REX - A Tool for Discovering Evolution Trends in Ontology Regions

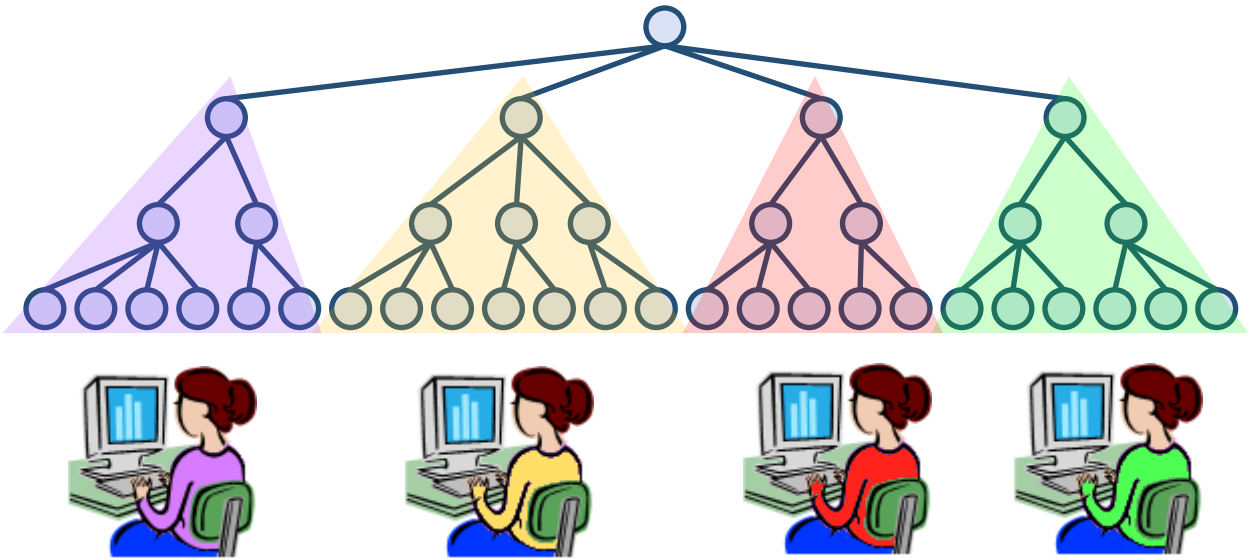
Victor Christen, Anika Gross, Michael Hartung

UNIVERSITÄT LEIPZIG

18TH July 2014, DILS, Lisboa

Ontology Evolution

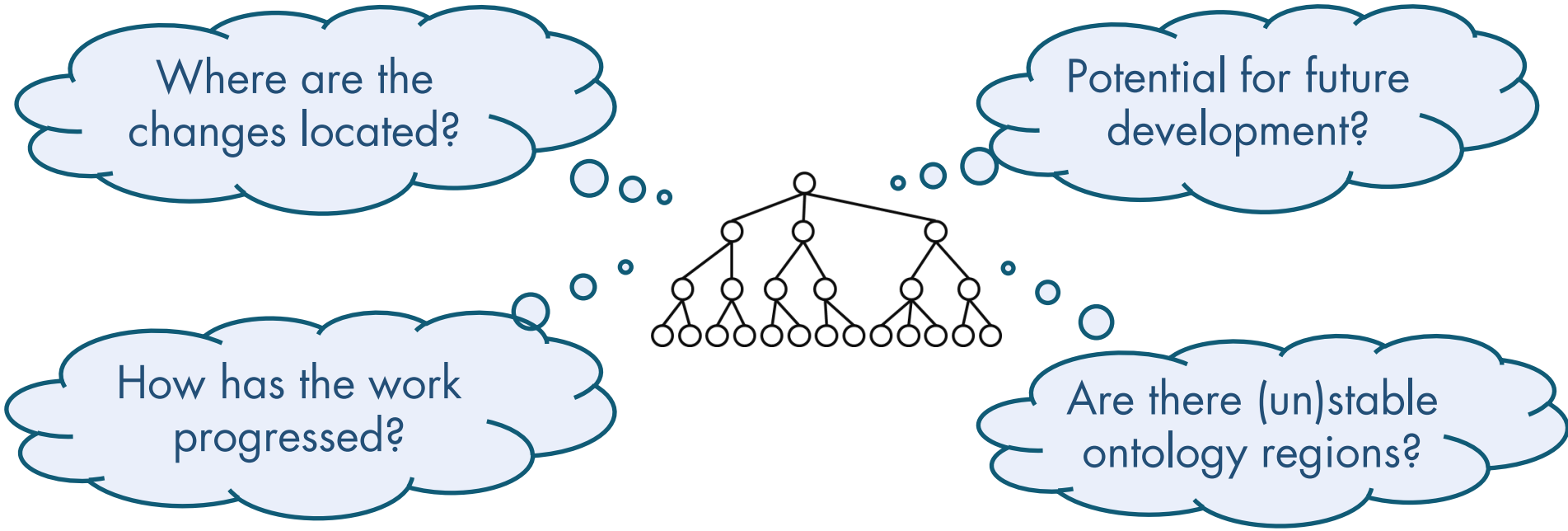
- Heavy usage of ontologies in the life sciences, e.g. annotations, semantic search, …
- Ongoing research, new findings → Ontology changes
- Continuous release of new versions 
- Collaborative, distributed development of large ontologies



Ontology developers with different fields of expertise

Coordination by project leader

Ontology Evolution



- Unstable ontology regions
  - → Many modifications → focus of recent development
  - → Impact of changes on ontology-based algorithms or applications → redo analyses?
- Stable ontology regions
  - → Already completed?
  - → Low interest so far → further changes necessary?

Current State & Contributions

- BioPortal, OBO Foundry , …
  - Provide life science ontologies with versions
- Diff computation (PromptDiff, COntoDiff, …)
- Web tools to analyze ontology evolution (GOChase, OnEX, …)

→ **No tool to analyze and compare evolution in different ontology parts**

Web application **R**egion **E**volution E**x**plorer (REX)

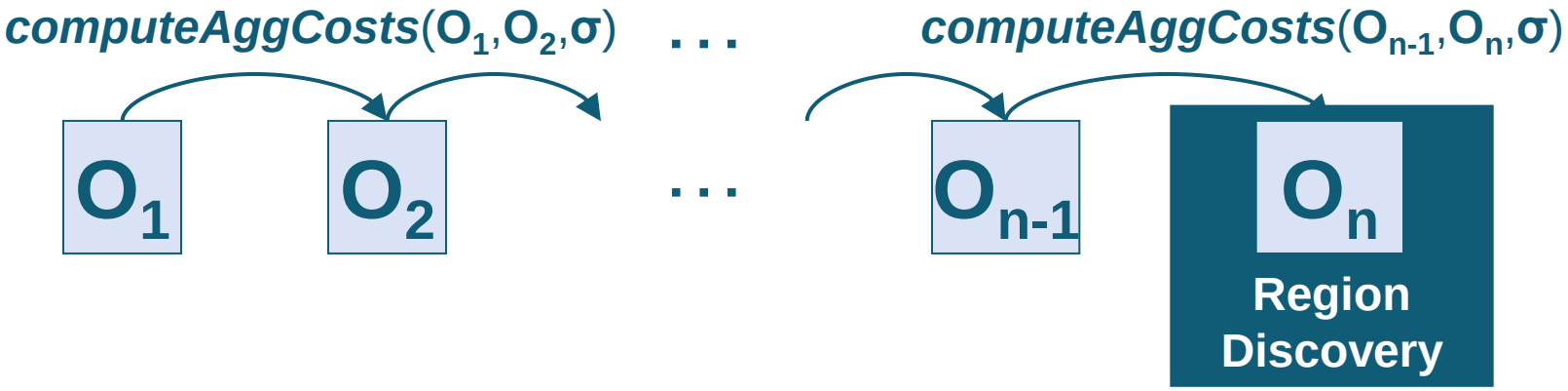
- Determination of differently changing ontology regions
- Interactive exploration of regions w.r.t. their change intensity
- Comparative trend analysis
  → Monitor long-term evolution for regions of interest
  → Track the work or coordinate future development

Agenda

- Region Discovery Method
- Trend Discovery
- REX
  - Infrastructure
  - Analysis workflows
- Conclusions and Outlook

Region Discovery Method

- **Basic idea**: Compute change intensities for regions based on changes between succeeding ontology versions O_1 … O_n using change cost model σ



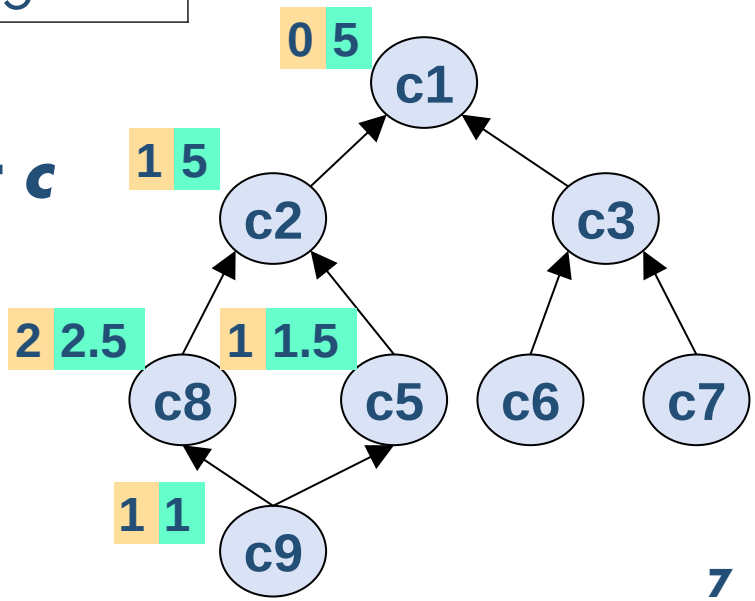
1) Compute changes between succeeding versions
2) Propagate change costs within the is-a hierarchy
3) Transfer aggregated costs from the first to the last version
4) Compute change intensities to discover differently evolving regions

Ontology Changes and Costs

	Change operation	Change costs
Concepts	*addC*	1
	delC	2
Relationships	*addR*	0.5/0.5
	delR	1.0/1.0
Attributes	*addA*	0.5
	delA	0.5
	chgAttValue	0.5

Cost model can be adapted!

- **Costs of an ontology concept *c***
  - Local costs ***lc(c)***
    - Changes that directly affect ***c***
  - Aggregated costs ***ac(c)***
    - Changes on *is-a* descendants of ***c***



Compute Aggregated Costs for Two Versions

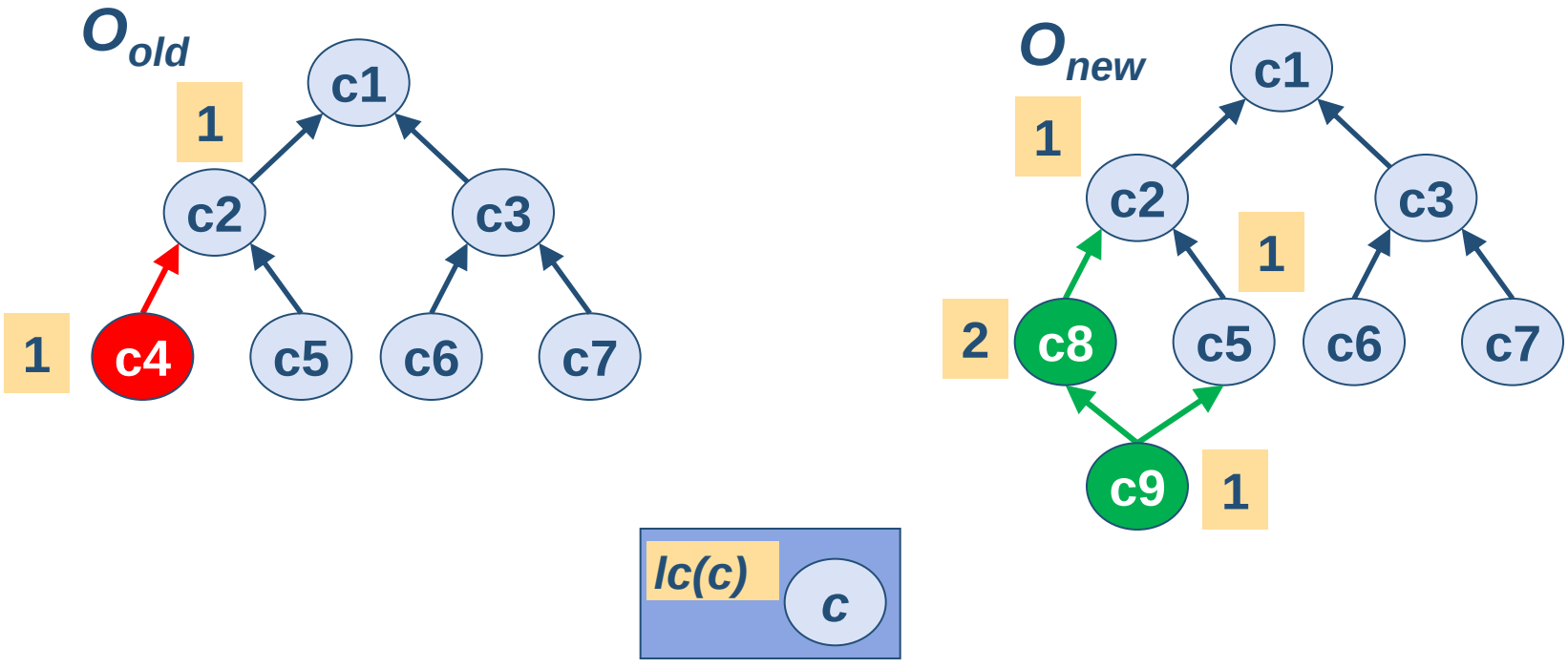
Algorithm: *computeAggregatedCosts*(O_{old}, O_{new}, σ)

- **Input:** Two ontology versions O_{old} and O_{new}, change costs σ
- **Output:** O_{new} with computed aggregated costs (*ac*)

1. Detect changes: diff (O_{old}, O_{new})
2. Assign local costs to O_{old} and O_{new}
3. Aggregate costs in O_{old} and in O_{new}
4. Transfer costs $O_{old} \rightarrow O_{new}$

Change Detection And Local Costs

- **Simple diff computation:** based on accession numbers used in ontology elements → *addC, delC, addR, delR, …*
- Example
  - Uniform change costs of 1
  - Relationship costs to target

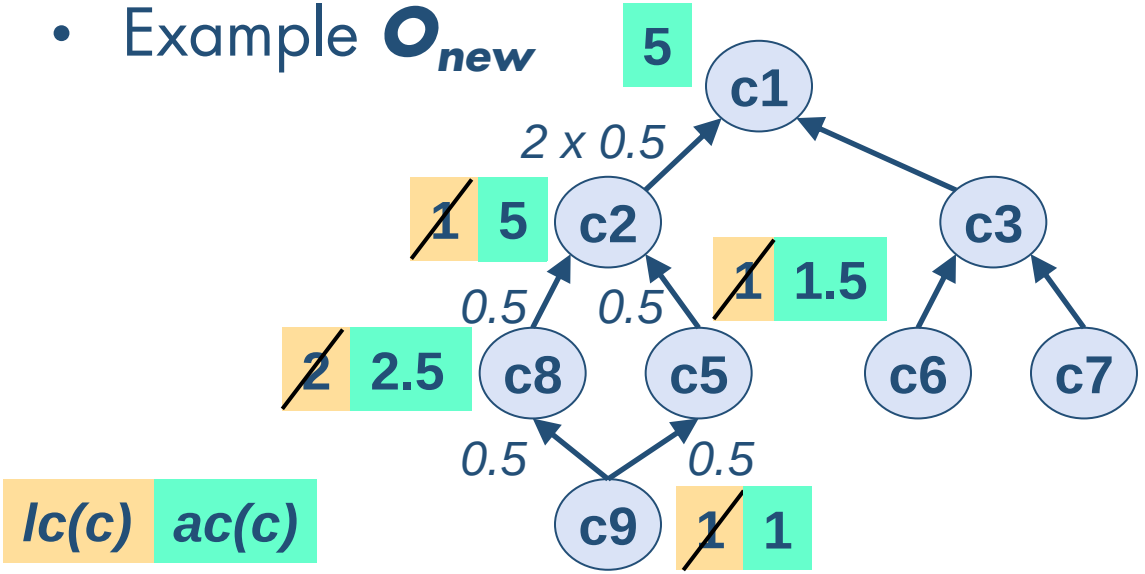


Cost Propagation and Aggregation

- Propagation of local costs ***lc***
- Aggregated costs ***ac(c)*** of concept *c*:
  Weighted sum of all children *ac*'s plus own local costs *lc*(*c*)

$$ac(c) = \sum_{direct\ children\ c'\ of\ c} \frac{ac(c')}{|parents(c')|} + lc(c)$$

- Example O_{new}



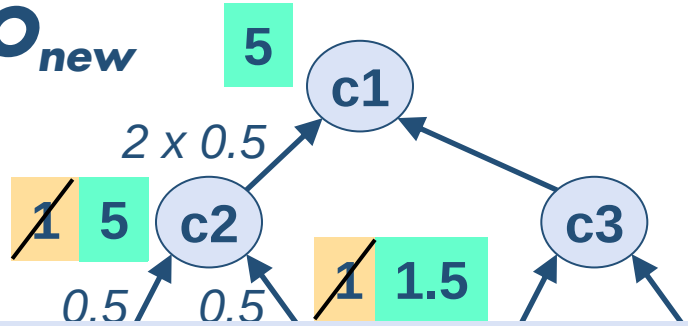
propagation of *lc*(c9)
propagation of *lc*(c8)
propagation of *lc*(c5)
propagation of *lc*(c2)

Cost Propagation and Aggregation

- Propagation of local costs ***lc***
- Aggregated costs ***ac(c)*** of concept *c*:
  Weighted sum of all children *ac*'s plus own local costs *lc*(*c*)

$$ac(c) = \sum_{direct\ children\ c'\ of\ c} \frac{ac(c')}{|parents(c')|} + lc(c)$$

- Example O_{new}



propagation of *lc*(c9)
propagation of *lc*(c8)
propagation of *lc*(c5)

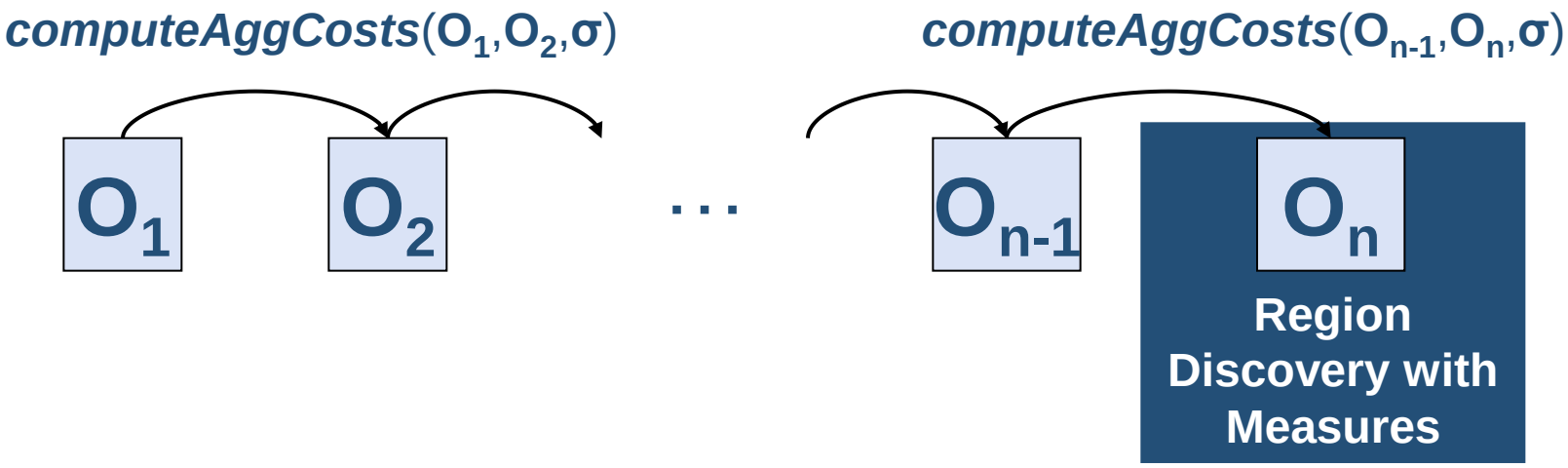
- Transfer aggregated costs $O_{old} \rightarrow O_{new}$

→ Costs of del changes reflected in new(er) version(s)

- Costs of equal concepts are summed up

Discovery Algorithm for Multiple Versions

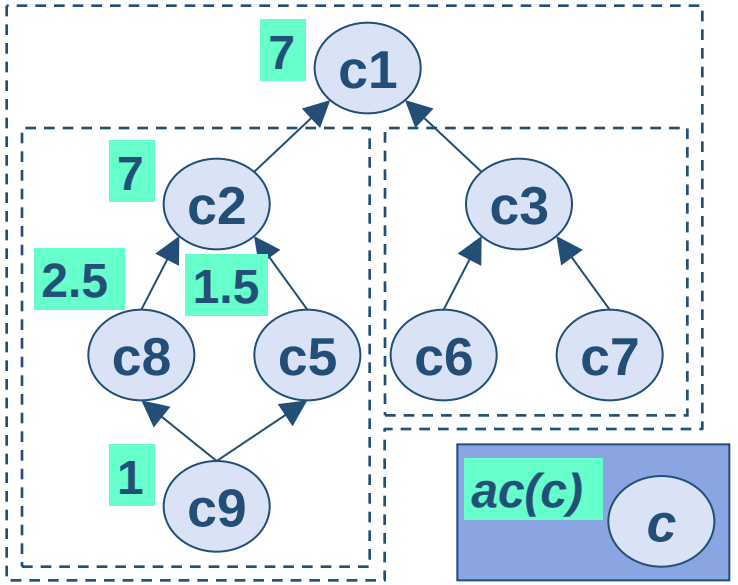
- Reuse of ***computeAggregatedCosts*** algorithm
  - Successive computation and transfer of aggregated costs
  - Apply region discovery on latest version



Ontology Regions and Measures

- **Ontology region *OR***
  - Subgraph of an ontology consisting of a single root concept and all concepts in its *is-a* subgraph

- **Region measures**
  to quantify change intensity
  - Absolute change costs
  - Absolute / relative region size
  - Average change costs
  - …



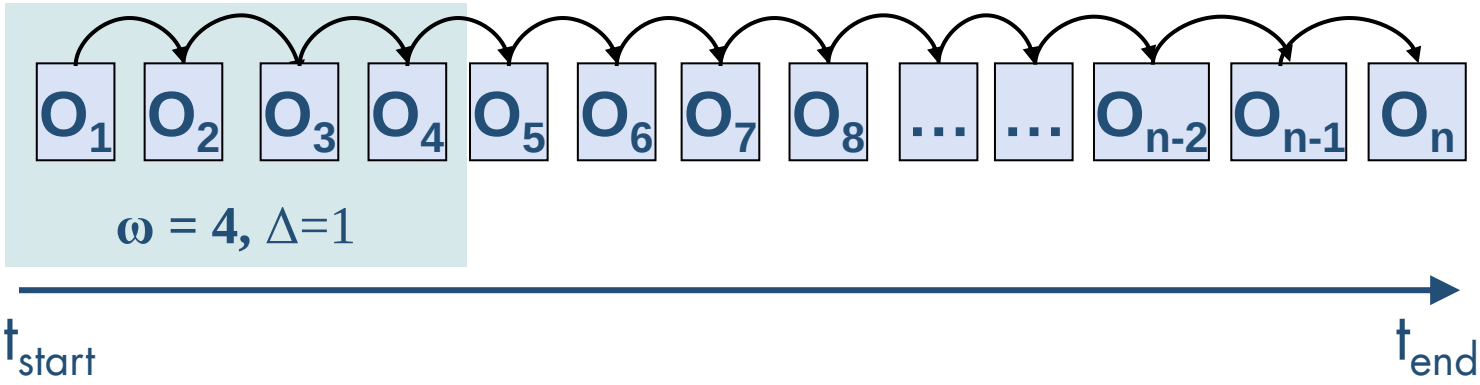
region	*abs_costs*	*rel_size*	*avg_costs*
c1	7	8/8=1	7/8=0.875
c2	7	4/8=0.5	7/4=1.75
c3	0	3/8=0.375	0/3=0

Trend Discovery

- Trend discovery based on sliding windows
- Monitor region changes over long periods of time
  - Ontology O, region of interest OR
  - Time interval (t_{start}, t_{end})
  - Sliding window of size ω
  - Step width Δ

→ Call region discovery algorithm within ω

→ Collect change intensities for region of interest over time

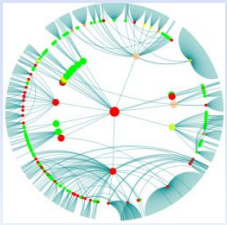


Agenda

- Region Discovery Method
- Trend Discovery
- REX
  - Infrastructure
  - Analysis workflows
- Conclusions and Outlook

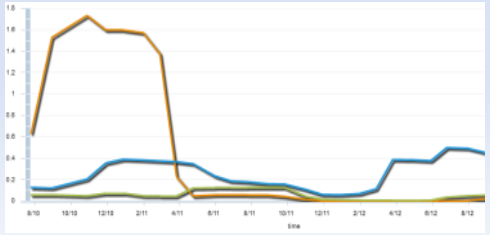
INFRASTRUCTURE

Client



heart

accession	name	avg...	children
MA-0000488	apex of heart	2.0	[]
MA-0003017	apex of heart vent...	1.5	[]
MA-0002131	great vein of heart	1.5	[]
MA-0000072	heart	11.0	["MA-0000488","M...
MA-0000073	heart atrium	3.0	["MA-0002955","M...



request **response**

Communication Interface

Function Calls

Server

Trend Analysis Service → Region Algorithm ← Graph Data Service

Quantitative Service

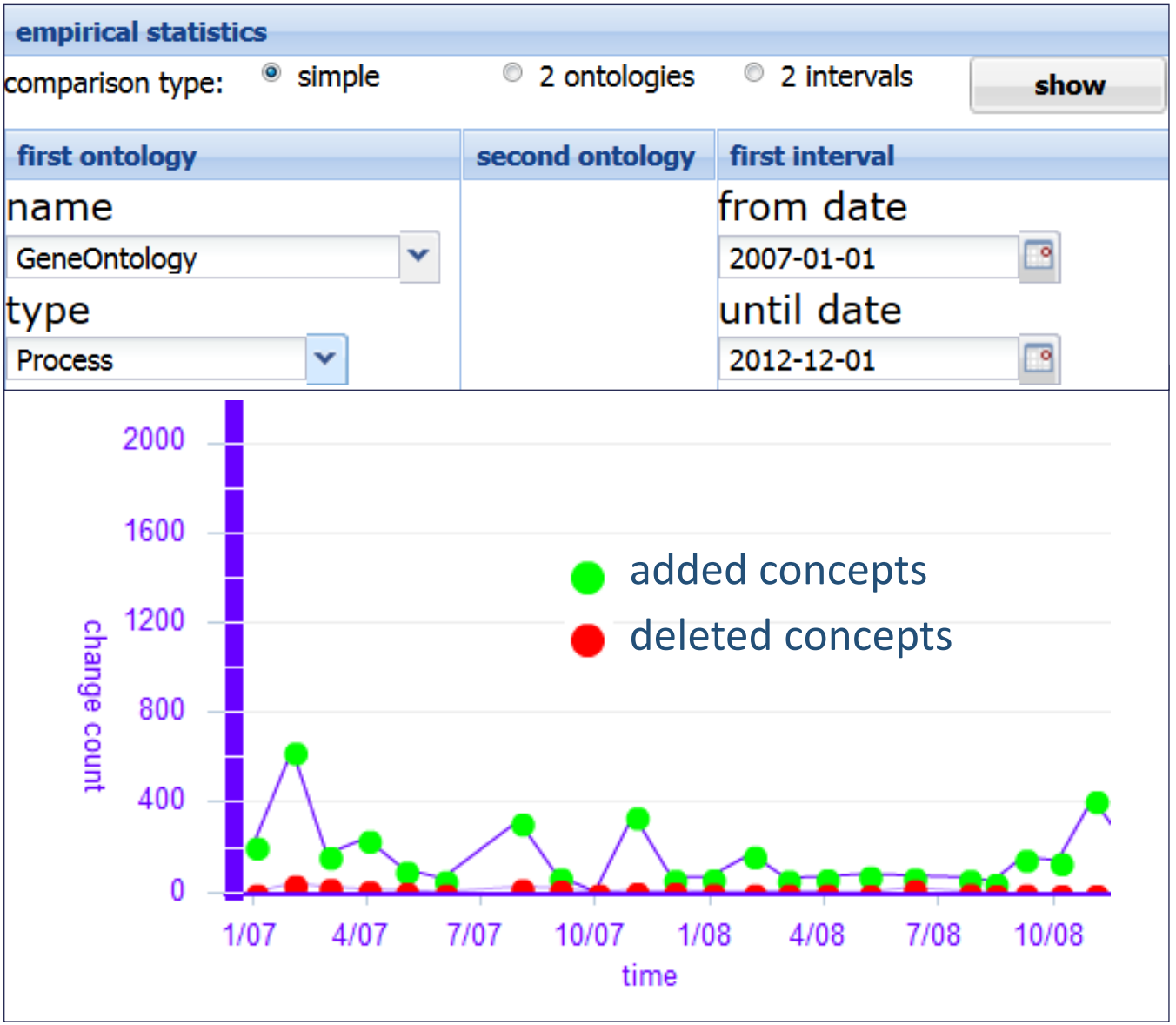
SQL

Knowledge Base

OnEX

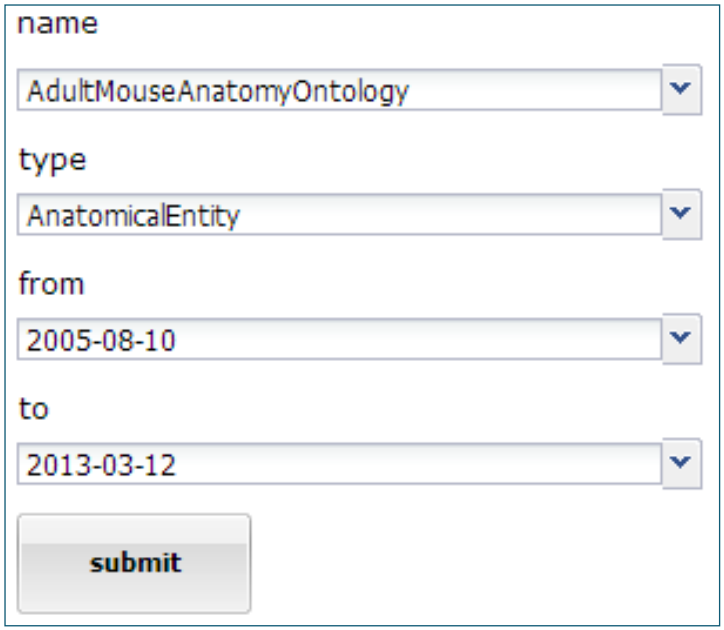
Ontology versions from the life sciences

Quantitative Change Analysis

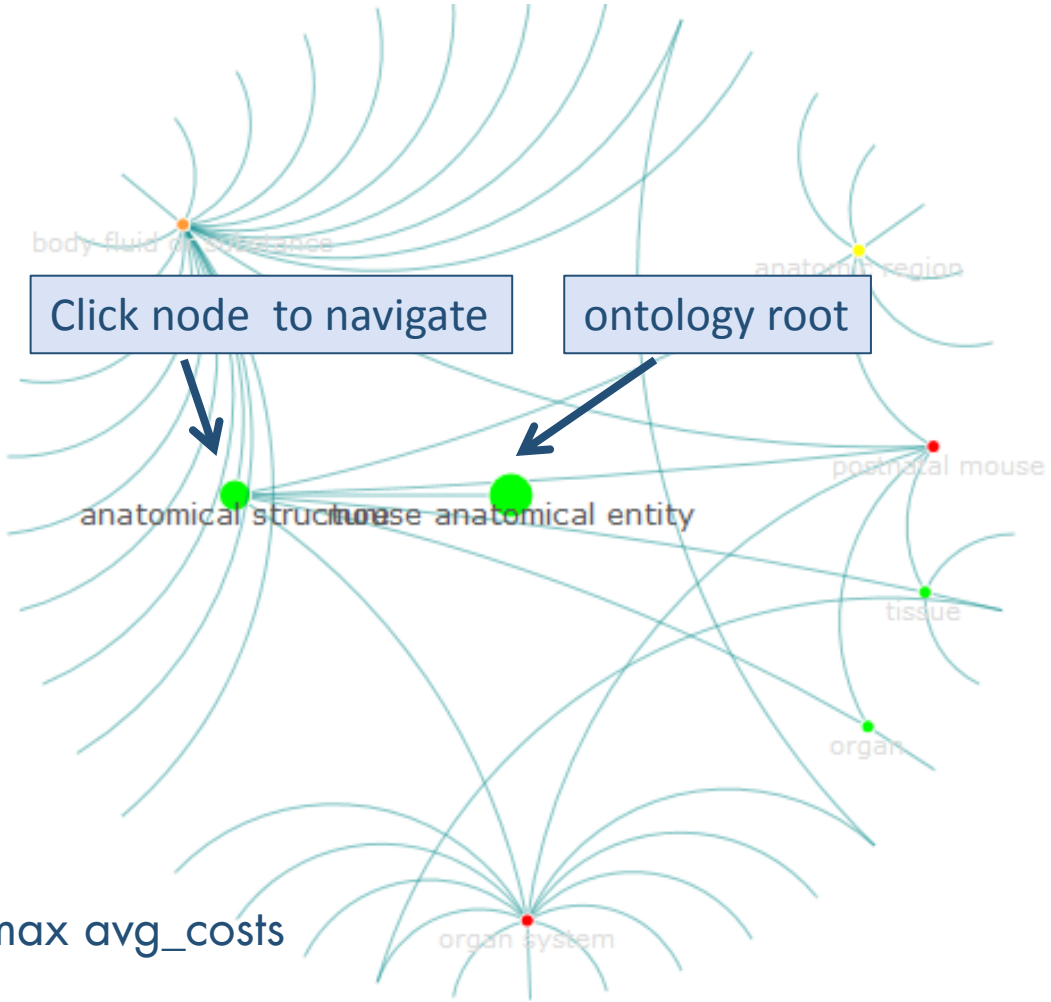


Structural Analysis

Input



Browser View



stable unstable

Coloring schemes

- Interval based grouping
- Equal distribution between min/max avg_costs

Structural Analysis

Input

name

AdultMouseAnatomyOntology

type

AnatomicalEntity

from

2005-08-10

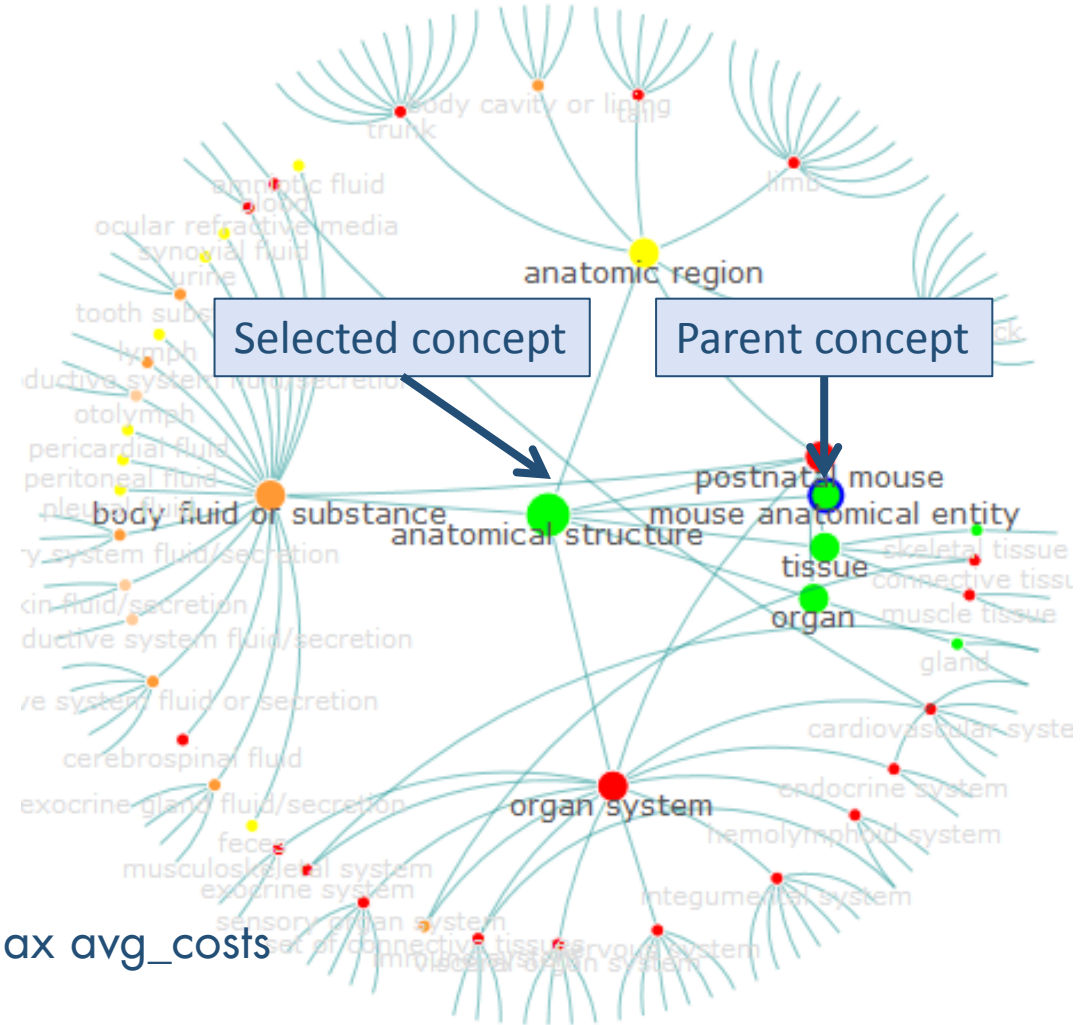
to

2013-03-12

submit

stable unstable

Browser View



Coloring schemes

- Interval based grouping
- Equal distribution between min/max avg_costs

Structural Analysis

Input

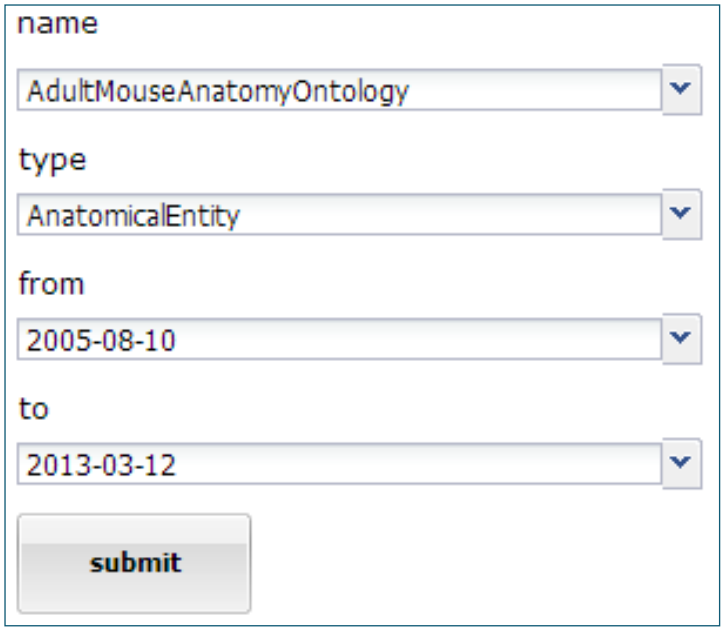


Table View

heart

accession	name	avgCo...	children
MA-0000072	heart	11	["MA-0000488","MA-00
MA-0000098	heart mesentery	7.5	["MA-0000482"]
MA-0000085	heart interventri...	5.5	["MA-0002939","MA-00

Change History

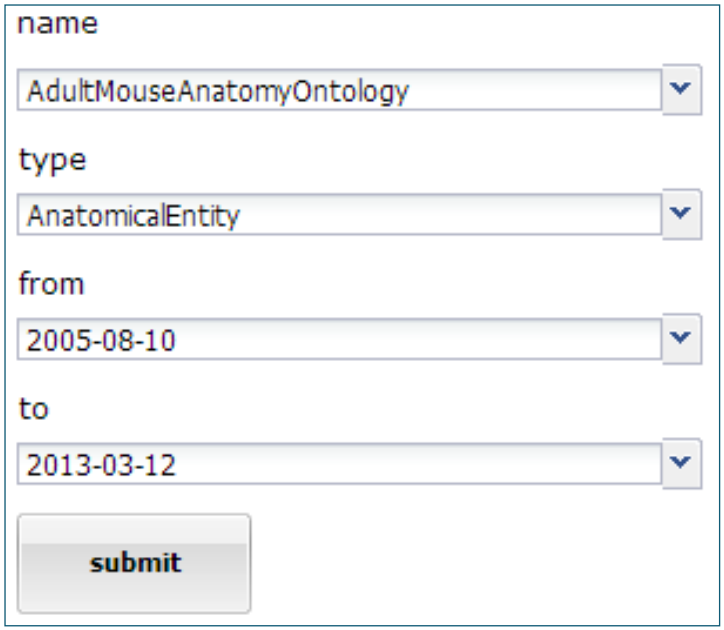
Change history of concept MA-0000072

date of cha...	type of ch...	type of i...	changed i...	new value(s)	old v
2005-08	addition	attribute	name	heart	
2005-08	addition	relation	is_a	MA-0000557	
2005-08	addition	relation	part_of	MA-0000010	
2005-12	addition	relation	part_of	MA-0002449	
2005-12	deletion	relation	is_a		MA-0

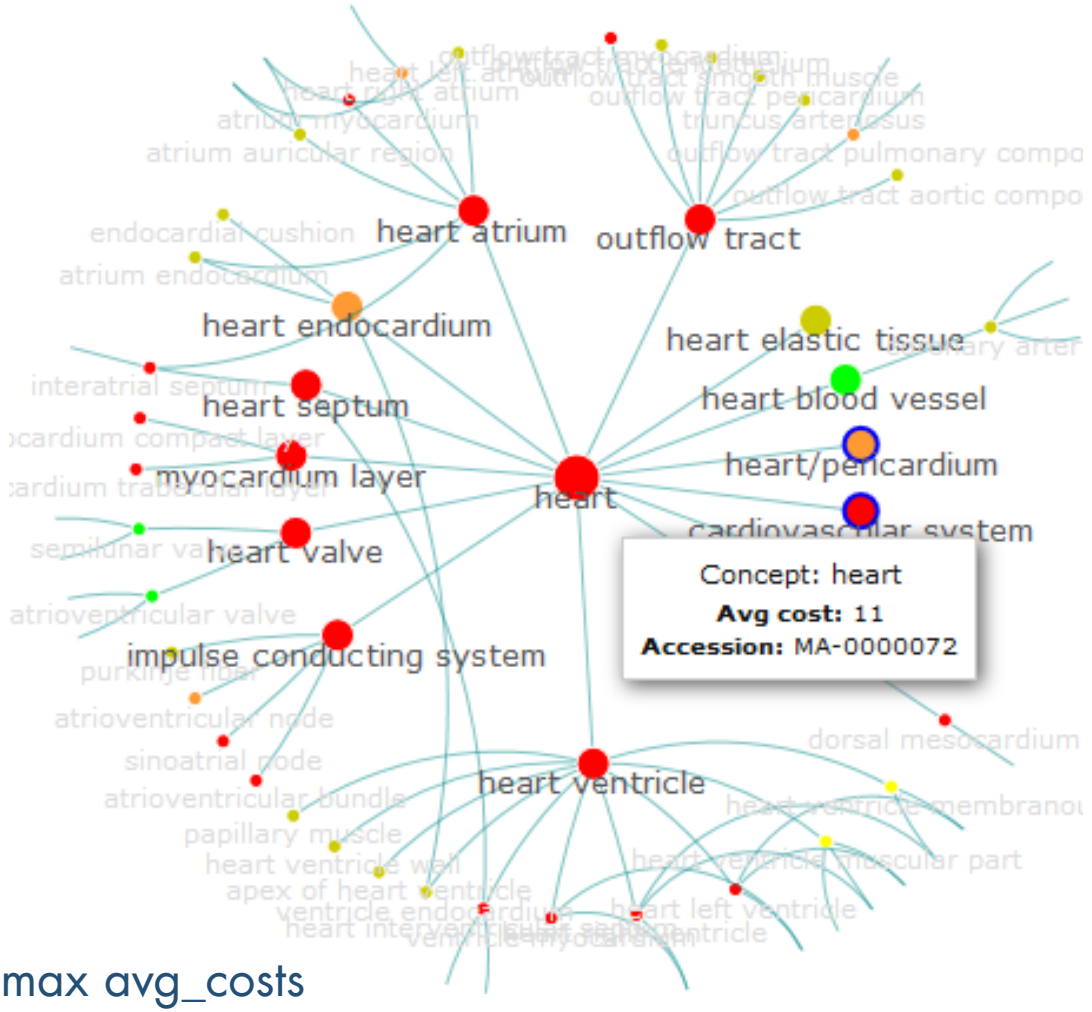
- Filter / sort by accession number, name, avg_costs …
- View individual change history of concepts

Structural Analysis

Input



Browser View



stable unstable

Coloring schemes

- Interval based grouping
- Equal distribution between min/max avg_costs

Trend Analysis

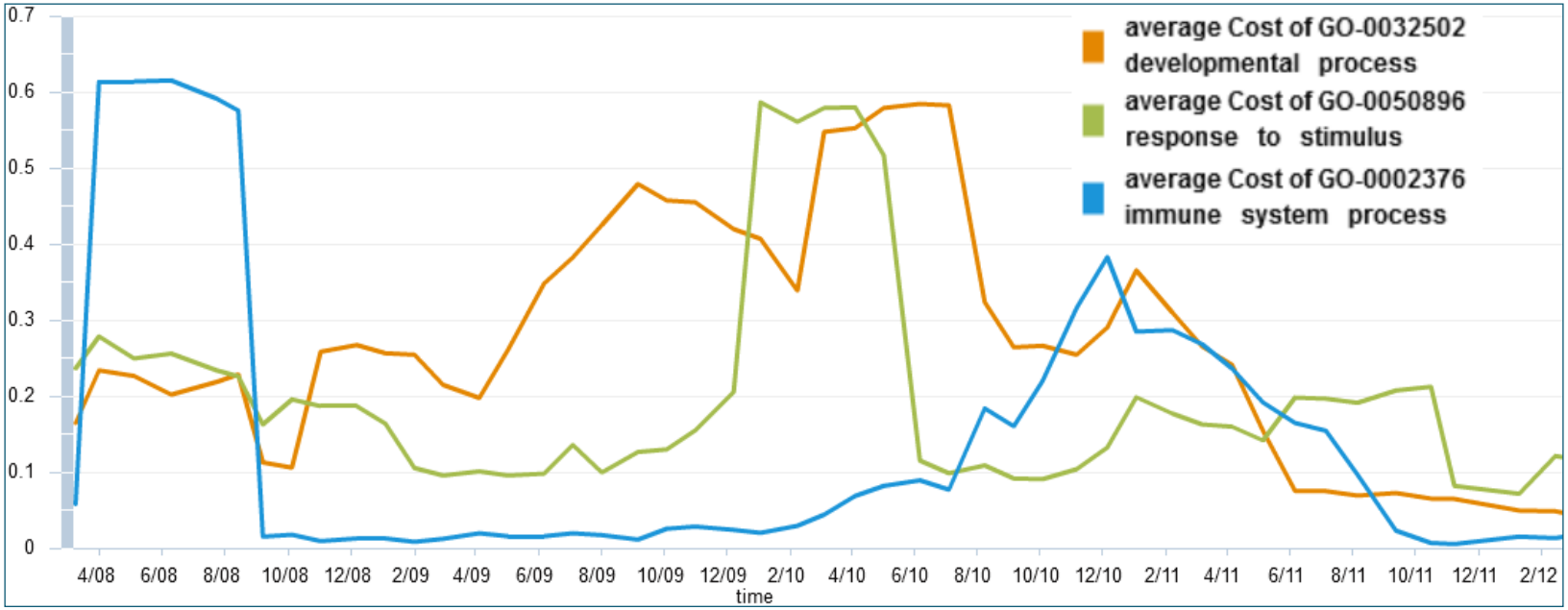
specification

name: GeneOntology
type: Process
from: 2008-03-09
to: 2013-03-12
step size: 1
window size: 6
1. submit 2. calculate trends 3. export

immune system

accession	name
GO-0002376	immune system process
GO-0002520	immune system development
GO-0002682	regulation of immune system process
GO-0002683	negative regulation of immune system process
GO-0002684	positive regulation of immune system process

accession	name
GO-0032502	developmental process
GO-0050896	response to stimulus
GO-0002376	immune system process



Conclusions And Future Work

- Explore evolution of life science ontologies to discover (un)stable regions and trends
- Monitor development of large ontologies (developers, project coordinators)
- Support ontology-based algorithms and analysis

- Support versioning of individual ontologies (upload of own versions)
- Include user-based adaptation of cost model
- Web service interface