

Ten Challenges for Ontology Matching

Pavel Shvaiko



Informatica Trentina, TasLab, Trento, Italy
Pavel.Shvaiko@infotn.it



Jérôme Euzenat



Montbonnot Saint-Martin, France
Jerome.Euzenat@inrialpes.fr

November 21, 2008

Outline

- 1 Introduction
- 2 Applications and use cases
- 3 Challenges for ontology matching
- 4 Conclusions

Outline

- 1 Introduction
- 2 Applications and use cases
- 3 Challenges for ontology matching
- 4 Conclusions

The semantic heterogeneity problem

... heterogeneity can be reduced in 2 steps:



o_1

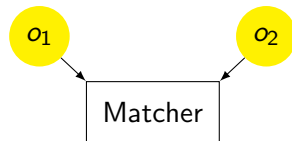


o_2

The semantic heterogeneity problem

... heterogeneity can be reduced in 2 steps:

(i) match,

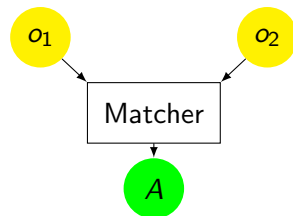


The semantic heterogeneity problem

... heterogeneity can be reduced in 2 steps:

(i) match,

thereby determine an alignment;



The semantic heterogeneity problem

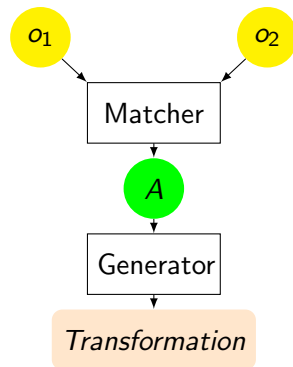
... heterogeneity can be reduced in 2 steps:

(i) match,

thereby determine an alignment;

(ii) generate

a processor (for merging, transforming, etc.)



The semantic heterogeneity problem

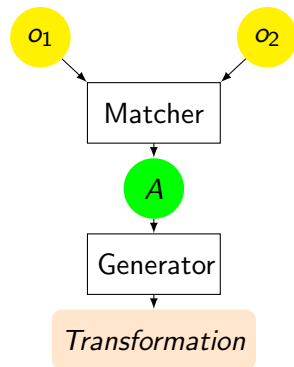
... heterogeneity can be reduced in 2 steps:

(i) match,

thereby determine an alignment;

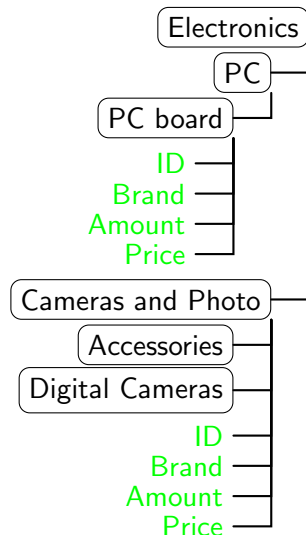
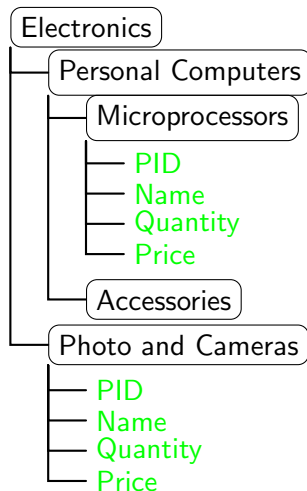
(ii) generate

a processor (for merging, transforming, etc.)

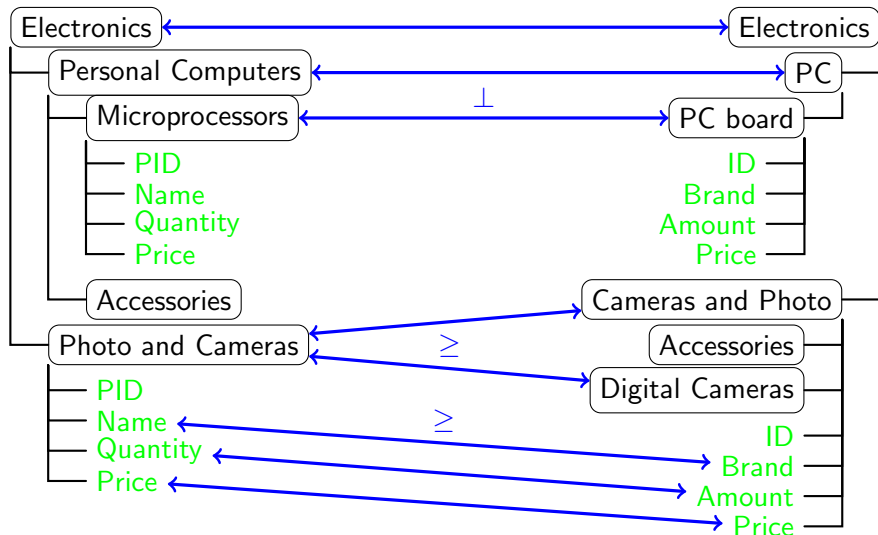


Matching can be performed at design time or at run time

Motivation: two XML schemas



Motivation: two XML schemas



Outline

- 1 Introduction
- 2 Applications and use cases
- 3 Challenges for ontology matching
- 4 Conclusions

Applications

► **Traditional**

- Ontology evolution
- Schema integration
- Catalog integration
- Data integration

Applications

▶ **Traditional**

- ▶ Ontology evolution
- ▶ Schema integration
- ▶ Catalog integration
- ▶ Data integration

▶ **Emergent**

- ▶ P2P information sharing
- ▶ Agent communication
- ▶ Web service composition
- ▶ Query answering on the web

Role of final users

- It is rarely the case that final users, i.e., final consumers of a product, are directly involved in the definition of use cases in ontology matching research projects (some exceptional examples include OpenKnowledge, STITCH, ...)

Role of final users

- ▶ It is rarely the case that final users, i.e., final consumers of a product, are directly involved in the definition of use cases in ontology matching research projects (some exceptional examples include OpenKnowledge, STITCH, ...)
- ▶ Involving final users into R&D cycles requires addressing a social challenge of integrating relevant actors: research centers, technology providers and user institutions ...

Role of final users

- ▶ It is rarely the case that final users, i.e., final consumers of a product, are directly involved in the definition of use cases in ontology matching research projects (some exceptional examples include OpenKnowledge, STITCH, ...)
- ▶ Involving final users into R&D cycles requires addressing a social challenge of integrating relevant actors: research centers, technology providers and user institutions ...
- ▶ In order to foster an early practical exploitation of the matching prototypes, it is necessary to directly involve final users in R&D cycles, e.g., by using such user-oriented open innovation methodologies as Living Labs

Outline

- 1 Introduction
- 2 Applications and use cases
- 3 Challenges for ontology matching
- 4 Conclusions

Ten challenges

1. Large-scale evaluation
2. Performance of ontology matching techniques
3. Discovering missing background knowledge
4. Uncertainty in ontology matching
5. Matcher selection and self-configuration
6. User involvement
7. Explanation of matching results
8. Social and collaborative ontology matching
9. Alignment management: infrastructure and support
10. Reasoning with alignments

Ten challenges

1. Large-scale evaluation
2. Performance of ontology matching techniques
3. Discovering missing background knowledge
4. Uncertainty in ontology matching
5. Matcher selection and self-configuration
6. User involvement
7. Explanation of matching results
8. Social and collaborative ontology matching
9. Alignment management: infrastructure and support
10. Reasoning with alignments

Large-scale evaluation

- ▶ OAEI campaigns gave only some preliminary evidence of the scalability characteristics of the ontology matching technology
 - ▶ Larger tests involving 10.000, 100.000, and 1.000.000 entities per ontology are to be designed and conducted
 - ▶ Issues of a wider automation for acquisition of reference alignments are to be tackled

Large-scale evaluation

- ▶ OAEI campaigns gave only some preliminary evidence of the scalability characteristics of the ontology matching technology
 - ▶ Larger tests involving 10.000, 100.000, and 1.000.000 entities per ontology are to be designed and conducted
 - ▶ Issues of a wider automation for acquisition of reference alignments are to be tackled
- ▶ Application specific measures

Large-scale evaluation

- ▶ OAEI campaigns gave only some preliminary evidence of the scalability characteristics of the ontology matching technology
 - ▶ Larger tests involving 10.000, 100.000, and 1.000.000 entities per ontology are to be designed and conducted
 - ▶ Issues of a wider automation for acquisition of reference alignments are to be tackled
- ▶ Application specific measures
- ▶ Semi-automatic test generation methods of desired test hardness

Large-scale evaluation

- ▶ OAEI campaigns gave only some preliminary evidence of the scalability characteristics of the ontology matching technology
 - ▶ Larger tests involving 10.000, 100.000, and 1.000.000 entities per ontology are to be designed and conducted
 - ▶ Issues of a wider automation for acquisition of reference alignments are to be tackled
- ▶ Application specific measures
- ▶ Semi-automatic test generation methods of desired test hardness
- ▶ Interoperability benchmarks

Discovering missing background knowledge

- ▶ Various strategies have been used so far:
 - ▶ Declaring the missing axioms manually as a pre-match effort
 - ▶ Reusing previous match results
 - ▶ Querying the web
 - ▶ Using domain specific corpus or ontologies
 - ▶ Using ontologies available on the semantic web
 - ▶ ...

Discovering missing background knowledge

- ▶ Various strategies have been used so far:
 - ▶ Declaring the missing axioms manually as a pre-match effort
 - ▶ Reusing previous match results
 - ▶ Querying the web
 - ▶ Using domain specific corpus or ontologies
 - ▶ Using ontologies available on the semantic web
 - ▶ ...
- ▶ Still, these approaches have to be systematically investigated, combined in a complementary fashion and improved

Discovering missing background knowledge

- ▶ Various strategies have been used so far:
 - ▶ Declaring the missing axioms manually as a pre-match effort
 - ▶ Reusing previous match results
 - ▶ Querying the web
 - ▶ Using domain specific corpus or ontologies
 - ▶ Using ontologies available on the semantic web
 - ▶ ...
- ▶ Still, these approaches have to be systematically investigated, combined in a complementary fashion and improved
- ▶ In dynamic settings the matching input is often shallow, ... it is useful to compute the minimal background knowledge necessary to resolve a particular problem

User involvement

- ▶ In traditional applications semi-automatic matching/design-time interaction is a promising way to improve quality of the results

User involvement

- ▶ In traditional applications semi-automatic matching/design-time interaction is a promising way to improve quality of the results
- ▶ Burdenless to the user interaction schemes
 - ▶ Usability
 - ▶ Scalability of visualization

User involvement

- ▶ In traditional applications semi-automatic matching/design-time interaction is a promising way to improve quality of the results
- ▶ Burdenless to the user interaction schemes
 - ▶ Usability
 - ▶ Scalability of visualization
- ▶ Exploit the user feedback
 - ▶ to adjust matcher parameters
 - ▶ to take it as (partial) input alignment to a matcher
 - ▶ ...

User involvement

- ▶ In traditional applications semi-automatic matching/design-time interaction is a promising way to improve quality of the results
- ▶ Burdenless to the user interaction schemes
 - ▶ Usability
 - ▶ Scalability of visualization
- ▶ Exploit the user feedback
 - ▶ to adjust matcher parameters
 - ▶ to take it as (partial) input alignment to a matcher
 - ▶ ...
- ▶ In dynamic settings, agents involved in the matching process can negotiate the mismatches in a fully automated way

Social and collaborative ontology matching

- ▶ The network effect:
 - ▶ Each person has to do a small amount of work
 - ▶ Each person can improve on what has been done by others
 - ▶ Errors remain in minority

Social and collaborative ontology matching

- ▶ The network effect:
 - ▶ Each person has to do a small amount of work
 - ▶ Each person can improve on what has been done by others
 - ▶ Errors remain in minority

- ▶ A community of people can share alignments and argue about them by using annotations

Social and collaborative ontology matching

- ▶ The network effect:
 - ▶ Each person has to do a small amount of work
 - ▶ Each person can improve on what has been done by others
 - ▶ Errors remain in minority

- ▶ A community of people can share alignments and argue about them by using annotations

- ▶ The key issues are to:
 - ▶ Provide adequate annotation support and description units
 - ▶ Handle adequately contradictory and incomplete alignments
 - ▶ Incentivise active user participation
 - ▶ Handle adequately the malicious users

Outline

- 1 Introduction
- 2 Applications and use cases
- 3 Challenges for ontology matching
- 4 Conclusions

Summary

- ▶ We outlined various matching applications and emphasized the role of final users in R&D cycles
- ▶ We discussed several challenges for ontology matching
- ▶ We believe that these challenges are on the critical path, hence, addressing them should accelerate the progress of ontology matching

Acknowledgments

- ▶ The first author appreciates support from the **Trentino as a Lab** (TasLab) project of the **European Network of the Living Labs**. The second author has been partially supported by the European integrated project **NeOn** (IST-2005-027595) and the **RNTL project WebContent**.
- ▶ We are thankful to the TasLab group members: Isabella Bressan, Ivan Pilati, Valentina Ferrari, Luca Mion and Marco Combetto for many fruitful discussions on the Living Labs methodology for innovation.
- ▶ We are grateful to Fausto Giunchiglia, Maurizio Marchese, Mikalai Yatskevich, Roberta Cuel (University of Trento), Lorenzo Vaccari (Urban Planning and Environment Protection department of the Autonomous Province of Trento), Paolo Besana (University of Edinburgh), Marta Sabou (Open University), Antoine Zimmermann (INRIA), Zharko Aleksovski (Vrije Universiteit Amsterdam), and Malgorzata Mochol (Free University of Berlin) for their support and insightful comments.

Ontology matching the book

Jérôme Euzenat, Pavel Shvaiko

Ontology matching

1. Applications
2. Problem definition
3. Classification
4. Basic techniques
5. Strategies
6. Systems
7. Evaluation
8. Representation
9. Explanation
10. Processing

<http://book.ontologymatching.org>



Thank you
for your attention and interest!

Questions?

Pavel.Shvaiko@infotn.it

Jerome.Euzenat@inrialpes.fr