DEXA 2009

GLOBE 2009

A Semantic-based Ontology Matching Process for PDMS

Carlos Eduardo Pires[1], Damires Souza[2], Thiago Pachêco[1], and Ana Carolina Salgado[1]

[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

VIRTUS IMPAVIDA

Centro de Informática U·F·P·E

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA PARAÍBA

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

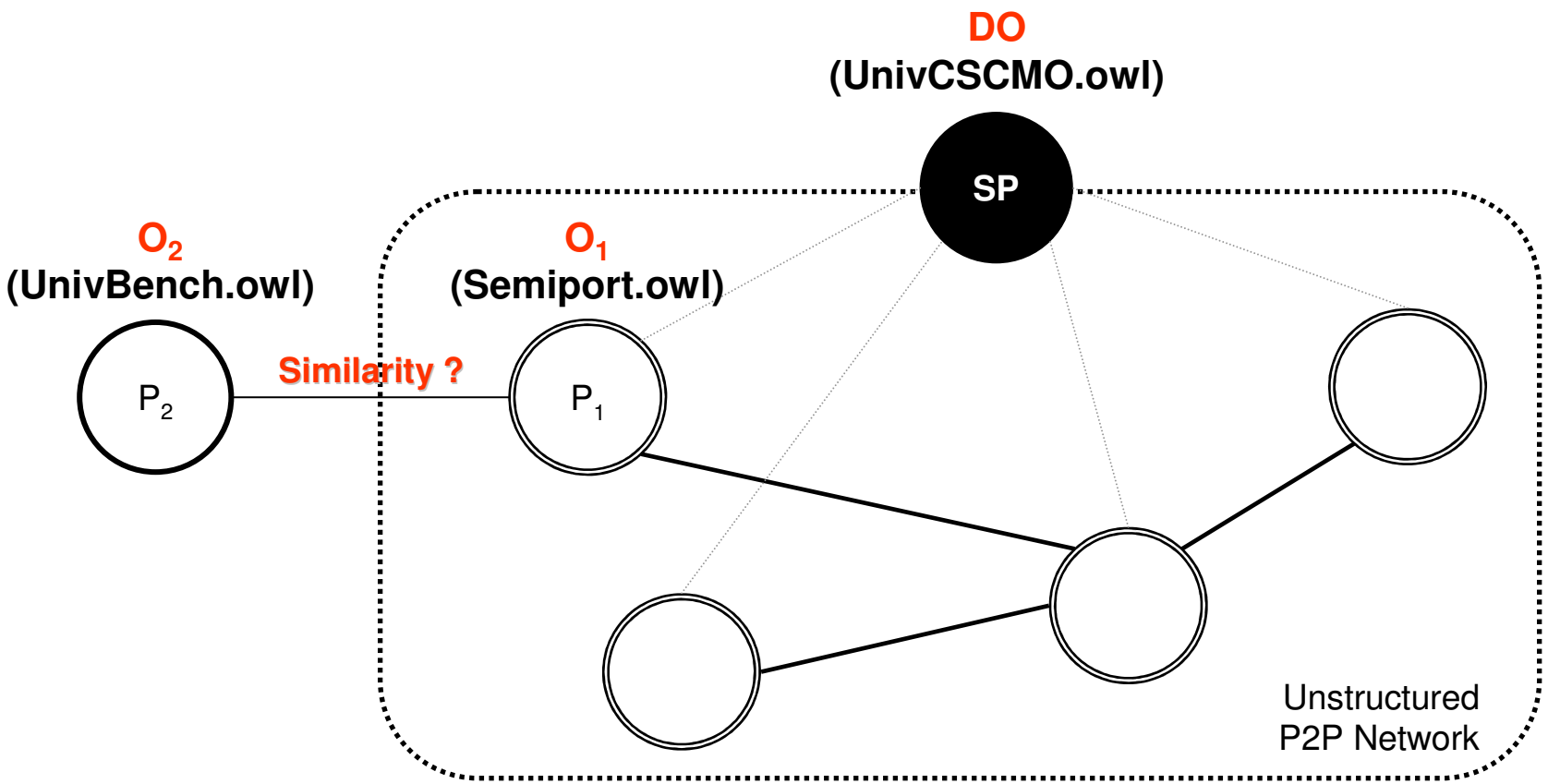
Outline

- 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

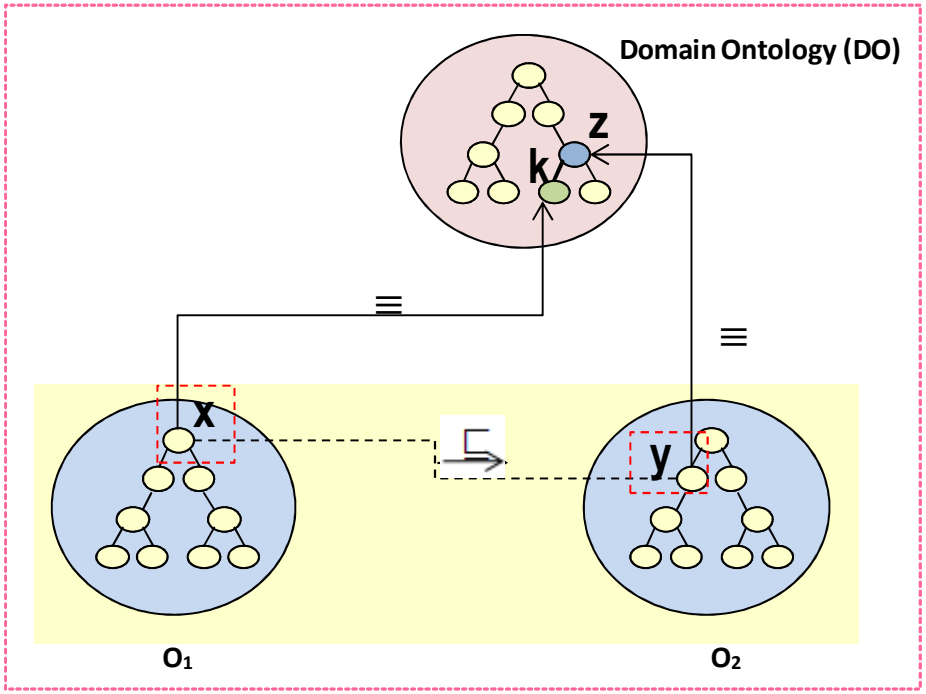
**** Education Domain ****

Working Scenario

- 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 to Define Semantic Correspondences

- Domain Ontology (DO)
  - Reliable reference available on the Web
  - Used as Background Knowledge
    - Bridge the conceptual differences or similarities between two peer ontologies



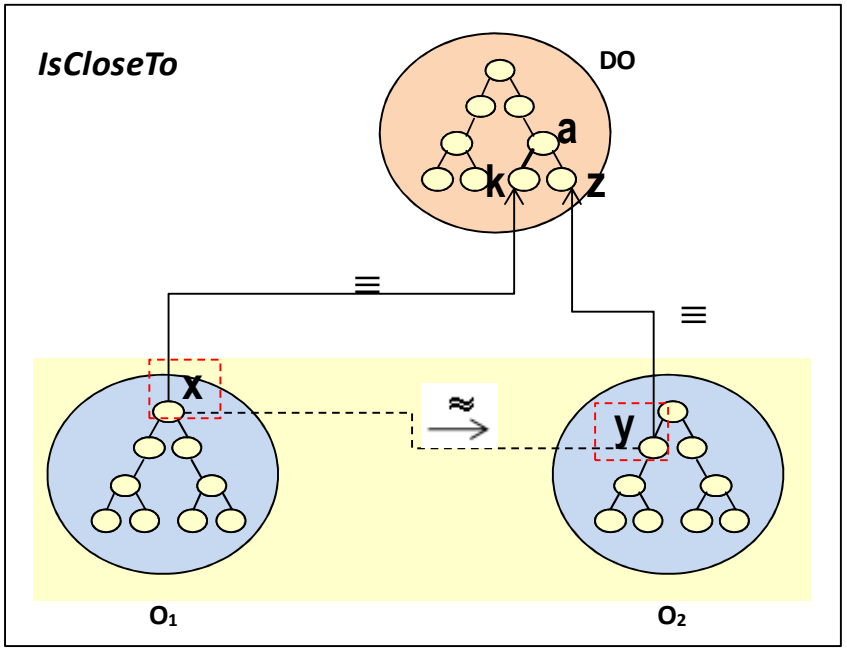
Using a Domain Ontology to Define Semantic Correspondences

- **Definition.** A semantic correspondence is represented by one of the following expressions:
  - O_1:x $\equiv\!\!\Rightarrow$ O_2:y, an *isEquivalentTo* correspondence
  - O_1:x $\sqsubseteq\!\!\Rightarrow$ O_2:y, an *isSubConceptOf* correspondence
  - O_1:x $\sqsupseteq\!\!\Rightarrow$ O_2:y, an *isSuperConceptOf* correspondence
  - O_1:x $\triangleright\!\!\Rightarrow$ O_2:y, an *isPartOf* correspondence
  - O_1:x $\triangleleft\!\!\Rightarrow$ O_2:y, an *isWholeOf* correspondence
  - O_1:x $\approx\!\!\Rightarrow$ O_2:y, an *isCloseTo* correspondence
  - O_1:x $\perp\!\!\Rightarrow$ O_2:y, an *isDisjointWith* correspondence

(*) *x* and *y* are elements belonging to the peer ontologies

Using a Domain Ontology to Define Semantic Correspondences

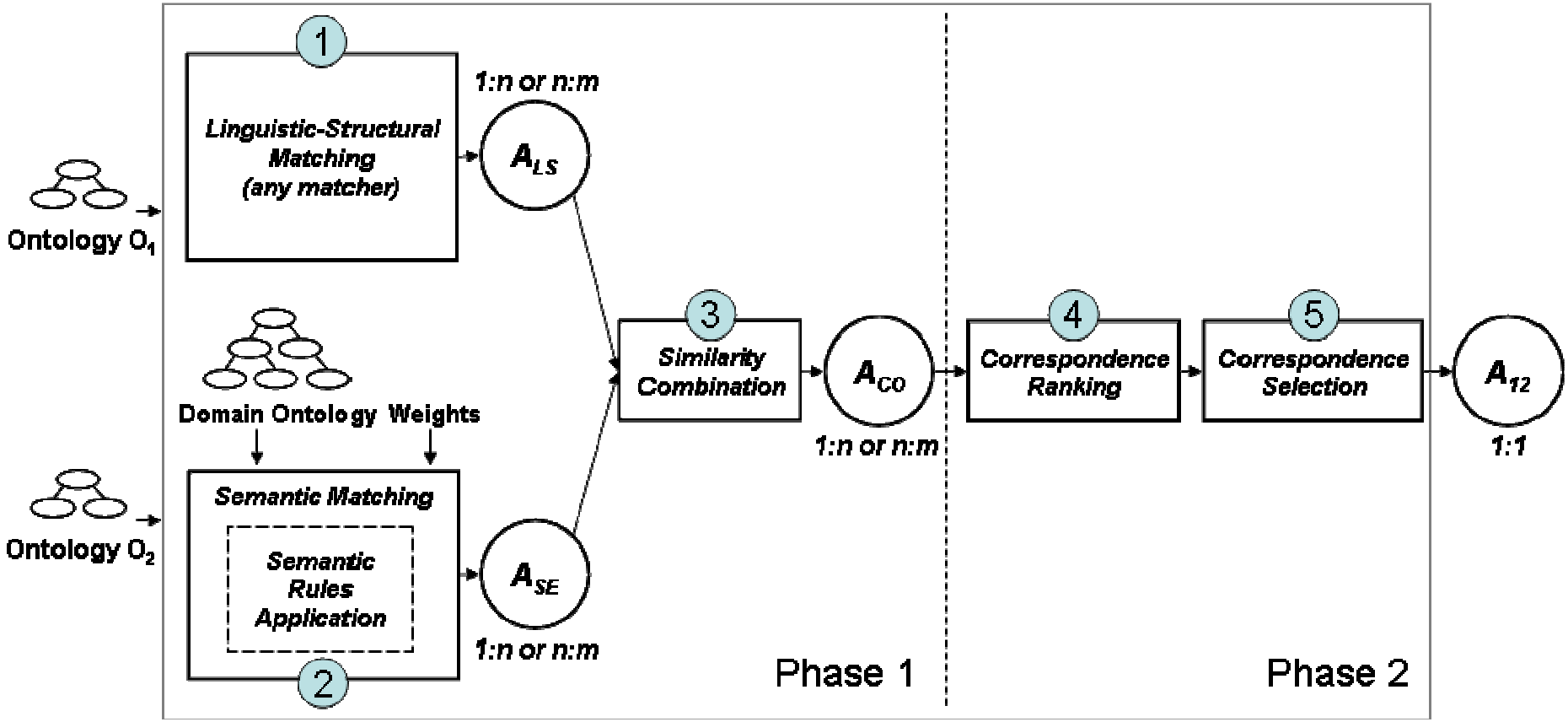
- O_1:k and O_2:z are **close** if
  - They share a common ancestor in the DO
  - The common ancestor is not the root ($\top$)
  - The concepts do not hold neither subsumption nor disjointness
  - The measured depths (*thresholdRoot* and *thresholdCommonAncestor*) are evaluated to true



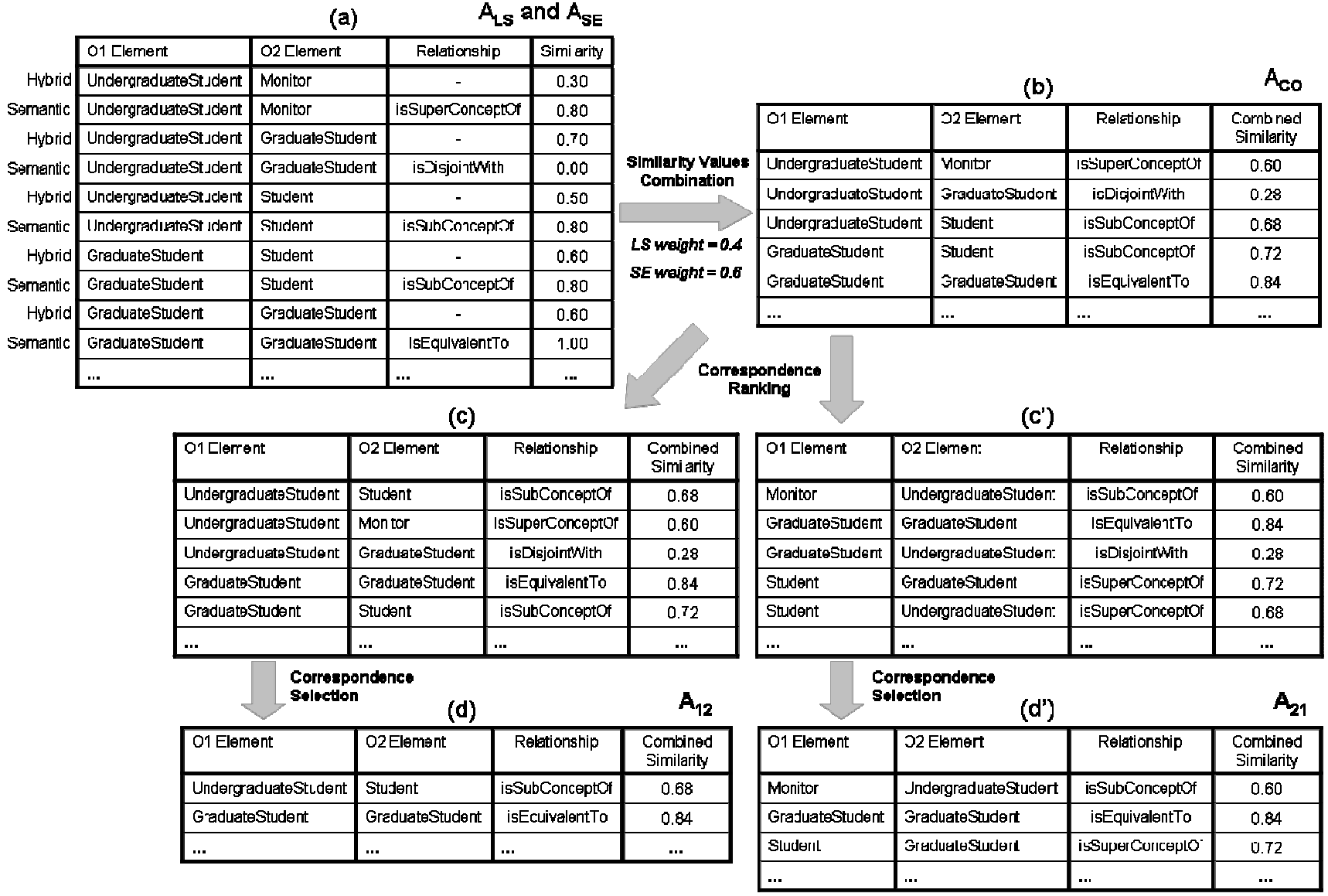
Example:

O_1.Notebook $\approx$ O_2.MacintoshPC

Semantic-based Ontology Matching Process



Semantic-based Ontology Matching Process

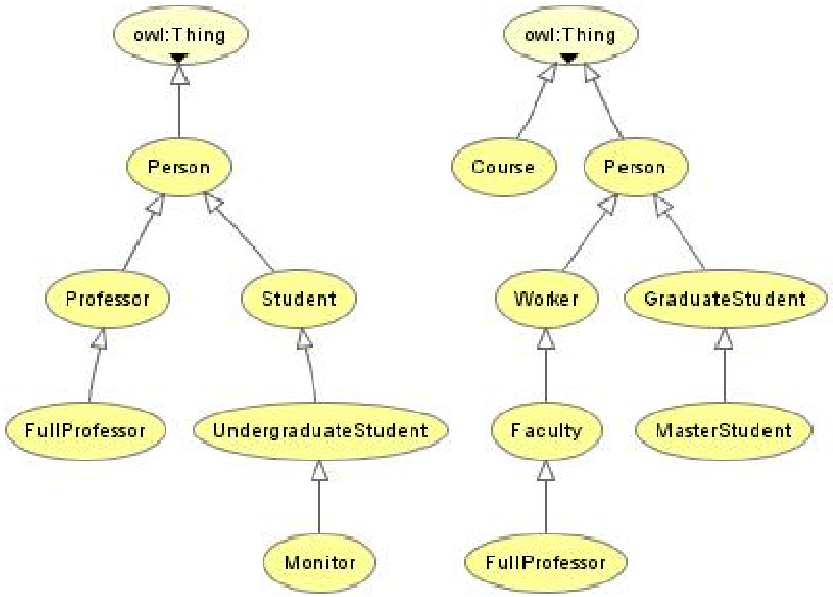


Calculating the Global Similarity Measure

- 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

$$Weighted\ Average(O_1, O_2) = \frac{\sum_{i=1}^{|A12|} n + \sum_{j=1}^{|A21|} n}{|O_1| + |O_2|}$$



Ontology O_1 **Ontology O_2**

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)

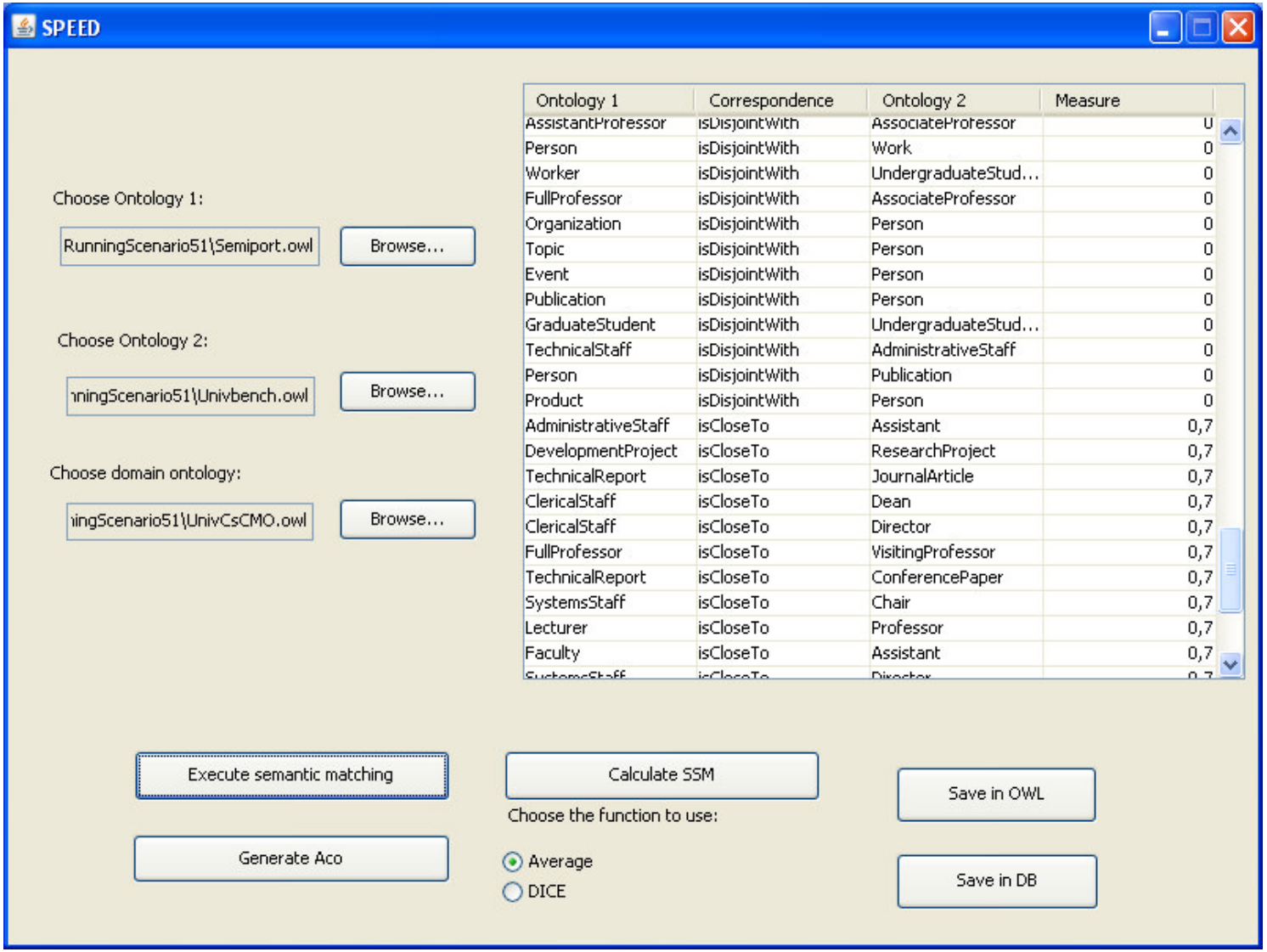
Alignment A_{21}
(1, Person, Person, isEquivalentTo, 1.0)
(2, FullProfessor, FullProfessor, isEquivalentTo, 1.0)
(3, Course, UndergraduateStudent, isWholeOf, 0.3)
(4, Worker, Person, isSubConceptOf, 0.8)
(5, GraduateStudent, UndergraduateStudent, isDisjointWith, 0.0)
(6, Faculty, Professor, isSuperConceptOf, 0.8)
(7, MasterStudent, Student, isSubConceptOf, 0,8)

$$Weighted\ Average(O_1, O_2) = \frac{(1.0+1.0+0.3+0.8+0.8)+(1.0+1.0+0.3+0.8+0.0+0.8+0.8)}{|6|+|7|} = 0.66$$

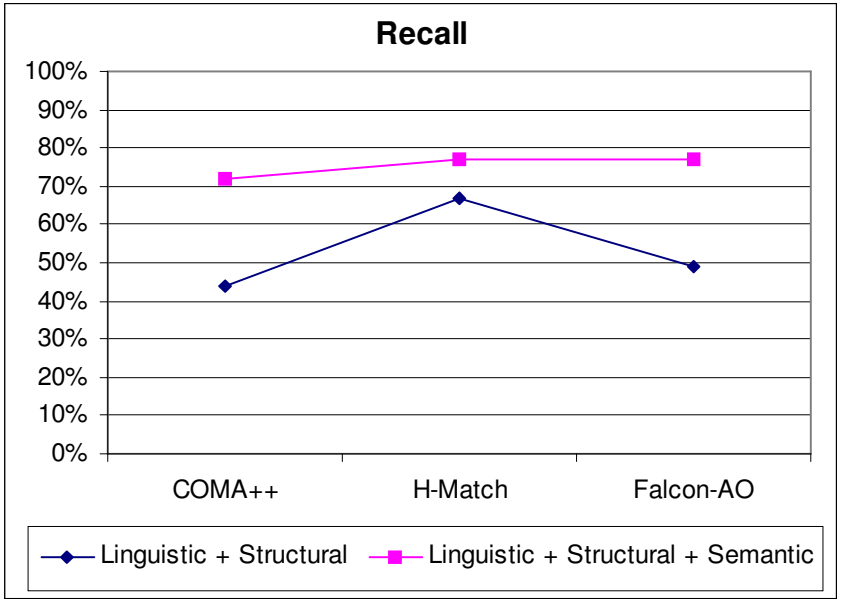
Experiments and Results

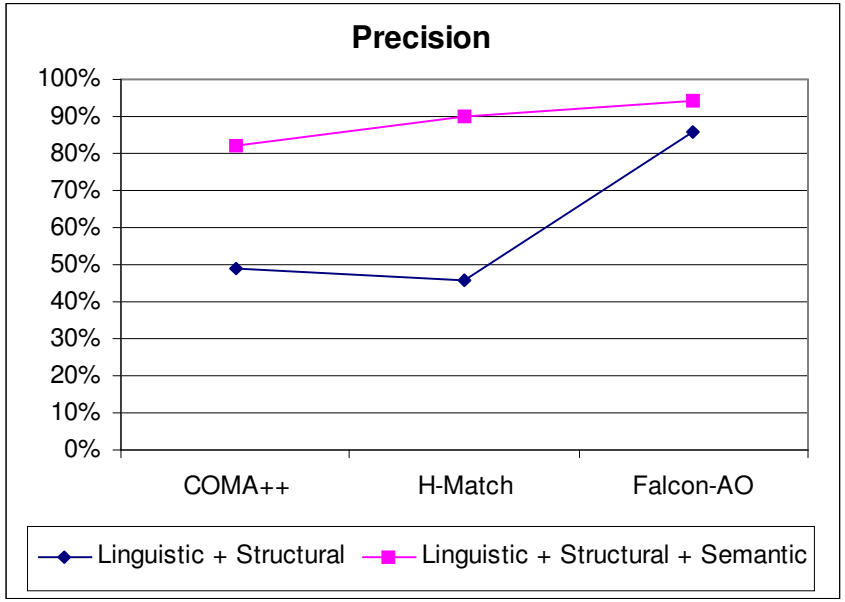
- The semantic-based ontology matching tool
  - Implemented in Java
  - Jena has been used to provide ontology manipulation and reasoning
  - H-Match has been used as the hybrid matcher
- Correspondence identification has been restricted to concepts
  - Properties are not included
- http://www.cin.ufpe.br/~speed/SemMatch/index.htm

Experiments and Results



Experiments and Results





Correspondences for O_1:Faculty	
O_1:Faculty $\equiv$ O_2:Faculty	O_1:Faculty $\sqsupset$ O_2:PostDoc
O_1:Faculty $\sqsubset$ O_2:Worker	O_1:Faculty $\approx$ O_2:Assistant
O_1:Faculty $\sqsupset$ O_2:Professor	O_1:Faculty $\approx$ O_2:AdministrativeStaff

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

Related Work

- 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

DEXA 2009

GLOBE 2009

A Semantic-based Ontology Matching Process for PDMS

Carlos Eduardo Pires[1], Damires Souza[2], Thiago Pachêco[1], and Ana Carolina Salgado[1]

[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

VIRTUS IMPAVIDA

Centro de Informática U·F·P·E

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA PARAÍBA