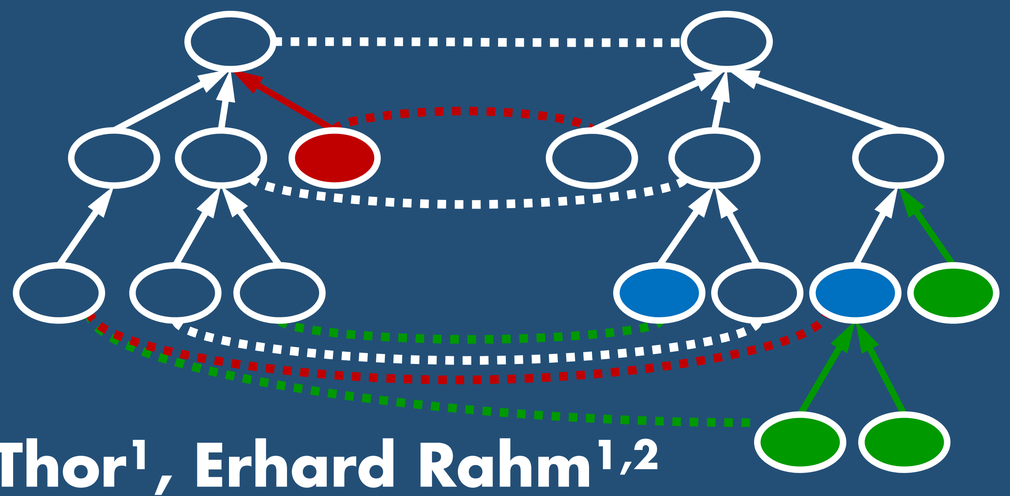
How do Ontology Mappings Change in the Life Sciences?



Anika Groß[1,2], Michael Hartung[1,2], Andreas Thor[1], Erhard Rahm[1,2]

[1] Department of Computer Science, University of Leipzig
[2] Interdisciplinary Centre for Bioinformatics, University of Leipzig
contact: gross@informatk.uni-leipzig.de

http://dbs.uni-leipzig.de
http://www.izbi.de

Motivation

Ontology Evolution

- Ongoing research, new findings → continuous modifications
- Periodical release of new ontology versions

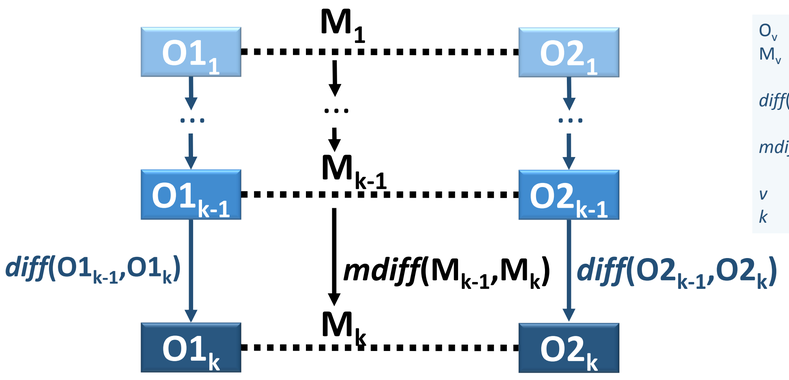
Ontology Mappings

- Set of semantic correspondences between concepts of different ontologies
- Possible invalidation of previously determined ontology mappings due to ontology evolution

Aims

- Investigate evolution of life science ontology mappings
- Study impact of ontology evolution on mapping evolution
- Future Work: Use known ontology changes to semi-automatically adapt ontology mappings

General evolution scheme



O_v	Ontology version
M_v	Ontology Mapping version ($M_{O1_v,O2_v}$)
$diff(O_v,O_{v+1})$	Difference between two ontology versions
$mdiff(M_v,M_{v+1})$	Difference between two mapping versions
v	Version number
k	Current version

Change Operations

Ontologies Changes

Extension set $Ext(O_{v\to v+1})$
Add new concept, subgraph, relationship, attribute, …

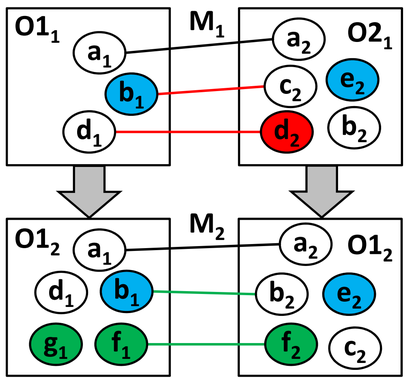
Reduction set $Red(O_{v\to v+1})$
Delete existing concept, subgraph, relationship, …

Revision set $Rev(O_{v\to v+1})$
Split, merge, substitute, move concept, …

Ontology Mapping Changes

Addition set $Add(M_{v\to v+1}) = M_{v+1} \backslash M_v$ **Deletion set** $Del(M_{v\to v+1}) = M_v \backslash M_{v+1}$

Change Ratios



Mapping / Ontologies		ADD	DEL
		$\{(b_1,b_2),(f_1,f_2)\}$	$\{(b_1,c_2),d_1,d_2)\}$
EXT	$\{f_1, g_1\}\cup\{f_2\}$	2/3	0
RED	$\varnothing\cup\{d_2\}$	0	1
REV	$\{b_1\}\cup\{e_2\}$	1/2	1/2

Ontology Change Ratio *(OCR)*

$$OCR(O_{v\to v+1}) = \frac{|Ext(O_{v\to v+1}) \cup Red(O_{v\to v+1}) \cup Rev(O_{v\to v+1})|}{|O_v \cup O_{v+1}|}$$

Mapping Change Ratio *(MCR)*

$$MCR(M_{v\to v+1}) = \frac{|Add(M_{v\to v+1}) \cup Del(M_{v\to v+1})|}{|M_v \cup M_{v+1}|}$$

$$OCR(O1_{1\to 2}) = \frac{3}{5} \qquad OCR(O2_{1\to 2}) = \frac{3}{6} \qquad MCR(M_{1\to 2}) = \frac{4}{5}$$

Impact ratio *(IR)*

$$IR(O_{Ch}, M_{Ch}) = \frac{|\{c \in O_{Ch} | \exists c' : (c,c') \in M_{Ch} \vee (c',c) \in M_{Ch}\}|}{|O_{Ch}|}$$

Fraction of additive ontology changes that lead to new correspondences:

$O_{Ch} = Ext(O1_{1\to 2}) \cup Ext(O2_{1\to 2})$
$M_{Ch} = Add(M_{1\to 2})$
⇒ $IR(Ext, Add) = \frac{2}{3}$

Evaluation

Three Life Science Match Problems

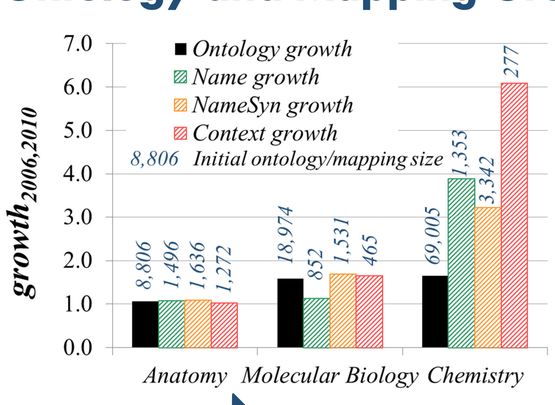
- Analyze versions between 2006 and 2010

Anatomy	Adult Mouse Anatomical Dictionary (MA)	NCI Thesaurus Anatomy part (NCITa)	
Molecular Biology	Molecular Functions (MF)	Biological Processes (BP)	
Chemistry	Chemical Entities of Biological Interest (ChEBI)	NCI Thesaurus (NCIT)	

3 Meta-data based Matchers

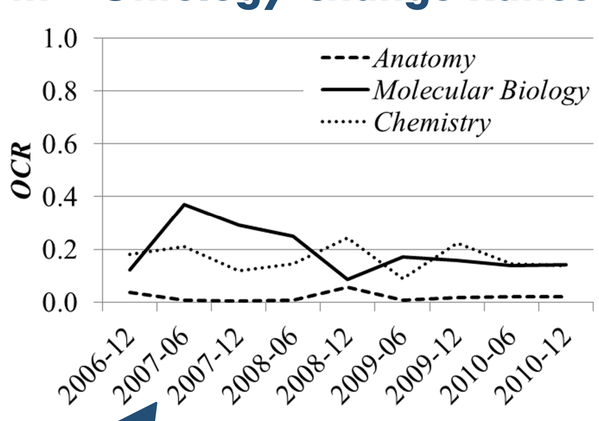
Preprocessing Normalization → *Name* / *NameSyn* / *Context* → Selection (t=0.6, Max1)

Ontology and Mapping Growth



Anatomy relatively stable

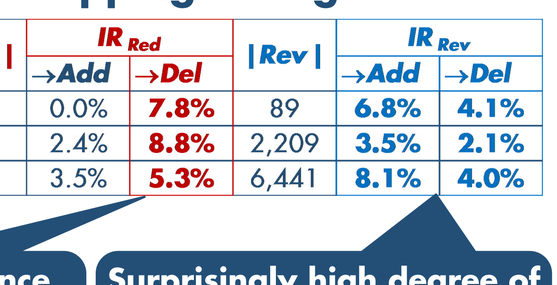
Ontology Change Ratios



Heavy changes for *Molecular Biology* and *Chemistry*

Mapping Changes

More correspondence additions + High degree of deletions



Ontology And Mapping Change Ratios

Correlation between ontology and mapping change factors

Different stability for different match techniques
Context → most unstable

Impact of Ontology Changes on Mapping Changes

	|Ext|	IR_{Ext} →Add	IR_{Ext} →Del	|Red|	IR_{Red} →Add	IR_{Red} →Del	|Rev|	IR_{Rev} →Add	IR_{Rev} →Del
Anatomy	95	**18.7%**	0.1%	7	0.0%	**7.8%**	89	**6.8%**	**4.1%**
Molecular Biology	2,359	**4.6%**	0.7%	223	2.4%	**8.8%**	2,209	**3.5%**	**2.1%**
Chemistry	8,377	**11.7%**	1.2%	366	3.5%	**5.3%**	6,441	**8.1%**	**4.0%**

Most correspondence additions are caused by ontology extensions

Most correspondence deletions are caused by ontology reductions

Surprisingly high degree of mapping changes caused by ontology revisions