

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

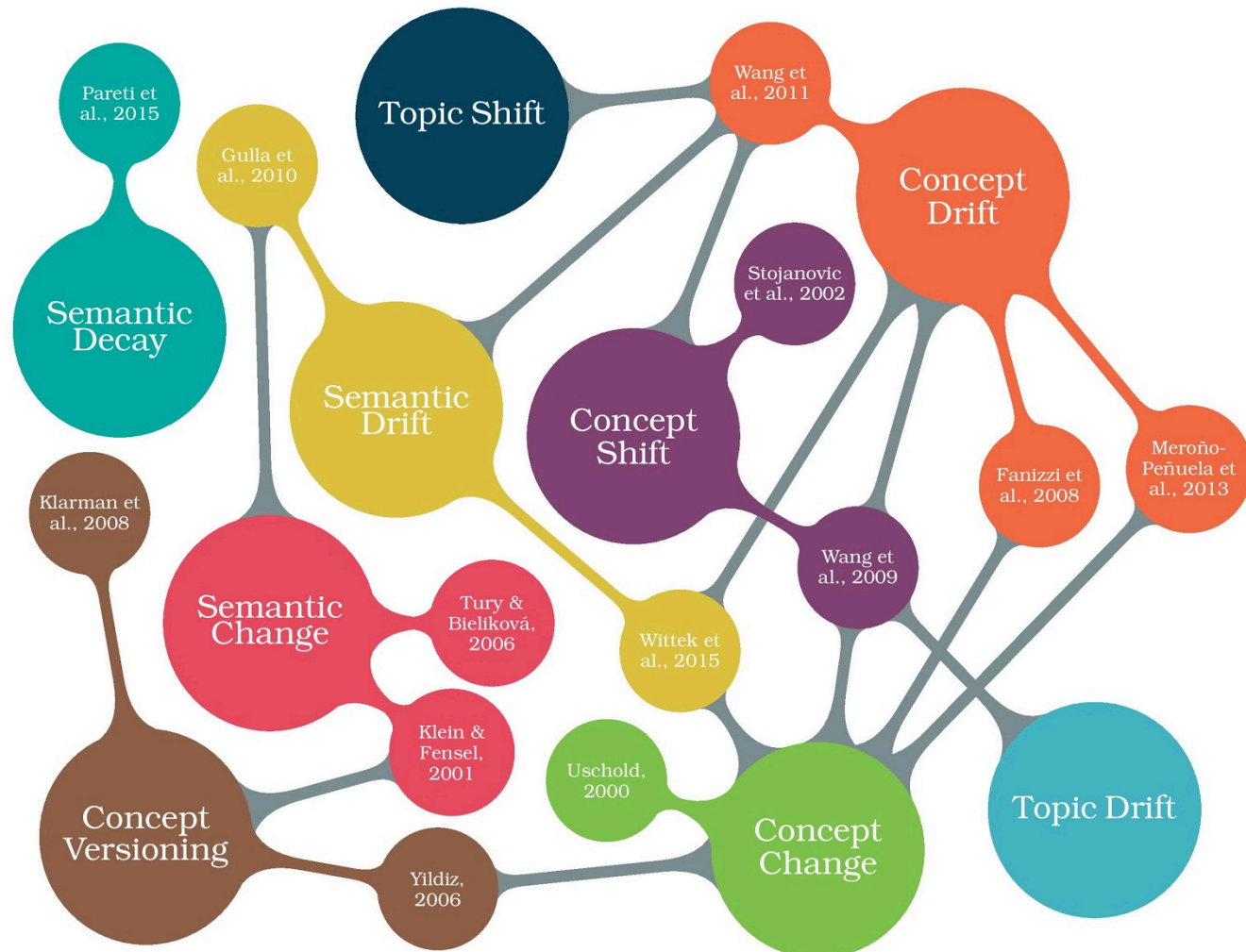
► Aim

- 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 ...

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

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

tate_2011 (http://www.semanticweb.org/ontologies/mklab/2016/4/tate_2011) : [C:\Users\andreadis\Drop...

File Edit View Reasoner Tools Refactor Window Help

tate_2011 (http://www.semanticweb.org/ontologies/mklab/2016/4/tate_2011) Search for entity

Active Ontology Entities Individuals SemaDrift

Class hierarchy:

- Thing
 - Calendar \equiv TimeCalendar
 - Datatype
 - Duration \equiv TimeDuration
 - Instant \equiv TimeInstant
 - Interval
 - Reference \equiv TimeReference
 - Resource
 - AbstractResource
 - Agent
 - AggregatedResource
 - 'Concrete Resource'
 - 'Digital Resource'
 - 'Digital Artwork'
 - 'Computer Based Artwork'
 - 'Mixed Media Artwork'
 - 'Software Based Artwork'
 - SoftwareTool
 - Text
 - 'Versioned Digital Resource'
 - Equipment
 - 'Human Agent'
 - ExogenousResource
 - ImmutableResource
 - Irm:VersionedResource
 - MutableResource
 - TemporalMeasurement
 - TimeCalendar \equiv Calendar
 - TimeInterval \equiv TimeInterval
 - TimeReference \equiv Reference
 - Uncertainty

SemaDrift:

Find the semantic drift between the active ontology and a subsequent version by calculating the stability between their concepts.

Please load an ontology

2

Browse

Measure Drift

1

No Reasoner set. Select a reasoner from the Reasoner menu ☒ Show Inferences

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

tate_2011 (http://www.semanticweb.org/ontologies/mklab/2016/4/tate_2011) : [C:\Users\andreadisst\Drop...

File Edit View Reasoner Tools Refactor Window Help

www.semanticweb.org/ontologies/mklab/2016/4/tate_2011) Search for entity

SemaDrift

SemaDrift: Find the semantic drift between the active ontology and a subsequent version by calculating the stability between their concepts.

...st\Documents\versions\2012\tate_2012.owl Browse

Measure Drift

Average Concept Stability

Label	Intensional	Extensional	Whole	
	0.782	0.667	0.163	0.537

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

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

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

No Reasoner set. Select a reasoner from the Reasoner menu ☒ Show Inferences

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

Average Concept Stability

Label	Intensional	Extensional	Whole
	0.782	0.667	0.537

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

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

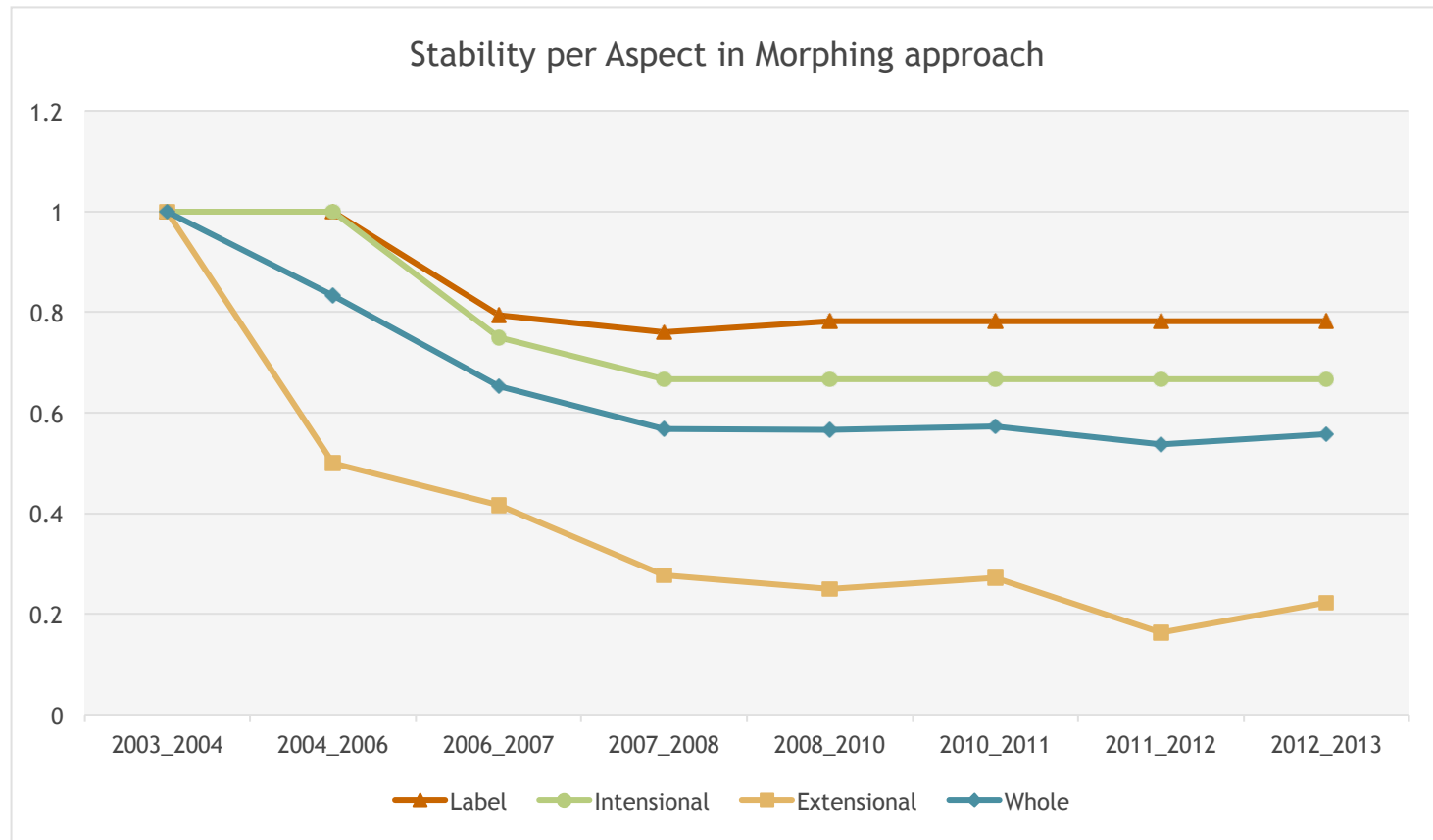
Whole

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

Similarities but no migrations

Instances migrated

Plotting drift across versions



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

Precondition became Condition

Parameter, Process and Profile have properties but remained the same

Similar due to no properties

Concept-per-Concept Stability

Rows: Profile 1.0 Columns: Profile 1.2

Intension

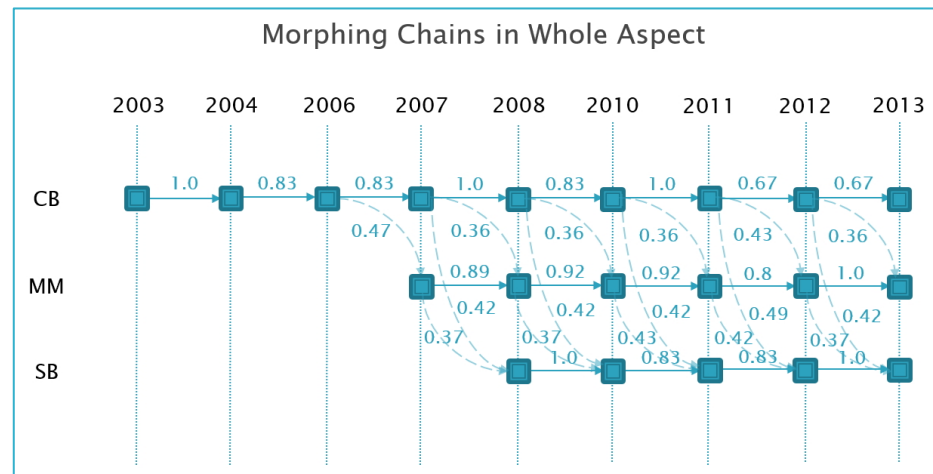
	Condition	Input	Output	Parameter	Process	Profile	Result	Service profile
ConditionalEffect	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ConditionalOutput	0.000	1.000	1.000	0.000	0.000	0.000	0.000	1.000
Input	0.000	1.000	1.000	0.000	0.000	0.000	0.000	1.000
Parameter	0.000	0.000	0.000	1.000	0.000	0.000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

SemaDrift FX

SeeDrift 0.1



A Tool to Measure Semantic Concept Drift

► Preview

Open Ontology1

Reload Ontology1

Open Ontology2

Reload Ontology2

- owl:Thing
 - SoftwareAgent
 - Identity
 - Description
 - Dependency
 - DigitalResource
 - DigitalArtwork
 - ComputerBased
 - wasAttributedTo
 - identification
 - hasProgrammingLanguage
 - hasSize
 - to
 - hasStatusFlag
 - describedBy
 - date
 - HumanAgent
 - ConcreteResource

owl:Thing

Identity

Description

HumanAgent

SoftwareAgent

Dependency

ConcreteResource

DigitalResource

DigitalArtwork

ComputerBased

wasAttributedTo

identification

hasProgrammingLanguage

hasSize

to

hasStatusFlag

describedBy

date

HumanAgent

ConcreteResource



Measure Drift

Average Concept Stability

Label	Intentsional	Extensional	Whole
0,348	0,038	0,003	0,130

Concept-per-Concept Stability

Label Aspect

%	Dependency	Identity	Size	Computer	ConcreteR...	SoftwareB...	Artist	Illuminator	HardwareD...	DigitalRes...	Sta
Monitor	0,086	0,124	0,083	0,348	0,143	0,095	0,089	0,443	0,086	0,152	0,12
HardwareD...	0,188	0,087	0,083	0,079	0,096	0,154	0,094	0,063	0,375	0,081	0,083
Equipment	0,119	0,125	0,083	0,050	0,067	0,067	0,056	0,096	0,119	0,050	0,056
Size	0,083	0,131	0,622	0,083	0,083	0,133	0,083	0,083	0,083	0,083	0,083
OperatingS...	0,120	0,119	0,083	0,268	0,122	0,093	0,128	0,259	0,120	0,097	0,11
Dependency	0,404	0,133	0,083	0,092	0,108	0,080	0,067	0,060	0,333	0,086	0,06
Artist	0,067	0,111	0,083	0,089	0,089	0,200	0,566	0,089	0,122	0,089	0,20
DataDepen...	0,233	0,158	0,067	0,138	0,146	0,090	0,100	0,138	0,233	0,151	0,16
StatusFlag	0,058	0,097	0,067	0,086	0,111	0,137	0,153	0,139	0,083	0,128	0,566
Description	0,067	0,139	0,131	0,067	0,133	0,107	0,122	0,048	0,067	0,083	0,12
DisplayDevice	0,127	0,083	0,100	0,073	0,094	0,107	0,106	0,071	0,127	0,103	0,11
Programmi...	0,070	0,067	0,083	0,178	0,067	0,080	0,044	0,197	0,070	0,090	0,17
ComputerB...	0,067	0,082	0,072	0,172	0,124	0,508	0,149	0,106	0,138	0,119	0,12
ConcreteRe...	0,075	0,067	0,083	0,108	0,333	0,127	0,089	0,087	0,087	0,219	0,11
SoftwareDe...	0,188	0,087	0,108	0,079	0,096	0,207	0,078	0,063	0,263	0,081	0,08

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