Interoperability

Introduction to the Semantic Web Tutorial at ISWC 2010

Jérôme Euzenat

INRIA & LIG

Montbonnot, France
Jerome.Euzenat@inrialpes.fr

Thanks to Natasha Noy for our collaboration on a former version of these slides

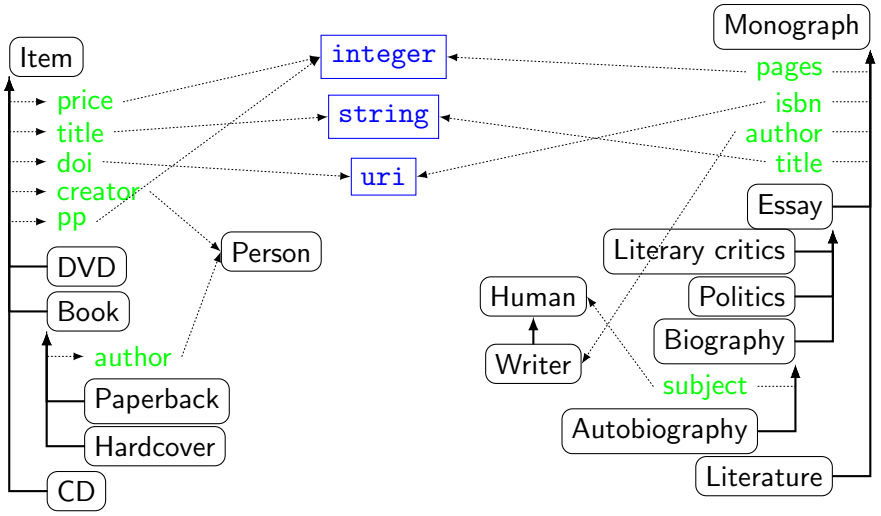
What you have learned so far

- Data can be expressed in RDF
- Linked through URIs
- Modelled with OWL ontologies
- Retrieved through SPARQL queries

Being serious about the semantic web

- It is not one person's ontology
- It is not several people common ontology
- It is many people's many ontologies
- So it is a mess, but a meaningful mess.

Ontology heterogeneity

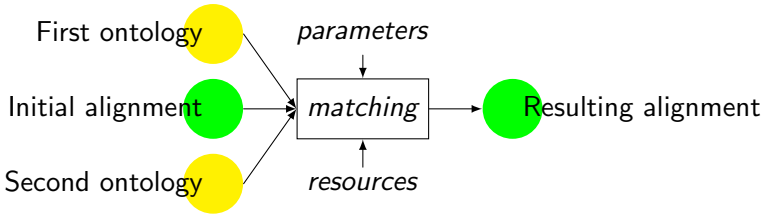


Heterogeneity problem

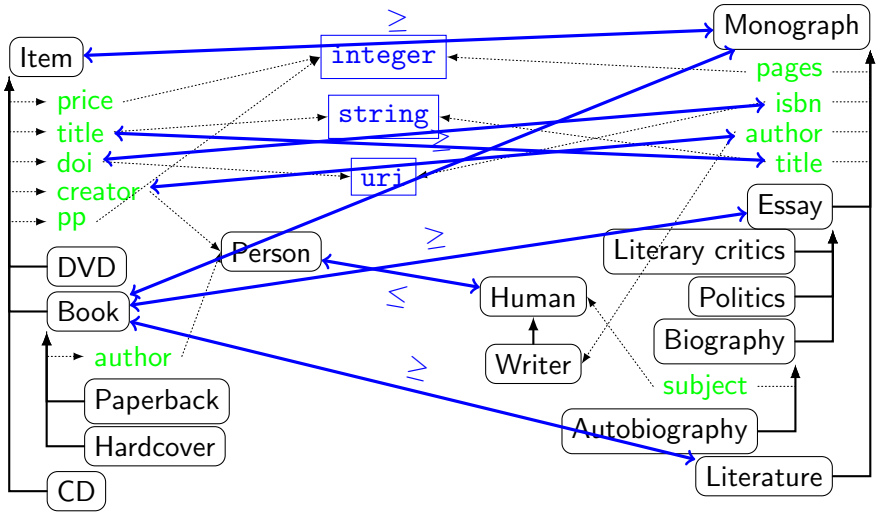
Resources being expressed in different ways must be reconciled before being used.
Mismatch between formalized knowledge can occur when:

- different languages are used (OWL vs. Topic maps);
- different terminologies are used:
  - English vs. Chinese;
  - Book vs. Monograph.
- different models are used:
  - different classes: Autobiography vs. Paperback;
  - classes vs. property: Essay vs. literarygenre;
  - classes vs. instances: One physical book as an instance vs. one work as an instance.
- different scopes and granularity are used.
  - Only books vs. cultural items vs. any product;
  - Books detailed to the print and translation level vs. books as works.

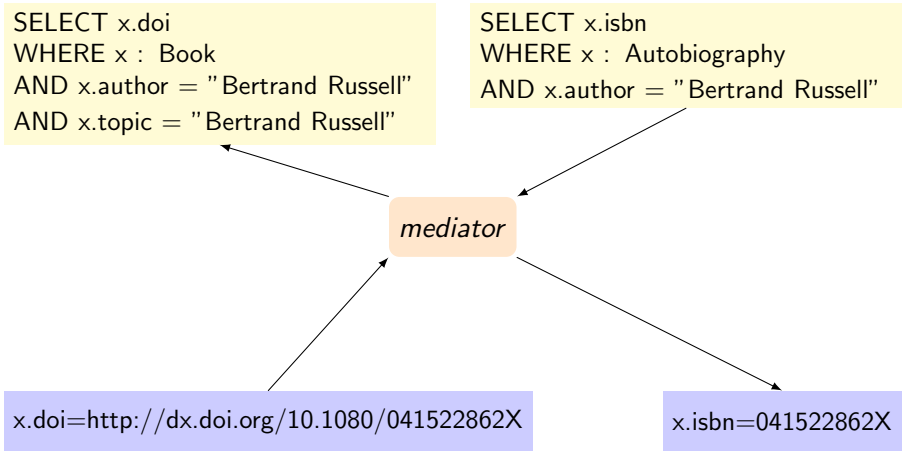
How can we address the problem?



Ontology alignment



Transformation and mediation

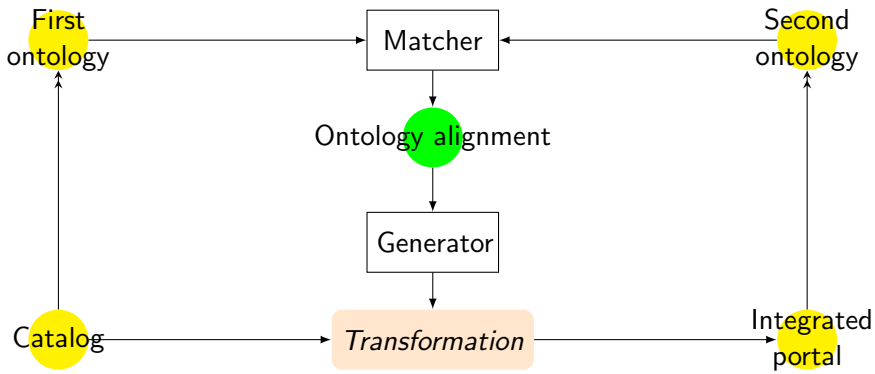


Why should we deal with this?

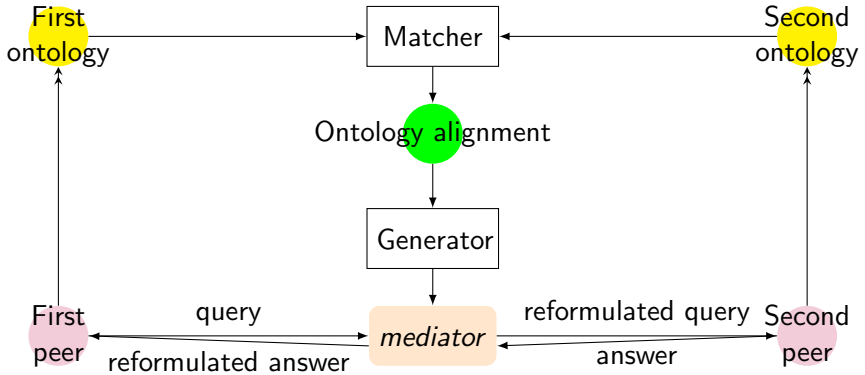
Applications of semantic integration

- Catalogue integration
- Schema and data integration
- Query answering
- Peer-to-peer information sharing
- Web service composition
- Agent communication
- Data transformation
- Ontology evolution

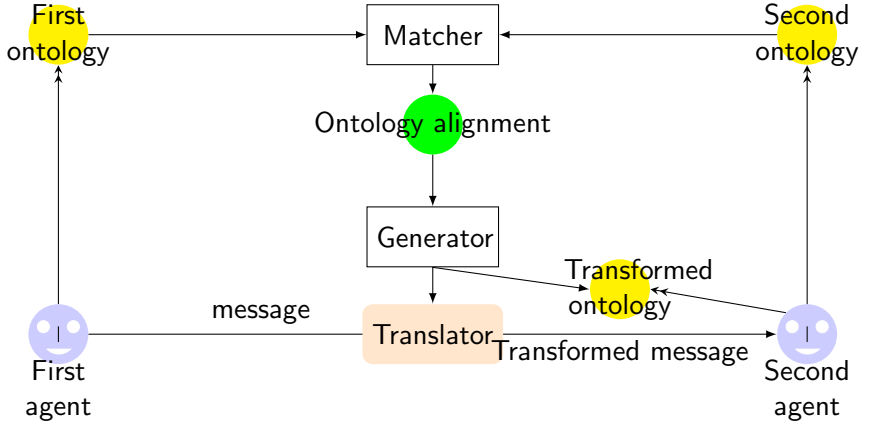
Application: Catalog integration



Applications: Query answering



Applications: Agent communication



Ontology matching in three steps

Reconciliation can be performed in 3 steps

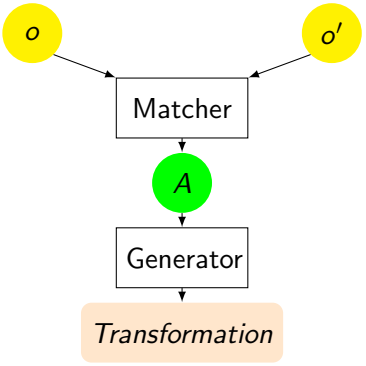
Match,

thereby determines the alignment

Generate

a processor (for merging, transforming, etc.)

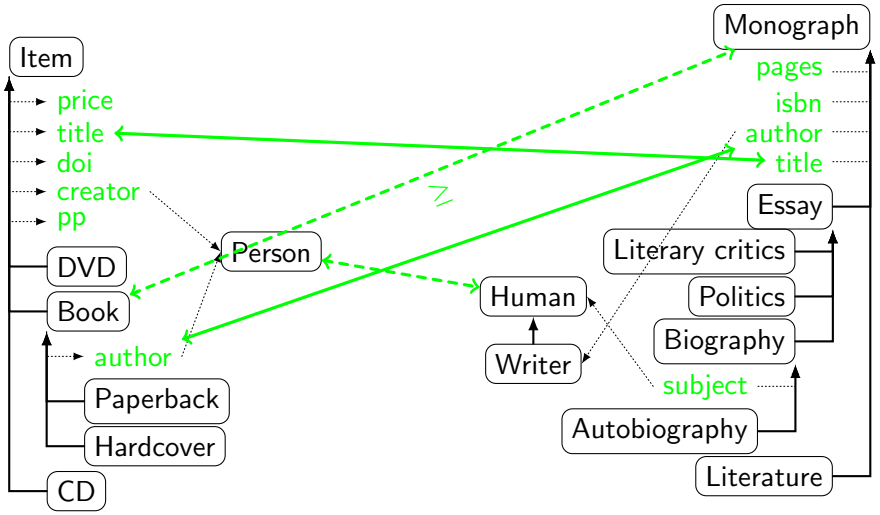
Apply



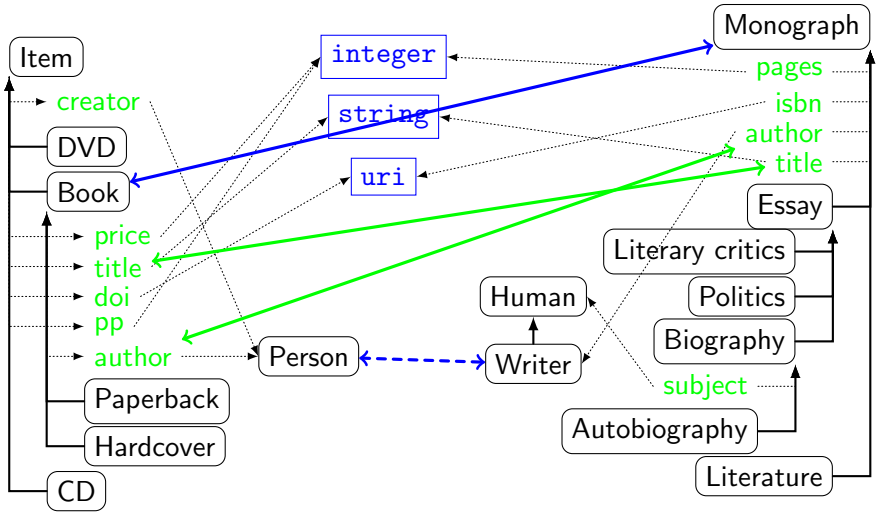
On what basis can we match?

- Content: relying on what is inside the ontology
  - Name, comments, alternate names, names of related entities: NLP, IR, etc.
  - Internal structure: constraints on relations, typing
  - External structure: relations between entities: Data mining, Discrete mathematics
  - Extension: Statistics, data analysis, data mining, machine learning
  - Semantics (models): Reasoning techniques
- Context: the relations of the ontology with the outside
  - Annotated resources:
  - The web
  - External ontologies: dbpedia, etc.
  - External resources: wordnet, etc.

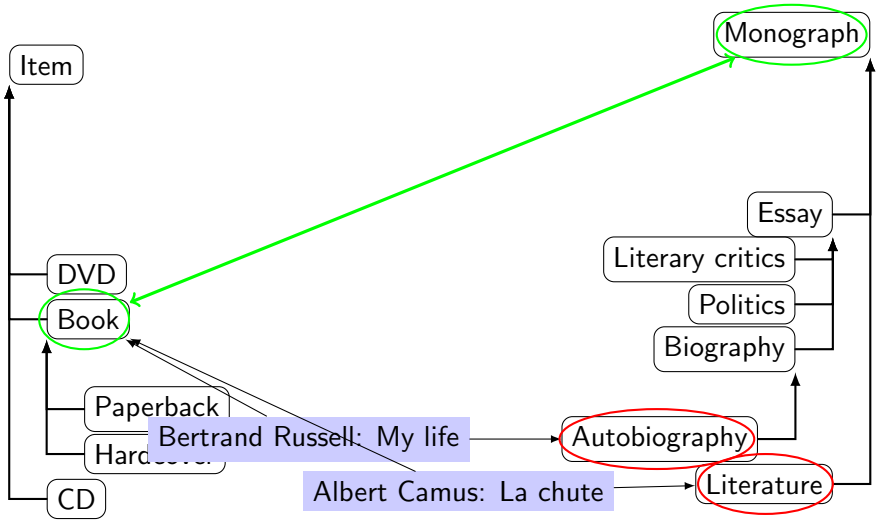
Name similarity



Structure similarity

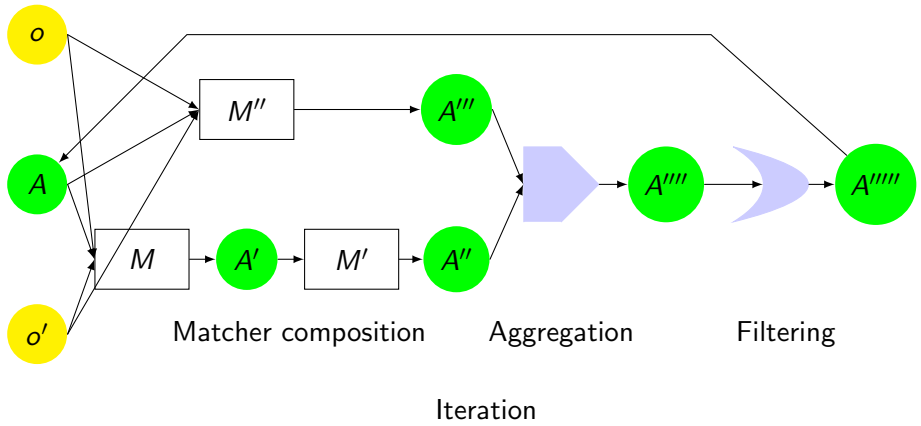


Instance similarity



Combining different techniques

Basic matchers provide candidate correspondences, most of the systems use several such matchers and further combine and filter their results.



How well do these approaches work?

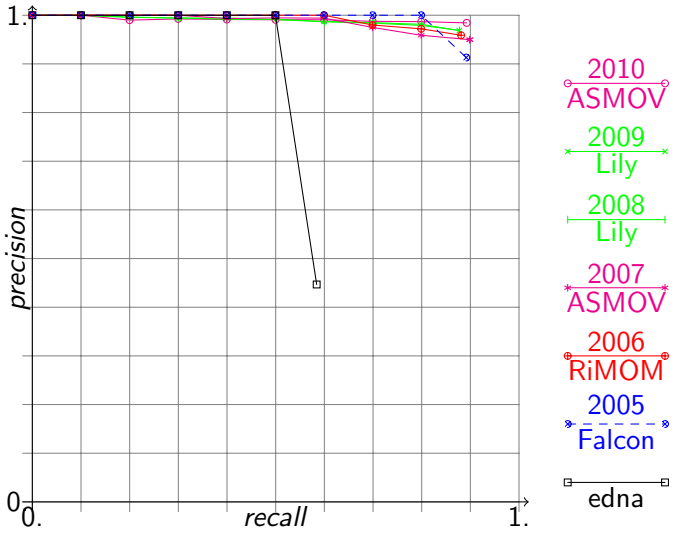
Ontology Alignment Evaluation Initiative (OAEI)

- Formal comparative evaluation of different ontology-matching tools;
- Run every year since 2004;
- Variety of test cases (in size, in formalism, in content);
- Results consistent across test cases;
- Results very dependent on the tasks and the data (from under 50% of precision and recall to well over 80% if ontologies are relatively similar)
- Progress every year!

`http://oaei.ontologymatching.org`

Now involved in the SEALS (Semantics Evaluation At Large Scale) project.

Benchmark results (precision and recall curves)



Tools you should be aware of

- Frameworks
  - PROMPT (a Protégé plug-in): includes a user interface and a plug-in architecture.
  - Alignment API: used by many tools; provides an exchange format and evaluation tools for OAEI.
  - COMA++: oriented toward database integration (many basic algorithms implemented).
- Matching systems
  - OAEI best performers (Falcon, RiMOM, ASMOV, etc.)
  - Available systems (FOAM, Falcon, COMA++, Aroma, etc.)

Selected challenges

- Scalability and efficiency
  - Current matchers can be fast, scale and accurate, but not all at once.
- New sources of matching
  - Context-based matching,
- General purpose matching (vs. special purpose matching)
  - Matcher combination,
  - Matcher selection and self-configuration,
- User involvement,
  - Matching (serendipitously) while working,
  - How to explain alignments?
  - Social and collaborative ontology matching,
- Alignment management: infrastructure and support,
  - How do we maintain alignments when ontologies evolve?
  - Reasoning with alignments,
  - Being robust to incorrect alignments.

and, of course, many others,

Further reading

- "Ontology Matching" by Euzenat and Shvaiko
- Proceedings of ISWC, ASWC, ESWC, WWW conferences, etc.
- Journal of web semantics, Journal on data semantics, etc.
- `http://www.ontologymatching.org`

Jerome.Euzenat@inria.fr

http://exmo.inrialpes.fr