Opening the Black Box of Ontology Matching

Hoa Ngo, Zohra Bellahsene, Konstantin Todorov

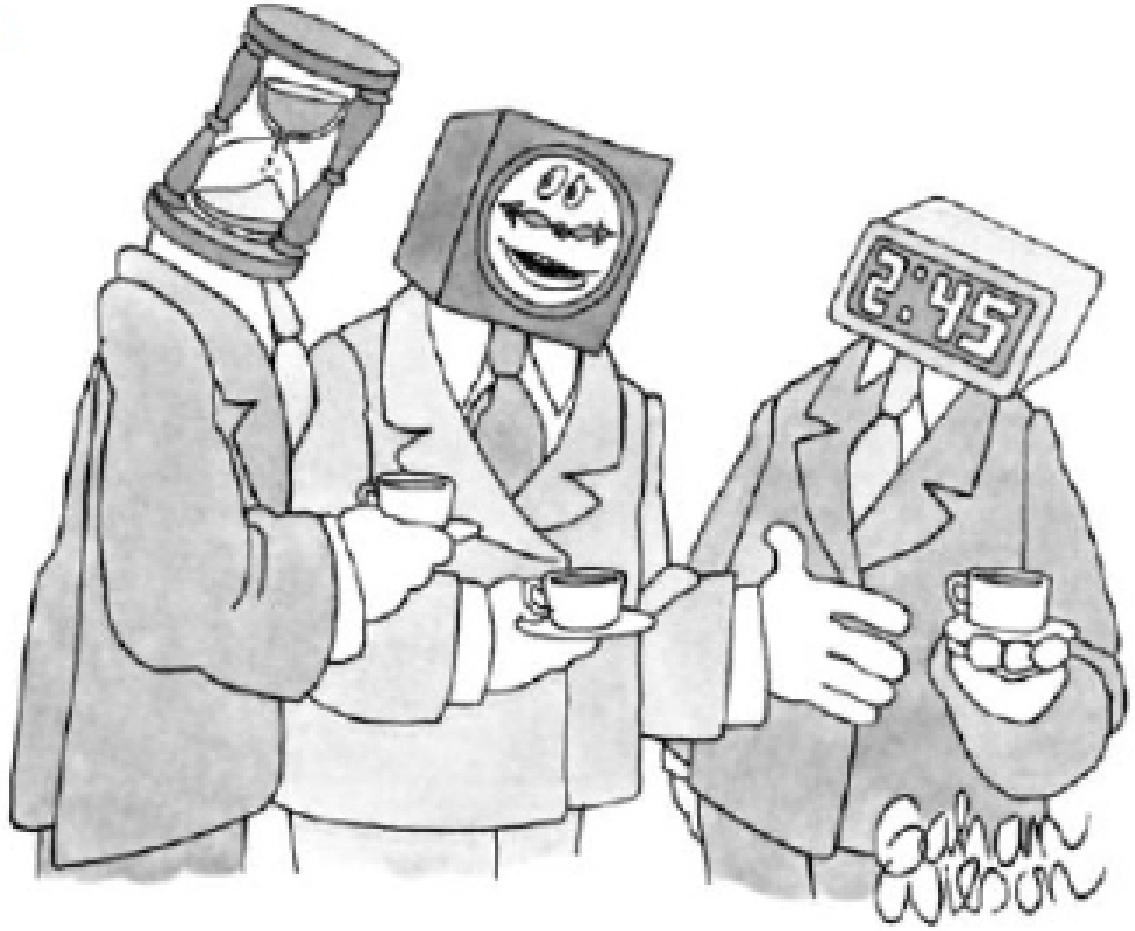
{firstname.lastname}@lirmm.fr

LIRMM / University of Montpellier 2

ESWC, Montpellier 2013

Outline

Outline



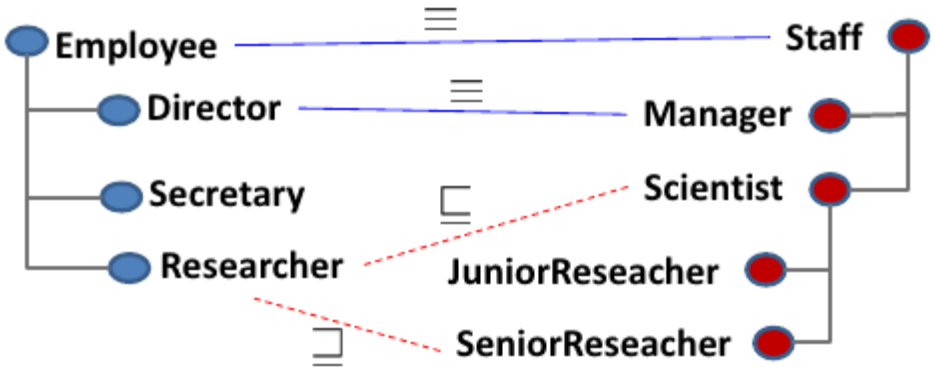
"Basically, we're all trying to say the same thing."

Borrowed by a tutorial by S. Staab and A. Hotho.

Ontologies are created in a **decentralized**, strongly **human biased** manner.
Many ontologies describing the same domain of interest

=> **ontology heterogeneity:**

- syntactic
- terminological
- conceptual / structural



=> **Ontology Matching:** detect the semantic correspondences between the elements of two ontologies.

A Generic Framework for Ontology Matching and Evaluation

Matching and Evaluation Framework

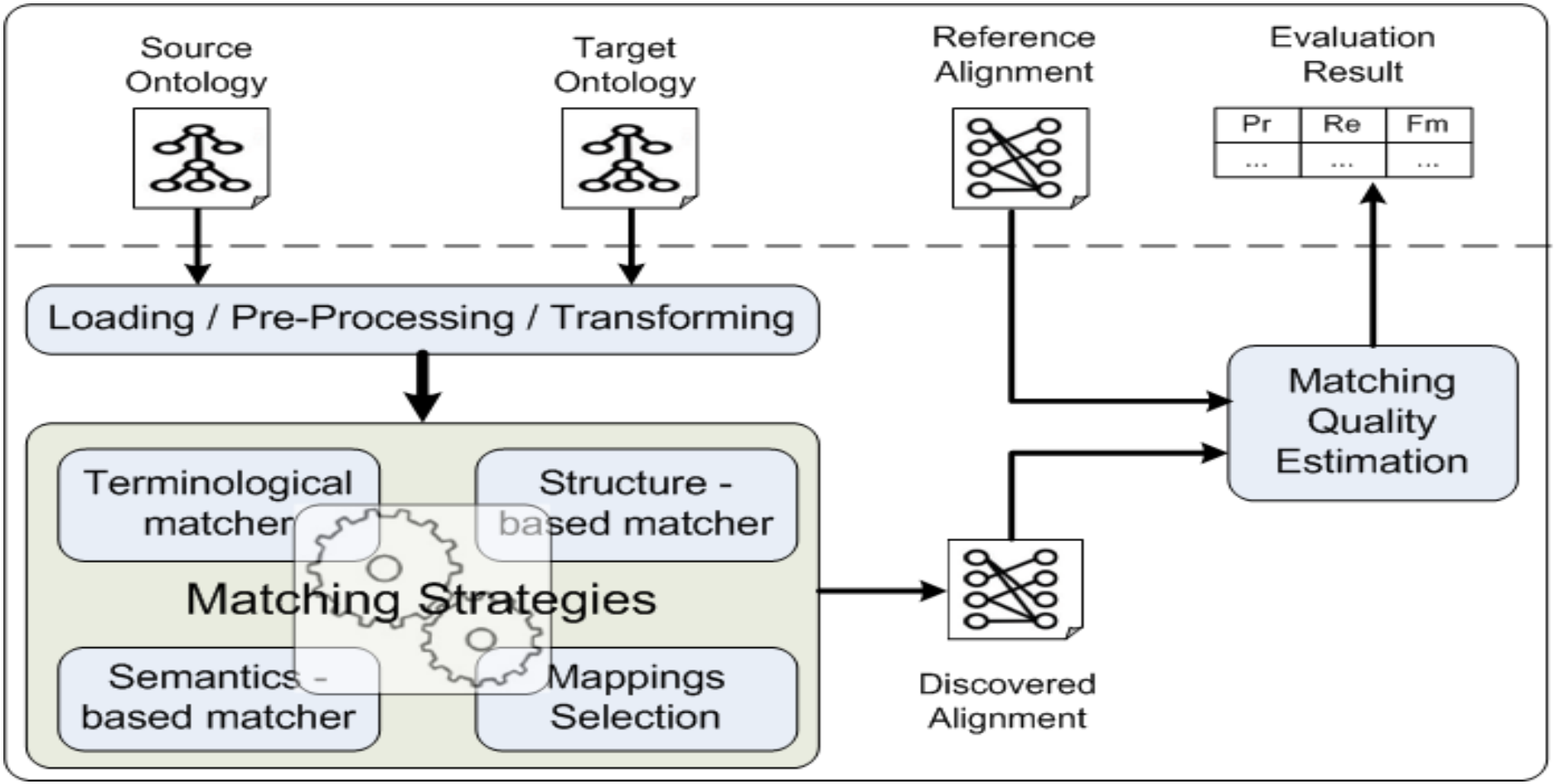


Figure : Ontology Matching: System Architecture and Evaluation Scenario

On n tests, we compute:

$$H(p) = \frac{\sum_{i=1}^{n} |C_i|}{\sum_{i=1}^{n} |A_i|}; \quad H(r) = \frac{\sum_{i=1}^{n} |C_i|}{\sum_{i=1}^{n} |R_i|}; \quad H(fm) = \frac{2 * H(p) * H(r)}{H(p) + H(r)}.$$

For the ith test:

- $|A_i|$ – the total number of mappings discovered by a matching system,
- $|C_i|$ – the number of correct mappings,
- $|R_i|$ – the number of reference mappings (expert).

Outline

Terminological Matchers and Mapping Selection

Methods and Evaluation Strategy

Goal:

- Study the interaction between the mapping selection module and terminological matchers
- Compare global vs. local methods

Dataset:

- Conference dataset from OAEI[1], 21 test cases
- Moderate-size real world ontologies, terminologically highly heterogeneous, describe the same domain

---

Interaction scheme: Terminological matchers <-> mapping selection

[1] The Ontology Alignment Evaluation Initiative

Terminological Matchers and Mapping Selection

Methods and Evaluation Strategy

Local methods (similarity of individual entities)

- *Edit distance-based methods*. Levenstein and ISUB
- *Token-based methods*. QGrams and TokLev (using Levestein to compare tokens)
- *Hybrid methods.* HybLinISUB and HybJCLev

Global methods (combination of local methods)

- *Weighted Average with Local Confidence (LC)*
- *Harmony-based Adaptive Similarity Aggregation (HADAPT)*
- *Machine Learning-Based Approach (ML)*
  (training data: OAEI Benchmark 2009 and I3CON)
- *Information Retrieval-Based Approach (IR)*

Terminological Matchers and Mapping Selection

Results

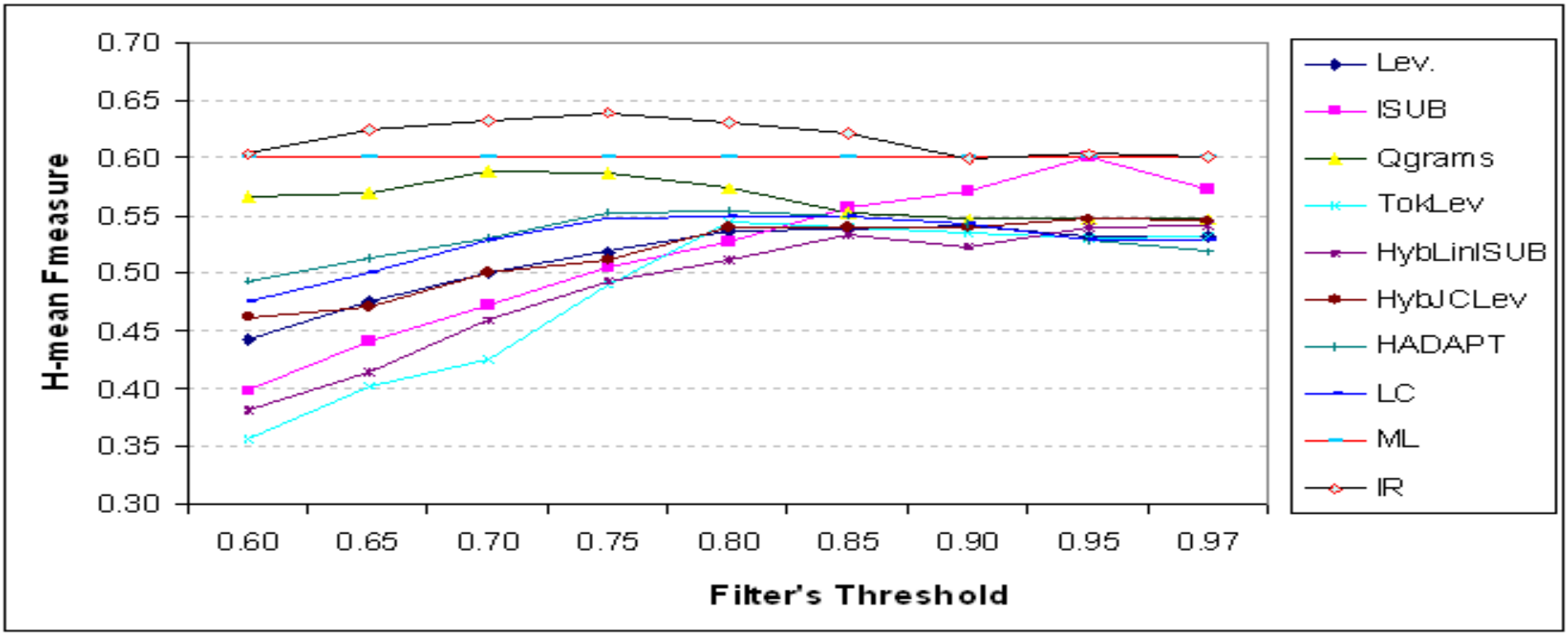


Figure : Mapping Selection for the Terminological Matcher Module

Outline

Structural Matchers and Mapping Selection

Methods and Evaluation Strategy

Goal:

- Study the behavior of structural matchers with respect to different settings of the mapping selection module

Dataset:

- Benchmark 2011 dataset from the OAEI campaign, 103 test cases
- Dataset construction: modification of label names and ontology structure

---

Interaction scheme: Terminological matcher (identical metric) ->
structural matcher <-> mapping selection

Structural Matchers and Mapping Selection

Methods and Evaluation Strategy

- Standard structural methods
  - exploring standard structural relations between entities within the ontologies:
    descendants, ancestors, leaves, adjacent, etc.
  - relying on already discovered similarities

- SP (Similarity Propagation),
  - extends the similarity flooding algorithm
  - relies on directed relations in an ontology

Structural Matchers and Mapping Selection

Results

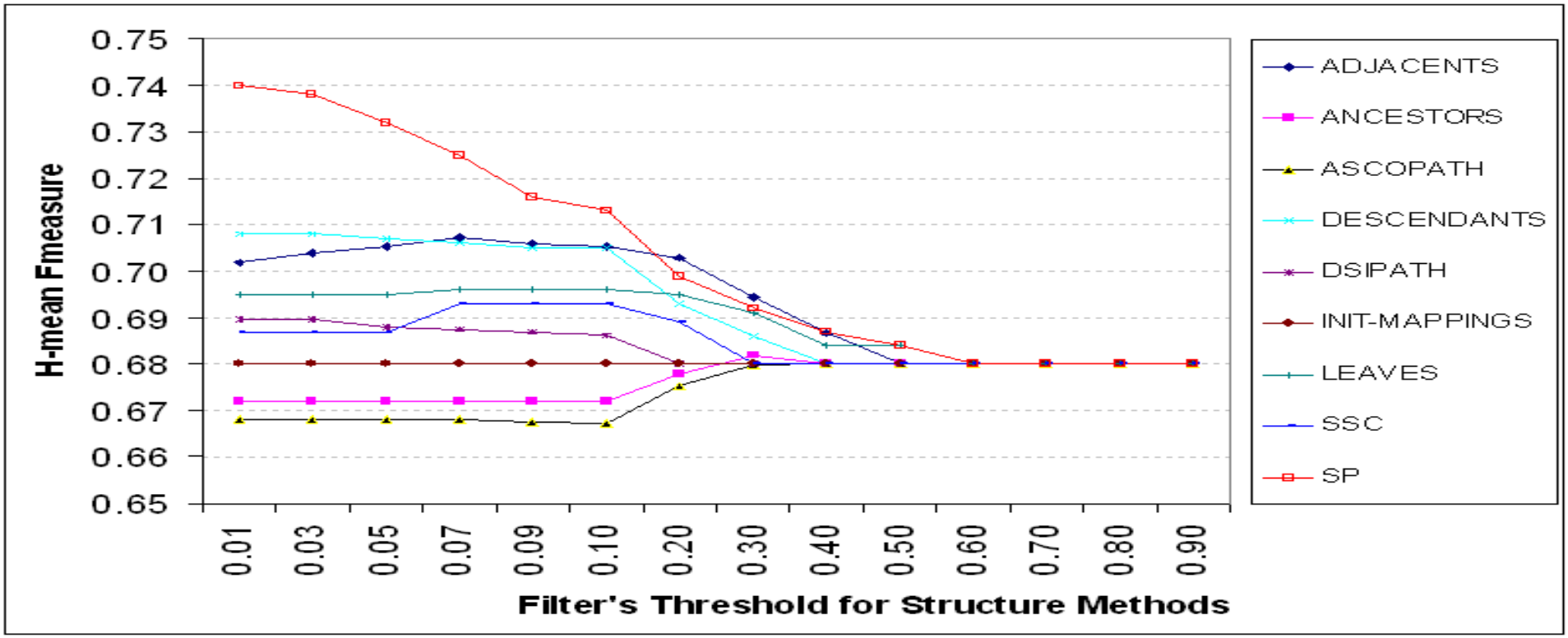


Figure : Mapping Selection for Structural Methods

Outline

Impact of Noisy Input on Structural Matchers

Methods and Evaluation Strategy

Goal:

- Evaluate the behavior of different structural matchers when we add noise into the input mappings (coming from a terminological matcher)

Noise:

- a pair of entities falsely labeled as a "match".

Interaction scheme: Noise -> Terminological Matcher -> Structural Matcher

Impact of Noisy Input on Structural Matchers

Methods and Evaluation Strategy

- At ***terminological level***:
  - Similarity measure: identical metric
  - Adding noise: a number of random incorrect mappings, a portion the original init mappings
- At ***structural level***:
  - take input from the terminological matcher
  - select the best threshold filter for each structural method (according to previous experiments).
- ***Dataset:***
  - Benchmark 2011 dataset, 103 test cases
  - At each iteration, count the total number of correct mappings and the total number of incorrect mappings

Impact of Noisy Input on Structural Matchers

Results

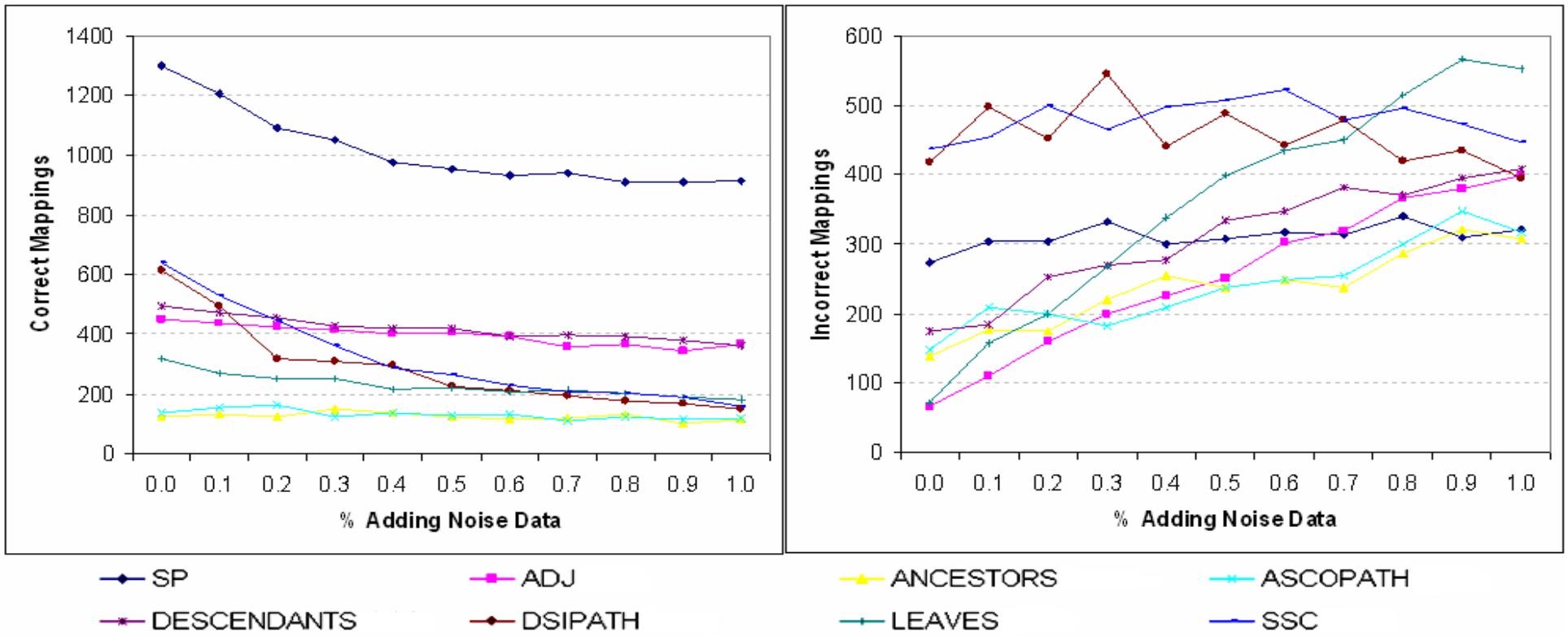


Figure : Impact of input noise on structural matchers.

Outline

Terminological vs. Structural and Semantic Matchers

Methods and Evaluation Strategy

Goal:

- Study the performance of terminological methods when used alone and when used as an input for structural and semantic methods.
- Identify the terminological matchers which provide best performance of the structural and the semantic methods for a given mapping selection threshold

Dataset:

- Conference from OAEI

Interaction scheme: Terminological Matcher -> Structural Matcher
Terminological Matcher -> Semantic Matcher

Terminological vs. Structural and Semantic Matchers

Methods and Evaluation Strategy

- At ***terminological level*** – three different methods to produce initial mappings:
  - QGrams (representing token-based methods); ISUB (for edit-based methods); IR (for global methods).
- At ***structural level*** – the SP method
  - Best performing among the structural matchers (previous exp.)
- At ***semantic level*** – the global diagnosis optimization method
  - refines input terminological mappings in order to remove inconsistent ones

Terminological vs. Structural and Semantic Matchers

Results I

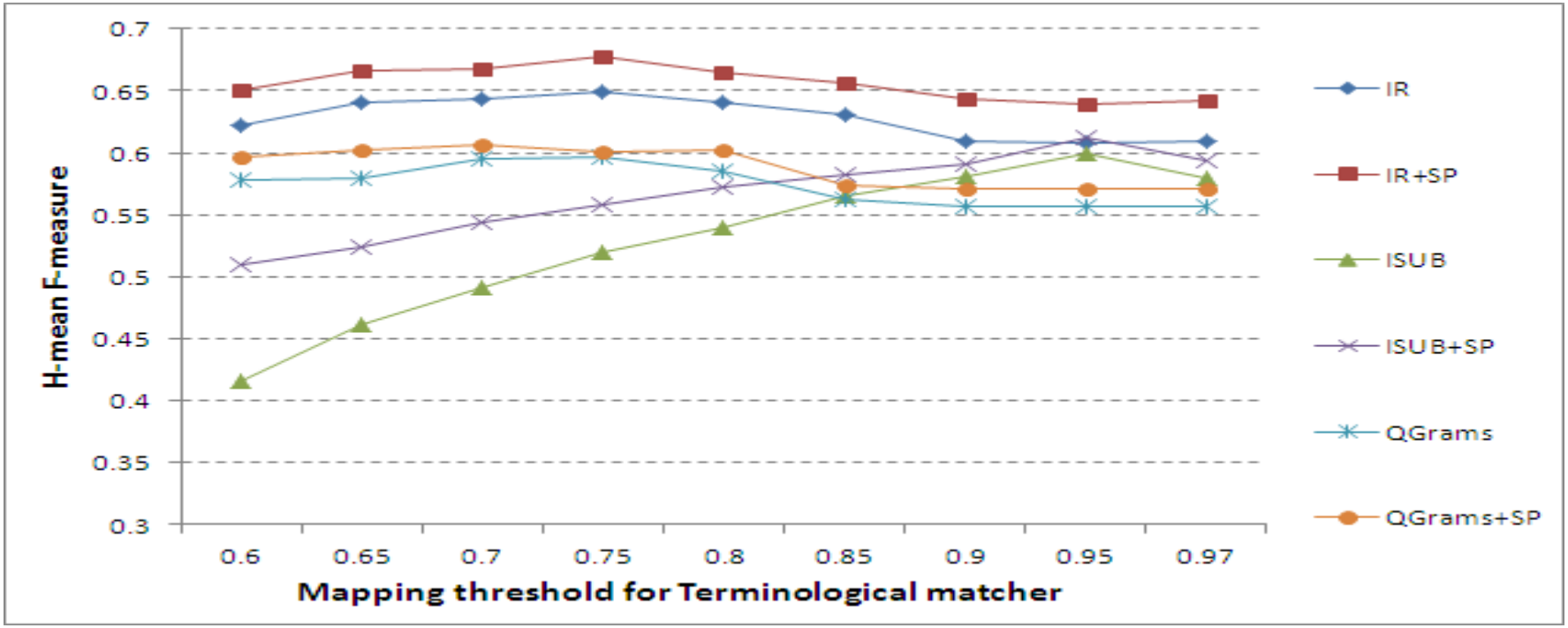


Figure : Interaction of terminological methods with a structural matcher (SP) w.r.t. different values of the mapping selection filters.

Terminological vs. Structural and Semantic Matchers

Results II

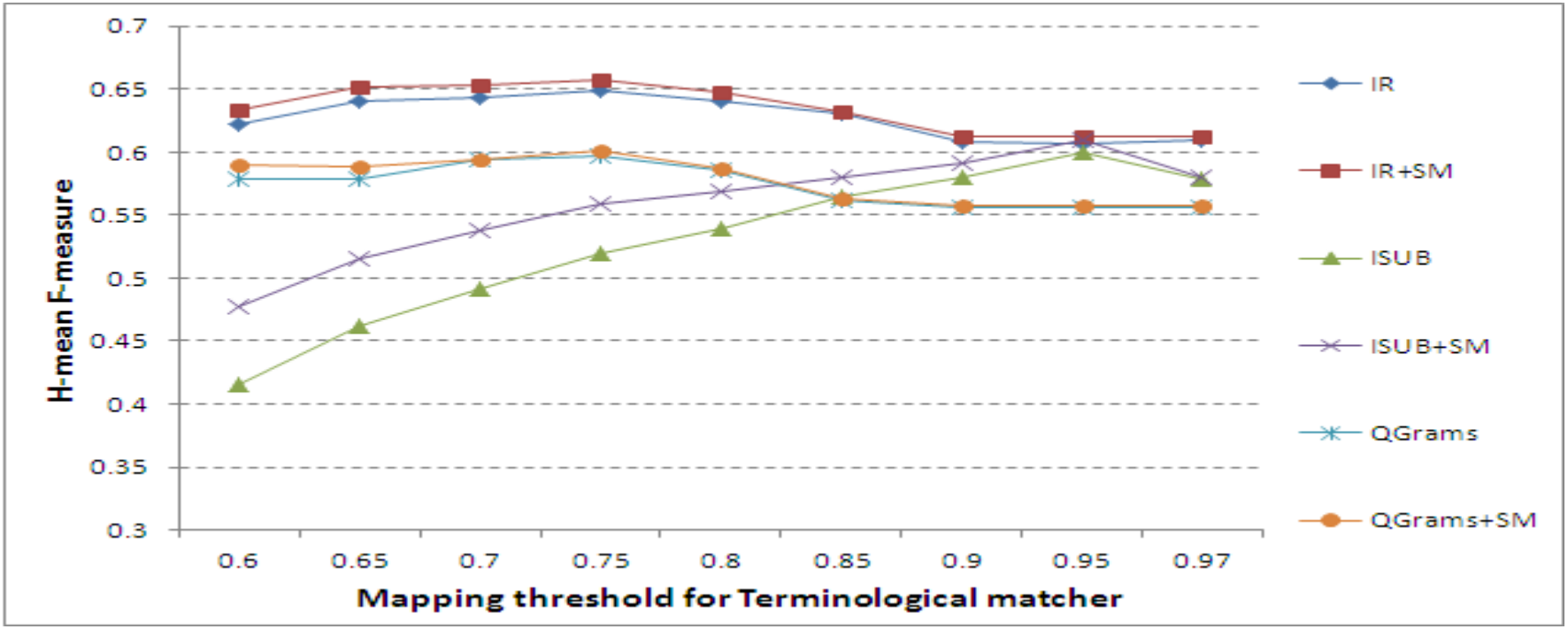
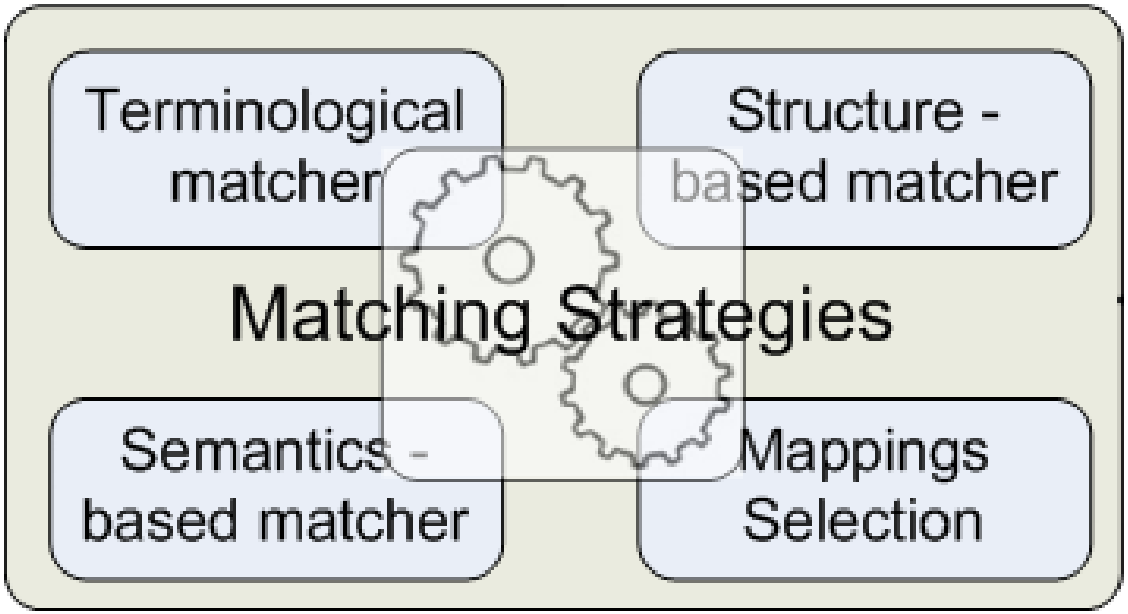


Figure : Interaction of terminological methods with a semantic matcher (SM) w.r.t. different values of the mapping selection filters.

Outline

Conclusion

- An OM system: combines several matching components
- These components interact with one another
- Understanding these interactions =>
  matcher selection and combination, parametrizing the OM tool



Thank you for listening.