Evaluation of Automatic Ontology Matching for Materials Sciences and Engineering

Supervisor: M. Sc. Martin Huschka
First Examiner: Prof. Dr. Hannah Bast
Second Examiner: Prof. Dr.-Ing. habil. Stefan Hiermaier

Engy Nasr
Department of Computer Science
Chair for Algorithms and Data Structure
Data .. Data .. Everywhere ..

Data Volume in Zettabytes

- Years [1]
  - 2010: 0
  - 2011: 5
  - 2012: 10
  - 2013: 15
  - 2014: 20
  - 2015: 25
  - 2016: 30
  - 2017: 35
  - 2018: 40
  - 2019: 45
  - 2020: 50
  - 2021: 55
  - 2022: 60
  - 2023: 65
  - 2024: 70
  - 2025: 75

- Structured Data
- Unstructured Data

- 2010: 10
- 2011: 15
- 2012: 20
- 2013: 25
- 2014: 30
- 2015: 35
- 2016: 40
- 2017: 45
- 2018: 50
- 2019: 55
- 2020: 60
- 2021: 65
- 2022: 70
- 2023: 75
- 2024: 80
- 2025: 85

- 2010: 15
- 2011: 20
- 2012: 25
- 2013: 30
- 2014: 35
- 2015: 40
- 2016: 45
- 2017: 50
- 2018: 55
- 2019: 60
- 2020: 65
- 2021: 70
- 2022: 75
- 2023: 80
- 2024: 85
- 2025: 90

- 2010: 20
- 2011: 25
- 2012: 30
- 2013: 35
- 2014: 40
- 2015: 45
- 2016: 50
- 2017: 55
- 2018: 60
- 2019: 65
- 2020: 70
- 2021: 75
- 2022: 80
- 2023: 85
- 2024: 90
- 2025: 95

12.03.2020
Evaluation of Automatic Ontology Matching for Materials Sciences and Engineering
**Materials Modelling:** is the mathematical representation of a material’s behavior under certain applied actions and conditions.

PE: Physics/chemistry Equation  
MR: Material Relation  
RoMM: Review of Materials Modelling
Interoperability

➢ **Interoperability**: is when systems, actors or applications provide/accept services to/from other systems, actors or applications and collaboratively use the exchanged services.
Content

Introduction

Theoretical Background

Problem Statement & Approach

Methodology

Results & Discussion

Conclusion & Future Work
**Balaenoptera musculus** $\subseteq$ **Animal** $\cap$ **maximum length** $\leq$ 33.6 m
„People can’t share knowledge if they don’t speak a common language“

*Thomas Davenport (1997)*
Knowledge Representation Systems

To speak a common language:

Weak Semantics

Strong Semantics

List

Weak Semantics
To speak a common language:

**Weak Semantics**

- Informal Hierarchy (Weak Taxonomy)
  - List

**Strong Semantics**

- Domain Eukarya
- Kingdom Animalia
- Phylum Chordata
- Class Mammalia
- Order Carnivora
- Family Canidae
- Genus *Vulpes*
- Species *Vulpes vulpes*
Knowledge Representation Systems

To speak a common language:

- Informal Hierarchy (Weak Taxonomy)
- List
- Thesaurus
- Taxonomy
- Strong Semantics

Diagram showing a hierarchy from weak to strong semantics.
Knowledge Representation Systems

To speak a common language:
Ontology Small Example [2]

Animal \subseteq Living\,Organism

Plant \subseteq Living\,Organism

Balaenoptera musculus \subseteq Animal \land \text{maximum length} \leq 33.6\,m
Ontologies can be Huge
Upper Level Ontologies

- Most General Things
  - Process
  - Location
- Airspace
- Target Area of Interest

Upper Ontology

Domain Ontology

BFO

EMMO
 Ontology Matching

- **Ontology Matching**: is where actual relations & alignments between entities of different ontologies are found.

\[ O_1: \text{1st Ontology (source)} \quad O_2: \text{2nd Ontology (target)} \quad A: \text{output alignment} \]
Ontology Matching

1. Living Organism

2. Human

3. Animal

1. Living Thing

2. Plant

OAEI: Ontology Alignment Evaluation Initiative

O1: 1st Ontology (source)  O2: 2nd Ontology (target)
Theoretical Background

- Semantics
- Computers
- Logic
- Ontologies
- Ontology Matching
Content

Introduction

Theoretical Background

Problem Statement & Approach

Methodology

Results & Discussion

Conclusion & Future Work
Problem Statement

Automatic ontology matching tools are yearly evaluated, updated and adapted to the fields of biomedicine, bioinformatics and medicine

BUT NOT

to the Materials Sciences field
Approach

- **Adapt**: Existing evaluation workflow for automatic ontology matching
- **Identify**: Existing automatic ontology matchers
- **Develop**: Materials Sciences Benchmark
- **Evaluate**: Performance of selected matchers using the adapted evaluation workflow along with the developed benchmark
Identification of Matchers

- Ontology Alignment Evaluation Initiative (OAEI)
- Active
- Openly Available on the Web

AML
LogMap
Participating Ontology Matchers
Agreement Maker Light

- AML

AML's Ontology Matchers

Primary
- Lexical
- Word
- Property
- Background Knowledge
- Structure

Secondary
- Obsolete
- Cardinality
- Coherence
- String
LogMap

- It has a high repair algorithm

- A highly scalable ontology matching system, which can handle large sized ontologies.
Materials Sciences Benchmark Test Cases
| No of Alignments $|R|$ | 1st Test Case | 2nd Test Case | 3rd Test Case |
|-----------------|-----------------|-----------------|-----------------|
| $|R|$ | 23 | 302 | 11 |

For 1st Test Case: subclass ($\subseteq$), superclass ($\supseteq$) and equivalence ($=$), no property and no instances.

For 2nd Test Case: equivalence ($=$), no property and no instances.

For 3rd Test Case: equivalence ($=$), no property and no instances.
Periodic Table

European Materials Modelling Ontology (EMMO)
Evaluation of Automatic Ontology Matching for Materials Sciences and Engineering

**Evaluation Workflow**

**O1**: 1st Ontology (source)  
**O2**: 2nd Ontology (target)  
**A**: output alignment  
**R**: reference alignment  
**m**: performance measurements

- **parameter**
- **resources**
Recall = \frac{|R \cap A|}{|R|}

Precision = \frac{|R \cap A|}{|A|}

F-measure = \frac{\text{Precision} \times \text{Recall}}{(1 - \alpha) \times \text{Precision} + \alpha \times \text{Recall}}
## Experiments

Experiments from 1 to 5 designed using AML matchers:

<table>
<thead>
<tr>
<th>Experiment No.</th>
<th>Experiment ID</th>
<th>AML Property Matcher</th>
<th>AML Background Knowledge Matcher</th>
<th>Background Knowledge Ontology Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AML - L</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>AML - LP</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>AML - LB&lt;sub&gt;pt&lt;/sub&gt;</td>
<td></td>
<td>✓</td>
<td>Periodic Table</td>
</tr>
<tr>
<td>4</td>
<td>AML - LPB&lt;sub&gt;pt&lt;/sub&gt;</td>
<td>✓</td>
<td>✓</td>
<td>Periodic Table</td>
</tr>
<tr>
<td>5</td>
<td>AML - LPB&lt;sub&gt;EMMO&lt;/sub&gt;</td>
<td>✓</td>
<td>✓</td>
<td>EMMO</td>
</tr>
</tbody>
</table>
The sixth experiment designed using LogMap:

<table>
<thead>
<tr>
<th>Experiment No.</th>
<th>Experiment ID</th>
<th>LogMap Ontology Matcher</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>LogMap - L</td>
<td>✅</td>
</tr>
</tbody>
</table>
Evaluation of Automatic Ontology Matching for Materials Sciences and Engineering
Total No. of Alignments per Experiment

<table>
<thead>
<tr>
<th>EXPERIMENT NO.</th>
<th>1st Test Case</th>
<th>2nd Test Case</th>
<th>3rd Test Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp 1</td>
<td>AML - L</td>
<td>AML - LP</td>
<td>AML - LB_{pt}</td>
</tr>
<tr>
<td>Exp 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp 6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NO. OF ALIGNMENTS

0 50 100 150 200 250 300
1st Test Case Results

<table>
<thead>
<tr>
<th>Experiment No.</th>
<th>Exp 1</th>
<th>Exp 2</th>
<th>Exp 3</th>
<th>Exp 4</th>
<th>Exp 5</th>
<th>Exp 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AML - L</td>
<td>AML - LP</td>
<td>AML - LB_{pt}</td>
<td>AML - LPB_{pt}</td>
<td>AML - LPB_{EMMO}</td>
<td>LogMap - L</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PERCENTAGE OF PERFORMANCE EVALUATION</th>
<th>100.0 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPERIMENT NO.</td>
<td></td>
</tr>
<tr>
<td>Precision</td>
<td></td>
</tr>
<tr>
<td>Recall</td>
<td></td>
</tr>
<tr>
<td>F-measure</td>
<td></td>
</tr>
</tbody>
</table>
2\textsuperscript{nd} Test Case Results

<table>
<thead>
<tr>
<th>EXPERIMENT NO.</th>
<th>EXP 1</th>
<th>EXP 2</th>
<th>EXP 3</th>
<th>EXP 4</th>
<th>EXP 5</th>
<th>EXP 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>AML - L</td>
<td>AML - LP</td>
<td>AML - L\textsubscript{B}\textsubscript{pt}</td>
<td>AML - L\textsubscript{P}\textsubscript{B}\textsubscript{pt}</td>
<td>AML - L\textsubscript{P}\textsubscript{B}\textsubscript{EMMO}</td>
<td>LogMap - L</td>
<td></td>
</tr>
</tbody>
</table>
3rd Test Case Results

PERCENTAGE OF PERFORMANCE EVALUATION

<table>
<thead>
<tr>
<th>EXPERIMENT NO.</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>88.9 %</td>
<td>80.0 %</td>
<td>88.9 %</td>
</tr>
<tr>
<td>2</td>
<td>72.7 %</td>
<td>72.7 %</td>
<td>76.2 %</td>
</tr>
<tr>
<td>3</td>
<td>80.0 %</td>
<td>72.7 %</td>
<td>72.7 %</td>
</tr>
<tr>
<td>4</td>
<td>80.0 %</td>
<td>72.7 %</td>
<td>72.7 %</td>
</tr>
<tr>
<td>5</td>
<td>80.0 %</td>
<td>76.2 %</td>
<td>76.2 %</td>
</tr>
<tr>
<td>6</td>
<td>100.0 %</td>
<td>77.8 %</td>
<td>63.6 %</td>
</tr>
</tbody>
</table>

Exp 1 | Exp 2 | Exp 3 | Exp 4 | Exp 5 | Exp 6  
AML - L | AML - LP | AML - LB_{pt} | AML - LPB_{pt} | AML - LPB_{EMMO} | LogMap - L
Content

- Introduction
- Theoretical Background
- Problem Statement & Approach
- Methodology
- Results & Discussion
- Conclusion & Future Work
Conclusion

- The more **number of matcher** used the more number of alignments found.
- The **LogMap** achieves high precision value.
- Best evaluation results are observed using the right **Background Knowledge ontology**.
- **Property matcher** lowers the precision value.
- No fully automatic ontology matching presents.
Future Work

- Improvement of Materials Sciences test cases based on more developed Materials Sciences ontologies.

- Other Materials Sciences background knowledge ontologies have to be created.

- The examined ontology matching tools have to be improved in terms of the property matcher.

- Other logically related alignments from the ontology matchers are essential, and accordingly improved benchmark will include them too.
References


Thank you