The dynamic knowledge medium

Jérôme Euzenat

Montbonnot, France
Jerome.Euzenat@inria.fr
http://exmo.inria.fr

November 29, 2013

INRIA

(French) National Research Institute in Computer Science and Control

- Governmental non academic research institute;
- 1500 employees — 8 centers;
- Bordeaux - Grenoble - Lille - Nancy - Orsay - Rocquencourt - Rennes - Sophia-Antipolis;
- 170 teams — 2900 scientists (3700 in total).

LIG

Grenoble Computer Science Laboratory

- Laboratory associated to three universities of Grenoble, CNRS and INRIA;
- Axes:
  - Software and information system engineering
  - Formal methods, models, and languages
  - Interactive and cognitive systems
  - Distributed systems, parallel computing, and networks
  - Data and knowledge processing at large scale
- Network — Software — Interaction — Knowledge;
- 22 teams — 165 scientists (470 in total).

Exmo

- Small team (three permanent researchers, 1 post doc, 3 PhD students);
- Recognised by both in INRIA and LIG;
- Started in 2000; formally in 2003;
- Computer-mediated exchanges of formalised knowledge (Échanges de connaissance formalisée médiatisés par ordinateur);
- Focussed on Ontology matching.

Outline

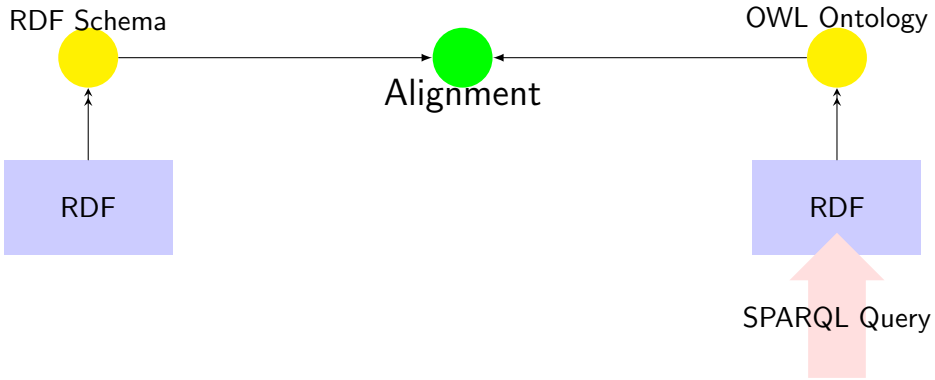
Our goal

Sharing knowledge at web scale

⇓

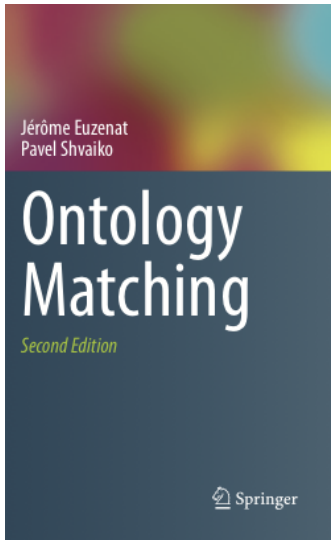
Dealing with heterogeneity

Semantic web technologies



W3C Recommendations

Research topics



- Ontology matching
- Alignment semantics
- Alignment composition (inference)
- Data interlinking and (link)key inference
- Query evaluation and containment

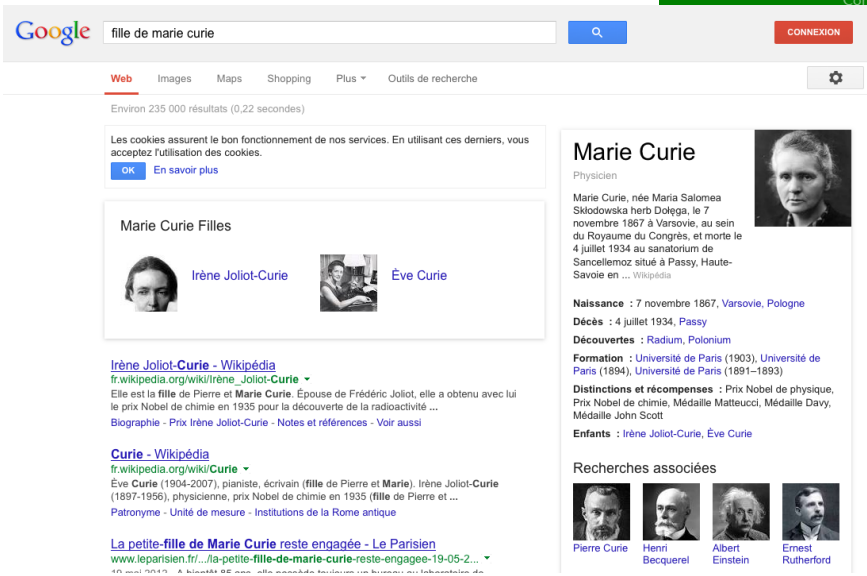
Theory — Software — Experiments

Message of the talk

"The semantic web may be seen as a medium through which people communicate knowledge. Thanks to the vast amount of related knowledge it carries, it helps expressing meaning. However, due to its formal structure it may reveal as a difficult medium. We need to find ways to smoothly evolve it to ease communication. This means that such a medium is a dynamic structure. Fortunately."

Jérôme Euzenat, 29.11.2013

The state of the semantic web



Data, ontologies, alignments

Plenty of data

- LOD cloud (2011: 300 datasets, 31G triples, 500M links),
- CKAN/Datahub (2012: 4k datasets not all RDF, revision frequency increases)
- dpbedia.org (2012: 10M resources, 1.9G triples)
- Sindice (2011: 12+G triples)
- ODG, schema.org? (order of billions), GKG (2012: 500M resources, 3.5G facts)

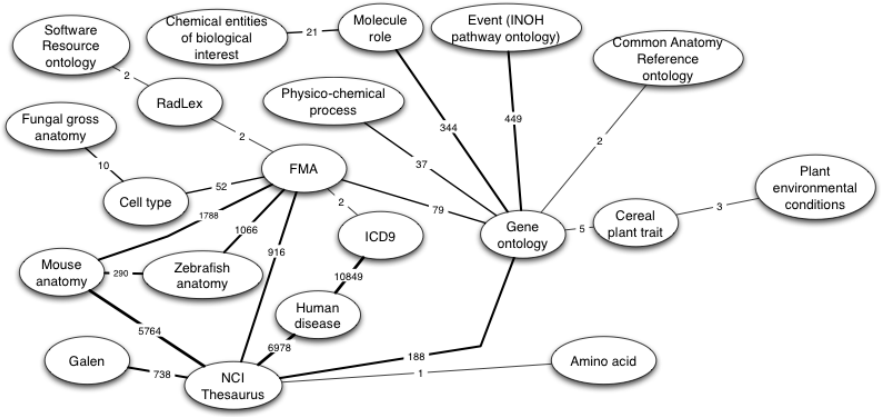
Many ontologies

- Swoogle (200?: 10k ontologies)
- Watson (2007: 9.4k ontologies, 1.1M terms)
- Falcon (2012: 2.9k ontologies, 455k terms)
- LOV (2012: 284 ontologies)

Less alignments

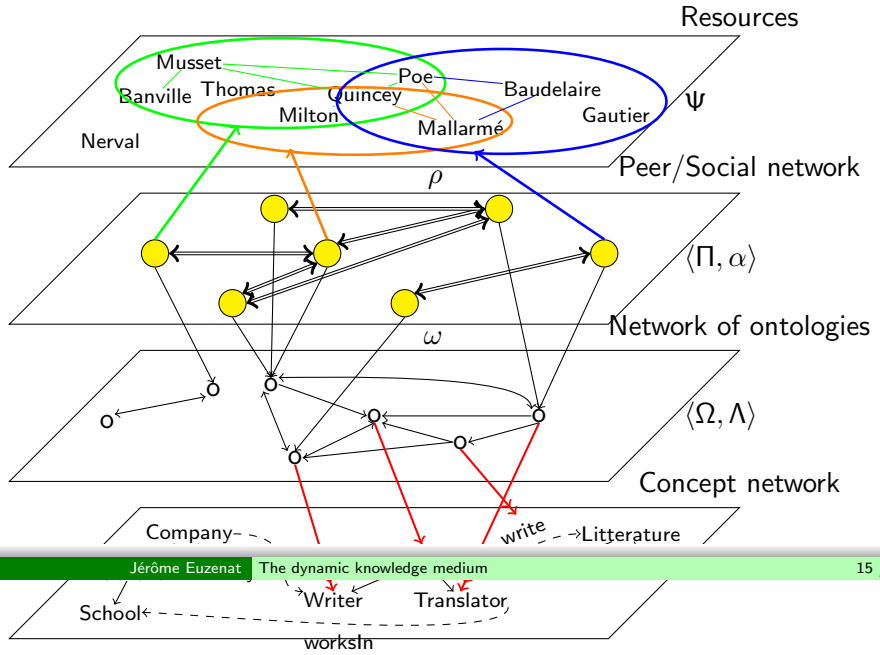
- Alignment server,
- Bioportal (322 ontologies, 5M terms, 10M mappings)

The Bioportal network



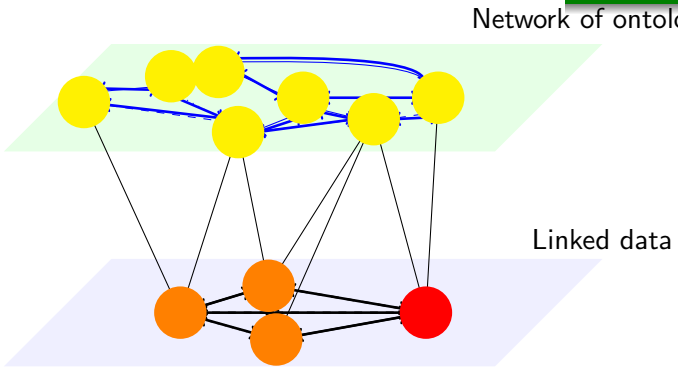
From: Natalya Noy, Nicholas Griffith, Mark Musen, Collecting Community-Based Mappings in an Ontology Repository, Proc. 7th International Semantic Web Conference (ISWC), pp371–386, 2008

Think globally: Anatomy of a semantic P2P network



The playground

It is alive!



- From ontology alignments to linked data;
- From linked data to ontology alignments;
- More local inference (composition);
- Reasoning;
- Repairing, trusting, learning;

Networks of ontologies are living things

- New ontologies appear or disappear;
- New alignments are created by different matchers;
- Alignments appear or disappear;
- Errors are corrected;
- Views of the world evolve;
- Data are created every millisecond.

We must be robust to that.

Conclusion: keep it alive

- The semantic web is like the web: heterogeneous and dynamic;
- Tools must be developed to deal with this;
- Not fight it.

A medium?

A medium is a space (milieu) which supports communication.

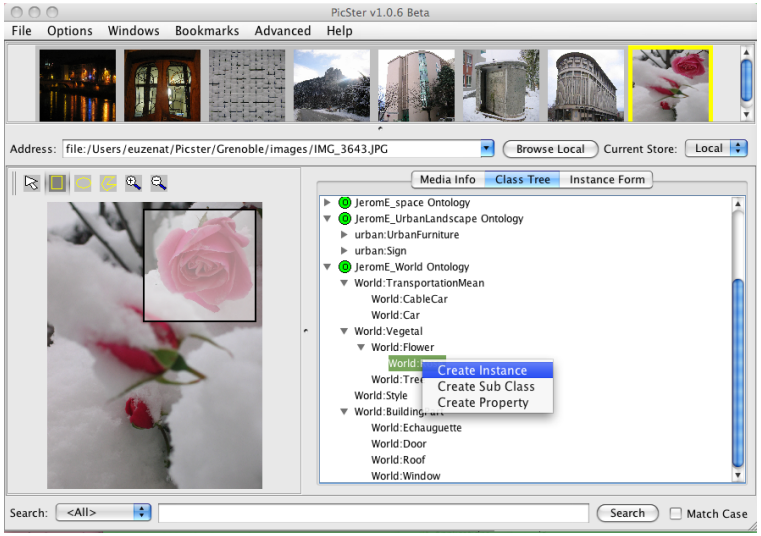
What kind of medium the semantic web is?

- an active medium, tries to adapt the message,
- a dynamic medium: it is always changing,
- a formal medium

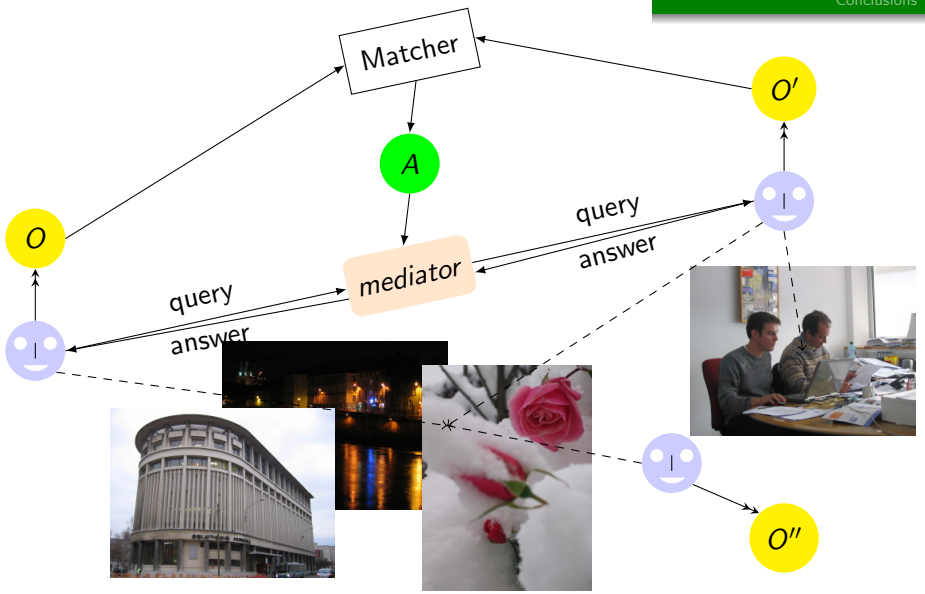
Examples

- The telephone is a transparent medium (delay, awkwardness) but does not touch much the message,
- The web started first as a literal medium, but went more and more
- . . . through search engines, first based on literal content, then on global opinion links, and through the mediation of complex algorithm, and now it uses the semantic web part...g,
- The thesis of Luca is about an active medium: it tries to present information in a way which it feels adapted to the situation (less awkwardness)

PicSter: an application



Interoperability in semantic P2P systems



Refinement

- How do we interact with the semantic web?
- How do we interact together through the semantic web?
- Can it be both flexible or rigid?

Communication problems

Here the communication is through people expressing ontologies, or publishing data.

- Inconsistency
- Unexpected consequences
- Incomplete consequences
- ≈ Misunderstandings

Semantic feedback

How to detect such problems?

- Direct feed back from users;
- Indirect usage feedback (not achieving –common– goals);
- Indirect feedback through data injection.

Recovery strategies/smooth operation

- Revise the ontology? the data? the aligments? Everything.
- Select a (consistent) subset of the semantic web which is sufficient for current communication?
- Prototype
- Maintain parallel alternative views?
- Find a way to negotiate ontology refinement so that it becomes consistent?

Conclusions

- The semantic web is a medium;
- Its formal structure mandates to develop evolution techniques;
- Much work has to be done.
- Fortunately.

If the medium is the message

What is the message of the semantic web?

Answers?

`http://exmo.inria.fr`

`Jerome . Euzenat @ inria . fr`