A Semantic-based Ontology Matching Process for PDMS

Carlos Eduardo Pires¹, Damires Souza², Thiago Pachêco¹, and Ana Carolina Salgado¹

¹ Federal University of Pernambuco (UFPE), Center for Informatics, Brazil
{cesp,tpap,acs}@cin.ufpe.br

² Federal Institute of Education, Science and Technology of Paraíba (IFPB), Brazil
damires@ifpb.edu.br

http://www.cin.ufpe.br/~speed/SemMatch/index.htm
Motivation

- **Peer Data Management System (PDMS)**
  [Adjiman et al., 2007]
  - Each peer is an autonomous data source that makes available a local schema
  - Schema mappings (correspondences between schema elements) are generated to allow information exchange between peers

- **Ontologies**
  - Make explicit the content of data sources (peer ontologies)
  - Enhance information integration
Motivation

- **Peer ontologies**
  - Designed and developed *autonomously*
  - Contain several forms of *heterogeneity*

- **Ontology matching techniques** [Euzenat and Shvaiko, 2007]
  - Deal with the *diverse concept meanings* existing in peer ontologies
  - Reconcile peer ontologies and find correspondences between their elements
Goal

- Propose a **semantic-based ontology matching process** which has been instantiated in a PDMS

**Contributions**

- Identification of **semantic correspondences** between two peer ontologies
  - Taking into account a **domain ontology** as background knowledge
- Determination of the **global similarity** between two peer ontologies
Outline

- Ontology Matching
- Using a Domain Ontology to Define Semantic Correspondences
- Semantic-based Ontology Matching Process
- Calculating the Global Similarity Measure
- Experiments and Results
- Related Work
- Conclusions and Further Work
Ontology Matching

- Process of finding correspondences between elements of different ontologies [Euzenat and Shvaiko, 2007]
  - Normally describing the same or similar domains

- An element is a concept, property or instance

- Ontology Alignment
  - Set of correspondences indicating which elements of two ontologies logically correspond to each other
  - Produced by one or more matchers which are executed sequentially or in parallel
Working Scenario

Semantic Community of Peers
* Education Domain *

DO
(UnivCSCMO.owl)

SP

O₁
(Semiport.owl)

P₁

O₂
(UnivBench.owl)

P₂

Unstructured P2P Network

Similarity?
Our focus

- Identify semantic correspondences between $O_1$ and $O_2$ elements
- Determine if $P_1$ and $P_2$ are semantic neighbors
  - Two peers are semantic neighbors if their global similarity is higher than a certain threshold
Using a **Domain Ontology (DO)**

- Reliable reference available on the Web
- Used as Background Knowledge
  - Bridge the conceptual differences or similarities between two peer ontologies
Definition. A semantic correspondence is represented by one of the following expressions:

- $O_1:x \equiv O_2:y$, an isEquivalentTo correspondence
- $O_1:x \sqsubseteq O_2:y$, an isSubConceptOf correspondence
- $O_1:x \sqsupseteq O_2:y$, an isSuperConceptOf correspondence
- $O_1:x \sqsupset O_2:y$, an isPartOf correspondence
- $O_1:x \sqsubseteq O_2:y$, an isWholeOf correspondence
- $O_1:x \not\equiv O_2:y$, an isCloseTo correspondence
- $O_1:x \perp O_2:y$, an isDisjointWith correspondence

(*) $x$ and $y$ are elements belonging to the peer ontologies
Using a Domain Ontology to Define Semantic Correspondences

- $O_1:k$ and $O_2:z$ are close if
  - They share a common ancestor in the DO
  - The common ancestor is not the root ($\top$)
  - The concepts do not hold neither subsumption nor disjointness
  - The measured depths ($\text{thresholdRoot}$ and $\text{thresholdCommonAncestor}$) are evaluated to true

Example:
$O_1$.Notebook $\simeq$ $O_2$.MacintoshPC
Semantic-based Ontology Matching Process

Phase 1:
1. Linguistic-Structural Matching (any matcher) → 1:n or m:m → A_{LS}
2. Semantic Matching → 1:n or m:m → A_{SE}
3. Similarity Combination → A_{CO} (1:n or m:m)
4. Correspondence Ranking
5. Correspondence Selection

Phase 2:
A_{12} (1:1)

A Semantic-Based Ontology Matching Process for PDMS
# Semantic-based Ontology Matching Process

## (a) $A_{Lb}$ and $A_{wb}$

<table>
<thead>
<tr>
<th>O1 Element</th>
<th>O2 Element</th>
<th>Relationship</th>
<th>Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>UndergraduateStudent</td>
<td>Monitor</td>
<td></td>
<td>0.50</td>
</tr>
<tr>
<td>UndergraduateStudent</td>
<td>GraduateStudent</td>
<td></td>
<td>0.70</td>
</tr>
<tr>
<td>UndergraduateStudent</td>
<td>Student</td>
<td></td>
<td>0.50</td>
</tr>
<tr>
<td>Student</td>
<td>GraduateStudent</td>
<td></td>
<td>0.50</td>
</tr>
<tr>
<td>GraduateStudent</td>
<td>GraduateStudent</td>
<td></td>
<td>0.80</td>
</tr>
<tr>
<td>GraduateStudent</td>
<td>GraduateStudent</td>
<td>isEquivalentTo</td>
<td>1.00</td>
</tr>
</tbody>
</table>

## (b) $A_{CD}$

### Similarity Values Combination
- LB weight = 0.4
- SE weight = 0.6

<table>
<thead>
<tr>
<th>O1 Element</th>
<th>O2 Element</th>
<th>Relationship</th>
<th>Combined Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>UndergraduateStudent</td>
<td>Monitor</td>
<td>isSuperConceptOf</td>
<td>0.60</td>
</tr>
<tr>
<td>UndergraduateStudent</td>
<td>GraduateStudent</td>
<td>isDisjointWith</td>
<td>0.26</td>
</tr>
<tr>
<td>UndergraduateStudent</td>
<td>Student</td>
<td>isSubConceptOf</td>
<td>0.88</td>
</tr>
<tr>
<td>GraduateStudent</td>
<td>Student</td>
<td>isSubConceptOf</td>
<td>0.72</td>
</tr>
<tr>
<td>Student</td>
<td>GraduateStudent</td>
<td>isEquivalentTo</td>
<td>0.84</td>
</tr>
</tbody>
</table>

## (c) Correspondence Ranking

<table>
<thead>
<tr>
<th>O1 Element</th>
<th>O2 Element</th>
<th>Relationship</th>
<th>Combined Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>UndergraduateStudent</td>
<td>Student</td>
<td>isSubConceptOf</td>
<td>0.88</td>
</tr>
<tr>
<td>UndergraduateStudent</td>
<td>Monitor</td>
<td>isSuperConceptOf</td>
<td>0.80</td>
</tr>
<tr>
<td>UndergraduateStudent</td>
<td>GraduateStudent</td>
<td>isDisjointWith</td>
<td>0.28</td>
</tr>
<tr>
<td>GraduateStudent</td>
<td>GraduateStudent</td>
<td>isEquivalentTo</td>
<td>0.84</td>
</tr>
<tr>
<td>GraduateStudent</td>
<td>Student</td>
<td>isSubConceptOf</td>
<td>0.72</td>
</tr>
<tr>
<td>Student</td>
<td>GraduateStudent</td>
<td>isSuperConceptOf</td>
<td>0.84</td>
</tr>
</tbody>
</table>

## (d) Correspondence Selection

<table>
<thead>
<tr>
<th>O1 Element</th>
<th>O2 Element</th>
<th>Relationship</th>
<th>Combined Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>UndergraduateStudent</td>
<td>Student</td>
<td>isSubConceptOf</td>
<td>0.88</td>
</tr>
<tr>
<td>GraduateStudent</td>
<td>GraduateStudent</td>
<td>isEquivalentTo</td>
<td>0.84</td>
</tr>
</tbody>
</table>

## (d') Correspondence Selection

<table>
<thead>
<tr>
<th>O1 Element</th>
<th>O2 Element</th>
<th>Relationship</th>
<th>Combined Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monitor</td>
<td>UndergraduateStudent</td>
<td>isSubConceptOf</td>
<td>0.60</td>
</tr>
<tr>
<td>GraduateStudent</td>
<td>GraduateStudent</td>
<td>isEquivalentTo</td>
<td>0.84</td>
</tr>
<tr>
<td>Student</td>
<td>GraduateStudent</td>
<td>isSuperConceptOf</td>
<td>0.72</td>
</tr>
</tbody>
</table>
Calculating the Global Similarity Measure

- Uses the alignment sets $A_{12}$ and $A_{21}$
- Existing similarity measures can be adapted
  - *Dice* [Aümuller et al., 2005], *weighted average* [Castano et al., 1998] and *overlap* [Rijsbergen, 1979]
- All of them consider the size of the input ontologies
- The size of an ontology ($|O|$) is determined by the number of its elements
Calculating the Global Similarity Measure

\[
\text{Weighted Average}(O_1, O_2) = \frac{\sum_{n=1}^{\mid A_{12} \mid} n + \sum_{j=1}^{\mid A_{21} \mid} n}{\mid O_1 \mid + \mid O_2 \mid}
\]

**Alignment A_{12}**
1. (Person, Person, isEquivalentTo, 1.0)
2. (FullProfessor, FullProfessor, isEquivalentTo, 1.0)
3. (UndergraduateStudent, Course, isPartOf, 0.3)
4. (Student, Person, isSubConceptOf, 0.8)
5. (Professor, Faculty, isSubConceptOf, 0.8)

**Alignment A_{21}**
1. (Person, Person, isEquivalentTo, 1.0)
2. (FullProfessor, FullProfessor, isEquivalentTo, 1.0)
3. (Course, UndergraduateStudent, isWholeOf, 0.3)
4. (Worker, Person, isSubConceptOf, 0.8)
5. (GraduateStudent, UndergraduateStudent, isDisjointWith, 0.0)
6. (Faculty, Professor, isSuperConceptOf, 0.8)
7. (MasterStudent, Student, isSubConceptOf, 0.8)

\[
\text{Weighted Average}(O_1, O_2) = \frac{(1.0 + 1.0 + 0.3 + 0.8 + 0.8) + (1.0 + 1.0 + 0.3 + 0.8 + 0.0 + 0.8 + 0.8)}{\mid 16 \mid + \mid 17 \mid} = 0.66
\]
Experiments and Results

- The semantic-based ontology matching tool
  - Implemented in Java
  - Jena has been used to provide ontology manipulation and reasoning
  - H-Match has been used as the hybrid matcher
- Correspondence identification has been restricted to concepts
  - Properties are not included
Experiments and Results

<table>
<thead>
<tr>
<th>Ontology 1</th>
<th>Correspondence</th>
<th>Ontology 2</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>AssistantProfessor</td>
<td>isDisjointWith</td>
<td>AssociateProfessor</td>
<td>0</td>
</tr>
<tr>
<td>Person</td>
<td></td>
<td>Work</td>
<td>0</td>
</tr>
<tr>
<td>Worker</td>
<td>isDisjointWith</td>
<td>UndergraduateStudent</td>
<td>0</td>
</tr>
<tr>
<td>FullProfessor</td>
<td>isDisjointWith</td>
<td>AssociateProfessor</td>
<td>0</td>
</tr>
<tr>
<td>Organization</td>
<td>isDisjointWith</td>
<td>Person</td>
<td>0</td>
</tr>
<tr>
<td>Topic</td>
<td>isDisjointWith</td>
<td>Person</td>
<td>0</td>
</tr>
<tr>
<td>Event</td>
<td>isDisjointWith</td>
<td>Person</td>
<td>0</td>
</tr>
<tr>
<td>Publication</td>
<td>isDisjointWith</td>
<td>Person</td>
<td>0</td>
</tr>
<tr>
<td>GraduateStudent</td>
<td>isDisjointWith</td>
<td>UndergraduateStudent</td>
<td>0</td>
</tr>
<tr>
<td>TechnicalStaff</td>
<td>isDisjointWith</td>
<td>AdministrativeStaff</td>
<td>0</td>
</tr>
<tr>
<td>Person</td>
<td>isDisjointWith</td>
<td>Publication</td>
<td>0</td>
</tr>
<tr>
<td>Product</td>
<td>isDisjointWith</td>
<td>Person</td>
<td>0</td>
</tr>
<tr>
<td>AdministrativeStaff</td>
<td>isCloseTo</td>
<td>Assistant</td>
<td>0,7</td>
</tr>
<tr>
<td>DevelopmentProject</td>
<td>isCloseTo</td>
<td>ResearchProject</td>
<td>0,7</td>
</tr>
<tr>
<td>TechnicalReport</td>
<td>isCloseTo</td>
<td>JournalArticle</td>
<td>0,7</td>
</tr>
<tr>
<td>ClericalStaff</td>
<td>isCloseTo</td>
<td>Dean</td>
<td>0,7</td>
</tr>
<tr>
<td>ClericalStaff</td>
<td>isCloseTo</td>
<td>Director</td>
<td>0,7</td>
</tr>
<tr>
<td>FullProfessor</td>
<td>isCloseTo</td>
<td>VisitingProfessor</td>
<td>0,7</td>
</tr>
<tr>
<td>TechnicalReport</td>
<td>isCloseTo</td>
<td>ConferencePaper</td>
<td>0,7</td>
</tr>
<tr>
<td>SystemsStaff</td>
<td>isCloseTo</td>
<td>Chair</td>
<td>0,7</td>
</tr>
<tr>
<td>Lecturer</td>
<td>isCloseTo</td>
<td>Professor</td>
<td>0,7</td>
</tr>
<tr>
<td>Faculty</td>
<td>isCloseTo</td>
<td>Assistant</td>
<td>0,7</td>
</tr>
<tr>
<td>SystemsStaff</td>
<td>isCloseTo</td>
<td>Systems</td>
<td>0,7</td>
</tr>
</tbody>
</table>

A Semantic-Based Ontology Matching Process for PDMS
Experiments and Results

**Recall**

<table>
<thead>
<tr>
<th>COMA++</th>
<th>H-Match</th>
<th>Falcon-AO</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="#" alt="Graph" /></td>
<td><img src="#" alt="Graph" /></td>
<td><img src="#" alt="Graph" /></td>
</tr>
</tbody>
</table>

- **Linguistic + Structural**
- **Linguistic + Structural + Semantic**

**Precision**

<table>
<thead>
<tr>
<th>COMA++</th>
<th>H-Match</th>
<th>Falcon-AO</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="#" alt="Graph" /></td>
<td><img src="#" alt="Graph" /></td>
<td><img src="#" alt="Graph" /></td>
</tr>
</tbody>
</table>

- **Linguistic + Structural**
- **Linguistic + Structural + Semantic**

**Correspondences for \(O_1:\text{Faculty}\)**

<table>
<thead>
<tr>
<th>Correspondence</th>
<th>(O_1:\text{Faculty} \leftrightarrow O_2:\text{Faculty})</th>
<th>(O_1:\text{Faculty} \rightarrow O_2:\text{PostDoc})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(O_1:\text{Faculty} \leftrightarrow O_2:\text{Worker})</td>
<td>(O_1:\text{Faculty} \Rightarrow O_2:\text{Assistant})</td>
<td>(O_1:\text{Faculty} \Rightarrow O_2:\text{AdministrativeStaff})</td>
</tr>
</tbody>
</table>
Related Work

- Only a few semantic-based approaches consider the use of background knowledge to improve ontology matching
  - S-Match, TaxoMap, CTXMatch
- Correspondences are usually restricted to equivalence
  - CTXMatch considers other ones (specialization and generalization)
- We also identify other types of semantic correspondences
  - E.g., disjointness and closeness
Global Similarity Measure

- Not produced by the previous works
- [Castano et al., 1998] propose a kind of such measure
  - Concerned with ER schemas
- COMA++ [Aümuller et al., 2005] argues that calculates a global measure
  - Considering the version we performed our tests, we were not able to find out such feature explicitly
Conclusions and Further Work

- Our matching process tries to overcome limitations of linguistic and structural approaches by using domain ontologies as background knowledge.
  - A semantic matcher identifies, besides traditional types of correspondences, other ones (e.g., closeness and disjointness).
  - Determination of a global similarity measure between two ontologies (not only between their elements).
Conclusions and Further Work

- The combination of different matchers can improve the alignments produced by ontology matchings tools
  - Taking out incorrect or meaningless correspondences and including relevant ones

- Further work
  - Extend our tool to consider properties
  - Include an alignment-reuse matcher
A Semantic-based Ontology Matching Process for PDMS

Carlos Eduardo Pires¹, Damires Souza², Thiago Pachêco¹, and Ana Carolina Salgado¹

¹ Federal University of Pernambuco (UFPE), Center for Informatics, Brazil
   {cesp,tpap,acs}@cin.ufpe.br

² Federal Institute of Education, Science and Technology of Paraíba (IFPB), Brazil
   damires@ifpb.edu.br

http://www.cin.ufpe.br/~speed/SemMatch/index.htm