SemaDrift: A Protégé Plugin for Measuring Semantic Drift in Ontologies

T. G. Stavropoulos*, S. Andreadis, E. Kontopoulos, M. Riga, P. Mitzias, I. Kompatsiaris

“This project has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no601138.”
Problem & Aim

- **Background**
  - Semantic Web (also known as Web 3.0)
  - Knowledge representation via ontologies
  - Semantic change can have drastic consequences

- **Problem**
  - To detect & measure semantic change in ontologies across time and versions
  - Lack of practical methods & tools directly applicable to Semantic Web

- **Aim**
  - To develop a framework for measuring semantic drift in ontologies across multiple versions
Semantic Change – Intro

- Monitors & measures **changes in the meaning** of concepts along with their potential **replacement** by other meanings over time
- **Drastic consequences** on the use of knowledge representation models in applications
- Relates to various lines of research
  - Ontology change, evolution, management & versioning …
- **Diverse terminology**
  - Semantic drift/shift/decay, concept drift/shift …
Semantic Change in Literature

Semantic Decay
- Pareti et al., 2015
- Gulla et al., 2010

Semantic Drift
- Topic Shift
- Wang et al., 2011
- Stojačević et al., 2002
- Fanizzi et al., 2008
- Meron-point-Peres et al., 2013

Concept Drift
- Concept Shift
- Wang et al., 2009

Semantic Change
- Klarman et al., 2008
- Tury & Bielikova, 2006
- Wittek et al., 2015
- Uschold, 2000
- Klein & Pensi, 2001
- Yildiz, 2006

Concept Versioning
- Mark

Topic Drift
Concept Drift

- **Concept drift**: Change in concept’s meaning over time
  - Possibly also over location, culture, etc.

- Notions & metrics for concept drift in data mining transferred to semantic change/drift *

- Aspects
  - Label, Intension, Extension

- Correspondence of a concept across versions
  - Known (Identity-based approach)
  - Unknown (Morphing-based approach)

Semantic Drift Metrics

- **Label**
  - Description, name or title
  - `rdfs:label`
  - Drift: String similarity (Monge–Elkan)

- **Intension**
  - Characteristics
  - Set of OWL datatype or object property triples
  - Drift: Jaccard similarity

- **Extension**
  - Things a concept extends to
  - Set of instances
  - Drift: Jaccard similarity

- **Whole**
SemaDrift Library (API)

- Java, OWL-API, Simmetrics
- [http://mklab.iti.gr/results/tools](http://mklab.iti.gr/results/tools) –> SemaDrift
  - Apache V2 License
- Load an array of ontologies (URL/files) i.e. its multiple versions
- Get drift metrics:
  - Average concept stability for each metric (all ontologies)
  - Concept-per-concept stability (a pair of ontologies)
  - Morphing chains, ranking (all ontologies)
- Utilities for clients
  - e.g. get the ontology tree structure
  - avoid re-parsing ontologies e.g. at the front-end

A GUI for the SemaDrift Library (API)
Java, Apache V2 License
  - [http://mklab.iti.gr/results/tools](http://mklab.iti.gr/results/tools) → SemaDrift

Pros:
  - Popular Protégé Ontology Editor
  - Exploit the Protégé Environment

Cons:
  - Non-flexible development
    - Mandates use of Java Swing for GUI (outdated)
    - Different versions of Protégé use different versions of OWL-API
Currently an ontology pair
1: Protégé working ontology (available to view & edit)
2: Second external ontology
Measure Drift

- Average Concept Stability
  - Across all concepts

- Concept-per-Concept Stability
  - Concept pairs across the ontology pair
Use Case 1: Digital Preservation

- Synthesized ontologies for 2003–2013 (one per year) for software-based artworks of Tate London
- Pericles inspired (Tate partner), exploring similarities between CB, MM and SB

- Lowest average: Extensional

<table>
<thead>
<tr>
<th>Label</th>
<th>Intensional</th>
<th>Extensional</th>
<th>Whole</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB</td>
<td>0.667</td>
<td>0.163</td>
<td>0.537</td>
</tr>
<tr>
<td>MM</td>
<td>0.867</td>
<td>0.183</td>
<td>0.537</td>
</tr>
<tr>
<td>SB</td>
<td>0.833</td>
<td>0.638</td>
<td>0.537</td>
</tr>
</tbody>
</table>

- Similarities but no migrations
- Instances migrated
Plotting drift across versions

Stability per Aspect in Morphing approach
Use Case 2: Web Services

- **OWL-S Ontology**
  - A popular standard for Semantic Markup of Web Services

- **OWL-S Profile ontology version 1.0 vs 1.2**
  - [www.w3.org/Submission/OWL-S](http://www.w3.org/Submission/OWL-S)

- **Average Concept Drift points to Intensional**
  - No instances exist (Extension),
  - Slight change in labels

<table>
<thead>
<tr>
<th>Average Concept Stability</th>
<th>Intensional</th>
<th>Extensional</th>
<th>Whole</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>0.807</td>
<td>0.139</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Intension Concept-to-concept Drift

Were removed, not only by name but also similarity

Precondition became Condition

Parameter, Process and Profile have properties but remained the same

Similar due to no properties
Future Work

- Multiple ontology versions
  - Supported in the API but not visually

- Graphs
  - ... from tables

- Identity based
  - Using user input

- Hybrid method
  - First find identity, then measure
SemaDrift FX

A Tool to Measure Semantic Concept Drift

Concept-drift-Concept Stability
Label Aspect

<table>
<thead>
<tr>
<th></th>
<th>Dependency</th>
<th>Identity</th>
<th>Size</th>
<th>Computer</th>
<th>Concrete</th>
<th>Software</th>
<th>Artist</th>
<th>Illuminator</th>
<th>Hardware</th>
<th>DigitalRes.</th>
<th>Sta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monitor</td>
<td>0.086</td>
<td>0.124</td>
<td>0.083</td>
<td>0.348</td>
<td>0.143</td>
<td>0.095</td>
<td>0.089</td>
<td>0.443</td>
<td>0.066</td>
<td>0.152</td>
<td>0.12</td>
</tr>
<tr>
<td>HardwareD.</td>
<td>0.188</td>
<td>0.087</td>
<td>0.083</td>
<td>0.079</td>
<td>0.096</td>
<td>0.154</td>
<td>0.094</td>
<td>0.063</td>
<td>0.375</td>
<td>0.081</td>
<td>0.08</td>
</tr>
<tr>
<td>Equipment</td>
<td>0.119</td>
<td>0.125</td>
<td>0.083</td>
<td>0.050</td>
<td>0.076</td>
<td>0.067</td>
<td>0.056</td>
<td>0.096</td>
<td>0.119</td>
<td>0.050</td>
<td>0.09</td>
</tr>
<tr>
<td>Size</td>
<td>0.083</td>
<td>0.131</td>
<td>0.622</td>
<td>0.083</td>
<td>0.133</td>
<td>0.083</td>
<td>0.083</td>
<td>0.083</td>
<td>0.083</td>
<td>0.083</td>
<td>0.08</td>
</tr>
<tr>
<td>OperatingS.</td>
<td>0.120</td>
<td>0.119</td>
<td>0.083</td>
<td>0.268</td>
<td>0.122</td>
<td>0.098</td>
<td>0.128</td>
<td>0.259</td>
<td>0.120</td>
<td>0.087</td>
<td>0.11</td>
</tr>
<tr>
<td>Dependency</td>
<td>0.404</td>
<td>0.133</td>
<td>0.083</td>
<td>0.092</td>
<td>0.108</td>
<td>0.080</td>
<td>0.067</td>
<td>0.060</td>
<td>0.333</td>
<td>0.086</td>
<td>0.08</td>
</tr>
<tr>
<td>Artist</td>
<td>0.067</td>
<td>0.111</td>
<td>0.083</td>
<td>0.089</td>
<td>0.089</td>
<td>0.200</td>
<td>0.566</td>
<td>0.089</td>
<td>0.122</td>
<td>0.089</td>
<td>0.20</td>
</tr>
<tr>
<td>DataDepend.</td>
<td>0.230</td>
<td>0.158</td>
<td>0.067</td>
<td>0.138</td>
<td>0.146</td>
<td>0.090</td>
<td>0.100</td>
<td>0.136</td>
<td>0.233</td>
<td>0.151</td>
<td>0.16</td>
</tr>
<tr>
<td>StatusFlag</td>
<td>0.098</td>
<td>0.097</td>
<td>0.067</td>
<td>0.086</td>
<td>0.111</td>
<td>0.137</td>
<td>0.113</td>
<td>0.208</td>
<td>0.122</td>
<td>0.048</td>
<td>0.56</td>
</tr>
<tr>
<td>Description</td>
<td>0.067</td>
<td>0.139</td>
<td>0.131</td>
<td>0.067</td>
<td>0.133</td>
<td>0.107</td>
<td>0.100</td>
<td>0.047</td>
<td>0.070</td>
<td>0.000</td>
<td>0.17</td>
</tr>
<tr>
<td>DisplayDevice</td>
<td>0.127</td>
<td>0.083</td>
<td>0.100</td>
<td>0.073</td>
<td>0.094</td>
<td>0.070</td>
<td>0.106</td>
<td>0.071</td>
<td>0.127</td>
<td>0.103</td>
<td>0.11</td>
</tr>
<tr>
<td>Programmi...</td>
<td>0.070</td>
<td>0.067</td>
<td>0.083</td>
<td>0.178</td>
<td>0.067</td>
<td>0.080</td>
<td>0.044</td>
<td>0.197</td>
<td>0.070</td>
<td>0.000</td>
<td>0.17</td>
</tr>
<tr>
<td>ComputerB...</td>
<td>0.067</td>
<td>0.062</td>
<td>0.072</td>
<td>0.172</td>
<td>0.114</td>
<td>0.508</td>
<td>0.149</td>
<td>0.106</td>
<td>0.138</td>
<td>0.119</td>
<td>0.12</td>
</tr>
<tr>
<td>ConcreteRes...</td>
<td>0.075</td>
<td>0.067</td>
<td>0.083</td>
<td>0.108</td>
<td>0.333</td>
<td>0.127</td>
<td>0.089</td>
<td>0.087</td>
<td>0.219</td>
<td>0.021</td>
<td>0.11</td>
</tr>
<tr>
<td>SoftwareDe...</td>
<td>0.188</td>
<td>0.087</td>
<td>0.108</td>
<td>0.079</td>
<td>0.096</td>
<td>0.207</td>
<td>0.078</td>
<td>0.063</td>
<td>0.263</td>
<td>0.081</td>
<td>0.08</td>
</tr>
</tbody>
</table>
Conclusions

- Semantic Drift tools to visually captivate semantic concept change in ontologies across versions
- SemaDrift Protégé Plugin brings and visualizes SemaDrift metrics API to a popular ontology development platform
- Use case scenarios in digital preservation and web service markup show insights previously not so easily accessible
Thank you!

- Contact: athstavr@iti.gr
- mklab.iti.gr