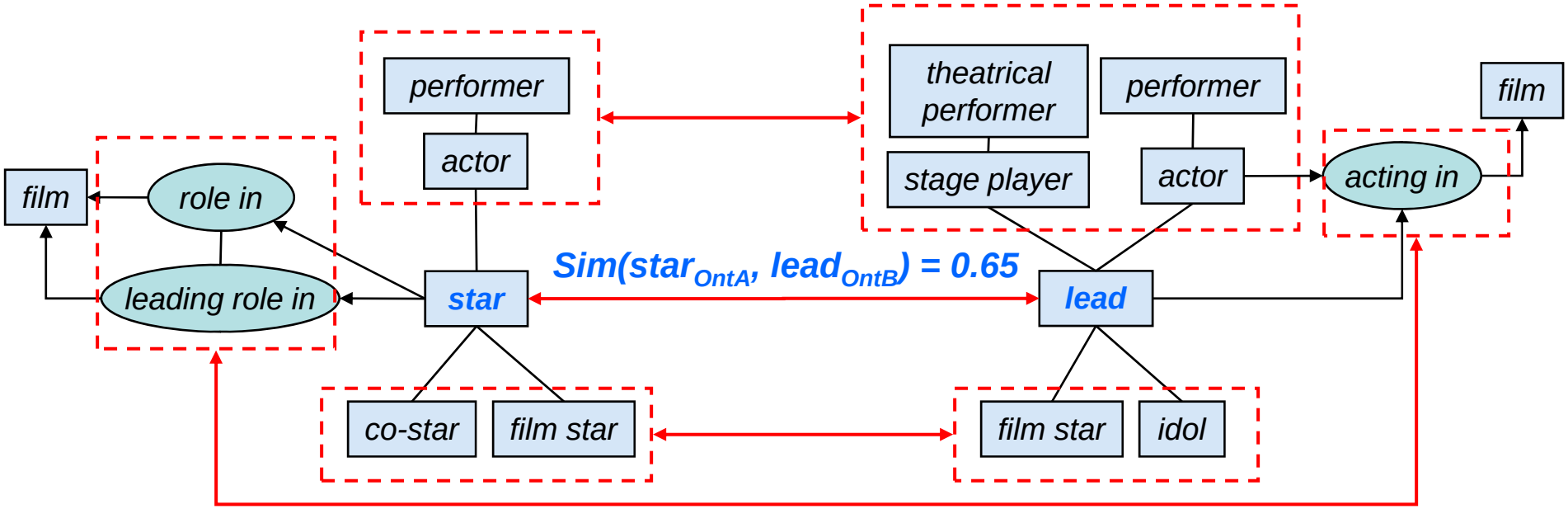
- **$lexical\_similarity(synonyms_1, synonyms_2)$**
- **$VSM(description_1, description_2)$**
- **$VSM(properties_1, properties_2)$**
- **$Graph\_similarity(graph_1, graph_2)$**

**$Graph\_similarity(graph_1, graph_2)$ = Linear Combination of**

- **$VSM(hyper_1 \uplus directHyper_1, hyper_2 \uplus directHyper_2)$**
- **$VSM(hypo_1 \uplus directHypo_1, hypo_2 \uplus directHypo_2)$**

# Semantic Measures: Similarity

- Analogously to compute similarity between properties and between individuals.
- Example of Semantic Similarity between Classes:



# *Semantic Measures: Relatedness*

- Hypothesis [Cilibrasi and Vitànyi 07] : For each $\mathbf{x, y} \in \mathbb{S}$

> **x, y are semantically related ⇔ x, y appear on the same web page**

- Semantic relatedness between two words

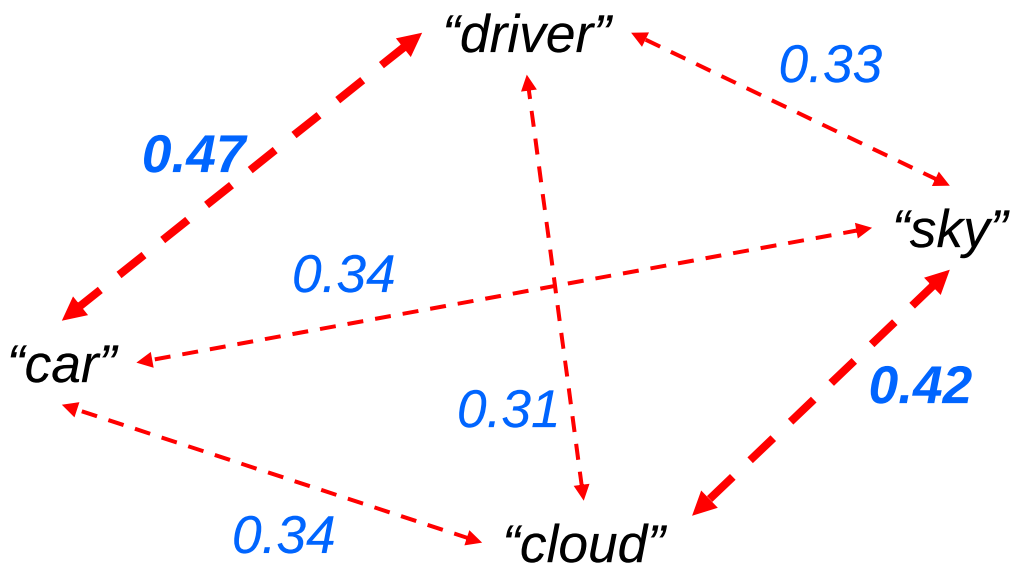
$$\mathbf{relWeb(x, y) = e^{-2NWD(x,y)}}$$

where NWD is Cilibrasi and Vitànyi's Google distance

# *Semantic Measures: Relatedness*

- An example of web-based relatedness computation

http://horus.cps.unizar.es:28080/Relatedness/Relatedness.html



# Semantic Measures: Relatedness

- Semantic relatedness between two ontology terms
  - Level 0: Labels and synonyms

**$rel_0(a, b)$ = average of relWeb between synonyms of a and b**

  - Level 1: $OC^m$ Minimal Ontological Context

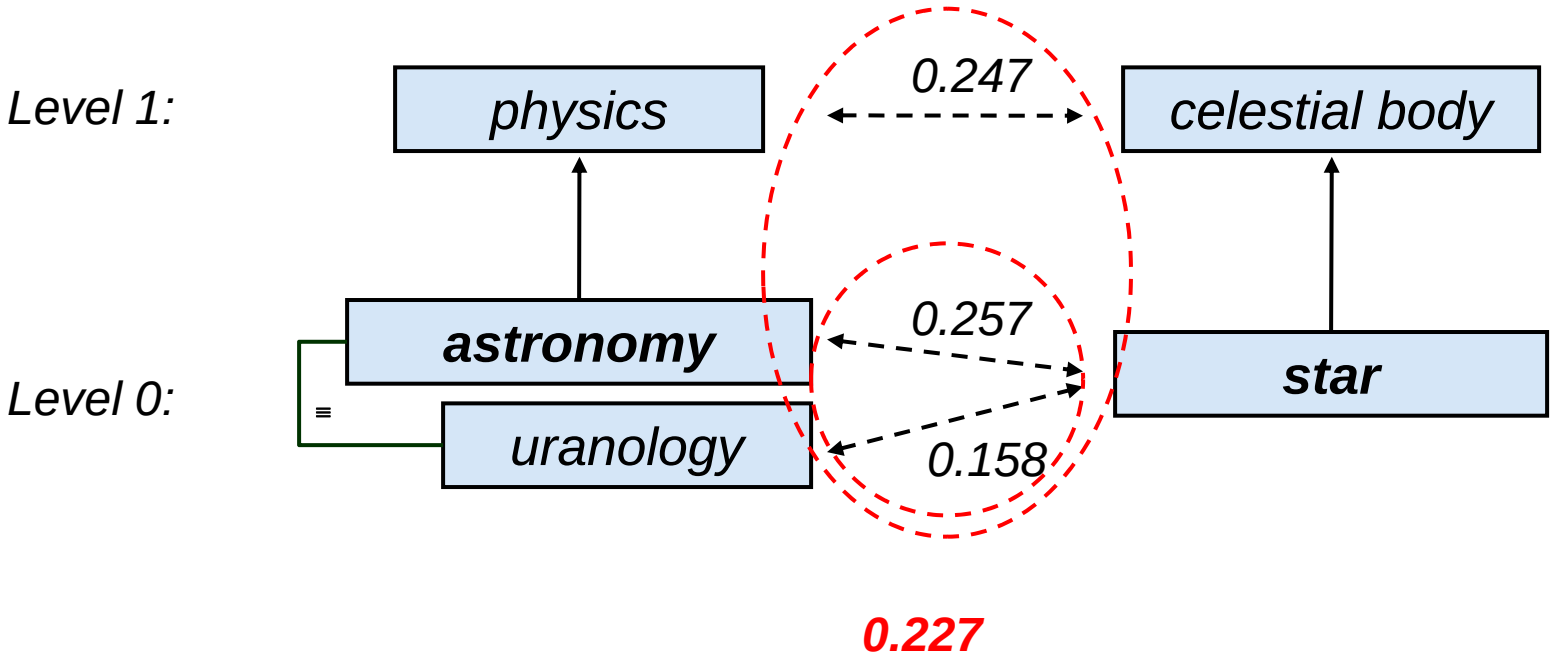
E.g.: Java → Island → Land → Thing

**$rel_1(a, b)$ = average of $rel_0$ between the elements of the $OC^m_a$ and $OC^m_b$**

**rel (a, b) = linear combination of $rel_0(a, b)$ and $rel_1(a, b)$**

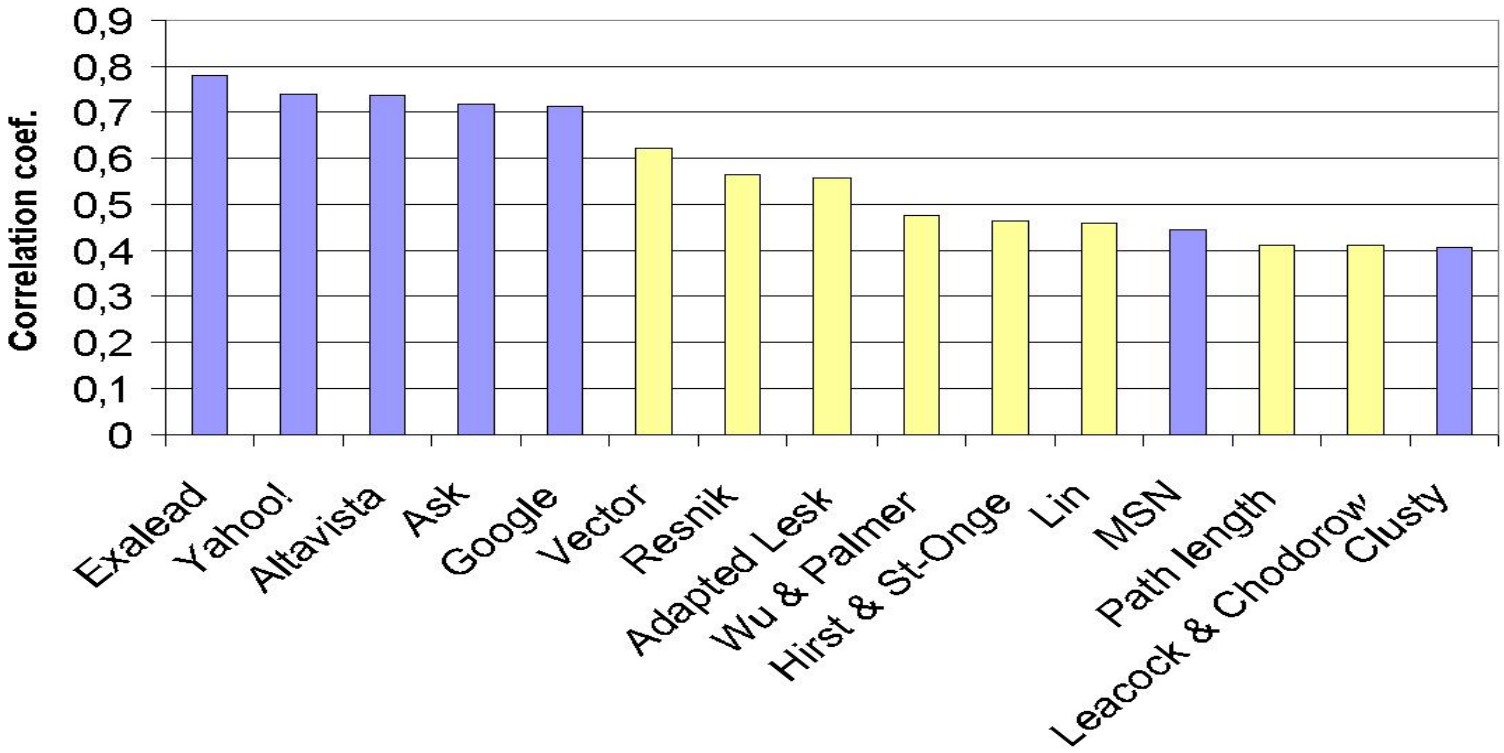
# Semantic Measures: Relatedness

- Example: relatedness between ontology terms



# *Semantic Measures: Relatedness*

- Experiment: relatedness between words
  - Web-based (blue), WordNet-based (yellow) measures




  - This supports the use of web frequencies for relatedness computation

# *Semantic Measures*

- Semantic Measures: other approaches
  - Based on Thesauri and Lexical Resources [Jiang and Conrath, Banerjee and Pedersen, Resnik, etc.]
  - Based on Wikipedia [Gabrilovich and Markovitch, Strube and Ponzetto, etc.]
  - Based on the Web [Bollegala et al., Chen et al., Sahami et al., OntoNL, etc.]
  - Based on DL constructs [D'Amato, et. al., Borgida et al., Janowicz et al., Hu et. al., etc.]

# Ontology Matching

# Ontology Matching

- A couple of definitions [borrowed from Euzenat & Shvaiko's OM tutorial]

**Definition (Correspondence)**

Given two ontologies $o$ and $o'$, a **correspondence** between $o$ and $o'$ is a 5-uple: $\langle id, e, e', r, n \rangle$ such that:

- $id$ is a unique identifier of the correspondence
- $e$ and $e'$ are entities of $o$ and $o'$ (e.g., XML elements, classes)
- $r$ is a relation (e.g., equivalence ($=$), more general ($\sqsupseteq$), disjointness ($\perp$))
- $n$ is a confidence measure in some mathematical structure (typically in the [0 1] range)

# Ontology Matching

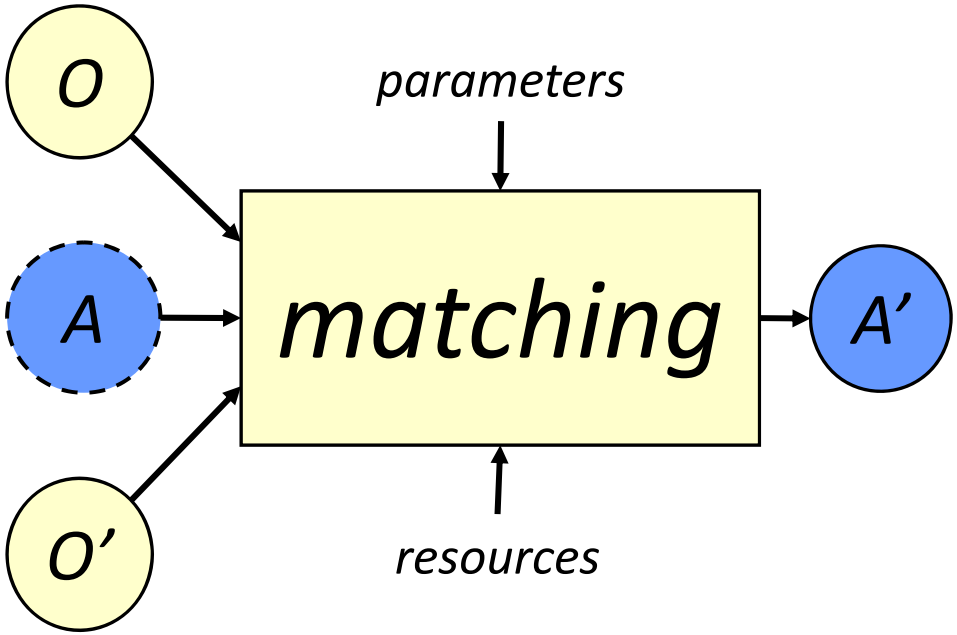
**Definition (Alignment)**

Given two ontologies $o$ and $o'$, an **alignment** ($A$) between $o$ and $o'$:

- is a set of correspondences on $o$ and $o'$
- with some additional metadata (multiplicity: 1-1, 1-*, method, date, properties, etc.)

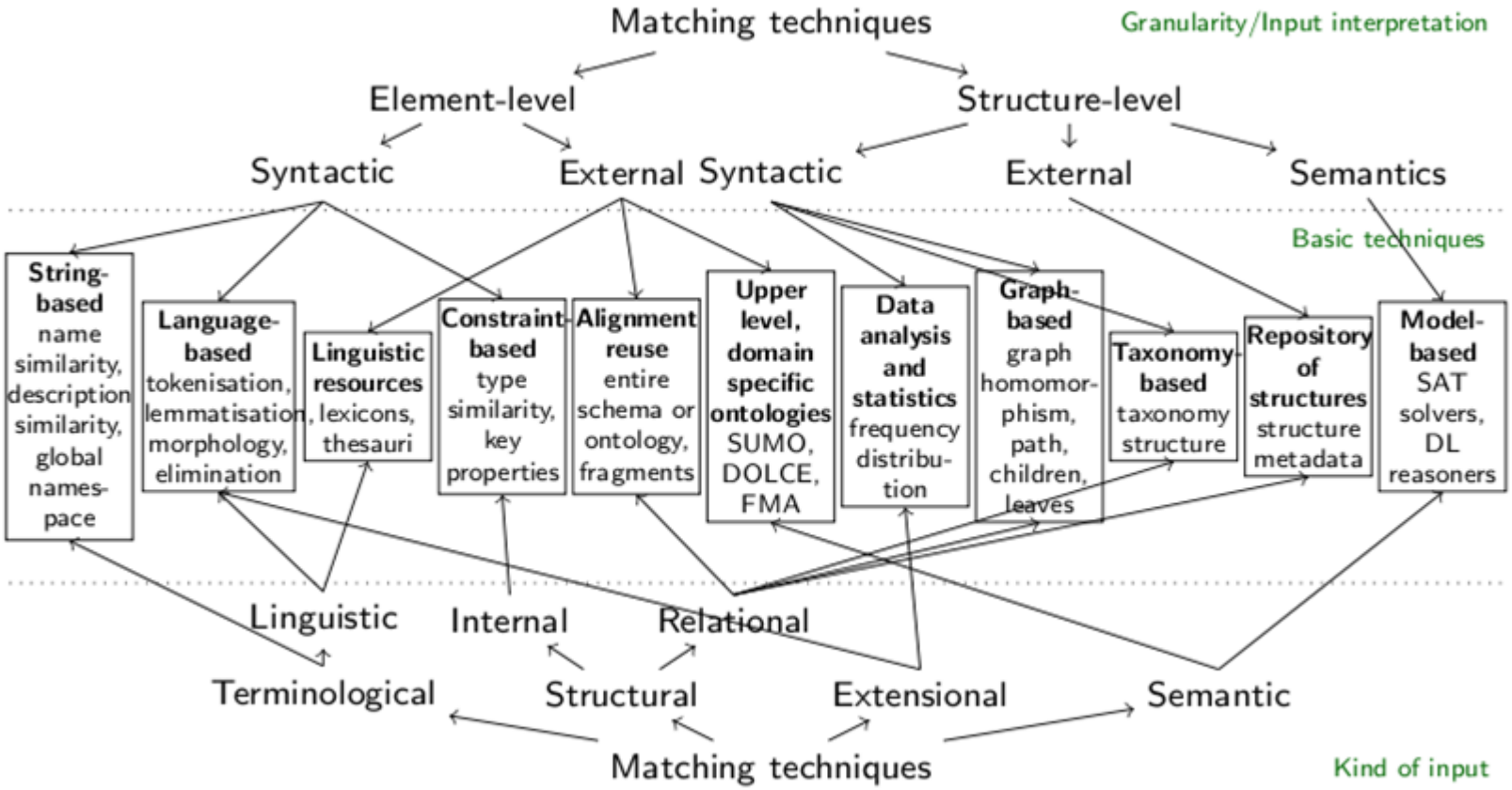
# Ontology Matching

- General scheme of an Ontology Matching process, being O and O' are the input ontologies, A an initial set of known correspondences (optional), and A' the resultant alignment (the found correspondences):



# Ontology Matching

- Matching techniques [Euzenat & Shvaiko, 2007]



# *Ontology Matching*

- Matching techniques [borrowed from Euzenat & Shvaiko's OM tutorial]:

  - Name of the entities
    - comments, alternate names, names of related entities

    $\Longrightarrow$ NLP, IR, etc.
  - Structure
    - internal structure: constraints on relations, typing
    - external structure: relations between entities

    $\Longrightarrow$ Data mining, Discrete mathematics
  - Extension
    - Instances themselves
    - Related resources: annotated documents, exchanges messages or queries

    $\Longrightarrow$ Statistics, data analysis, data mining, machine learning
  - Semantics (models)

    $\Longrightarrow$ Reasoning techniques
  - Background knowledge
    - the web
    - ontologies
    - wordnet, etc.

# Ontology Matching

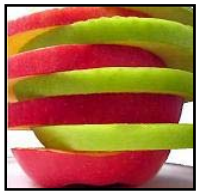
- How does an alignment looks like? E.g. (in RDF):

```
<Alignment>
<xml>yes</xml>
<level>0</level>
<type>11</type>
<onto1>./models/Nitrogen cycling.owl</onto1>
<onto2>./ models/Phosphorus cycling.owl</onto2>
....
<map>
<Cell>
    <entity1 rdf:resource='http://www.dynalearn.eu/models/Nitrogen#plant'/>
    <entity2 rdf:resource=' http://www.dynalearn.eu/models/Phosphorus#plantas'/>
    <relation>=</relation>
    <measure rdf:datatype='http://www.w3.org/2001/XMLSchema#float'>0.8</measure>
</Cell>
</map>
....
</Alignment>
```

# *Ontology Matching*

Two examples of ontology matching systems, designed for its use on the Semantic Web:

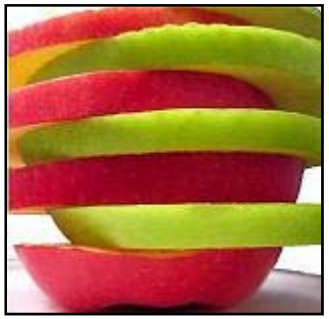
- CIDER *[Gracia & Mena 08b]* : Context and Inference baseD alignER
  - Based on our semantic similarity measure
  - Identifies semantic equivalences

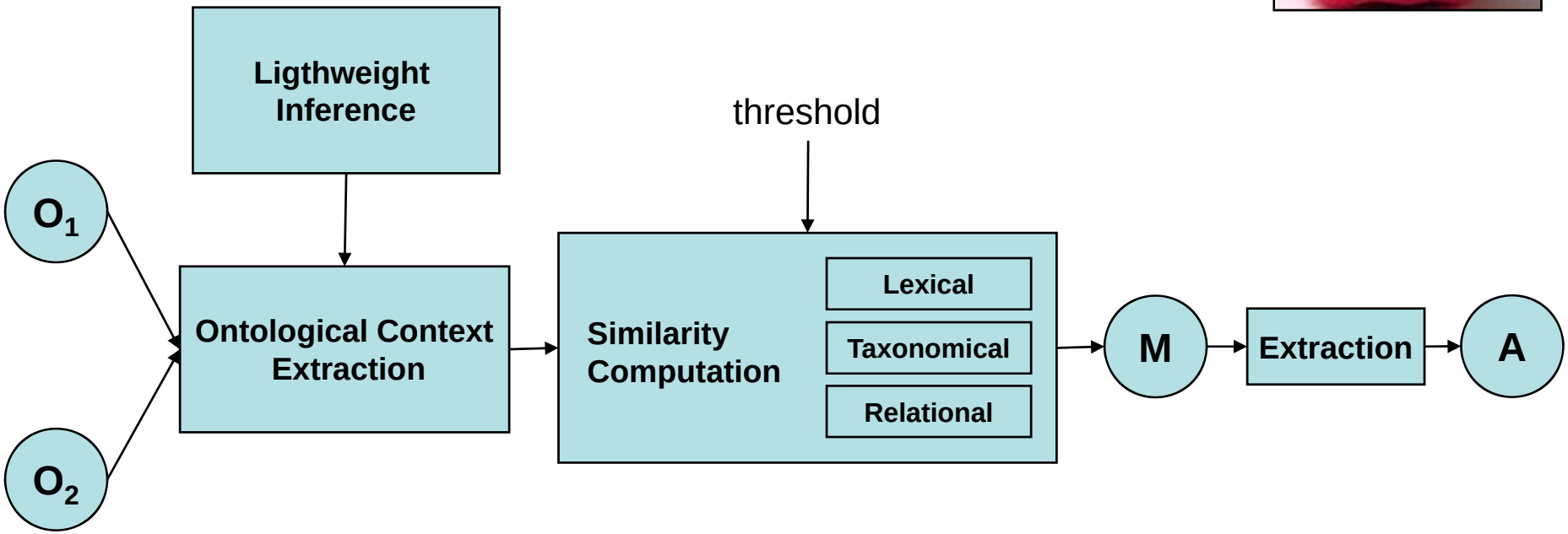


- Scarlet *[Sabou et al. 08]*
  - Background knowledge-based ontology alignment
  - Identifies various types of semantic relationships
  - Hampered by semantic ambiguity problems

# Ontology Matching: CIDER

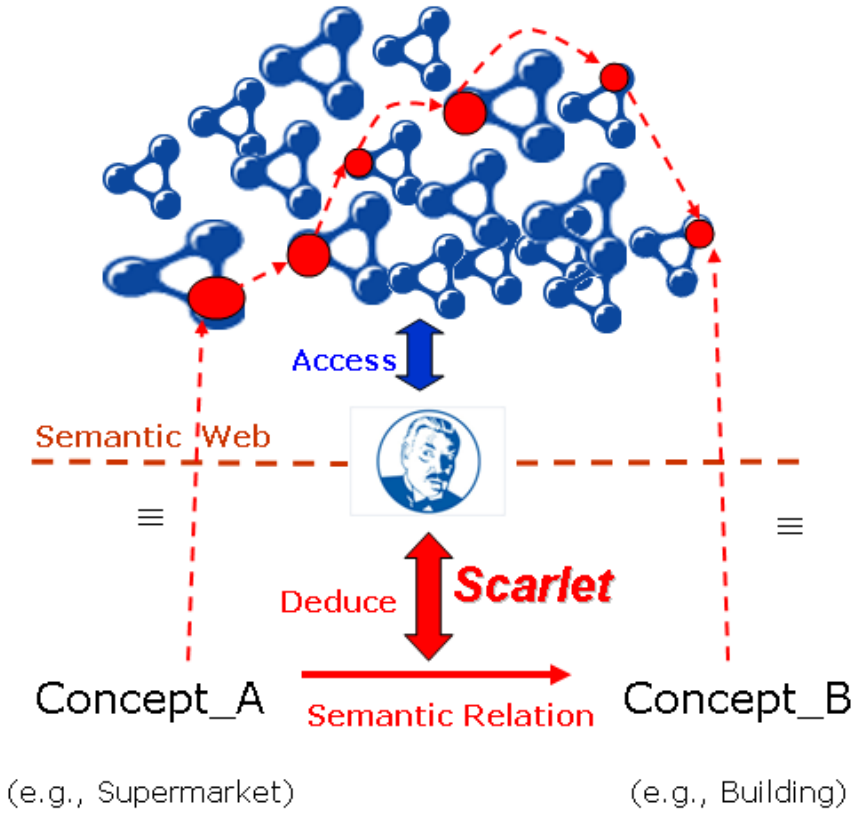
- CIDER: Context and Inference baseD alignER





# Ontology Matching: Spider

- Scarlet system (KMi, Open University, United Kingdom): uses online ontologies as background knowldege for ontology matching



# *Ontology Matching*

- Ontology Matching. Other approaches [Euzenat & Shvaiko 07]
  - Falcon-AO
  - ASMOV
  - Lily
  - Glue
  - S-Match
  - etc.

# References


| | |
|---|---|
| Cilibrasi & Vitányi 07 | Rudi L. Cilibrasi and Paul M.B. Vitányi. The Google Similarity Distance. IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. 3, pages 370{383, March 2007. |
| Euzenat & Shvaiko 07 | Jérôme Euzenat and Pavel Shvaiko. Ontology Matching. Springer-Verlag, 2007. |
| Gracia 09 | Jorge Gracia, “Integration and Disambiguation Techniques for Semantic Heterogeneity Reduction on the Web”, PhD Thesis, October 2009 |
| Gracia & Mena 08a | Jorge Gracia and Eduardo Mena, "Web-based Measure of Semantic Relatedness", Proc. of 9th International Conference on Web Information Systems Engineering (**WISE 2008**), Auckland (New Zealand), Springer Verlag LNCS, volume 5175, pp. 136-150, September 2008. |
| Gracia & Mena 08b | Jorge Gracia and Eduardo Mena, "Ontology Matching with CIDER: Evaluation Report for the OAEI 2008", Proc. of 3rd Ontology Matching Workshop (OM'08), at 7th International Semantic Web Conference (ISWC'08), Karlsruhe (Germany), CEUR-WS, volume 431, pp. 140-146, October 2008. |

# References


| | |
|---|---|
| Sabou et al. 08 | Marta Sabou, Mathieu d'Aquin and Enrico Motta. Exploring the Semantic Web as Background Knowledge for Ontology Matching. Journal of Data Semantics, vol. 11, pages 156-190, 2008. |
| Trillo et al. 07 | Raquel Trillo, Jorge Gracia, Mauricio Espinoza and Eduardo Mena, "Discovering the Semantics of User Keywords", Journal on Universal Computer Science (**JUCS**). Special Issue: Ontologies and their Applications, vol. 13(12):, pages 908-1935, Springer Verlag, December 2007 |

# *Further Readings*


Alexander Budanitsky & Graeme Hirst. Evaluating WordNet-based measures of semantic distance. Computational Linguistics, vol. 32, no. 1, pages 13-47, March 2006.

Zharko Aleksovski, Michel C. A. Klein, Warner ten Kate & Frank van Harmelen. Matching Unstructured Vocabularies Using a Background Ontology. In EKAW, volume 4248 of LNSC, pages 182-197, October 2006.

Danushka Bollegala, Yutaka Matsuo & Mitsuru Ishizuka. Measuring semantic similarity between words using web search engines. In Proceedings of the 16th International Conference on World Wide Web (WWW 2007) Ban, Alberta, Canada, pages 757-766. ACM, May 2007.

Ted Pedersen, Satanjeev Banerjee & Siddharth Patwardhan. Maximizing Semantic Relatedness to Perform Word Sense Disambiguation. Technical report, University of Minnesota Supercomputing Institute Research Report UMSI 2005/25, 2005.

Yuzhong Qu, Wei Hu & Gong Cheng. Constructing Virtual Documents for Ontology Matching. In Proc. of 15th International World Wide Web Conference (WWW'06), Edinburgh, UK, May 2006.

END of presentation