Semantic web evolution

tectonic quake or gentle drift?

Jéréme Euzenat

Laboratoire d’Informatique de Grenoble
Montbonnot, France
Jerome.Euzenat@inria.fr

November 19, 2016

Outline

The semantic web is a success Gloria allegro (ma no troppo)

Why is evolution a problem? Lamento andante

Addressing evolution Vivace moderato

Who am I

1992 Researcher at INRIA
2000 Started creating the semantic web :-) 
2012 Realised what we have done
2016 Fearing that it could break
Semantic web technologies

- RDF Schema
- Alignment
- OWL Ontology

Application of semantic web technologies:
- **Semantic web services** in which web services are semantically annotated;
- **Semantic P2P systems** in which shared resources are semantically annotated;
- **Semantic social networks** in which social relationships are semantically annotated;
- **Ambient intelligence** in which sensors, devices and information are semantically annotated;
- **Linked data** in which data is published with semantic web technologies;
- **Smart cities** in which city data is exchanged would benefit in using semantic web technologies.

The semantic web is a success!

Such technologies are used every day (by yourself).
- Tens of billions of RDF triples and thousands of ontologies on the web;
- Governments and their agencies publish their data in RDF;
- Facebook (OG), Google (GKG), Yahoo, Microsoft (schema.org) produce and consume semantic markup.
- And you do not even have to notice it.

Big data without knowledge?

Big data = large data sets + maths

\[ R_z = \frac{n}{\frac{\theta + \psi}{2}} = \frac{1}{2} \log \left| \frac{B}{A} \right| \]

If this is not informed... the result will be: a plot!
The semantic web is a success
Why is evolution a problem?
Addressing evolution

Bio2RDF

Drugbank in Bio2RDF

I am a biologist
looking to compare various available insulin products on the market
The semantic web is a success
Why is evolution a problem?
Addressing evolution

Drugbank in Bio2RDF (cont’d)

Insulin receptor [drugbank_target:36] (receptor)

Drugbank in Bio2RDF (receptor)

Insulin receptor [drugbank_target:36] (receptor)

Taxa in Bio2RDF

man[taxon:3606] (receptor)

Jérôme Euzenat
Semantic web evolution

18 / 0

19 / 0

21 / 0
The semantic web is a success
Why is evolution a problem?
Addressing evolution

Jérôme Euzenat
Semantic web evolution

The semantic web is a success
Why is evolution a problem?
Addressing evolution

Jérôme Euzenat
Semantic web evolution

Uniprot in Bio2RDF

The semantic web is a success
Why is evolution a problem?
Addressing evolution

Jérôme Euzenat
Semantic web evolution

Drugbank in Bio2RDF (dereferencing)

The semantic web is a success
Why is evolution a problem?
Addressing evolution

Jérôme Euzenat
Semantic web evolution
The semantic web is a success
Why is evolution a problem?
Addressing evolution

Bio2RDF SPARQL Query

Bio2RDF SPARQL Answers

Bio2RDF Resources
Problems with links

- The web relies on a robust design: links can break (404), but Human can cope.
- Nowadays... it is not clear that this is true anymore (think API changes). Quite some web site do not fail safely.
- With the semantic web, made for machine, the problem worsen (404 in your city... they call it a traffic jam).

Problems with knowledge: ways it could break

- incorrect property in data;
- incorrect subclass relation in ontology;
- incorrect membership relation between instance and class;
- incorrect sameAs link between data;
- incorrect correspondence in alignments;
- ...

Hence, the question is not if it will break, but what to do then. Remember all this is decentralised and distributed.
The semantic web is a success
Why is evolution a problem?
Addressing evolution

Knowledge interpretation

An engineering approach

Define a communication protocol
Define a message format
← Implement the protocol
Implement the protocol →
It works

Solutions?

- Freeze everything
- Track everything (ensure that everything is correct beforehand)
- Let things fail and repair them

We have lost adaptation capabilities
A living web

- Preserving decentralisation, distribution and diversity
- Accepting changes
- i.e., Breaking with the engineering approach
- Taking inspiration from how societies evolve

From the ontology matching standpoint:
- Going from "Match first, then communicate"
- To: "Try to communicate, and if it breaks match"

Cultural evolution

Comes from anthropology (and population genetics)

Applies evolution theory to culture:
- Culture is an “intellectual artefact”
- which is transmitted (from generation to generation but not exclusively)
- which can be subject to selection.

Experimental cultural evolution

- Pioneered by Robert Axelrod
- Applies multi-agent simulation to cultural artifacts
- Successfully applied to natural language by Luc Steels and colleagues
- Offers a systematic experimentation framework in which agents play "games"
Combining

**Knowledge representation**

and

**Experimental cultural evolution**

for continuous knowledge evolution

---

### Knowledge game setting

- **Environment**: populated by objects characterised by $n$ dimensions: $\Box$, $\triangle$, $\blacksquare$, $\blacktriangle$, $\Diamond$, $\blacktriangleleft$.
- **Population**: $n$ agents with their own representations (ontologies)
- **Initialisation**: randomly generated alignments between their representations
- **Game**: an agent draws randomly an object and ask to another (randomly selected) agent to which class it belongs. The former agent uses the alignments for determining to which class the entity belongs in his own ontology.
- **Success**: the resulting class subsumes the class of the object
- **Failure**: the class is disjoint (exclusive)
- **Repair**: (a) suppress the correspondence; (b) replace it by a weaker correspondences; (c) add an entailed correspondence.

**Secondary measure**: (Semantic) F-measure

**External evaluation**: Compare to Alcomo, LogMap

---

### Ontology network
The semantic web is a success
Why is evolution a problem?
Addressing evolution

Alignment repair game

Experimental questions
1. Does the process converge?
2. What is the effect of repair modalities?
3. How does this compare to baselines?
4. Does it scale?

Convergence and modalities

modality=add; #agents=4; #games=2000; #runs=1
# Convergence and modalities

![Graph showing convergence and modalities](image)

- modality=del,repl,add; #agents=4; #games=2000; #runs=10

```plaintext
del  success rate  F-measure
replace  success rate  F-measure
add  success rate  F-measure
```

# Modalities and baselines

<table>
<thead>
<tr>
<th>Modality</th>
<th>Size</th>
<th>Success rate</th>
<th>Incoherence</th>
<th>Semantic F-measure</th>
<th>Syntactic F-measure</th>
<th>Convergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>reference</td>
<td>70</td>
<td>-</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>-</td>
</tr>
<tr>
<td>initial</td>
<td>54</td>
<td>-</td>
<td>[0.46-0.49]</td>
<td>0.20</td>
<td>(0.20)</td>
<td>-</td>
</tr>
<tr>
<td>delete</td>
<td>6</td>
<td>0.98</td>
<td>0.0</td>
<td>0.16</td>
<td>(0.16)</td>
<td>400</td>
</tr>
<tr>
<td>replace</td>
<td>6</td>
<td>0.95</td>
<td>0.0</td>
<td>0.16</td>
<td>(0.16)</td>
<td>1000</td>
</tr>
<tr>
<td>add</td>
<td>12.7</td>
<td>0.89</td>
<td>0.0</td>
<td>0.23</td>
<td>(0.16)</td>
<td>1330</td>
</tr>
<tr>
<td>Alcomo</td>
<td>25.5</td>
<td>-</td>
<td>0.0</td>
<td>0.26</td>
<td>(0.14)</td>
<td>-</td>
</tr>
<tr>
<td>LogMap</td>
<td>36.5</td>
<td>-</td>
<td>0.0</td>
<td>0.26</td>
<td>(0.14)</td>
<td>-</td>
</tr>
</tbody>
</table>

- modality=del,repl,add; #agents=4; #games=2000; #runs=10

# Problem solving vs. survival

- The number of games for converging (reaching perfect communication) grows fast (with $n$);
- Indeed the probability of finding, at random, the last failure is really low;
- It is possible to produce an algorithm that converges faster;
- But this is not the problem;
- Their goal is not to solve a problem, but to live;
- How many more do you think it will take you to reach perfect communication with your closest relatives?

```plaintext
# Scalability

<table>
<thead>
<tr>
<th># agents</th>
<th>Initial</th>
<th>LogMap</th>
<th>Alcomo</th>
<th>Final</th>
<th>Initial</th>
<th>LogMap</th>
<th>Alcomo</th>
<th>Final</th>
<th>Convergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.31</td>
<td>0.0</td>
<td>0.0</td>
<td>0.32</td>
<td>0.35</td>
<td>0.36</td>
<td>0.36</td>
<td>0.33</td>
<td>300</td>
</tr>
<tr>
<td>4</td>
<td>0.47</td>
<td>0.0</td>
<td>0.0</td>
<td>0.20</td>
<td>0.24</td>
<td>0.25</td>
<td>0.21</td>
<td>1670</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.58</td>
<td>0.0</td>
<td>0.0</td>
<td>0.11</td>
<td>0.18</td>
<td>0.17</td>
<td>0.24</td>
<td>5400</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.63</td>
<td>0.0</td>
<td>0.0</td>
<td>0.06</td>
<td>0.12</td>
<td>0.11</td>
<td>0.14</td>
<td>10.000+</td>
<td></td>
</tr>
</tbody>
</table>
```

- modality=add; #agents=3,4,5,6; #games=10000; #runs=10
The semantic web is a success (and you need it)

- It can easily break due to the world evolving
- But the world is evolving (you will not stop that)
- So our design should be prepared to that
- Cultural evolution seems and appropriate inspiration

Jerome.Euzenat@inria.fr

http://exmo.inria.fr