

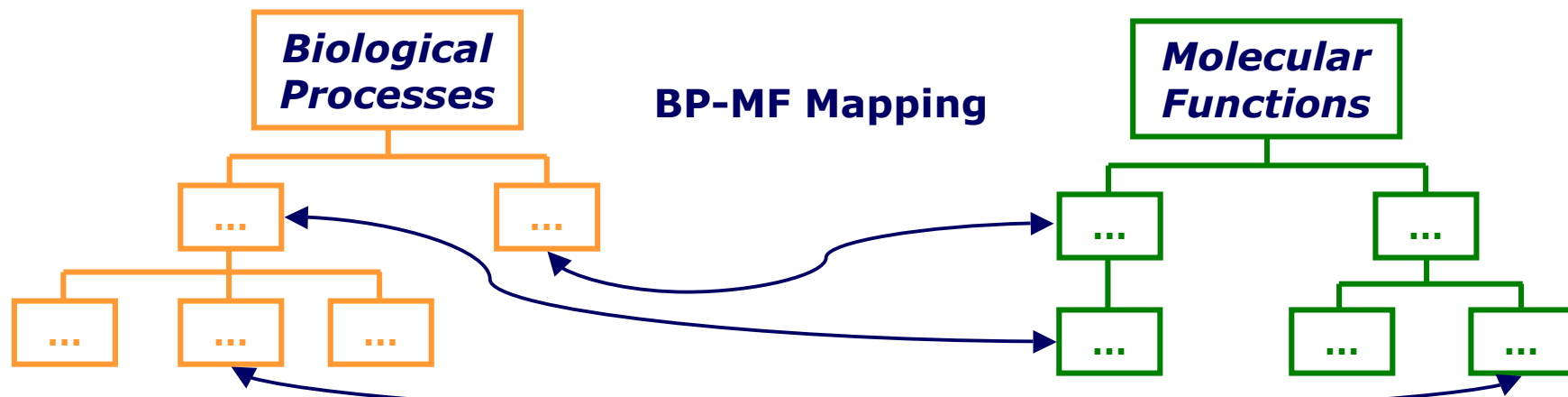
An Evolution-based Approach for Assessing Ontology Mappings – A Case Study in the Life Sciences

Andreas Thor, Michael Hartung, Anika Groß,
Toralf Kirsten, Erhard Rahm



Ontologies and Ontology Mappings

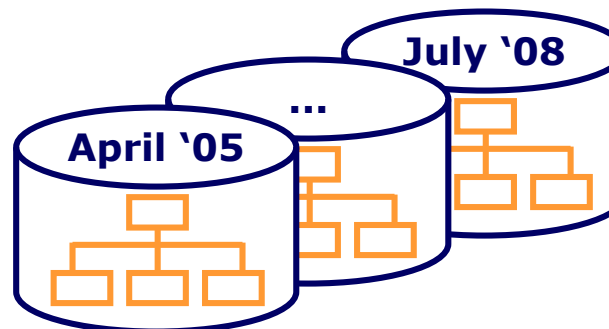
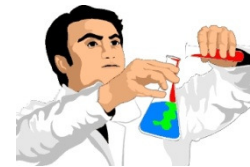
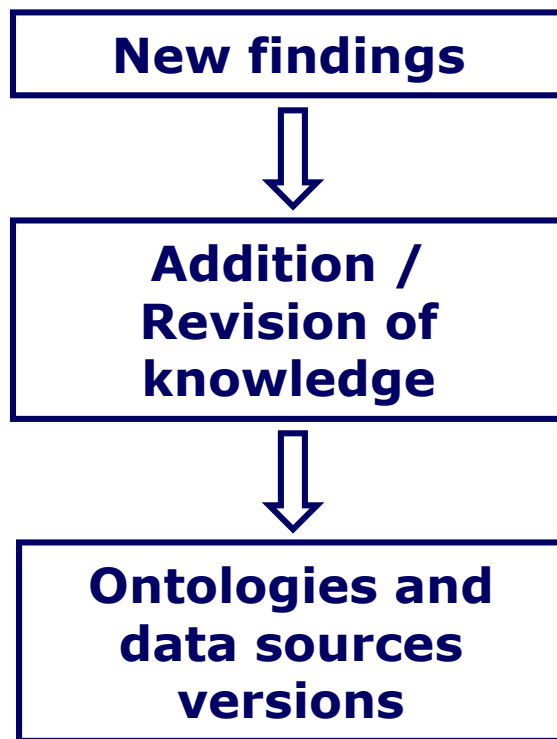
- **Ontologies** became widely accepted to represent knowledge
 - Life sciences: semantic descriptions of biological objects (annotations)
- **Ontology Mappings:** set of semantic correspondences between concepts of different ontologies
 - Crucial for data integration, enhanced data analysis, ...
 - Generated by ontology matching approaches



Evolution

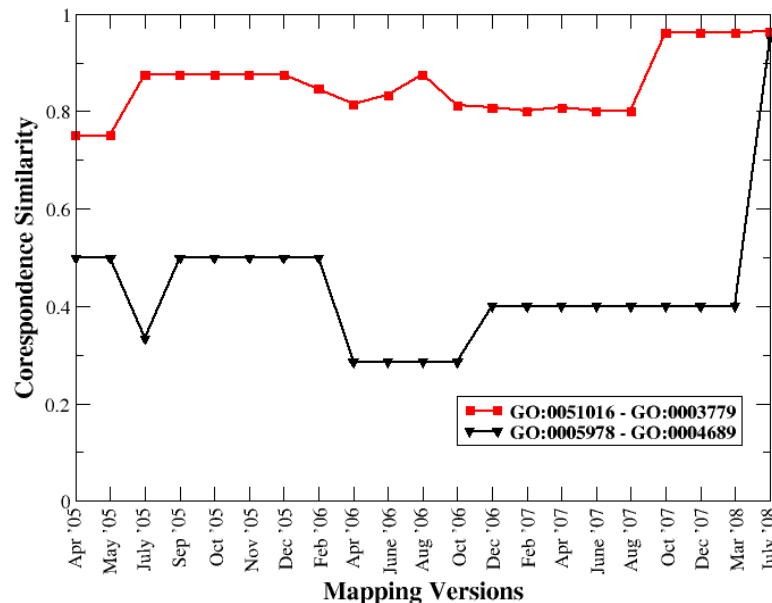
- **Domain knowledge changes**

- continuous and numerous modifications of ontologies and related data sources
- periodical releases



Impacts of Evolution on Ontology Mappings

- Current state Match approaches only consider information about latest (current) ontology versions



Is the **black** correspondence as good as the **red** one?

- Possible instabilities of match correspondences due to evolution of ontologies and/or related data source

- Assumption Reasonable to draw conclusions from **evolution** of a match correspondence to assess its **quality** in the current version

Contributions

- Generic approach to extend generated ontology mappings using information about its historic changes
 - Usable to classify correspondences w.r.t. their quality, e.g., in categories “good and stable” or “good & unstable”
- Definition of two stability measures
 - Quantify the evolution of similarity values for concept correspondences
- Example evaluation in the life sciences

Approach

- Matching process is executed for different ontology versions
- Sequence of match similarities **$sim_i(a,b,m)$**
 - Determined by (an arbitrary) matcher m
 - For correspondence (a,b)
 - In version i
- Characterize correspondence stability
 - For current version n
 - In comparison to last k versions

Stability Measures

- **Average Stability**

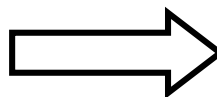
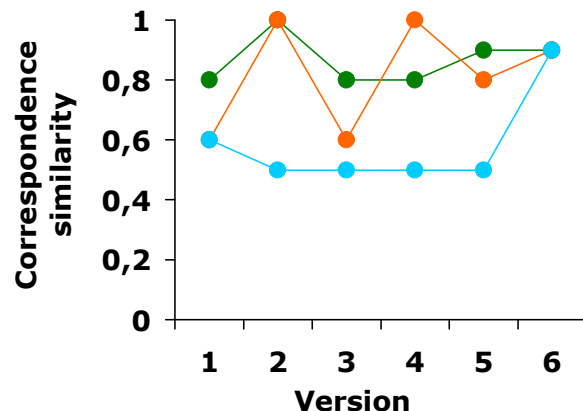
- Average similarity difference per version change

$$stabAvg_{n,k}(a,b,m) = 1 - \frac{1}{k} \cdot \sum_{i=n-k}^{n-1} |sim_{i+1}(a,b,m) - sim_i(a,b,m)| \in [0,1]$$

- **Weighted Maximum Stability**

- Proximity of similarities in the last versions compared to the current version

$$stabWM_{n,k}(a,b,m) = 1 - \max_{i=1\dots k} \left[\frac{|sim_n(a,b,m) - sim_{n-i}(a,b,m)|}{i} \right] \in [0,1]$$



	<i>stabAvg</i> _{6,5}	<i>stabWM</i> _{6,5}
(a ₁ , b ₁)	0.9	0.95
(a ₂ , b ₂)	0.7	0.9
(a ₃ , b ₃)	0.9	0.6

Evaluation

- Settings
 - Mapping GO *biological processes* to *molecular functions*
 - Instance based matching (using Ensembl source)
 - Result: 2497 correspondences (similarity ≥ 0.8)
- Classification of correspondences w.r.t. additional criteria

accepted

candidates

questionable

$\begin{matrix} \text{sim}_{26} > 0.9 \\ \text{sim}_{26} \leq 0.9 \end{matrix}$	$\begin{matrix} \text{stabWM} > 0.95 \\ \text{stabWM} \geq 0.85 \end{matrix}$	$\begin{matrix} 0.95 \geq \text{stabWM} \geq 0.85 \\ \text{stabWM} > 0.95 \end{matrix}$	$\begin{matrix} 0.85 > \text{stabWM} \\ \text{stabWM} \geq 0.85 \end{matrix}$	Σ
$\begin{matrix} \text{stabAvg} > 0.95 \\ \text{stabAvg} \geq 0.85 \end{matrix}$	424 55%	37 15%	11 25	596
$\begin{matrix} 0.95 \geq \text{stabAvg} \geq 0.85 \\ \text{stabAvg} > 0.95 \end{matrix}$	863 96	203 212	235 30%	1734
$\begin{matrix} 0.85 > \text{stabAvg} \\ \text{stabAvg} \geq 0.85 \end{matrix}$	17 5	13 16	85 31	167
Σ	1449	536	512	2497

Conclusions and Future Work

- Generic evolution-based approach for assessing ontology mappings
 - Independent from match approach
 - Based on stability measures
- Ranking of correspondences for the “match or non-match” decision
- Example evaluation in the life sciences
- Applicability of our approach in different domains
- Alternative stability measures by using further knowledge (e.g., ontology structure)