# Tutorial on Schema and Ontology Matching


Pavel Shvaiko — pavel@dit.unitn.it

Jérôme Euzenat — Jerome.Euzenat@inrialpes.fr

UNIVERSITÀ DEGLI STUDI DI TRENTO
Dipartimento di Informatica e Telecomunicazioni

INRIA RHÔNE-ALPES


knowledgeweb
realizing the semantic web

# Goals of the tutorial

- Illustrate the role of schema/ontology matching
- Provide an overview of the basic matching techniques
- Demonstrate the use of basic matching techniques in state of the art systems
- Motivate the future research

# Outline

- **Matching problem**
- Classification of schema-based matching techniques
- Basic techniques
- Matching process
- Review of the matching systems
- Conclusions

knowledgeweb
realizing the semantic web

# Motivations

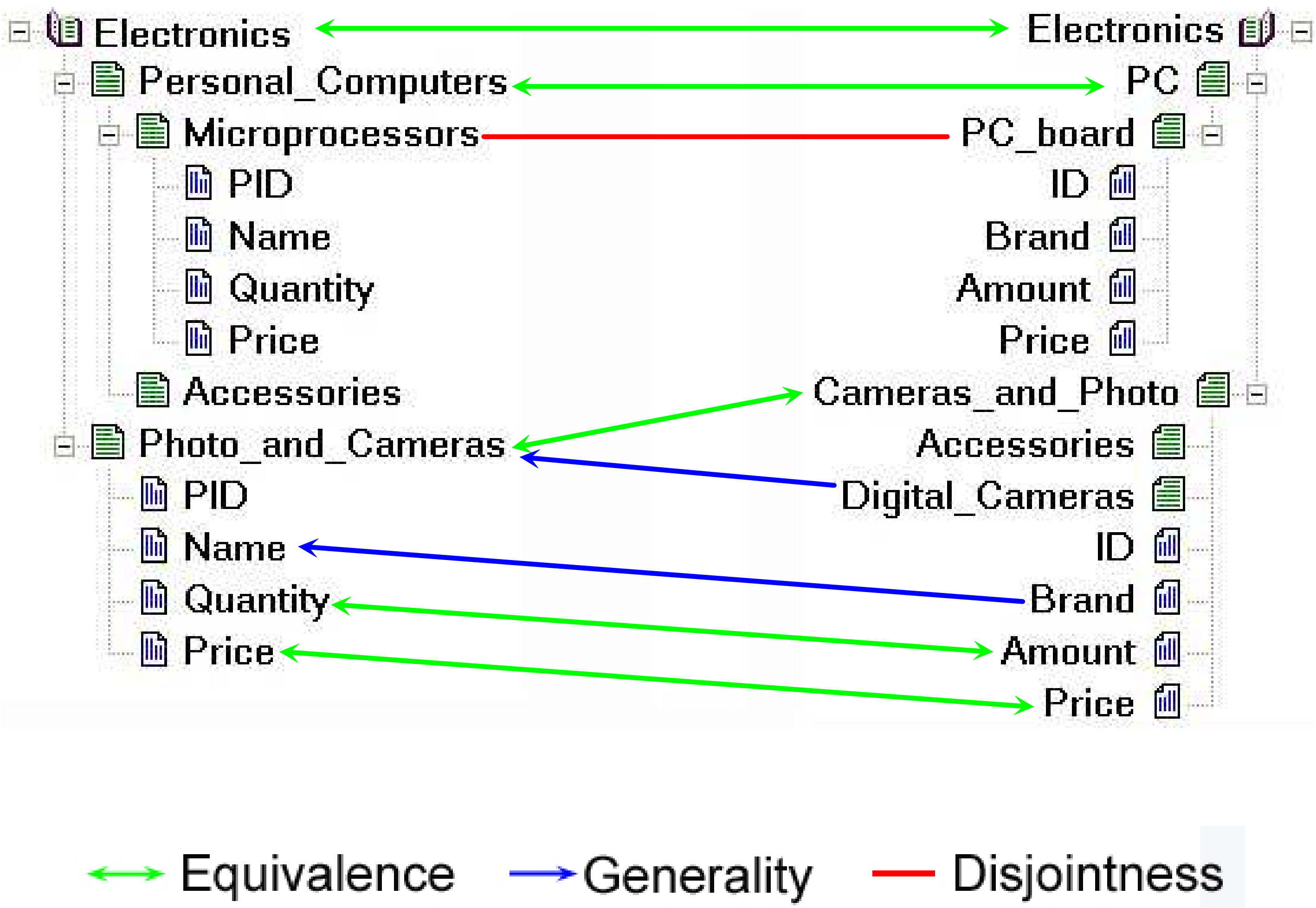
**Match operator**

takes two schemas/ontologies, each consisting of a set of discrete entities (e.g., tables, XML elements, classes, properties) as input and determines as output the relationships (e.g., equivalence, subsumption) holding between these entities

knowledgeweb
realizing the semantic web

# Motivations

## Two XML schemas




knowledgeweb
realizing the semantic web

# Motivations

## Two relational schemas

STAFF

| Name | Address | Tel_No | e-mail |
|---|---|---|---|
| John Dow | 12 Well St, Glasgow | 0141-123-4567 | john@aol.com |
| Mike O' Neill | 37 Achray St, Glasgow | 0141-987-6543 | mike@aol.com |

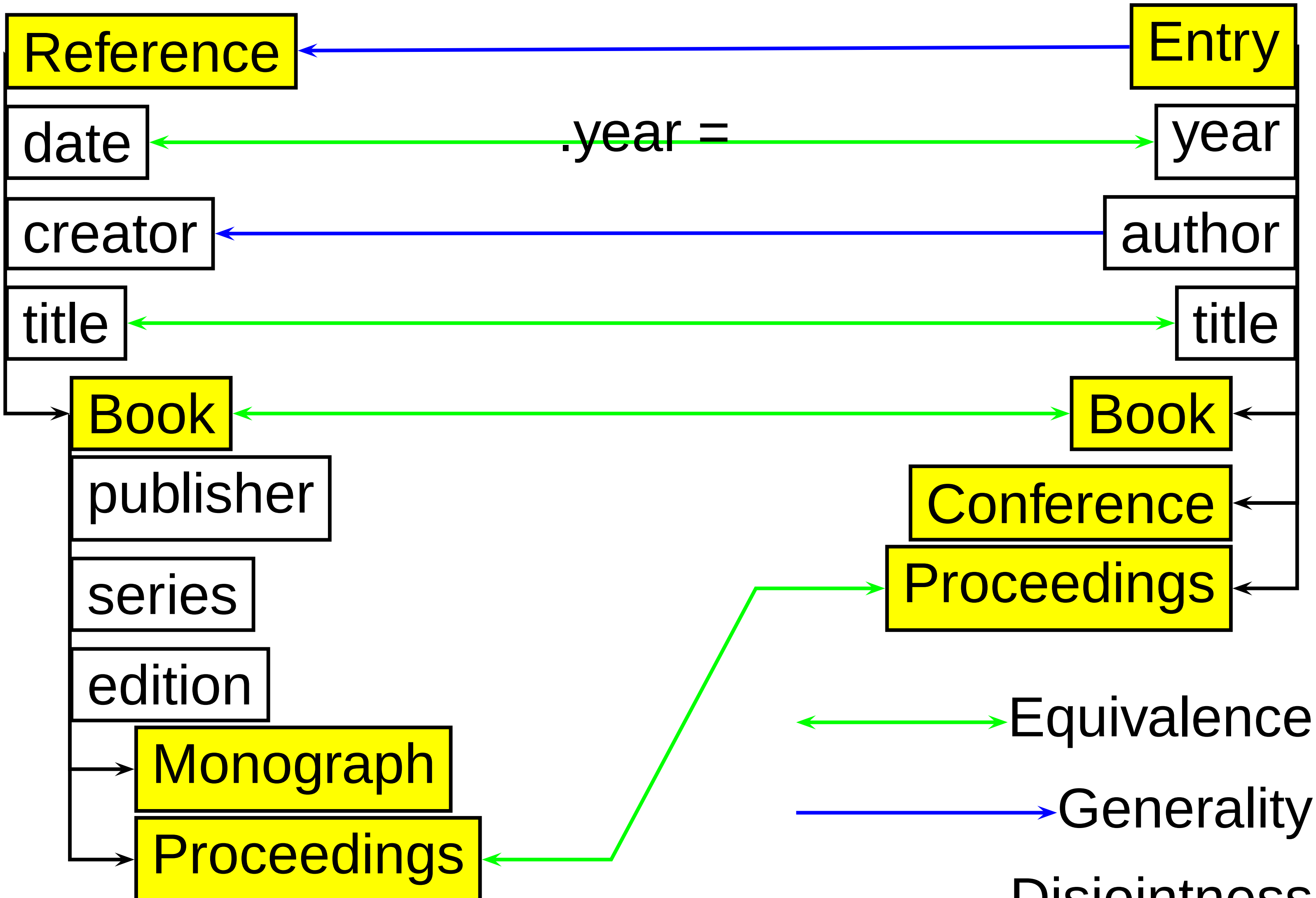
PERSONAL

| FirstName | LastName | Address | Telephone |
|---|---|---|---|
| Karen | Shaw | 31 High St, London | 0171-456-9876 |
| Tina | Craig | 12 Argyll St, London | 0171-664-5138 |

Equivalence Generality Disjointness

# Motivations

**Two ontologies**



# Schema matching vs. Ontology matching

**Differences:**

- Schemas often do not provide explicit semantics for their data
  - Relational schemas provide no generalization
- Ontologies are logical systems that constrain the meaning
  - Ontology definitions as a set of logical axioms

knowledgeweb
realizing the semantic web

# Schema matching vs. Ontology matching

**Commonalities:**

- Schemas and ontologies provide a vocabulary of terms that describes a domain of interest
- Schemas and ontologies constrain the meaning of terms used in the vocabulary

Techniques developed for both problems are of a mutual benefit

# Statement of the problem

## Scope

- Reducing heterogeneity can be performed in 2 steps
  - **Determine the alignment** (*matching*)
  - Process the alignment (merging, transforming, etc.)
- When do we match?
  - Design time
  - **Run time**

# Statement of the problem

**Mapping element $M$ is a 5-uple:** $\langle id, e, e', R, n \rangle$

- $id$ is a unique identifier of the given mapping element
- $e$ and $e'$ are entities (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)

knowledge**web**
realizing the semantic web

# Statement of the problem

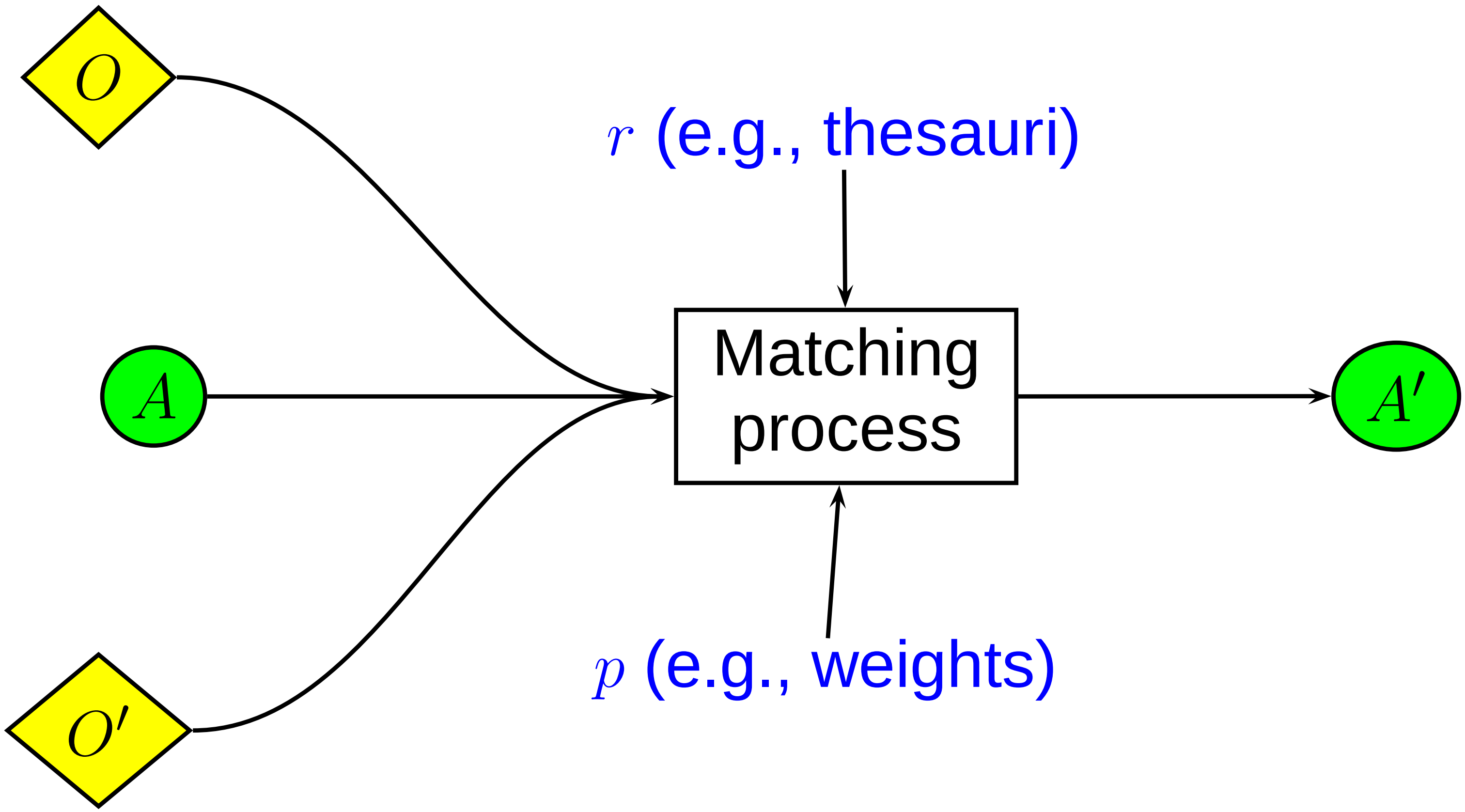
**Alignment ($A$)**

- is a set of mapping elements
- depending on the two schema/ontologies
- with some multiplicity: 1-1, 1-*, etc.
- and some other properties (complete)

knowledgeweb
realizing the semantic web

# Statement of the problem

## Matching process

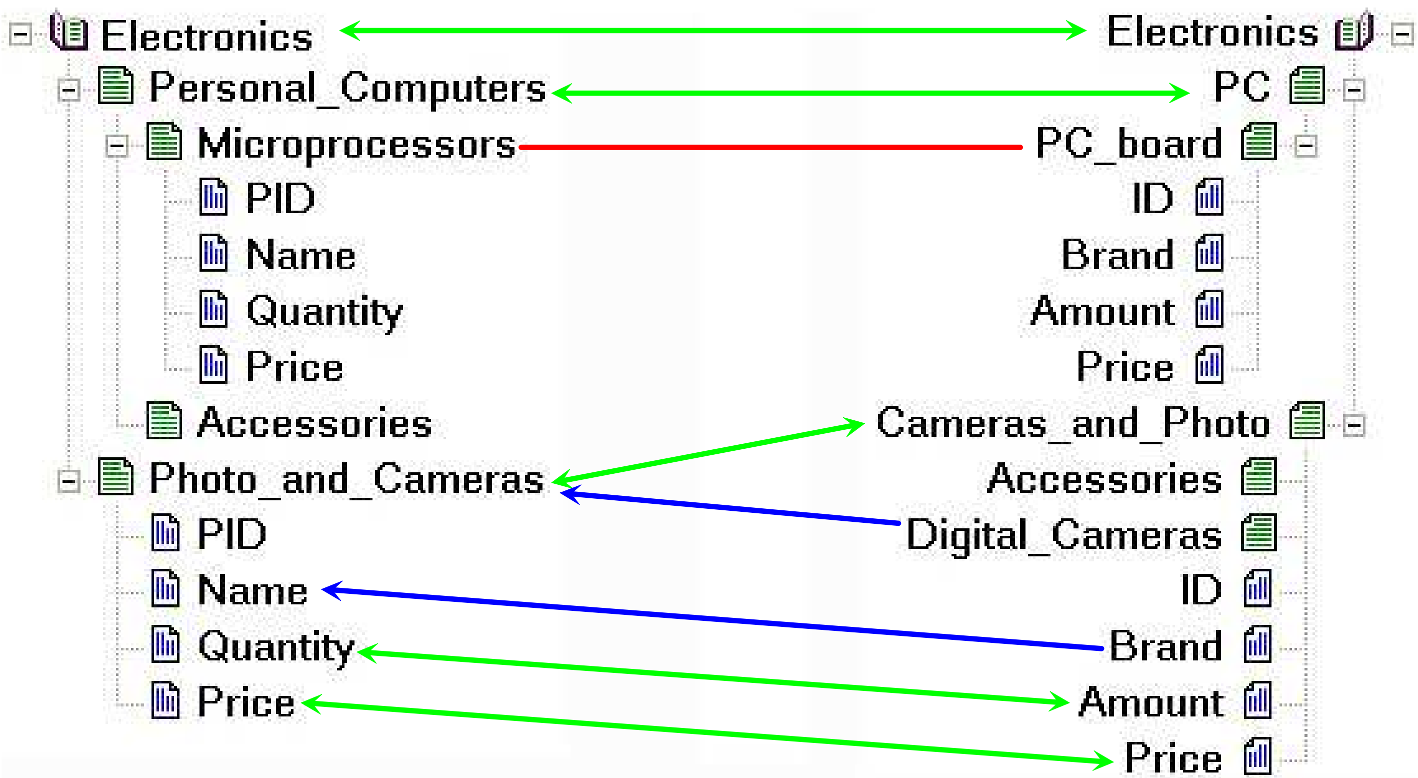


# Application domains

- **Traditional**
  - Schema integration
  - Data warehouses
  - Mediator generation
- **Emergent**
  - P2P databases
  - Agent communication
  - Web services integration

# Application domains

## Schema integration: catalog matching



# Application domains

## Schema integration: catalog matching

- In order for a private company to participate in the marketplace (e.g., eBay), it has to determine correspondences between entries of its catalogs and entries of a common catalog of a marketplace
- Once the correspondences between two schemas have been determined, the next step is to generate query expressions that automatically translate data instances of these catalogs under an integrated catalog
- Having aligned the catalogs, users of a marketplace have a unified access to the products which are on sale

knowledge**web**
realizing the semantic web

# Application domains

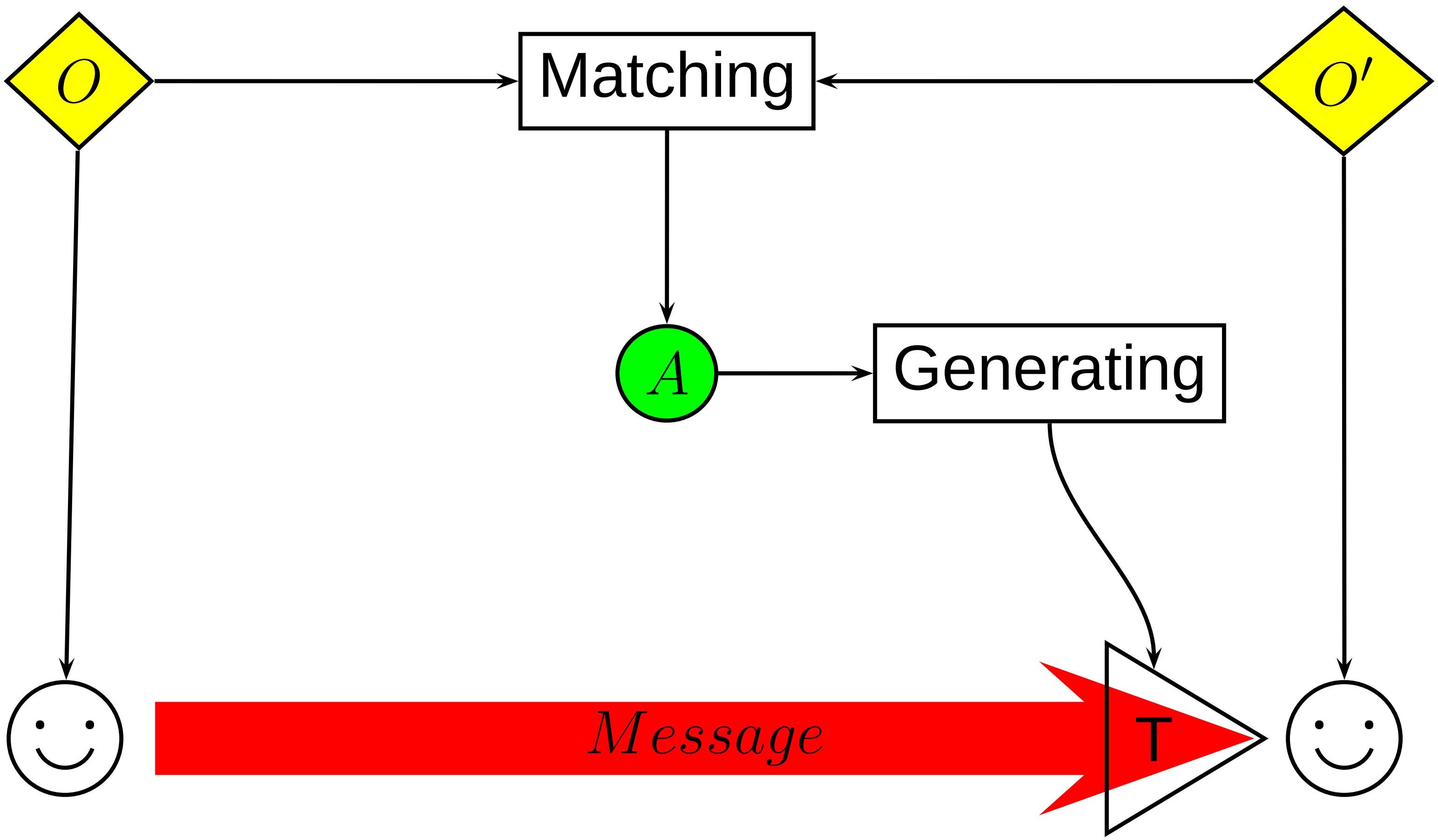
## P2P databases

- Peers are autonomous
  - They appear and disappear on the network
  - They use different terminology
- Matching (on-the-fly)
  - Determine the relationships between peer schemas
  - Use these relationships for query answering
  - An assumption that all peers rely on one global schema, as in data integration, can not be made, because the global schema might need to be updated any time the system evolves

knowledge**web**
realizing the semantic web

# Application domains

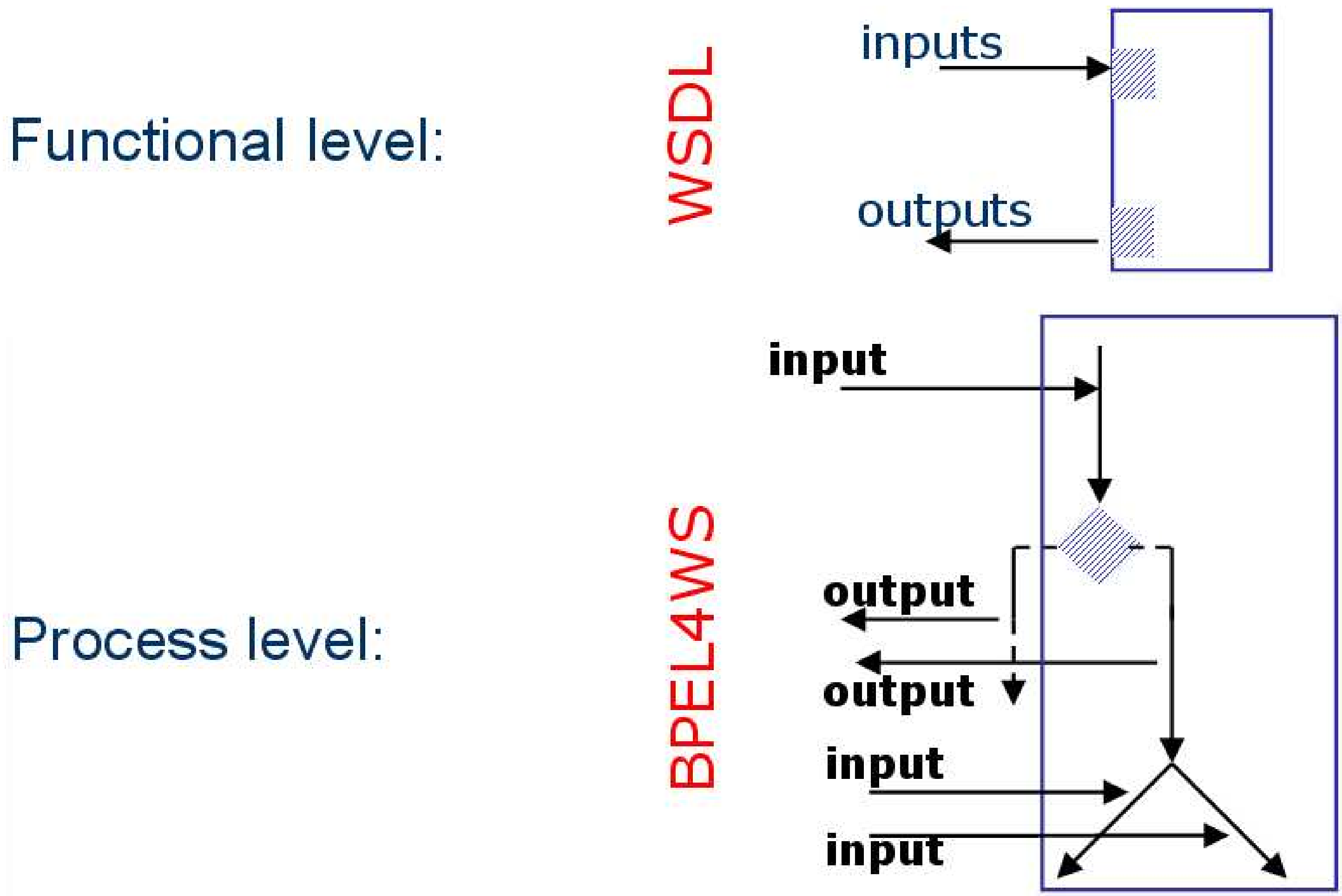
## Agent communication




knowledgeweb
realizing the semantic web

# Application domains

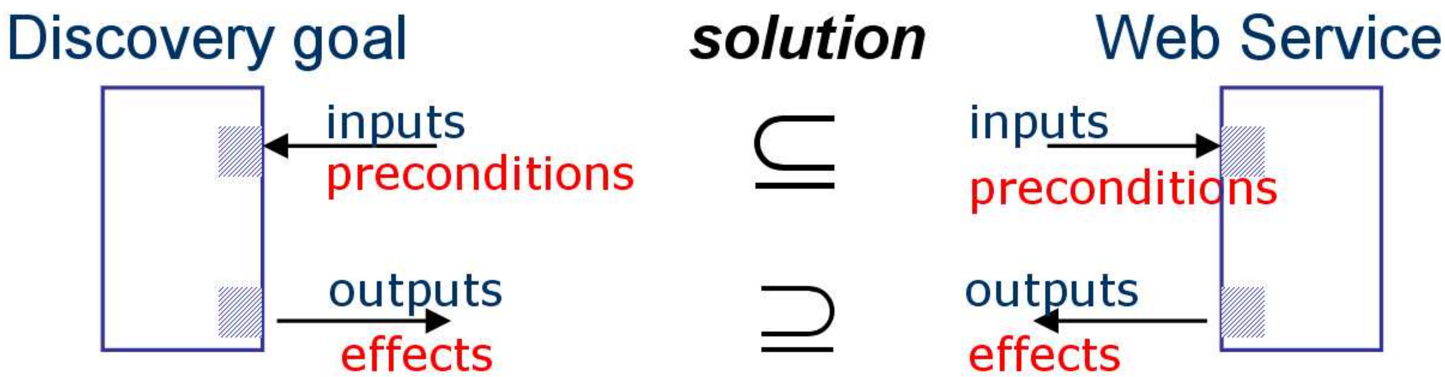
## Web services integration



# Application domains

## Web services integration

- Matching




- Executing the alignment
  - Generate a mediator able to transform the output of the first service in order to be input to the second one

# Outline

- Matching problem
- **Classification of schema-based matching techniques**
- Basic techniques
- Matching process
- Review of the matching systems
- Conclusions

# Matching dimensions

- Input dimensions
  - Underlying data models (e.g., XML, OWL)
  - **Schema-level** vs. Instance-level
- Process dimensions
  - Approximate vs. Exact
  - Interpretation of the input
- Output dimensions
  - Cardinality (e.g., 1:1, 1:m)
  - Equivalence vs. Diverse relations (e.g., subsumption)
  - Graded vs. Absolute confidence

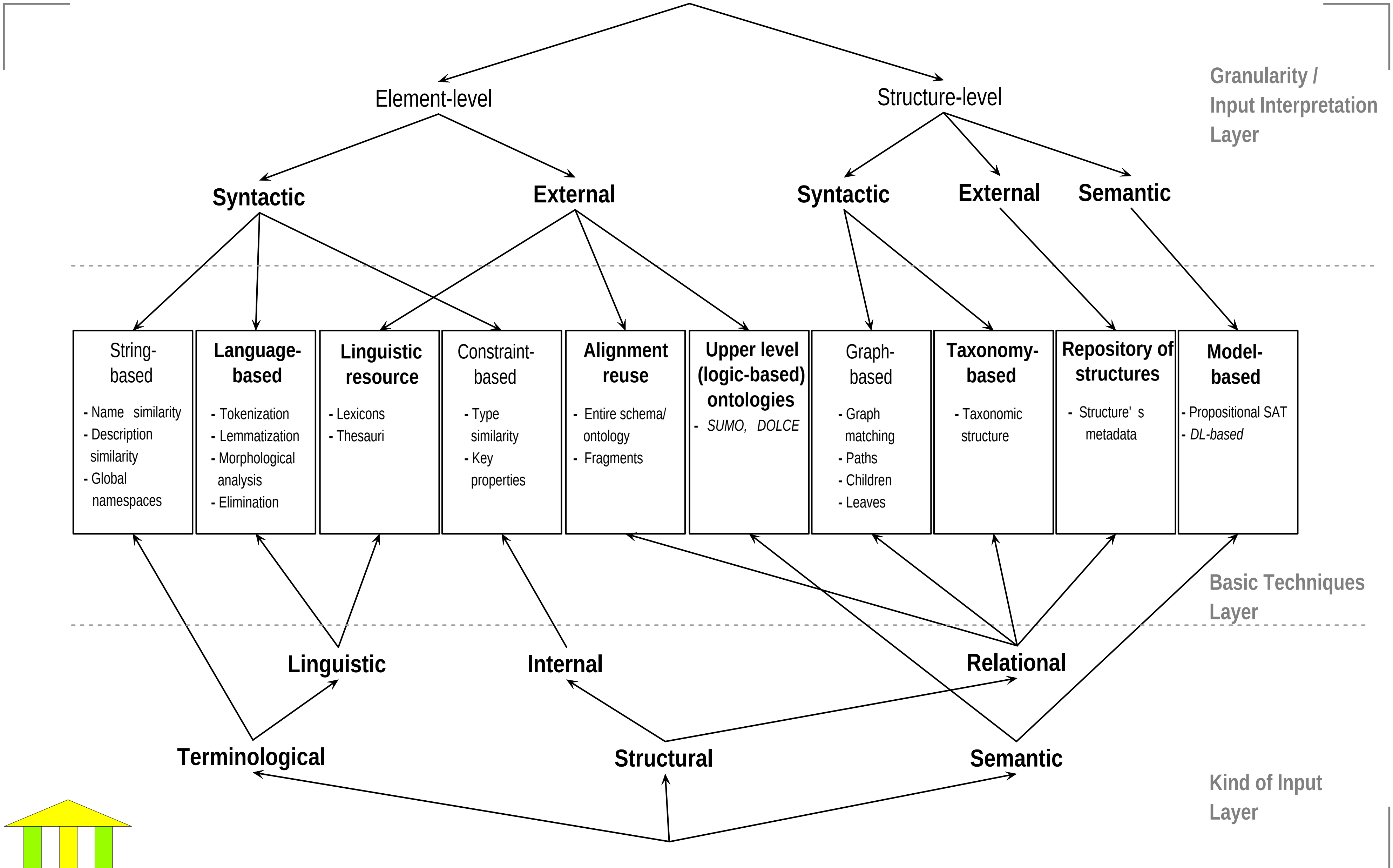
knowledgeweb
realizing the semantic web

# Classification of schema-based techniques

**Three layers**

- The upper layer
  - Granularity of match
  - Interpretation of the input information
- The middle layer represents classes of elementary (basic) matching techniques
- The lower layer is based on the kind of input which is used by elementary matching techniques

# Classification of schema-based techniques



# Outline

- Matching problem
- Classification of schema-based matching techniques
- **Basic techniques**
- Matching process
- Review of the matching systems
- Conclusions

knowledge**web**
realizing the semantic web

# Basic techniques

. . . techniques from the following systems have been taken into consideration:

- Anchor-PROMPT
- Artemis
- COMA, COMA++
- Cupid
- NOM, QOM, FOAM
- OLA
- SF, Rondo
- CtxMatch, S-Match

knowledgeweb
realizing the semantic web

# Element-level techniques

**String-based** (e.g., COMA, SF, S-Match, OLA)

- Prefix
  - It takes as input two strings and checks whether the first string starts with the second one
  - net = network; but also hot = hotel
- Suffix
  - It takes as input two strings and checks whether the first string ends with the second one
  - phone = telephone; but also word = sword

# Element-level techniques

**String-based** (e.g., S-Match, OLA, Anchor-Prompt)

- Edit distance
  - It takes as input two strings and calculates the number of *insertions*, *deletions*, and *substitutions* of characters required to transform one string into another, normalized by $max(length(string1), length(string2))$
  - EditDistance(NKN,Nikon) = 0.4

# Element-level techniques

**String-based** (e.g., COMA, S-Match)

- N-gram
  - It takes as input two strings and calculates the number of the same n-grams (i.e., sequences of $n$ characters) between them
  - $trigram(3)$ for the string nikon are nik, iko, kon

knowledgeweb
realizing the semantic web

# Element-level techniques

**Language-based** (e.g., COMA, Cupid, S-Match, OLA)

- Tokenization
  - Names are parsed into tokens by recognizing punctuation, cases
  - Hands-Free_Kits $\longrightarrow \langle$ hands, free, kits $\rangle$
- Lemmatization
  - Tokens are morphologically analyzed in order to find all their possible basic forms
  - Kits $\longrightarrow$ Kit

# Element-level techniques

**Language-based** (e.g., Cupid, S-Match)

- Elimination
  - Tokens that are articles, prepositions, conjunctions, and so on, are marked to be discarded
  - a, the, by, type of

# Element-level techniques

**Constraint-based** (e.g., OLA, COMA)

- Datatype comparison
  - $integer < real$
  - $date \in [1/4/2005\ 30/6/2005] < date[year = 2005]$
  - $\{a, c, g, t\}[1 - 10] < \{a, c, g, u, t\}+$
- Multiplicity comparison
  - $[1\ 1] < [0\ 10]$

knowledge**web**
realizing the semantic web

# Element-level techniques

**Linguistic resources** (e.g., Artemis, S-Match, OLA)

- Sense-based: WordNet
  - Relations between schema/ontology entities can be computed in terms of lexical relationships

knowledgeweb
realizing the semantic web

# Element-level techniques

**Linguistic resources** (e.g., Artemis, S-Match)

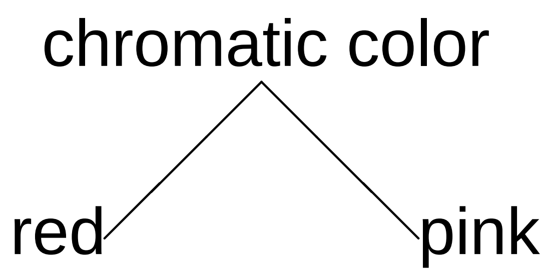
- Sense-based: WordNet
  - $A \sqsubseteq B$ if $A$ is a hyponym or meronym of $B$
    - Brand $\sqsubseteq$ Name
  - $A \sqsupseteq B$ if $A$ is a hypernym or holonym of $B$
    - Europe $\sqsupseteq$ Greece
  - $A = B$ if they are synonyms
    - Quantity $=$ Amount
  - $A \perp B$ if they are antonyms or the siblings in the part of hierarchy
    - Microprocessors $\perp$ PC_Board

knowledge**web**
realizing the semantic web

# Element-level techniques

**Linguistic resources** (e.g., S-Match)

- Sense-based: WordNet hierarchy distance
  - These return the equivalence relation if the distance between two input senses in the WordNet hierarchy is less than a given threshold
  - red = pink




knowledgeweb
realizing the semantic web

# Element-level techniques

**Linguistic resources** (e.g., S-Match)

- Gloss-based: WordNet gloss comparison
  - The number of the same words occurring in both input glosses increases the similarity value. The equivalence relation is returned if the resulting similarity value exceeds a given threshold
  - Maltese dog is a breed of toy dogs having a long straight silky white coat
    Afghan hound is a tall graceful breed of hound with a long silky coat

knowledgeweb
realizing the semantic web

# Element-level techniques

**Linguistic resources** (e.g., Cupid, COMA)

- Specific thesauri
  - These usually store specific domain knowledge
  - PO = Purchase Order
    uom = UnitOfMeasure
    line = item

knowledgeweb
realizing the semantic web

# Element-level techniques

**Alignment reuse** (e.g., COMA, COMA++, OLA)

- Entire schemas
- Schema fragments
- ... we need to match schema/ontology $o'$ and $o''$, given the alignments between $o$ and $o'$, and between $o$ and $o''$ from the external resource, storing previous match operations results

knowledgeweb
realizing the semantic web

# Structure-level techniques

**Taxonomy-based** (Anchor-Prompt, NOM, QOM)

. . . schemas/ontologies are viewed as graph-like structures containing terms and their inter-relationships

- Bounded path matching
  - These take two paths with links between classes defined by the hierarchical relations, compare terms and their positions along these paths, and identify similar terms
- Super(sub)-concepts rules
  - If super-concepts are the same, the actual concepts are similar to each other

knowledgeweb
realizing the semantic web

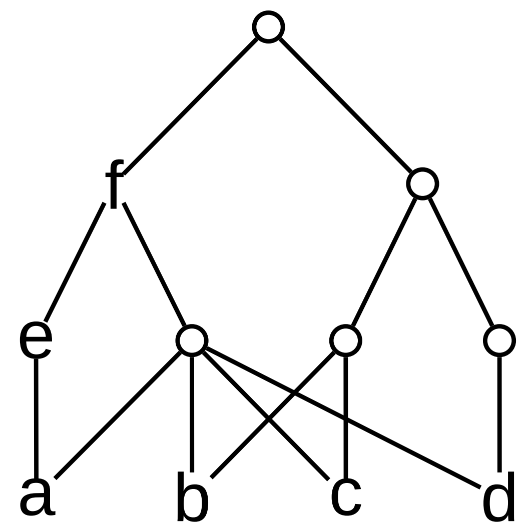
# Structure-level techniques

## Taxonomy-based

- Upward cotopic distance
  Measures the ratio of common superclasses.

$$\delta(c,c') = 1 - \frac{|UC(c,H) \cap UC(c',H)|}{|UC(c,H) \cup UC(c',H)|}$$

where $UC(c,H) = \{c' \in H; c \leq c'\}$ is the set of superclasses of $c$.




$\delta(a,a) = 1 - 1 = 0$

$\delta(a,e) = 1 - 3/5 = .4$

$\delta(a,f) = 1 - 2/5 = .6$

$\delta(d,a) = 1 - 3/8 \approx .625$

$\delta(b,c) = 1 - 5/7 \approx .286$

$\delta(c,d) = 1 - 4/8 = .5$

$\delta(a,b) = 1 - 3/8 \approx .625$

# Structure-level techniques

**Graph-based** (e.g., Cupid, COMA)

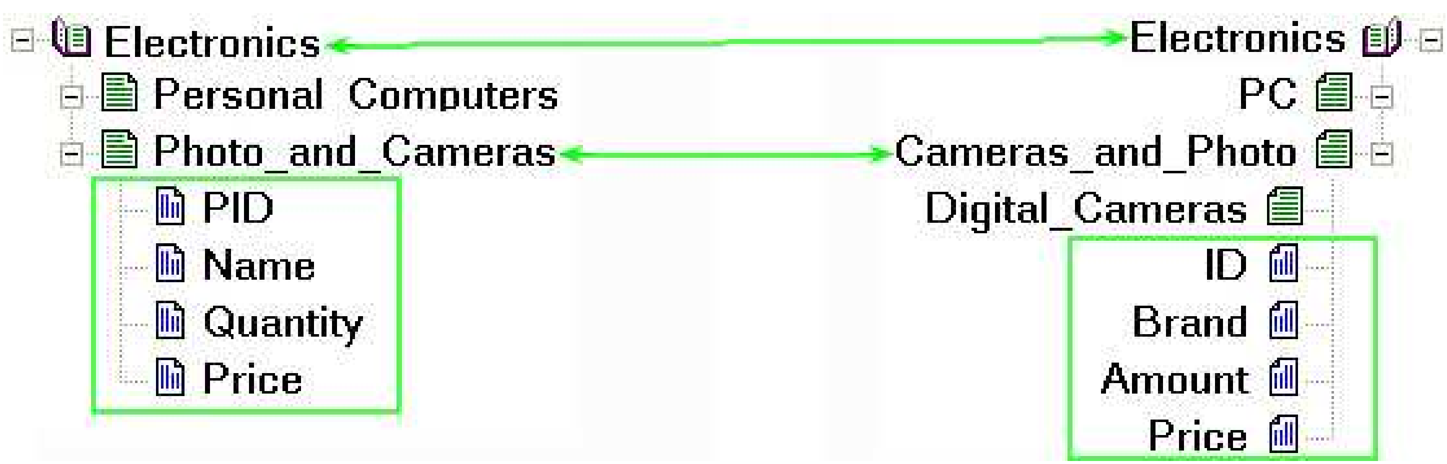
- Children
  - Two non-leaf schema elements are structurally similar if their immediate children sets are highly similar
- Leaves
  - Two non-leaf schema elements are structurally similar if their leaf sets are highly similar, even if their immediate children are not

knowledgeweb
realizing the semantic web

# Structure-level techniques

**Graph-based** (e.g., Cupid, COMA)

- Leaves



# Structure-level techniques

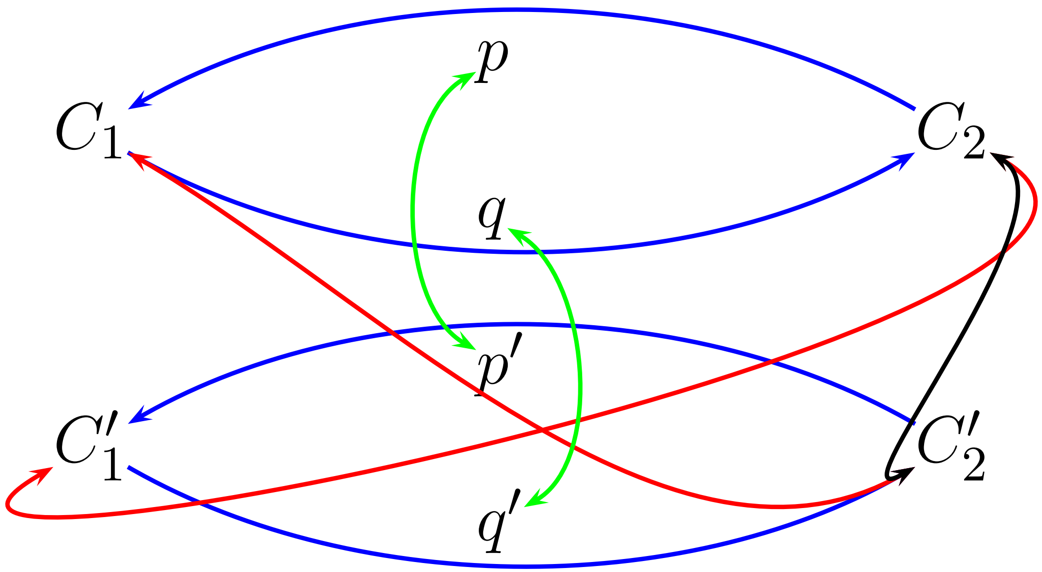
**Graph-based** (e.g., SF, OLA)

- Iterative fix point computation
  - If two nodes from two schemas/ontologies are similar, their neighbors might also be somehow similar

# Structure-level techniques

**Graph-based** (e.g., SF, OLA)

- Iterative fix point computation




| | $C_1$ | $p$ | $C_2$ | $q$ |
|---|---|---|---|---|
| $C'_1$ | .4 | | .6 | |
| $p'$ | | .8 | | .2 |
| $C'_2$ | .5 | | .6 | |
| $q'$ | | .4 | | .5 |

$$\sigma_C(c,c') = .6.\frac{1}{max(|A(c)|,|A(c')|)}.\sum_{\langle a,a'\rangle \in match(A(c),A(c'))} \sigma_A(a,a') + .4.\sigma(N(c),N(c'))$$

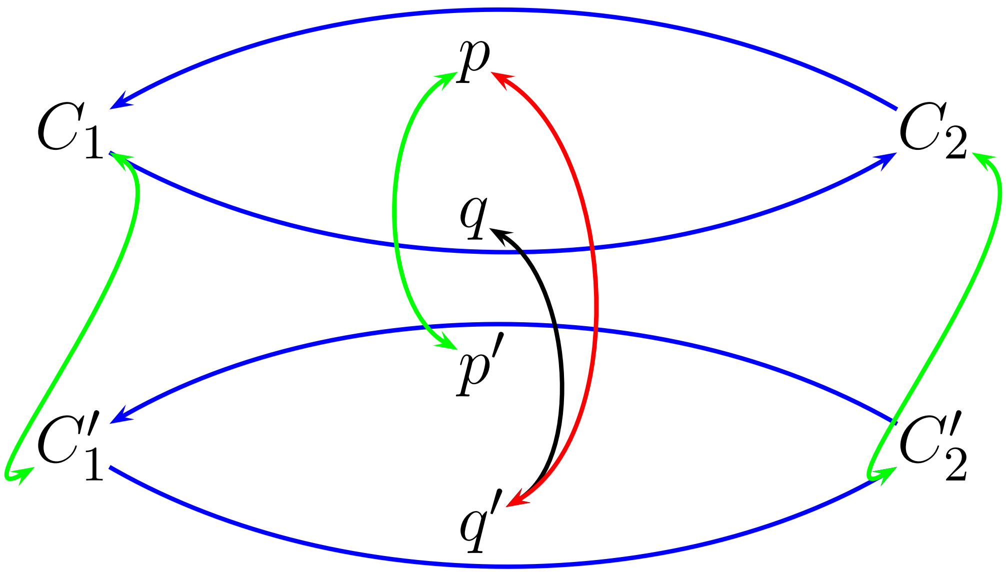
$$\sigma_A(a,a') = .6.\sigma_C(domain(a),domain(a')) + .4.\sigma(N(a),N(a'))$$

knowledge**web**
realizing the semantic web

# Structure-level techniques

**Graph-based** (e.g., SF, OLA)

- Iterative fix point computation




| | $C_1$ | $p$ | $C_2$ | $q$ |
|---|---|---|---|---|
| $C'_1$ | .64 | | .36 | |
| $p'$ | | .68 | | .38 |
| $C'_2$ | .32 | | .54 | |
| $q'$ | | .52 | | .44 |

$$\sigma_C(c,c') = .6.\frac{1}{max(|A(c)|,|A(c')|)}.\sum_{\langle a,a'\rangle \in match(A(c),A(c')} \sigma_A(a,a') + .4.\sigma(N(c),N(c'))$$

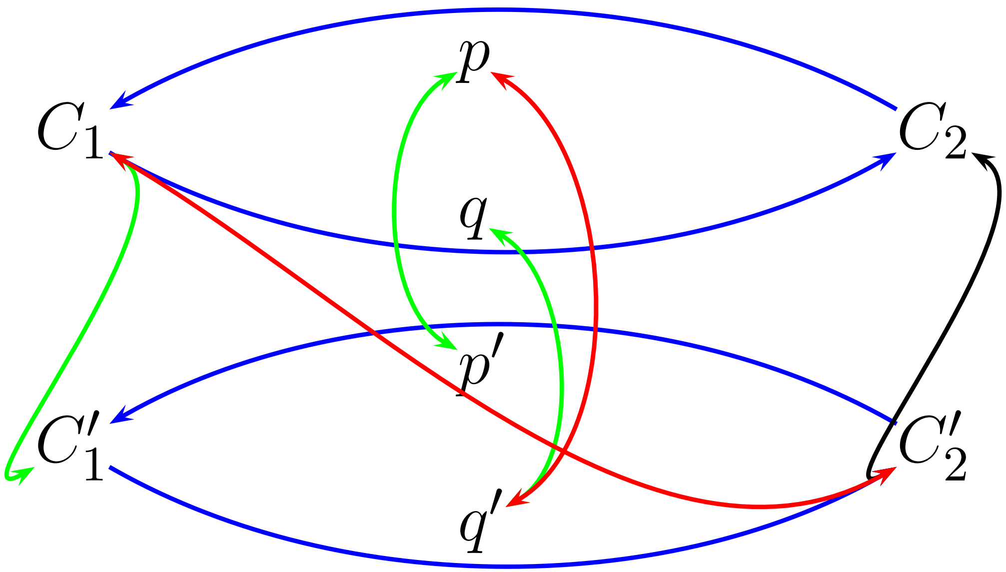
$$\sigma_A(a,a') = .6.\sigma_C(domain(a),domain(a')) + .4.\sigma(N(a),N(a'))$$

knowledge**web**
realizing the semantic web

# Structure-level techniques

**Graph-based** (e.g., SF, OLA)

- Iterative fix point computation




| | $C_1$ | $p$ | $C_2$ | $q$ |
|---|---|---|---|---|
| $C'_1$ | .57 | | .47 | |
| $p'$ | | .64 | | .27 |
| $C'_2$ | .51 | | .5 | |
| $q'$ | | .38 | | .58 |

$$\sigma_C(c,c') = .6.\frac{1}{max(|A(c)|,|A(c')|)}.\sum_{\langle a,a'\rangle \in match(A(c),A(c'))} \sigma_A(a,a') + .4.\sigma(N(c),N(c'))$$

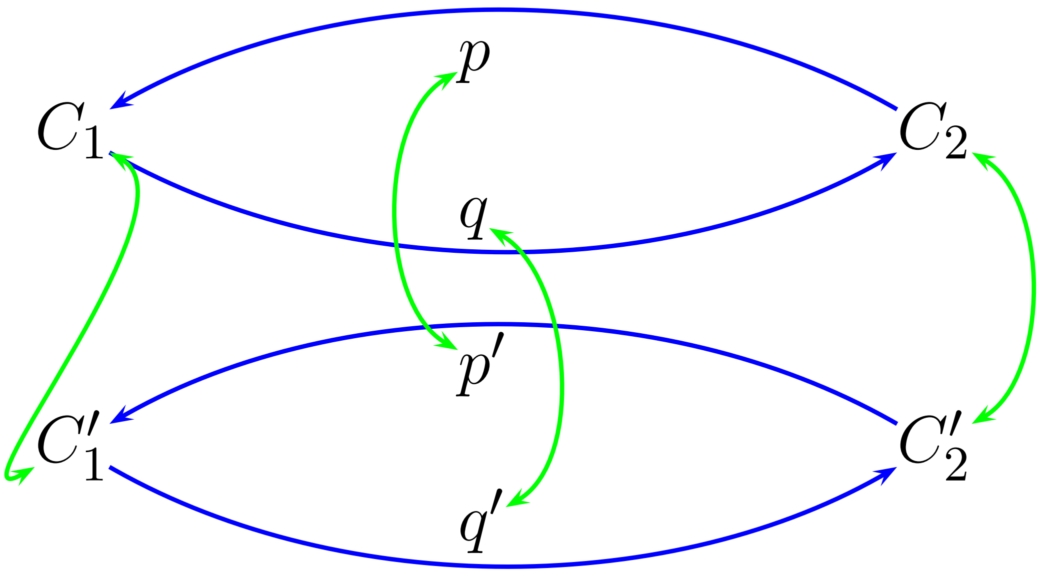
$$\sigma_A(a,a') = .6.\sigma_C(domain(a), domain(a')) + .4.\sigma(N(a),N(a'))$$

knowledge**web**
realizing the semantic web

# Structure-level techniques

**Graph-based** (e.g., SF, OLA)

- Iterative fix point computation




| | $C_1$ | $p$ | $C_2$ | $q$ |
|---|---|---|---|---|
| $C'_1$ | .54 | | .4 | |
| $p'$ | | .62 | | .39 |
| $C'_2$ | .43 | | .59 | |
| $q'$ | | .44 | | .54 |

$$\sigma_C(c,c') = .6.\frac{1}{max(|A(c)|,|A(c')|)}.\sum_{\langle a,a'\rangle \in match(A(c),A(c'))} \sigma_A(a,a') + .4.\sigma(N(c),N(c'))$$

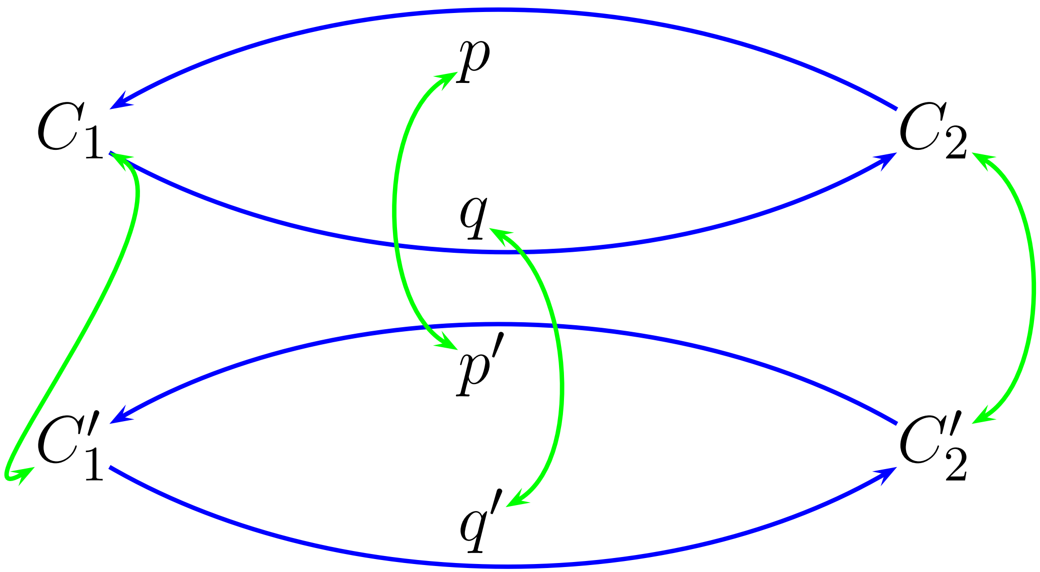
$$\sigma_A(a,a') = .6.\sigma_C(domain(a),domain(a')) + .4.\sigma(N(a),N(a'))$$

knowledge**web**
realizing the semantic web

# Structure-level techniques

**Graph-based** (e.g., SF, OLA)

- Iterative fix point computation




| | $C_1$ | $p$ | $C_2$ | $q$ |
|---|---|---|---|---|
| $C'_1$ | .53 | | .47 | |
| $p'$ | | .67 | | .34 |
| $C'_2$ | .46 | | .56 | |
| $q'$ | | .4 | | .52 |

Threshold reached: no $.1$ variation

$$\sigma_C(c,c') = .6.\frac{1}{max(|A(c)|,|A(c')|)}.\sum_{\langle a,a'\rangle \in match(A(c),A(c'))} \sigma_A(a,a') + .4.\sigma(N(c),N(c'))$$

$$\sigma_A(a,a') = .6.\sigma_C(domain(a),domain(a')) + .4.\sigma(N(a),N(a'))$$

# Structure-level techniques

**Model-based** (e.g., CtxMatch, S-Match)

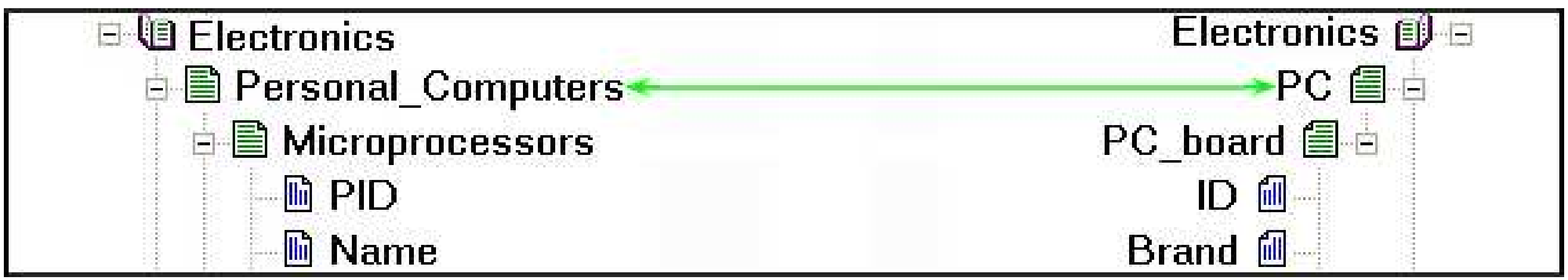
- Propositional satisfiability (SAT)
  - Decompose the graph (tree) matching problem into the set of node matching problems
  - Translate each node matching problem, namely pairs of nodes with possible relations between them, into a propositional formula
  - Check the propositional formula for validity

# Structure-level techniques

**Model-based** (e.g., CtxMatch, S-Match)

- Propositional satisfiability (SAT)

$$Axioms \rightarrow rel(context_1, context_2)$$




$$\overbrace{(Electronics_1 \leftrightarrow Electronics_2) \wedge (Personal\_Computers_1 \leftrightarrow PC_2)}^{Axioms} \longrightarrow$$

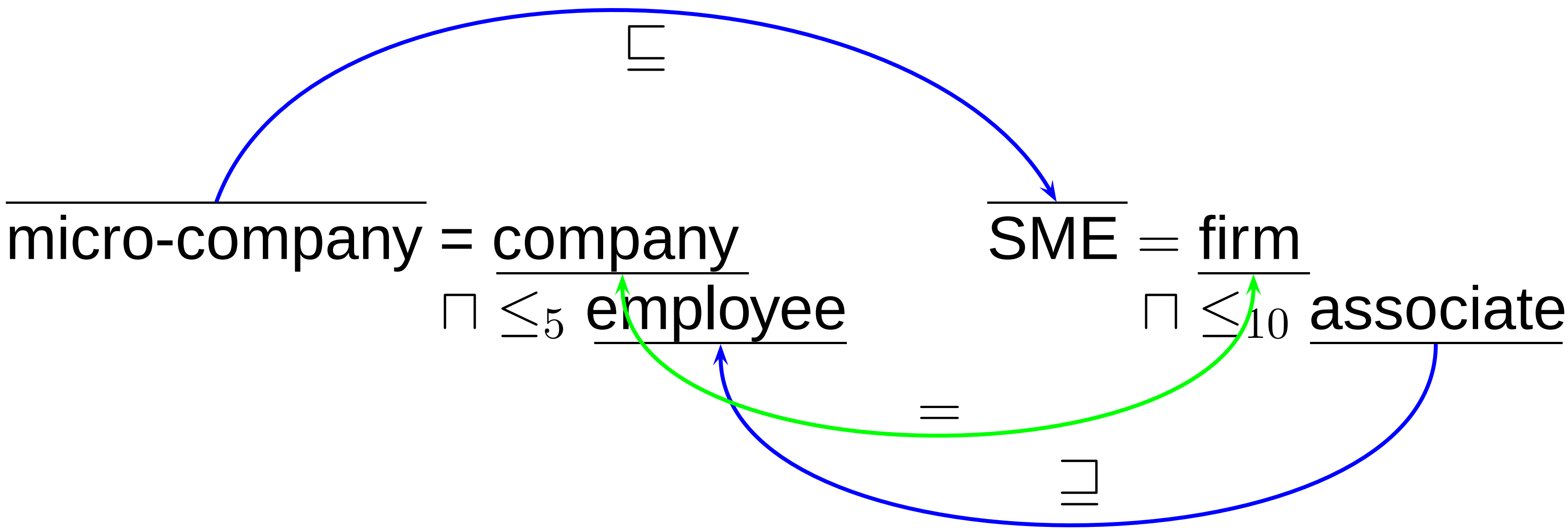
$$\overbrace{(Electronics_1 \wedge Personal\_Computers_1)}^{context_1} \longleftrightarrow \overbrace{(Electronics_2 \wedge PC_2)}^{context_2}$$

knowledgeweb
realizing the semantic web

# Structure-level techniques

## Model-based

- Description Logics (DL)-based




$$\frac{\text{company} = \text{firm} \;;\; \text{associate} \sqsubseteq \text{employee}}{\text{micro-company} \sqsubseteq \text{SME}}$$

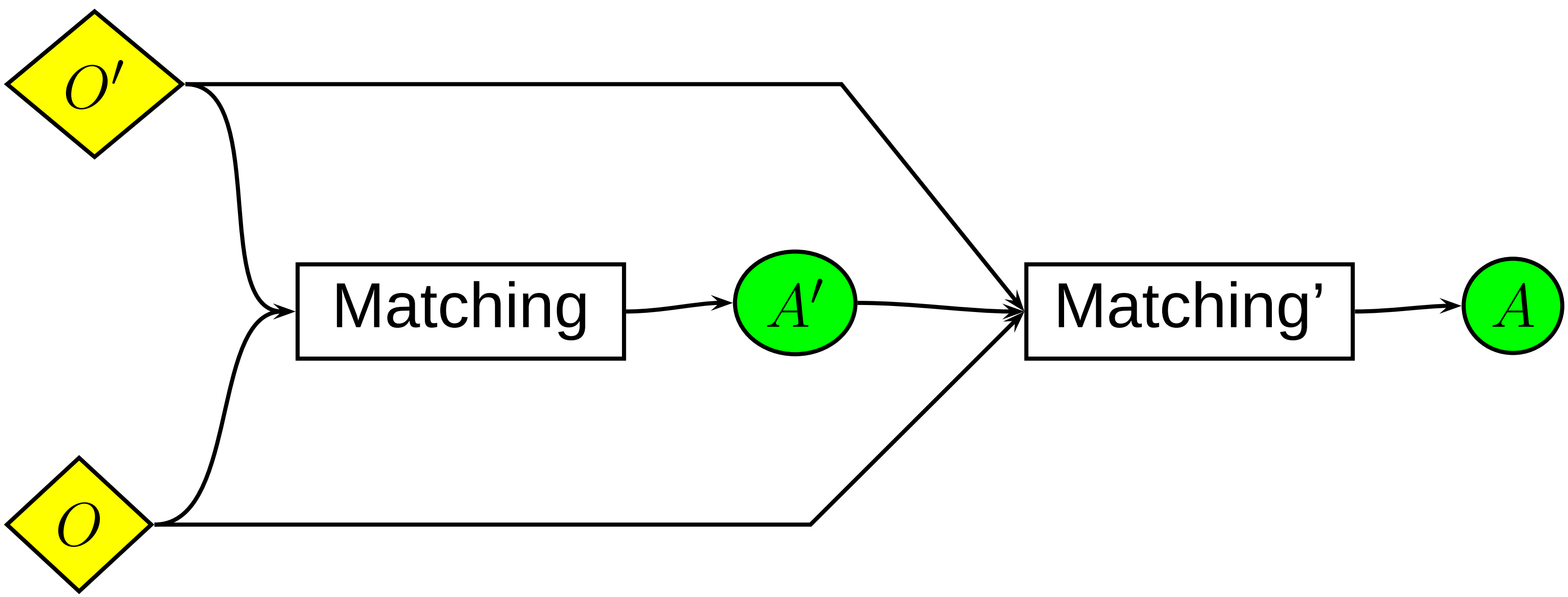
# Outline

- Matching problem
- Classification of schema-based matching techniques
- Basic techniques
- **Matching process**
- Review of the matching systems
- Conclusions

knowledge**web**
realizing the semantic web

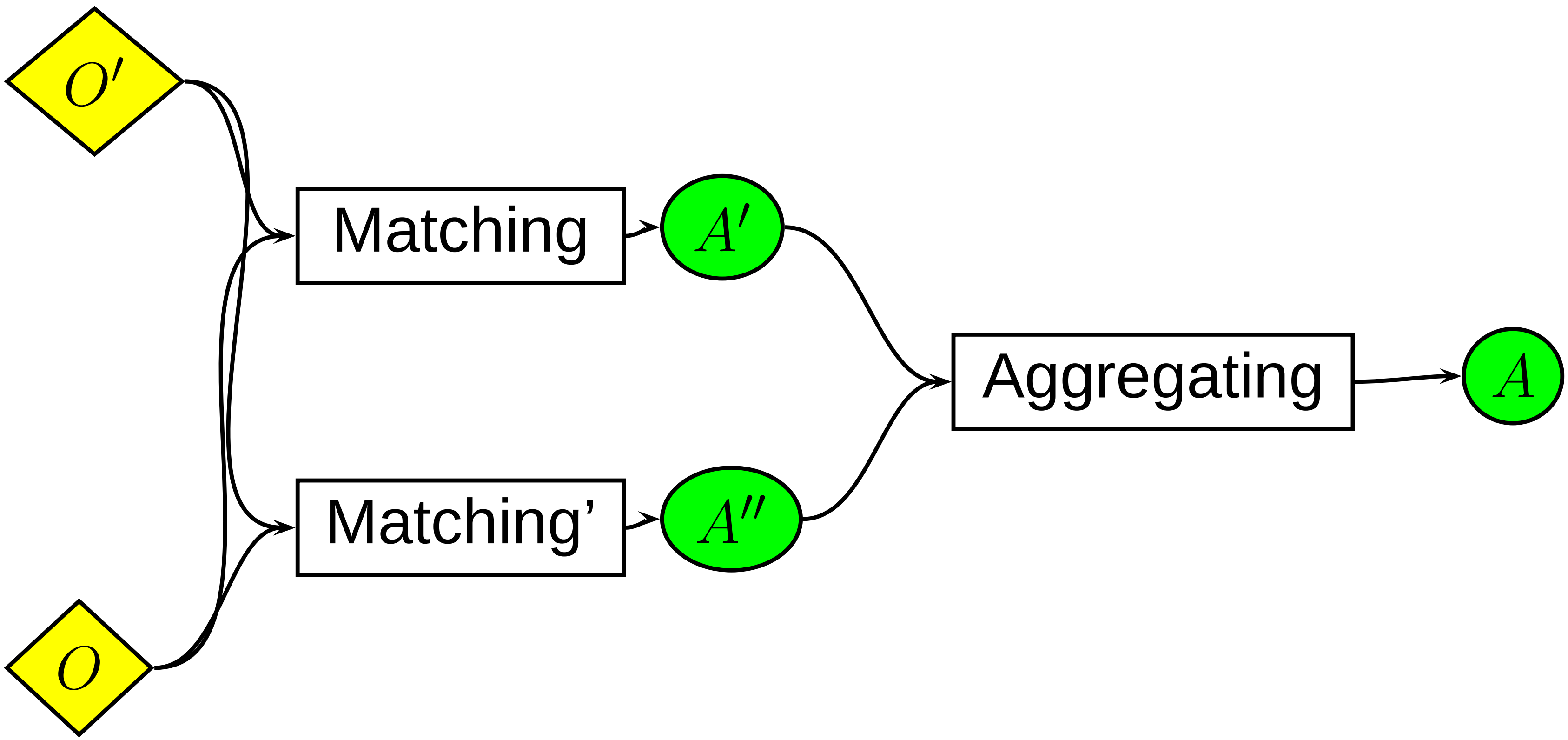
# Matching process

**Architectural perspective:** Sequential (hybrid) (e.g., Cupid, Artemis)



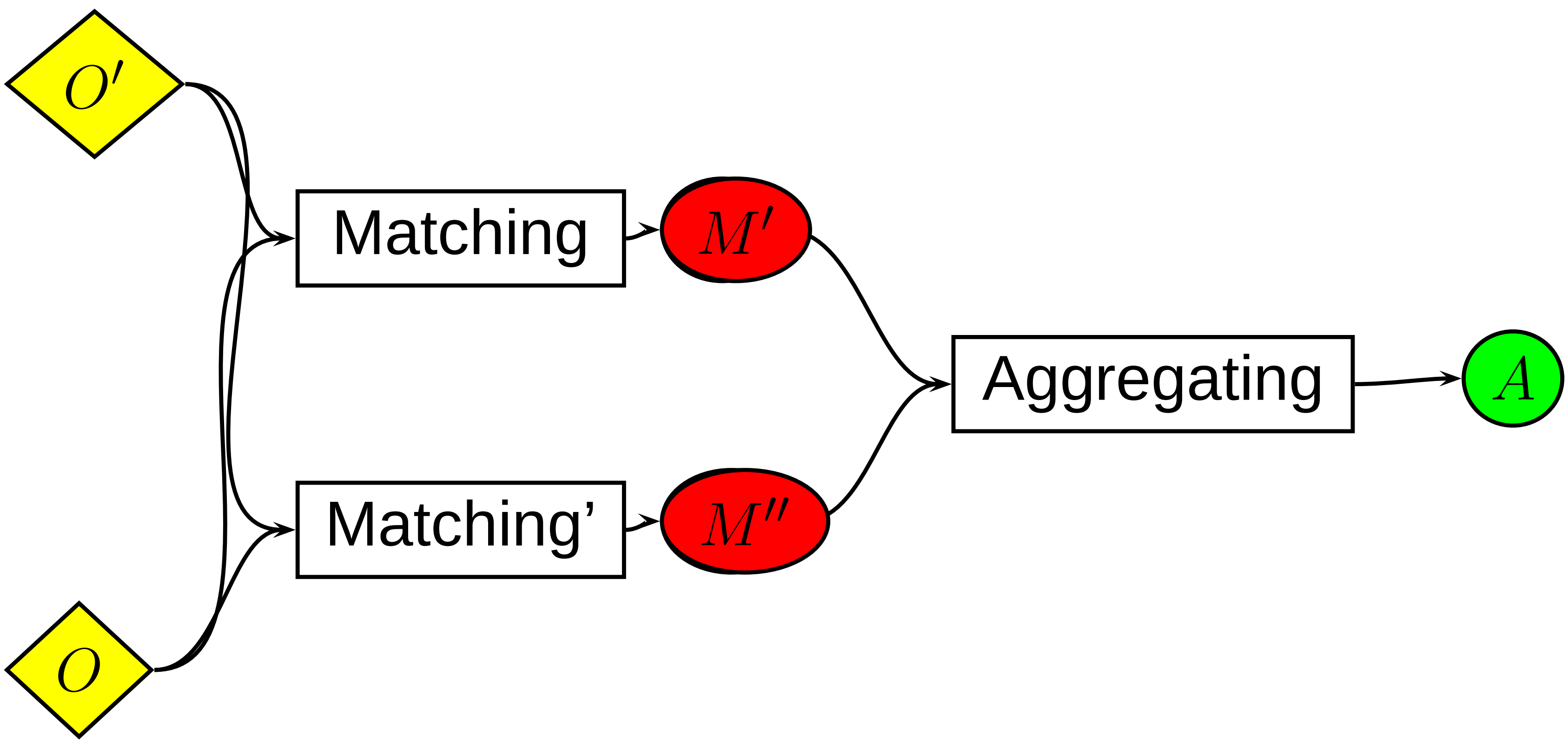
# Matching process

**Architectural perspective:** Parallel (composite) (e.g., COMA, QOM)



# Matching process

**Architectural perspective:** Parallel (composite) (e.g., COMA, QOM)




Aggregation (e.g., Min, Max, Weighted, Average)

# Matching process

## User-centric perspective

- Alignments as solutions (e.g., Rondo, OLA)
  - These consider the matching problem as an optimization problem and the alignment is a solution to it
- Alignments as theorems (e.g., S-Match)
  - These rely on semantics and require the alignment to satisfy it
- Alignments as likeness clues (e.g., Cupid)
  - These produce only reasonable indications to a user for selecting the alignment

# Matching process

## Selecting the final alignment

- Ranking strategies
  - Thresholds
  - MaxDelta
- Cardinalities
  - 1-1; 1-*; *-*
- Directionality
  - $O \rightarrow O'$; $O' \rightarrow O$ (SmallLarge, LargeSmall)
  - $O \rightarrow O'$ and $O' \rightarrow O$ (Both)

knowledge**web**
realizing the semantic web

# Outline

- Matching problem
- Classification of schema-based matching techniques
- Basic techniques
- Matching process
- **Review of the matching systems**
- Conclusions

# Review of the matching systems

**Some state of the art systems**

- Cupid (Microsoft Research, USA)
- FOAM/QOM (University of Karlsruhe, Germany)
- OLA (INRIA Rhône-Alpes/Université de Montréal, France/Canada)
- S-Match (University of Trento, Italy)
- ...

knowledgeweb
realizing the semantic web

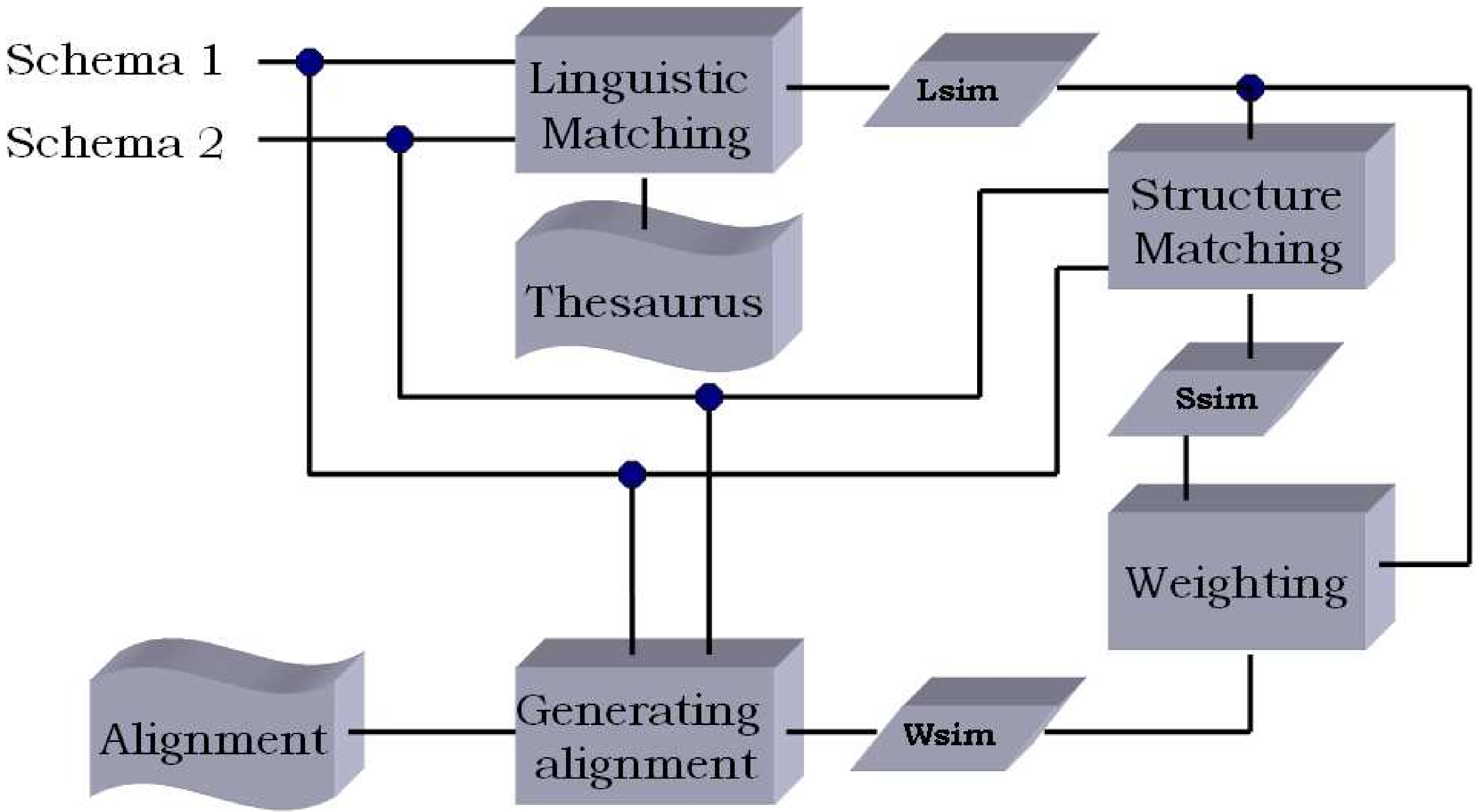
# Review of the matching systems

## Cupid

- Schema-based
- Computes similarity coefficients in the [0,1] range
- Performs linguistic and structure matching
- Sequential system
- Alignments as likeness clues

# Review of the matching systems

## Cupid




knowledge**web**
realizing the semantic web

# Review of the matching systems

## OLA

- Schema- and Instance-based
- Computes dissimilarities + extracts alignments (equivalences in the [0,1] range)
- Based on terminological (including linguistic) and structural (internal and relational) distances
- Neither sequential nor parallel
- Alignments as solutions (to an optimization problem)

# Review of the matching systems

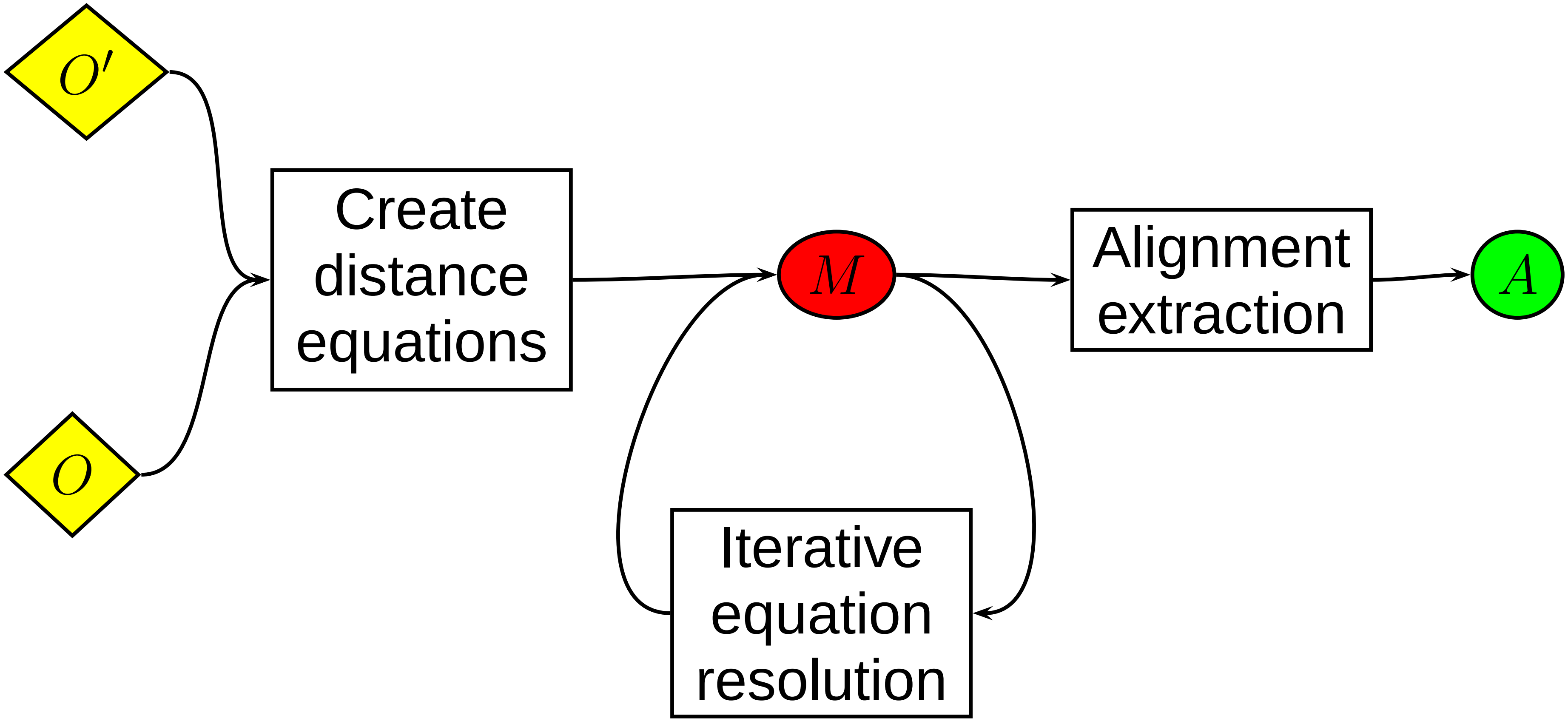
## QOM/FOAM

- Schema- and Instance-based
- Computes similarities + extracts alignments (equivalences in the [0,1] range)
- Based on terminological (including linguistic) and structural (internal and relational) distances
- Parallel with elaborated aggregation
- Alignments as likeness clues

knowledgeweb
realizing the semantic web

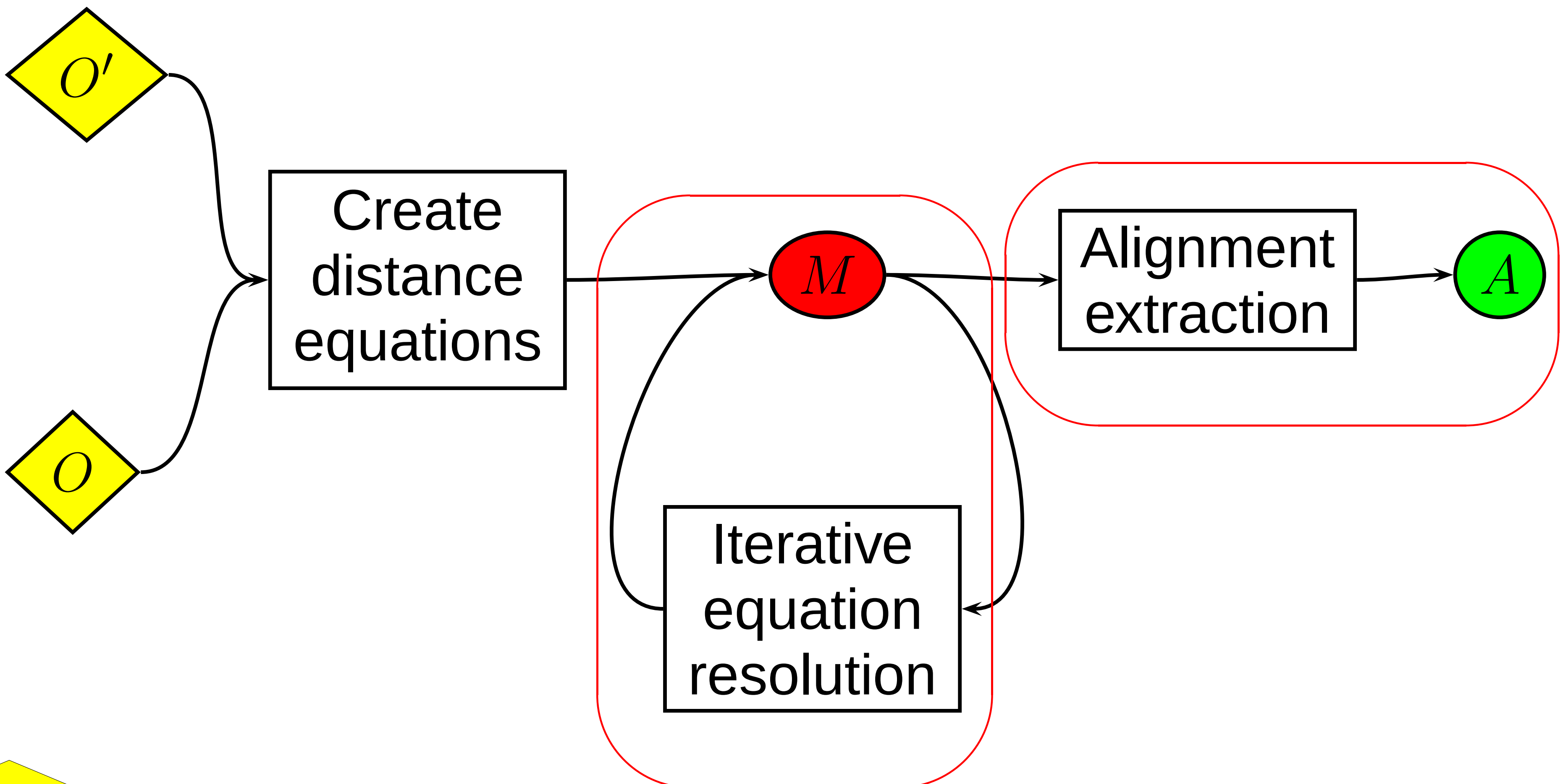
# Review of the matching systems

## OLA



# Review of the matching systems

## OLA



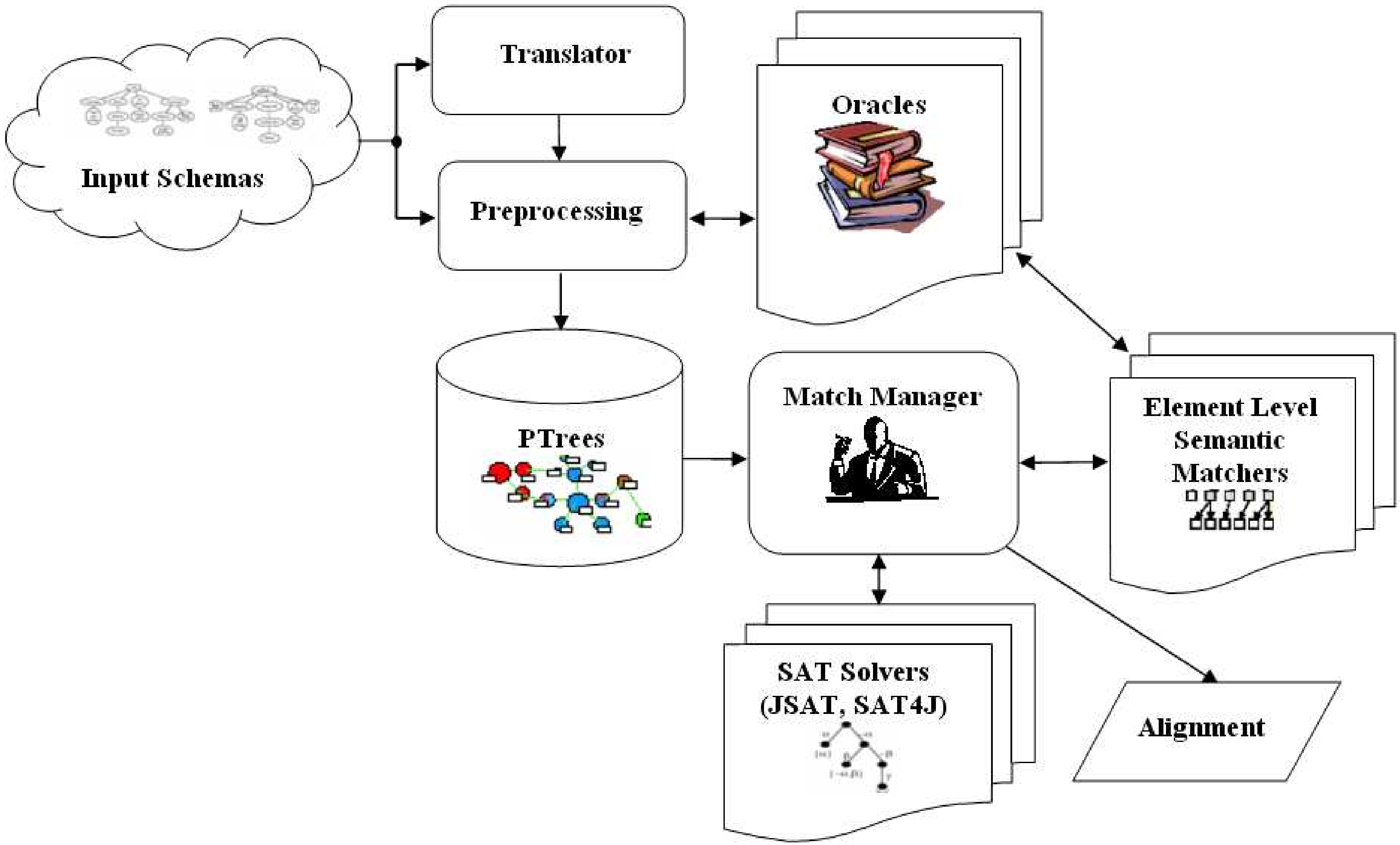
# Review of the matching systems

## S-Match

- Schema-based
- Computes equivalence (=); more general ($\sqsupseteq$); less general ($\sqsubseteq$); disjointness ($\perp$)
- Analyzes the meaning (concepts, not labels) which is codified in the elements and the structures of schemas/ontologies
- Sequential system with a "composition" at the element level
- Alignments as theorems

# Review of the matching systems

## S-Match




knowledgeweb
realizing the semantic web

# Review of the matching systems

## Analytical comparison

| | | SF | Artemis | Cupid | COMA | NOM | Anchor-PROMPT | OLA | S-Match |
|---|---|---|---|---|---|---|---|---|---|
| Element-level | Syntactic | string-based (2); data types; key properties | domain compatibility; language-based (1) | string-based (2); language-based (2); data types; key properties | string-based (4); language-based (1); data types | string-based (1); domains and ranges | string-based (1); domains and ranges | string-based (3); data types; language-based (1) | string-based (5); language-based (2) |
| | External | - | common thesaurus (CT): synonyms, broader terms, related terms | auxiliary thesaurus (synonyms, hypernyms, hyponyms, abbreviations ) | auxiliary thesaurus (synonyms, hypernyms, hyponyms, abbreviations ); alignment reuse (2) | application-specific vocabulary | - | WordNet(1) | WordNet: sense-based (2), gloss-based (6) |
| Structure-level | Syntactic | iterative fix-point computation | matching of neighbors via CT | tree matching weighted by leaves | DAG (tree) matching with a bias towards leaf or children structures (2); paths | matching of neighbors (2); taxonomic structure (4) | bounded paths matching (arbitrary links); bounded paths matching (processing *is-a* links separately) | Iterative fix-point computation; matching of neighbors; taxonomic structure | - |
| | Semantic | - | - | - | - | - | - | - | propositional SAT (2) |

knowledgeweb
realizing the semantic web

# Outline

- Matching problem
- Classification of schema-based matching techniques
- Basic techniques
- Matching process
- Review of the matching systems
- **Conclusions**

# Conclusions

## Summary

- We have discussed the schema/ontology matching problem and its application domains
- We have provided classificatory elements for approaching schema/ontology matching techniques
- We have presented a number of basic matching techniques as well as different strategies for building the matching process
- We have reviewed and compared (analytically) some existing matching systems

# Conclusions

## Uses of classifications

- They provide a common conceptual basis, and hence, can be used for comparing (analytically) different existing schema/ontology matching systems
- They can help in designing a new matching system, or an elementary matcher, taking advantages of state of the art solutions
- They can help in designing systematic benchmarks, e.g., by discarding features one by one from schemas/ontologies, namely, what class of basic techniques deals with what feature

# Conclusions

## Research Challenges

- Industry-strength schema/ontology matching
  - Scalability
- Interactive approaches
- Infrastructures (e.g., Rondo, Chimaera)
  - Representing the alignment
  - Executing the alignment
  - Explaining the alignment

# Conclusions

## Research Challenges

- Matching web services at the process level
- Lightweight ontology matching and emerging semantics
- Automatic partial alignment

knowledgeweb
realizing the semantic web

# Conclusions

## Research Challenges

- Evaluation
  - Testbed environment
    - Series of tests, each with a pre-defined problem
    - Real-world case studies
  - More accurate evaluation measures
  - Adequacy task / measure
  - Testing methodology which is able to estimate quality of the alignment between schemas/ontologies with thousands of entities

knowledgeweb
realizing the semantic web

Questions?

# Acknowledgments

We thank all the participants of the Heterogeneity workpackage of the Knowledge Web network of excellence

In particular, we are grateful to T.-L. Bach, J. Barrasa, P. Bouquet, J. Bo, R. Dieng-Kuntz, M. Ehrig, E. Franconi, R. García Castro, F. Giunchiglia, M. Hauswirth, P. Hitzler, M. Jarrar, M. Krötzsch, R. Lara, D. Maynard, A. Napoli, L. Serafini, G. Stamou, H. Stuckenschmidt, Y. Sure, S. Tessaris, P. Traverso, P. Valchev, S. van Acker, M. Yatskevich, and I. Zaihrayeu for their support and insightful comments

Thank You

for your attention and interest!

knowledgeweb
realizing the semantic web

# The ESWC'05 Tutorial on Schema and Ontology Matching BIBLIOGRAPHY

Pavel Shvaiko[1] and Jérôme Euzenat[2]

[1] University of Trento, Povo, Trento, Italy,
pavel@dit.unitn.it
[2] INRIA, Rhône-Alpes, France,
Jerome.Euzenat@inrialpes.fr

## 1 Surveys

Good surveys through the recent years are provided in [18, 24, 36, 40, 44, 47, 49].
Major contributions of the last decades were presented in [2, 26, 27, 43].

## 2 Schema-based matching systems

| **Name** | **Publications** | **Project web-site** |
|---|---|---|
| Artemis | [3, 4, 9] | - |
| COMA, COMA++ | [1, 11, 41] | http://dbs.uni-leipzig.de/Research/coma.html |
| CtxMatch | [7, 8] | - |
| Cupid | [29] | - |
| Naive Ontology Mapping (NOM) | [16] | http://www.aifb.uni-karlsruhe.de/WBS/meh/foam/ |
| OWL Lite Alignment (OLA) | [19, 20] | http://www.iro.umontreal.ca/∼owlola/alignment.html |
| PROMPT | [37–39] | http://protege.stanford.edu/plugins/prompt/prompt.html |
| Quick Ontology Mapping (QOM) | [15] | http://www.aifb.uni-karlsruhe.de/WBS/meh/foam/ |
| Similarity Flooding (SF) | [32, 33] | http://www-db.stanford.edu/ melnik/mm/sfa/ |
| S-Match | [21–23] | http://dit.unitn.it/∼accord/ |

## 3 Infrastructures

| **Name** | **Publications** | **Project web-site** |
|---|---|---|
| Chimaera | [30, 31] | http://www.ksl.stanford.edu/software/chimaera/ |
| OntoMerge | [14] | http://cs-www.cs.yale.edu/homes/dvm/daml/ontology-translation.html |
| Protoplasm | [5] | - |
| Rondo | [32, 34, 35] | http://www-db.stanford.edu/ melnik/mm/rondo/ |

## 4 Further Readings

– **Instance-based Matching**: [10, 13, 25];

– **Languages for the alignment representation**: [6, 42];
– **Executing the Alignment**: [28, 31, 48, 50];
– **Explaining the Alignment**: [10, 45];
– **Evaluation**: [12, 17, 46].

**Acknowledgments:** This work has been partly supported by the European Knowledge Web network of excellence (IST-2004-507482).

## References

1. D. Aumüller, H. H. Do, S. Massmann, and E. Rahm. Schema and ontology matching with COMA++. In *Proceedings of the International Conference on Management of Data (SIGMOD), Software Demonstration*, 2005.
2. C. Batini, M. Lenzerini, and S. B. Navathe. A comparative analysis of methodologies for database schema integration. *ACM Computing Surveys*, 18(4):323–364, 1986.
3. D. Beneventano, S. Bergamaschi, S. Lodi, and C. Sartori. Consistency checking in complex object database schemata with integrity constraints. *IEEE Transactions on Knowledge and Data Engineering*, (10(4)):576–598, 1998.
4. S. Bergamaschi, S. Castano, and M. Vincini. Semantic integration of semistructured and structured data sources. *SIGMOD Record*, (28(1)):54–59, 1999.
5. P. Bernstein, S. Melnik, M. Petropoulos, and C. Quix. Industrial-strength schema matching. *SIGMOD Record*, (33(4)):38–43, 2004.
6. P. Bouquet, F. Giunchiglia, F. van Harmelen, L. Serafini, and H. Stuckenschmidt. Contextualizing ontologies. *Journal of Web Semantics*, (26):1–19, 2004.
7. P. Bouquet, B. Magnini, L. Serafini, and S. Zanobini. A SAT-based algorithm for context matching. In *Proceedings of the International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT)*, pages 66–79, 2003.
8. P. Bouquet, L. Serafini, and S. Zanobini. Semantic coordination: A new approach and an application. In *Proceedings of the International Semantic Web Conference (ISWC)*, pages 130–145, 2003.
9. S. Castano, V. De Antonellis, and S. De Capitani di Vimercati. Global viewing of heterogeneous data sources. *IEEE Transactions on Knowledge and Data Engineering*, (13(2)):277–297, 2001.
10. R. Dhamankar, Y. Lee, A. Doan, A. Halevy, and P. Domingos. iMAP: Discovering complex semantic matches between database schemas. In *Proceedings of the International Conference on Management of Data (SIGMOD)*, pages 383–394, 2004.
11. H. H. Do and E. Rahm. COMA - a system for flexible combination of schema matching approaches. In *Proceedings of the Very Large Data Bases Conference (VLDB)*, pages 610–621, 2001.
12. H.H. Do, S. Melnik, and E. Rahm. Comparison of schema matching evaluations. In *Proceedings of the workshop on Web and Databases*, 2002.
13. A. Doan, J. Madhavan, P. Domingos, and A. Halevy. Learning to map ontologies on the semantic web. In *Proceedings of the International World Wide Web Conference (WWW)*, pages 662–673, 2003.
14. D. Dou, D. McDermott, and P. Qi. Ontology translation on the Semantic Web. *Journal on Data Semantics (JoDS)*, 2:35–57, 2005.
15. M. Ehrig and S. Staab. QOM: Quick ontology mapping. In *Proceedings of the International Semantic Web Conference (ISWC)*, pages 683–697, 2004.

16. M. Ehrig and Y. Sure. Ontology mapping - an integrated approach. In *Proceedings of the European Semantic Web Symposium (ESWS)*, pages 76–91, 2004.
17. J. Euzenat. An API for ontology alignment. In *Proceedings of the International Semantic Web Conference (ISWC)*, pages 698–712, 2004.
18. J. Euzenat, J. Barrasa, P. Bouquet, R. Dieng, M. Ehrig, M. Hauswirth, M. Jarrar, R. Lara, D. Maynard, A. Napoli, G. Stamou, H. Stuckenschmidt, P. Shvaiko, S. Tessaris, S. van Acker, I. Zaihrayeu, and T. L. Bach. D2.2.3: State of the art on ontology alignment. Technical report, NoE Knowledge Web project delivable, 2004. http://knowledgeweb.semanticweb.org/.
19. J. Euzenat and P. Valtchev. An integrative proximity measure for ontology alignment. In *Proceedings of the Semantic Integration workshop at the International Semantic Web Conference (ISWC)*, 2003.
20. J. Euzenat and P. Valtchev. Similarity-based ontology alignment in OWL-lite. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*, pages 333–337, 2004.
21. F. Giunchiglia and P. Shvaiko. Semantic matching. *The Knowledge Engineering Review Journal (KER)*, (18(3)):265–280, 2003.
22. F. Giunchiglia, P. Shvaiko, and M. Yatskevich. S-Match: an algorithm and an implementation of semantic matching. In *Proceedings of the European Semantic Web Symposium (ESWS)*, pages 61–75, 2004.
23. F. Giunchiglia and M. Yatskevich. Element level semantic matching. In *Proceedings of the Meaning Coordination and Negotiation workshop at the International Semantic Web Conference (ISWC)*, 2004.
24. Y. Kalfoglou and M. Schorlemmer. Ontology mapping: the state of the art. *The Knowledge Engineering Review Journal (KER)*, (18(1)):1–31, 2003.
25. J. Kang and J. F. Naughton. On schema matching with opaque column names and data values. In *Proceedings of the International Conference on Management of Data (SIGMOD)*, pages 205–216, 2003.
26. V. Kashyap and A. Sheth. Semantic and schematic similarities between database objects: a context-based approach. *The International Journal on Very Large Data Bases (VLDB)*, 5(4):276–304, 1996.
27. J. A. Larson, S. B. Navathe, and R. Elmasri. A theory of attributed equivalence in databases with application to schema integration. *IEEE Transactions on Software Engineering*, 15(4):449–463, 1989.
28. M. Lenzerini. Data integration: A theoretical perspective. In *Proceedings of the Symposium on Principles of Database Systems (PODS)*, pages 233–246, 2002.
29. J. Madhavan, P. Bernstein, and E. Rahm. Generic schema matching with Cupid. In *Proceedings of the Very Large Data Bases Conference (VLDB)*, pages 49–58, 2001.
30. D. L. McGuinness, R. Fikes, J. Rice, and S. Wilder. The chimaera ontology environment. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 1123–1124, 2000.
31. D. L. McGuinness, R. Fikes, J. Rice, and S. Wilder. An environment for merging and testing large ontologies. In *Proceedings of the International Conference on the Principles of Knowledge Representation and Reasoning (KR)*, pages 483–493, 2000.
32. S. Melnik. *Generic Model Management: Concepts and Algorithms*. LNCS-2967, 2004. Dissertation. University of Leipzig.
33. S. Melnik, H. Garcia-Molina, and E. Rahm. Similarity flooding: A versatile graph matching algorithm. In *Proceedings of the International Conference on Data Engineering (ICDE)*, pages 117–128, 2002.
34. S. Melnik, E. Rahm, and P. Bernstein. Developing metadata-intensive applications with Rondo. *Journal of Web Semantics*, 2003.

35. S. Melnik, E. Rahm, and P. Bernstein. Rondo: A programming platform for generic model management. In *Proceedings of the International Conference on Management of Data (SIGMOD)*, pages 193–204, 2003.
36. N. Noy. Semantic Integration: A survey of ontology-based approaches. *SIGMOD Record*, 33(4):65–70, 2004.
37. N. Noy and M. Musen. PROMPT: Algorithm and tool for automated ontology merging and alignment. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 450–455, 2000.
38. N. Noy and M. Musen. The PROMPT Suite: Interactive tools for ontology merging and mapping. *International Journal of Human-Computer Studies*, (59(6)):983–1024, 2003.
39. N. Noy and M. A. Musen. Anchor-prompt: Using non-local context for semantic matching. In *Proceedings of the workshop on Ontologies and Information Sharing at the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 63–70, 2001.
40. E. Rahm and P. Bernstein. A survey of approaches to automatic schema matching. *The International Journal on Very Large Data Bases (VLDB)*, (10(4)):334–350, 2001.
41. E. Rahm, H. H. Do, and S. Maßmann. Matching large XML schemas. *SIGMOD Record*, 33(4):26–31, 2004.
42. L. Serafini, H. Stuckenschmidt, and H. Wache. A formal investigation of mapping language for terminological knowledge. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2005.
43. A. Sheth and J. Larson. Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Computing Surveys*, 22(3):183–236, 1990.
44. P. Shvaiko and J. Euzenat. A survey of schema-based macthing approaches. *Journal on Data Semantics (JoDS)*, IV, 2005.
45. P. Shvaiko, F. Giunchiglia, P. Pinheiro da Silva, and D. L. McGuinness. Web explanations for semantic heterogeneity discovery. In *Proceedings of the European Semantic Web Conference (ESWC)*, pages 303–317, 2005.
46. Y. Sure, O. Corcho, J. Euzenat, and T. Hughes. *Evaluation of Ontology-based Tools*. Proceedings of the 3rd International Workshop on Evaluation of Ontology-based Tools (EON), 2004. http://CEUR-WS.org/Vol-128/.
47. M. Uschold and M. Gruninger. Ontologies and semantics for seamless connectivity. *SIGMOD Record*, 33(4):58–64, 2004.
48. Y. Velegrakis, R. J. Miller, and J. Mylopoulos. Representing and querying data transformations. In *Proceedings of the International Conference on Data Engineering (ICDE)*, pages 81–92, 2005.
49. H. Wache, T. Voegele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and S. Huebner. Ontology-based integration of information - a survey of existing approaches. In *Proceedings of the workshop on Ontologies and Information Sharing at the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 108–117, 2001.
50. L. Yan, R. Miller, L. Haas, and R. Fagin. Data driven understanding and refinement of schema mappings. *SIGMOD Record*, 30(2):485–496, 2001.