SemaDrift: A Protégé Plugin for Measuring Semantic Drift in Ontologies

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

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 …

Pericles
FP7 Digital Preservation

Semantic Change in Literature
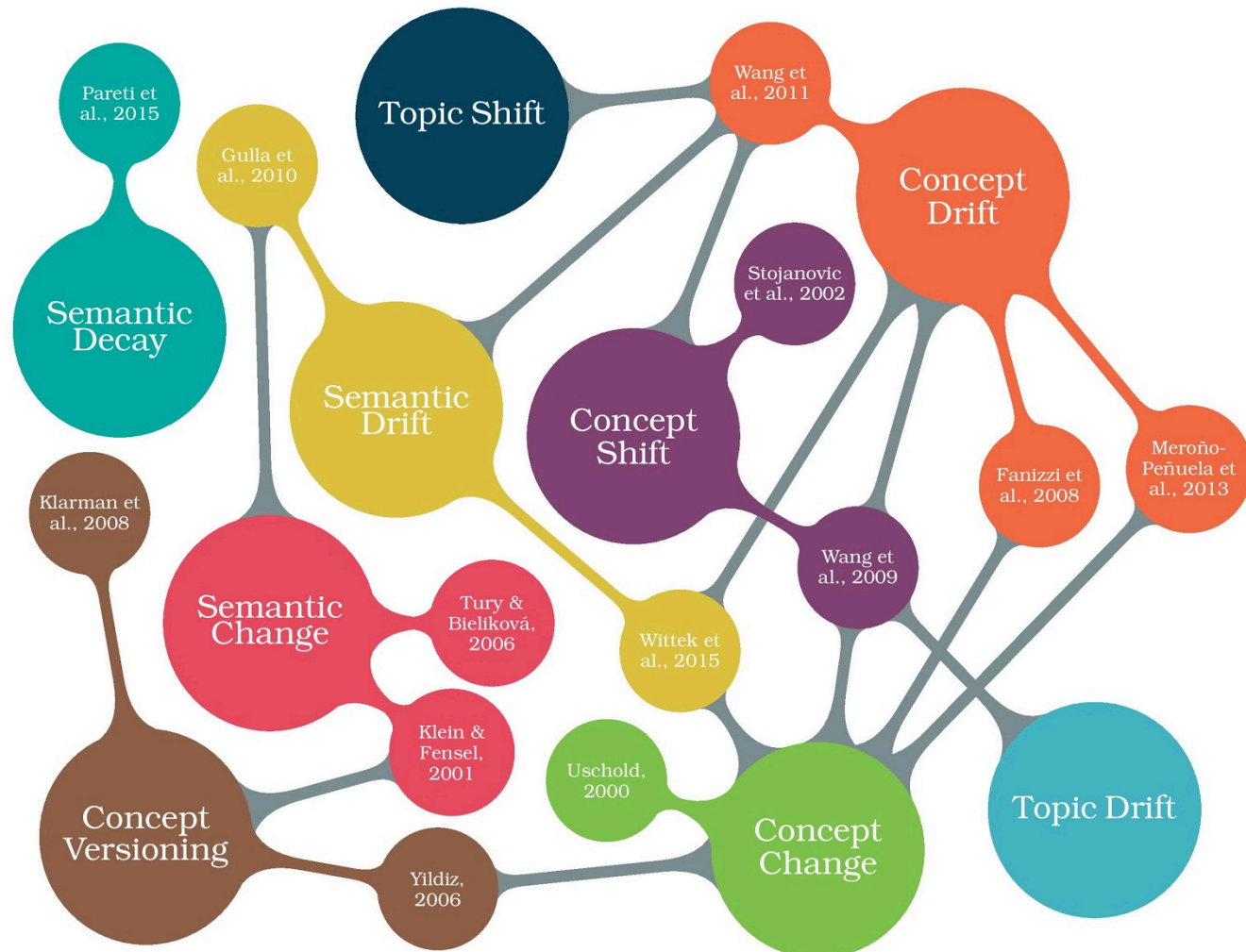
Concept Drift

- **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)

* Wang, S., Schlobach, S., & Klein, M. (2011). **Concept drift and how to identify it**. Web Semantics: Science, Services and Agents on the WWW, 9(3), 247–265.

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 -> 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

Pericles
FP7 Digital Preservation

SemaDrift Protégé Plugin

▸ A GUI for the SemaDrift Library (API)

▸ Java, Apache V2 License
  ◦ 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

Load Ontologies

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

Average Concept Stability

Label	Intensional	Extensional	Whole
0.782	0.667	0.163	0.537

- Lowest average: Extensional

Concept-per-Concept Stability
Rows: tate_2011 Columns: tate_2012

Label

	ComputerBased	MixedMedia	SoftwareBased
ComputerBased	1.000	0.587	0.773
MixedMedia	0.600	1.000	0.600
SoftwareBased	0.838	0.638	1.000

Intension

	ComputerBased	MixedMedia	SoftwareBased
ComputerBased	1.000	0.500	0.500
MixedMedia	0.500	1.000	0.500
SoftwareBased	0.500	0.500	1.000

Similarities but no migrations

Extension

	ComputerBased	MixedMedia	SoftwareBased
ComputerBased	0.000	0.200	0.200
MixedMedia	0.000	0.400	0.167
SoftwareBased	0.000	0.000	0.500

Instances migrated

Whole

	ComputerBased	MixedMedia	SoftwareBased
ComputerBased	0.667	0.429	0.491
MixedMedia	0.367	0.800	0.422
SoftwareBased	0.446	0.379	0.833

Pericles
FP7 Digital Preservation

Plotting drift across versions

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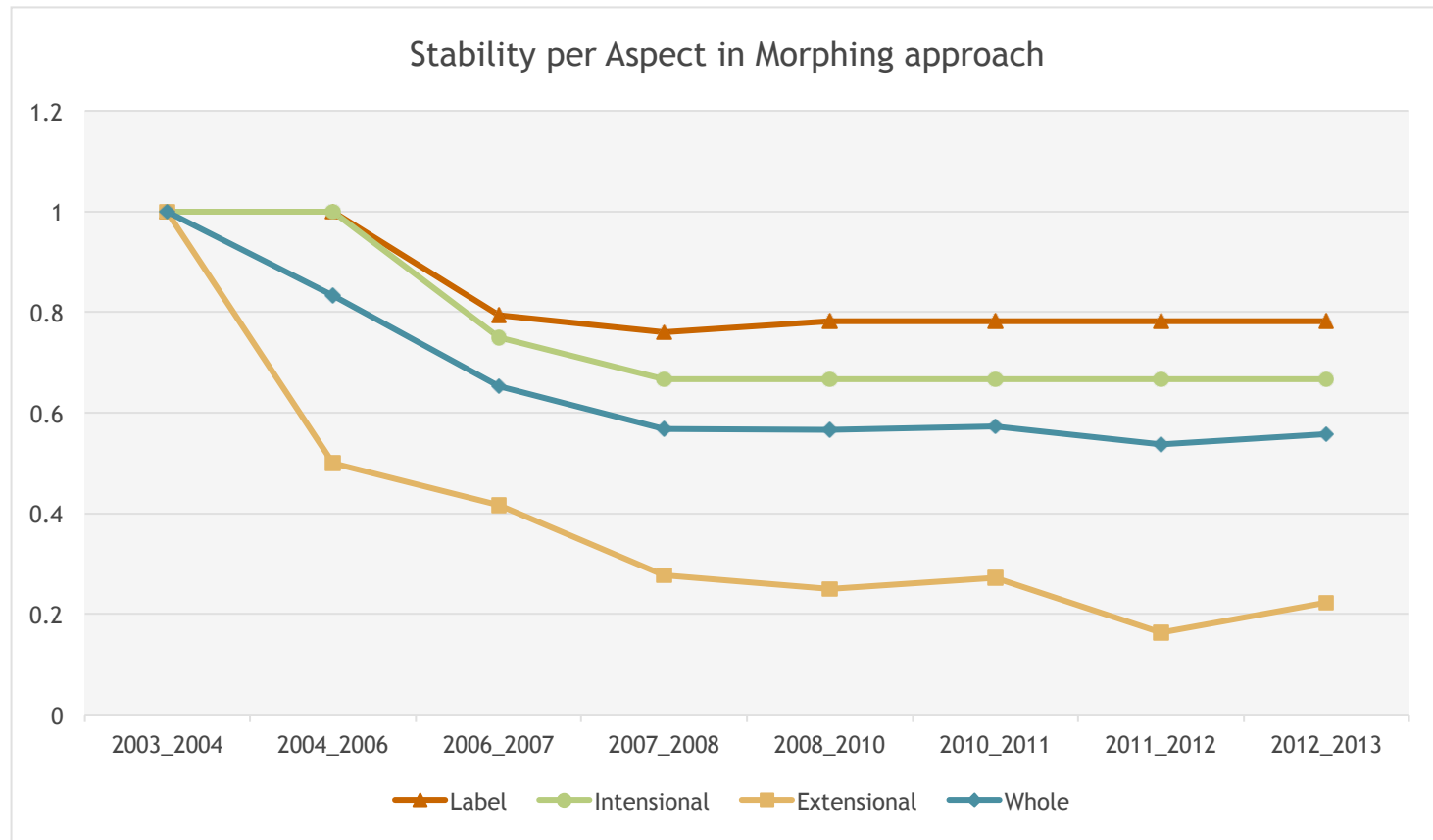
Use Case 2: Web Services

▸ 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

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

Average Concept Stability

Label	Intensional	Extensional	Whole
0.807	0.139	1.000	0.648

Intension
Concept–to–concept Drift

Were removed, not only by name but also similarity

Parameter, Process and Profile have properties but remained the same

Similar due to no properties

Precondition became Condition

Concept-per-Concept Stability
Rows: Profile 1.0 Columns: Profile 1.2
Intension

	Condition	Input	Output	Parameter	Process	Profile	Result	Service profile
ConditionalEffect	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ConditionalOutput	0.000	1.000	1.000	0.000	0.000	0.000	0.000	1.000
Input	0.000	1.000	1.000	0.000	0.000	0.000	0.000	1.000
Parameter	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
Precondition	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Process	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
Profile	0.000	0.000	0.000	0.000	0.000	0.200	0.000	0.000
ServiceCategory	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Future Work

▸ **Multiple ontology versions**
   ◦ Supported in the API but not visually

▸ **Graphs**
   ◦ … from tables

▸ **Identity based**
   ◦ Using user input
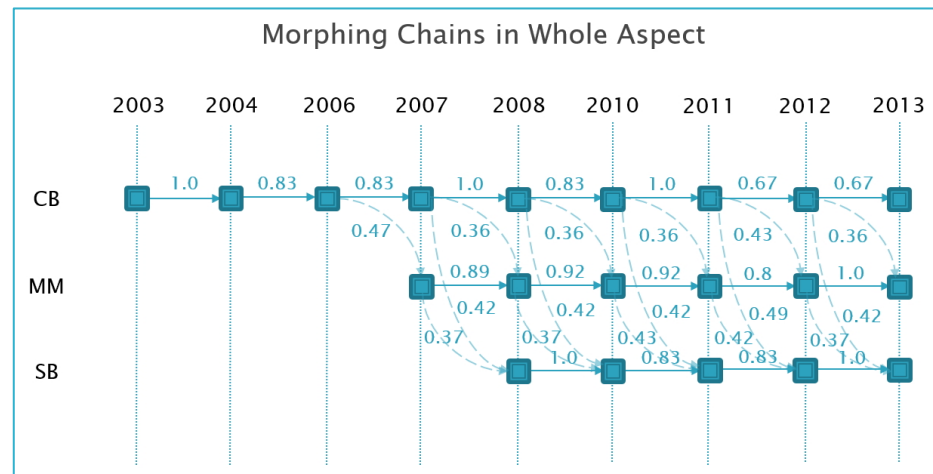
▸ **Hybrid method**
   ◦ First find identity, then measure

Morphing Chains in Whole Aspect

SemaDrift FX

▸ Preview

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