On the Feasibility of Using OWL 2 Reasoners in Ontology Alignment Repair Problems

Alessandro Solimando, Ernesto Jiménez-Ruiz, Giovanna Guerrini

DIBRIS, Università di Genova, Italy
Department of Computer Science, University of Oxford

OWL Reasoner Evaluation Workshop
Athens, June 6 2015
Outline

Preliminaries

Incoherence and Mapping Repair

Evaluation

Conclusions and Future Work
Ontology mappings

Mappings $\mathcal{M}$ are tuples $\langle e_1, e_2, n, \rho \rangle$

- $e_1, e_2$ are entities in the input ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$
- $n$ a confidence value between 0 and 1
- $\rho$ is the semantic relationship between $e_1$ and $e_2$ (e.g. subsumption, equivalence or disjointness)

Formalized as OWL 2 axioms

- Where the semantic relationship $\rho$ is one of $\{\equiv, \subseteq, \sqsubseteq, \perp\}$
- Confidence values $n$ are represented as axiom annotations
- No extra semantics
Ontology mappings

Mappings $\mathcal{M}$ are tuples $\langle e_1, e_2, n, \rho \rangle$

- $e_1, e_2$ are entities in the input ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$
- $n$ a confidence value between 0 and 1
- $\rho$ is the semantic relationship between $e_1$ and $e_2$ (e.g. subsumption, equivalence or disjointness)

Formalized as OWL 2 axioms

- Where the semantic relationship $\rho$ is one of $\{\equiv, \sqsubseteq, \sqsupseteq, \bot\}$
- Confidence values $n$ are represented as axiom annotations
- No extra semantics
Challenges in ontology integration

Challenges

• Ontologies are being developed by different groups, and
• use different classifications and naming schemas.
• Tool support required: ontologies may contain tens of thousands of entities: FMA (78,989 classes), NCI (66,724 classes) or SNOMED CT (306,591 classes) are prominent examples.
• Ontology integration is required in diverse scenarios: data migration, query answering, communication between agents, knowledge reuse, etc.
Ontology Matching System

- Given two input ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$ provides a set of Mappings $\mathcal{M}$ as output.

Quality of a set of mappings $\mathcal{M}$

- **Precision** and **recall** wrt reference mappings or gold standard $|\mathcal{M}_{GS}|$
  - Precision (Pre) = $|\mathcal{M} \cap \mathcal{M}_{GS}|/|\mathcal{M}|$
  - Recall (Rec) = $|\mathcal{M} \cap \mathcal{M}_{GS}|/|\mathcal{M}_{GS}|$
  - The F-score (F) = $(2 \times \text{Pre} \times \text{Rec})/(\text{Pre} + \text{Rec})$.
- **Coherence** of $\mathcal{M}$ wrt $\mathcal{O}_1$ and $\mathcal{O}_2$.
- Computation times are also considered.
Ontology Matching System

• Given two input ontologies $O_1$ and $O_2$ provides a set of Mappings $M$ as output.

Quality of a set of mappings $M$

• **Precision** and **recall** wrt reference mappings or gold standard $|M_{GS}|$
  - Precision (Pre) = $|M \cap M_{GS}|/|M|$
  - Recall (Rec) = $|M \cap M_{GS}|/|M_{GS}|$
  - The F-score (F) = $(2 \times \text{Pre} \times \text{Rec})/(\text{Pre} + \text{Rec})$.

• **Coherence** of $M$ wrt $O_1$ and $O_2$.
• Computation **times** are also considered.
Evaluation of Ontology Matching Systems: OAEI

Ontology Alignment Evaluation Initiative (OAEI)

- Typically collocated with ISWC-OM: http://www.ontologymatching.org/
- A annual campaign for the evaluation of matching systems.
- The matching problems are organized in several tracks.
- Each track involves different test ontologies and reference mapping sets.
Outline

Preliminaries

Incoherence and Mapping Repair

Evaluation

Conclusions and Future Work
Incoherent mappings

A set of Mappings $\mathcal{M}$ wrt $\mathcal{O}_1$ and $\mathcal{O}_2$ is incoherent if

- there exists a class $A$ such that
- $\mathcal{O}_1 \cup \mathcal{O}_2 \cup \mathcal{M} \models A \sqsubseteq \bot$, and
- $\mathcal{O}_1 \cup \mathcal{O}_2 \not\models A \sqsubseteq \bot$. 

Example: Incoherent Mappings
Mapping coherence in the OAEI

- Results from 2012 campaign (top systems largebio track)
- Unsat. provided by HermiT and ELK (in SNOMED-NCI)
- Systems Typically rely on lexical and structural heuristics.
  - Tools support required to solve incoherences

<table>
<thead>
<tr>
<th>System</th>
<th>FMA-NCI</th>
<th>FMA-SNOMED</th>
<th>SNOMED-NCI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unsat.</td>
<td>Ratio</td>
<td>Unsat.</td>
</tr>
<tr>
<td>LogMap</td>
<td>9</td>
<td>0.01%</td>
<td>10</td>
</tr>
<tr>
<td>ServOMap</td>
<td>48,743</td>
<td>27%</td>
<td>273,242</td>
</tr>
<tr>
<td>ServOMapL</td>
<td>50,334</td>
<td>28%</td>
<td>99,726</td>
</tr>
<tr>
<td>GOMMA</td>
<td>5,574</td>
<td>4%</td>
<td>10,752</td>
</tr>
<tr>
<td>GOMMA_{bk}</td>
<td>12,939</td>
<td>9%</td>
<td>119,657</td>
</tr>
<tr>
<td>YAM++</td>
<td>50,550</td>
<td>29%</td>
<td>106,107</td>
</tr>
</tbody>
</table>
Mapping Repair

$R$ is a repair for $M$ wrt $O_1$ and $O_2$

- if $M \setminus R$ is coherent.
- $R = M$ is a candidate repair.

$R$ is a diagnosis for $M$ wrt $O_1$ and $O_2$

- if $R' \subset R$ is not a repair for $M$.
- Diagnosis are (typically) expensive to obtain.
Mapping Repair

\( R \) is a **repair** for \( M \) wrt \( O_1 \) and \( O_2 \)

- if \( M \setminus R \) is coherent.
- \( R = M \) is a candidate repair.

\( R \) is a **diagnosis** for \( M \) wrt \( O_1 \) and \( O_2 \)

- if \( R' \subset R \) is not a repair for \( M \).
- Diagnosis are (typically) expensive to obtain.
Computing Mapping Repairs

Approximate Repairs $\mathcal{R} \approx$

- will not guarantee that $\mathcal{M} \setminus \mathcal{R} \approx$ is coherent;
- but aims at reducing the number of unsatisfiabilities,
- while preserving $\mathcal{M}$ as much as possible.

Complete Repairs $\mathcal{R}$

- require complete reasoning;
- rely on standard justification-based ontology debugging;
- they guarantee that $\mathcal{M} \setminus \mathcal{R}$ is coherent.
Computing Mapping Repairs

Approximate Repairs $R^\approx$

- will not guarantee that $\mathcal{M} \setminus R^\approx$ is coherent;
- but aims at reducing the number of unsatisfiabilities,
- while preserving $\mathcal{M}$ as much as possible.

Complete Repairs $R$

- require complete reasoning;
- rely on standard justification-based ontology debugging;
- they guarantee that $\mathcal{M} \setminus R$ is coherent
Outline

Preliminaries

Incoherence and Mapping Repair

Evaluation

Conclusions and Future Work
Datasets (I)

OAEI matching problems

- largebio, library and conference tracks.
- small fragments of FMA, NCI and SNOMED CT for largebio
- mappings from systems participating in 2013 and 2014
Datasets (II): ontologies

<table>
<thead>
<tr>
<th>Ontology</th>
<th>Track</th>
<th>#Concepts</th>
<th>#DataP</th>
<th>#ObjectP</th>
<th>DL</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMT</td>
<td>Conference</td>
<td>36</td>
<td>10</td>
<td>49</td>
<td>$ALCIN(D)$</td>
</tr>
<tr>
<td>CONFERENCE</td>
<td>Conference</td>
<td>60</td>
<td>18</td>
<td>46</td>
<td>$ALCHIF(D)$</td>
</tr>
<tr>
<td>CONFOF</td>
<td>Conference</td>
<td>38</td>
<td>23</td>
<td>13</td>
<td>$SIN(D)$</td>
</tr>
<tr>
<td>EKAW</td>
<td>Conference</td>
<td>74</td>
<td>0</td>
<td>33</td>
<td>$SHIN$</td>
</tr>
<tr>
<td>IASTED</td>
<td>Conference</td>
<td>140</td>
<td>3</td>
<td>38</td>
<td>$ALCIN(D)$</td>
</tr>
<tr>
<td>SIGKDD</td>
<td>Conference</td>
<td>49</td>
<td>11</td>
<td>17</td>
<td>$ALEI(D)$</td>
</tr>
<tr>
<td>FMA$_{NCI}$</td>
<td>Largebio</td>
<td>3,696</td>
<td>24</td>
<td>0</td>
<td>$ALCN(D)$</td>
</tr>
<tr>
<td>FMA$_{SNOMED}$</td>
<td>Largebio</td>
<td>10,157</td>
<td>24</td>
<td>0</td>
<td>$ALCN(D)$</td>
</tr>
<tr>
<td>NCI$_{FMA}$</td>
<td>Largebio</td>
<td>6,488</td>
<td>0</td>
<td>63</td>
<td>$ALC$</td>
</tr>
<tr>
<td>NCI$_{SNOMED}$</td>
<td>Largebio</td>
<td>23,958</td>
<td>0</td>
<td>82</td>
<td>$ALCH$</td>
</tr>
<tr>
<td>SNOMED$_{FMA}$</td>
<td>Largebio</td>
<td>13,412</td>
<td>0</td>
<td>18</td>
<td>$ALER$</td>
</tr>
<tr>
<td>SNOMED$_{NCI}$</td>
<td>Largebio</td>
<td>51,128</td>
<td>0</td>
<td>51</td>
<td>$ALER$</td>
</tr>
<tr>
<td>STW</td>
<td>Library</td>
<td>6,575</td>
<td>0</td>
<td>0</td>
<td>$AL$</td>
</tr>
<tr>
<td>TheSoz</td>
<td>Library</td>
<td>8,376</td>
<td>0</td>
<td>0</td>
<td>$AL$</td>
</tr>
</tbody>
</table>
Datasets (III): mapping sets

<table>
<thead>
<tr>
<th>Ontology 1</th>
<th>Ontology 2</th>
<th># Mappings</th>
<th>Matching System</th>
</tr>
</thead>
<tbody>
<tr>
<td>FMA</td>
<td>NCI</td>
<td>5960</td>
<td>MaasMatch\textsubscript{14}</td>
</tr>
<tr>
<td>SNOMED</td>
<td>NCI</td>
<td>2,500</td>
<td>IAMA\textsubscript{13}</td>
</tr>
<tr>
<td>SNOMED</td>
<td>NCI</td>
<td>3,040</td>
<td>OMReasoner\textsubscript{14}</td>
</tr>
<tr>
<td>SNOMED</td>
<td>NCI</td>
<td>13,270</td>
<td>YAM++\textsubscript{13}</td>
</tr>
<tr>
<td>SNOMED</td>
<td>NCI</td>
<td>13,582</td>
<td>AML\textsubscript{14}</td>
</tr>
<tr>
<td>FMA</td>
<td>SNOMED</td>
<td>21,110</td>
<td>GOMMA\textsubscript{13}</td>
</tr>
<tr>
<td>FMA</td>
<td>SNOMED</td>
<td>16,812</td>
<td>IAMA\textsubscript{13}</td>
</tr>
<tr>
<td>FMA</td>
<td>SNOMED</td>
<td>28,262</td>
<td>AML\textsubscript{14}</td>
</tr>
<tr>
<td>FMA</td>
<td>SNOMED</td>
<td>28,711</td>
<td>LogMapBio\textsubscript{14}</td>
</tr>
<tr>
<td>FMA</td>
<td>SNOMED</td>
<td>23,344</td>
<td>YAM++\textsubscript{13}</td>
</tr>
<tr>
<td>IASTED</td>
<td>SIGKDD</td>
<td>70</td>
<td>AOTL\textsubscript{14}</td>
</tr>
<tr>
<td>CONFOF</td>
<td>IASTED</td>
<td>10</td>
<td>AML\textsubscript{14}</td>
</tr>
<tr>
<td>CONFERENCE</td>
<td>EKAW</td>
<td>164</td>
<td>MaasMatch\textsubscript{14}</td>
</tr>
<tr>
<td>CMT</td>
<td>IASTED</td>
<td>32</td>
<td>MaasMatch\textsubscript{14}</td>
</tr>
<tr>
<td>CONFERENCE</td>
<td>IASTED</td>
<td>68</td>
<td>MaasMatch\textsubscript{14}</td>
</tr>
<tr>
<td>STW</td>
<td>TheSoz</td>
<td>7,254</td>
<td>AML\textsubscript{14}</td>
</tr>
<tr>
<td>STW</td>
<td>TheSoz</td>
<td>12,032</td>
<td>Hertuda\textsubscript{13}</td>
</tr>
<tr>
<td>STW</td>
<td>TheSoz</td>
<td>378</td>
<td>IAMA\textsubscript{13}</td>
</tr>
<tr>
<td>STW</td>
<td>TheSoz</td>
<td>5,684</td>
<td>LogMap\textsubscript{13}</td>
</tr>
<tr>
<td>STW</td>
<td>TheSoz</td>
<td>342</td>
<td>RSDLWB\textsubscript{14}</td>
</tr>
<tr>
<td>STW</td>
<td>TheSoz</td>
<td>80,686</td>
<td>XMapGen\textsubscript{13}</td>
</tr>
</tbody>
</table>
Evaluation settings

System Details

- Desktop computer
- Operating system: Ubuntu 12.04, 64-bit version
- Memory: 32GB DDR3 RAM at 1333MHz
- Java Runtime Environment (JRE): 1.8.0 45-b14
Tested reasoners

Reasones

- Konclude 0.6.0-408 64-bit
- ELK 0.4.1
- Pellet 2.3.1
- HermiT 1.3.8

Settings

- ELK, Pellet and HermiT used via the OWL-API
- Konclude is used via command line (invocation through OWLlink 1.2.1 is raising an error when parsing most of the ontologies in dataset)
- Note that, ELK is a reasoner for the OWL 2 EL profile.
Tested reasoners

Reasoners

- Konclude 0.6.0-408 64-bit
- ELK 0.4.1
- Pellet 2.3.1
- HermiT 1.3.8

Settings

- ELK, Pellet and HermiT used via the OWL-API
- Konclude is used via command line (invocation through OWLlink 1.2.1 is raising an error when parsing most of the ontologies in dataset)
- Note that, ELK is a reasoner for the OWL 2 EL profile.
**Evaluation tasks**

**Classification** (timeout of 60, 20 and 10 minutes for largebio, library and conference, respectively)

- Konclude
- ELK
- Pellet
- HermiT

**Black-box justification extraction** (timeout of 60 seconds per justification)

- ELK
- Pellet
- HermiT
Classification results: conference

<table>
<thead>
<tr>
<th>Reasoner-Dataset</th>
<th>CMT-IASTED</th>
<th>CONFERENCE-IASTED</th>
<th>CONFOF-IASTED</th>
<th>IASTED-SIGKDD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MaasMatch\textsubscript{14}</td>
<td>MaasMatch\textsubscript{14}</td>
<td>AML\textsubscript{14}</td>
<td>AOTL\textsubscript{14}</td>
</tr>
<tr>
<td>ELK</td>
<td>0.01</td>
<td>0.01</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>HERMIT</td>
<td>0.22</td>
<td>T/OUT</td>
<td>0.28</td>
<td>24</td>
</tr>
<tr>
<td>KONCLUDE</td>
<td>0.09</td>
<td>0.36</td>
<td>0.12</td>
<td>0.25</td>
</tr>
<tr>
<td>PELLET</td>
<td>T/OUT</td>
<td>10</td>
<td>4.76</td>
<td>23</td>
</tr>
</tbody>
</table>
Classification results: library

<table>
<thead>
<tr>
<th>Reasoner-Dataset</th>
<th>AML$^{14}$</th>
<th>Hertuda$^{13}$</th>
<th>IAMA$^{13}$</th>
<th>LogMp$^{13}$</th>
<th>RSDLWB$^{14}$</th>
<th>XMapGen$^{13}$</th>
<th>XMapSig$^{13}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELK</td>
<td>0.73</td>
<td>45</td>
<td>0.24</td>
<td>0.25</td>
<td>0.13</td>
<td>T/OUT</td>
<td>0.25</td>
</tr>
<tr>
<td>HERMIT</td>
<td>4.82</td>
<td>842</td>
<td>1.08</td>
<td>2.23</td>
<td>1.14</td>
<td>T/OUT</td>
<td>1.7</td>
</tr>
<tr>
<td>KONCLUDE</td>
<td>2.28</td>
<td>17</td>
<td>1.13</td>
<td>1.72</td>
<td>1.2</td>
<td>59</td>
<td>1.77</td>
</tr>
<tr>
<td>PELLET</td>
<td>8.7</td>
<td>T/OUT</td>
<td>0.21</td>
<td>1.42</td>
<td>0.45</td>
<td>T/OUT</td>
<td>0.92</td>
</tr>
</tbody>
</table>
# Classification results: largebio

<table>
<thead>
<tr>
<th>Reasoner-Dataset</th>
<th>FMA-NCI</th>
<th>FMA-SNOMED</th>
<th>SNOMED-NCI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MaasMtch\textsubscript{14}</td>
<td>LgMpBio\textsubscript{14}</td>
<td>YAM\textsubscript{13}</td>
</tr>
<tr>
<td>ELK</td>
<td>0.21</td>
<td>0.08</td>
<td>0.6</td>
</tr>
<tr>
<td>HERMIT</td>
<td>3.32</td>
<td>20.19</td>
<td>5.08</td>
</tr>
<tr>
<td>KONCLUDE</td>
<td>1.3</td>
<td>8.25</td>
<td>3.83</td>
</tr>
<tr>
<td>PELLET</td>
<td>T/OUT</td>
<td>30.46</td>
<td>T/OUT</td>
</tr>
</tbody>
</table>
Justification Extraction results (I)

FMA-NCI with MaasMatch Mappings

<table>
<thead>
<tr>
<th>Reasoner</th>
<th>Class.(s)</th>
<th>#Unsat</th>
<th>1Just.(s)</th>
<th>10Just.(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELK</td>
<td>0.21</td>
<td>7,377</td>
<td>15</td>
<td>162</td>
</tr>
<tr>
<td>HERMIT</td>
<td>3.32</td>
<td>8,767</td>
<td>43</td>
<td>1,206</td>
</tr>
<tr>
<td>PELLET</td>
<td>T/OUT</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- In Table, times to compute 1 justification or 10 justifications for 50 unsat concepts.
- Computing 1 justifications for all unsat concepts: >36m for ELK, >2h for HermiT.
- Computing 10 justifications for all unsat concepts: >6h for ELK, >58h for HermiT.
Justification Extraction results (II)

SNOMED-NCI with GOMMA Mappings

<table>
<thead>
<tr>
<th>Reasoner</th>
<th>Class.(s)</th>
<th>#Unsat</th>
<th>1Just.(s)</th>
<th>10Just.(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELK</td>
<td>2.91</td>
<td>50,189</td>
<td>45</td>
<td>259</td>
</tr>
<tr>
<td>HERMIT</td>
<td>49</td>
<td>53,448</td>
<td>39</td>
<td>1,350</td>
</tr>
<tr>
<td>PELLET</td>
<td>T/OUT</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- In Table, times to compute 1 justification or 10 justifications for 50 unsat
- Computing 1 justifications for all unsat concepts: >11h for ELK, >12h for HermiT
- Computing 10 justifications for all unsat concepts: >76h for ELK, >16days for HermiT
Outline

Preliminaries

Incoherence and Mapping Repair

Evaluation

Conclusions and Future Work
Conclusions

• Reasoning with two aligned ontologies still poses serious problems to OWL 2 reasoners
• Reasoning with a network of ontologies like BioPortal would even be more challenging
• Integration of ontologies via mappings seems ideal as reasoning benchmarks.

Future work

• Extend the number of justifications
• Increase time for time-outs
• Explore the use of glass-box techniques
• Integrate in a mapping repair system (with optimizations)
Conclusions

Reasoning with two aligned ontologies still poses serious problems to OWL 2 reasoners.

Reasoning with a network of ontologies like BioPortal would even be more challenging.

Integration of ontologies via mappings seems ideal as reasoning benchmarks.

Future work

- Extend the number of justifications
- Increase time for time-outs
- Explore the use of glass-box techniques
- Integrate in a mapping repair system (with optimizations)
Questions?

Thank you for your attention