# Ontology Alignment

Valentina Ivanova

most slides from Patrick Lambrix

### **Ontology Alignment**

- Ontology alignment
- Ontology alignment strategies
- Evaluation of ontology alignment strategies
- Ontology alignment challenges

### **Ontologies in Biomedical Research**

- many biomedical ontologies
   e.g. GO, OBO, SNOMED-CT
- practical use of biomedical ontologies
  - e.g. databases annotated with GO

### GENE ONTOLOGY (GO)

immune response i- acute-phase response i- anaphylaxis i- antigen presentation i- antigen processing i- cellular defense response i- cvtokine metabolism i- cytokine biosynthesis synonym cytokine production p-regulation of cytokine biosynthesis i-B-cell activation i- B-cell differentiation i- B-cell proliferation i- cellular defense response i- T-cell activation i- activation of natural killer cell activity . . .

## **Ontologies with Overlapping Information**

### **GENE ONTOLOGY (GO)** SIGNAL-ONTOLOGY (SigO) immune response **Immune Response** i- acute-phase response i- Allergic Response i- anaphylaxis Antigen Processing and Presentation i- antigen presentation i- B Cell Activation i- antigen processing • B Cell Development i- Complement Signaling i- cellular defense response • synonym complement activation i- cytokine metabolism i- cytokine biosynthesis i- Cytokine Response synonym cytokine production i- Immune Suppression i- Inflammation p- regulation of cytokine. i- Intestinal Immunity biosynthesis i- Leukotriene Response i- Leukotriene Metabolism \*Natural Killer Cell Response i-B-cell activation T Cell Activation i- B-cell differentiation i- T Cell Development i- T Cell Selection in Thymus i- B-cell proliferation i- cellular defense response i- T-cell activation i- activation of natural killer . . .

## **Ontologies with Overlapping Information**

- Use of multiple ontologies
  - custom-specific ontology + standard ontology
  - □ different views over same domain
  - overlapping domains
- Bottom-up creation of ontologies experts can focus on their domain of expertise

→ important to know the inter-ontology relationships

### GENE ONTOLOGY (GO)

### SIGNAL-ONTOLOGY (SigO)

i- Antigen Processing and Presentation

immune response
i- acute-phase response
i- anaphylaxis
i- antigen presentation
i- antigen processing
i- cellular defense response
i- cytokine metabolism
i- cytokine biosynthesis
synonym cytokine production
...
p- regulation of cytokine
biosynthesis
...
i- B-cell activation

i- B-cell differentiation i- B-cell proliferation i- cellular defense response ...

i- T-cell activationi- activation of natural killer cell activity

. . .

i- B Cell Development
i- Complement Signaling synonym complement activation
i- Cytokine Response
i- Immune Suppression
i- Inflammation
i- Intestinal Immunity
i- Leukotriene Response
i- Leukotriene Metabolism

Immune Response

i- Allergic Response

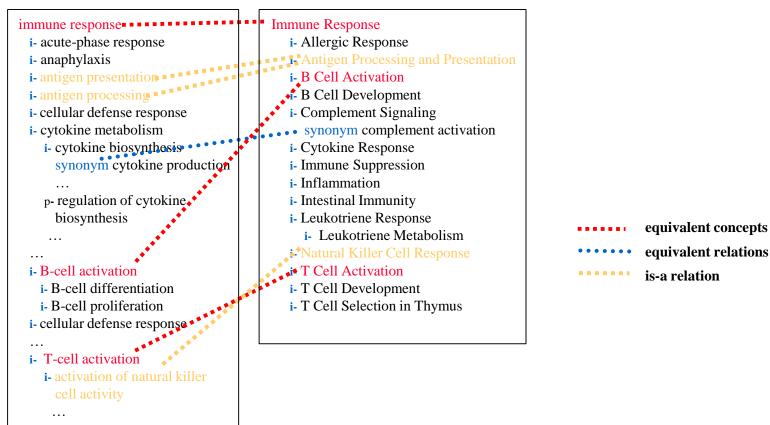
i- B Cell Activation

- i- Natural Killer Cell Response
- i- T Cell Activation
- i- T Cell Development
- i- T Cell Selection in Thymus

## **Ontology Alignment**

GENE ONTOLOGY (GO) S

SIGNAL-ONTOLOGY (SigO)

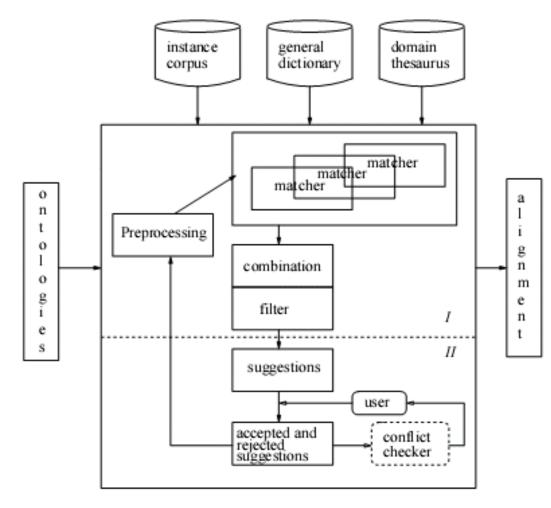


Defining the relations between the terms in different ontologies

### **Ontology Alignment**

- Ontology alignment
- Ontology alignment strategies
- Evaluation of ontology alignment strategies
- Ontology alignment challenges

### **An Alignment Framework**



alignment = a set of mappings

candidate mappings = mapping suggestions = mappings which have not been validated by a user

### Classification

- According to input
  - □ KR: OWL, UML, EER, XML, RDF, ...
  - □ components: concepts, relations, instance, axioms
- According to process
  - □ What information is used and how?
- According to output
  - □ 1-1, m-n
  - □ Similarity vs explicit relations (equivalence, is-a)
  - □ Confidence (similarity value)

# Preprocessing

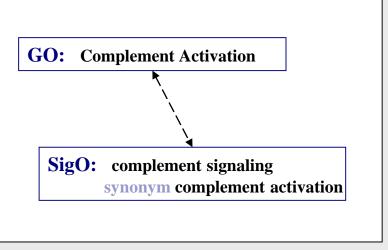
### Preprocessing

### For example,

- Selection of features
- Selection of search space

## Matchers

- Strategies based on linguistic matching
- Structure-based strategies
- Constraint-bas
- Instance-based
- Use of auxiliar



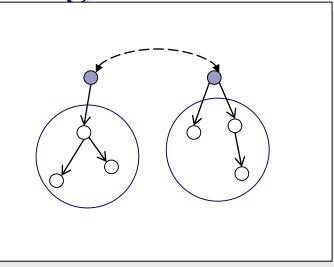
### Edit distance

- Number of deletions, insertions, substitutions required to transform one string into another
- $\Box$  aaaa  $\rightarrow$  baab: edit distance 2

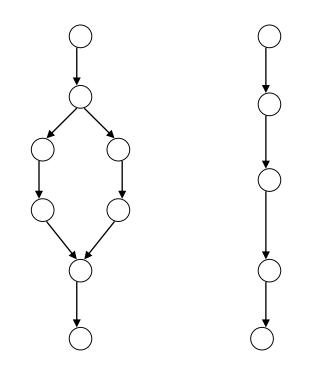
### N-gram

- N-gram : N consecutive characters in a string
- Similarity based on set comparison of n-grams
- aaaa : {aa, aa, aa}; baab : {ba, aa, ab}

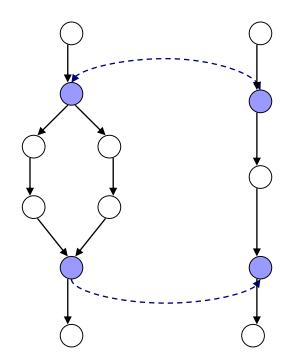
- Strategies based on linguistic matching
- Structure-based strategies
- Constraint-based
- Instance-based st
- Use of auxiliary



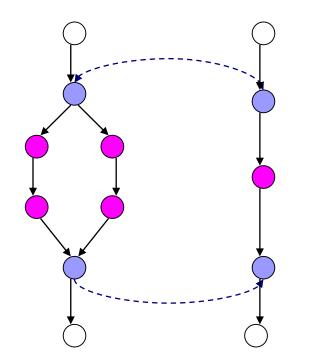
# Propagation of similarity valuesAnchored matching



# Propagation of similarity valuesAnchored matching

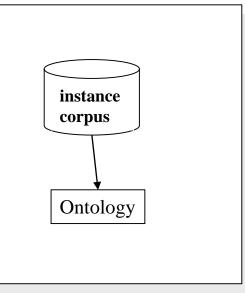


# Propagation of similarity valuesAnchored matching

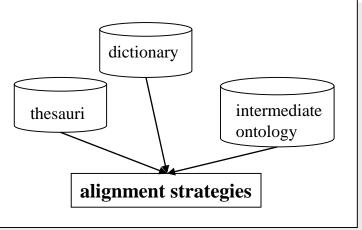


- Strategies based on linguistic matching
- Structure-based strategies
- Constraint-based approaches
- Instance-based strategies
- Use of auxiliary information

- Strategies based on linguisti
- Structure-based strategies
- Constraint-based approaches
- Instance-based strategies
- Use of auxiliary information



- Strategies based linguist
- Structure-based strategie
- Constraint-based approa
- Instance-based strategies
- Use of auxiliary information



### Use of WordNet

- Use WordNet to find synonyms
- Use WordNet to find ancestors and descendants in the isa hierarchy
- Use of Unified Medical Language System (UMLS)
  - Includes many ontologies
  - Includes many alignments (not complete)
  - Use UMLS alignments in the computation of the similarity values

# Combinations

### **Combination Strategies**

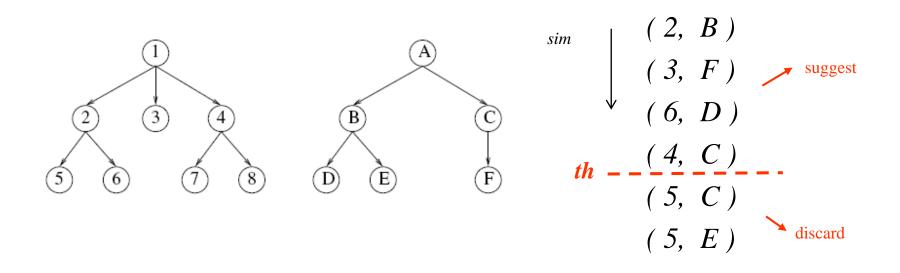
- Usually weighted sum of similarity values of different matchers
- Maximum of similarity values of different matchers



## **Filtering Techniques**

Threshold filtering

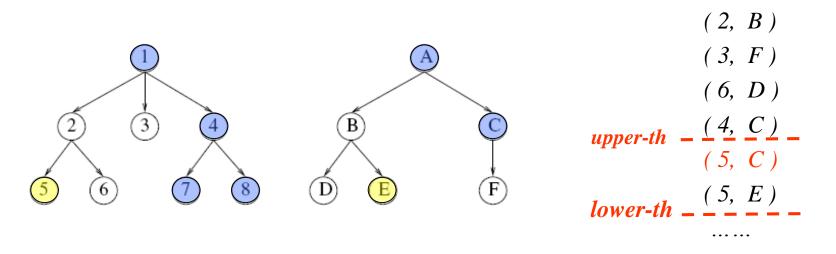
Pairs of concepts with similarity higher or equal than threshold are alignment suggestions



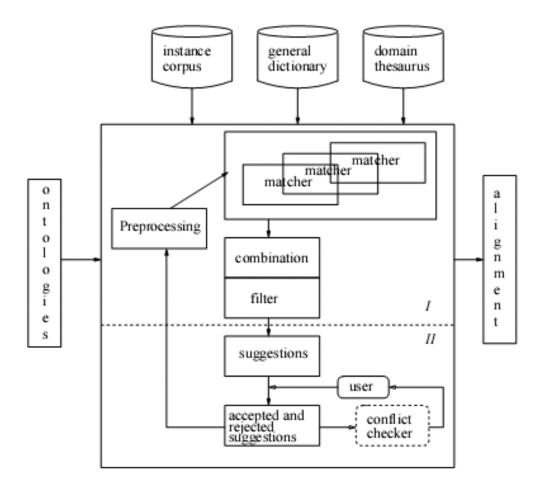
# **Filtering Techniques**

### Double threshold filtering

- (1) Pairs of concepts with similarity higher than or equal to **upper** threshold are alignment suggestions
- (2) Pairs of concepts with similarity between **lower** and **upper** thresholds are alignment suggestions if they make sense with respect to the structure of the ontologies and the suggestions according to (1)



### **An Alignment Framework**



	linguistic	structure	constraints	instances	auxiliary
ArtGen	name	parents, children		domain specific documents	WordNet
ASCO	name, label description	parents, children, siblings, path from root			WordNet
Chimaera	name	parents, children			
FCA-Merge	name			domain specific documents	
FOAM	name, label	parents, children	equivalence		
GLUE	name	neighborhood		instances	
HCONE	name	parents, children			WordNet
IF-Map				instances	a reference ontology
iMapper		leaf, non-leaf, children, related node	domain, range	instances	WordNet
OntoMapper		parents, children		documents	
(Anchor-) PROMPT	name	direct graphs			
SAMBO	name, synonym	is-a and part-of, descendants and ancestors		domain specific documents	WordNet, UMLS
S-Match	label	path from root	semantic relations codified in labels		WordNet

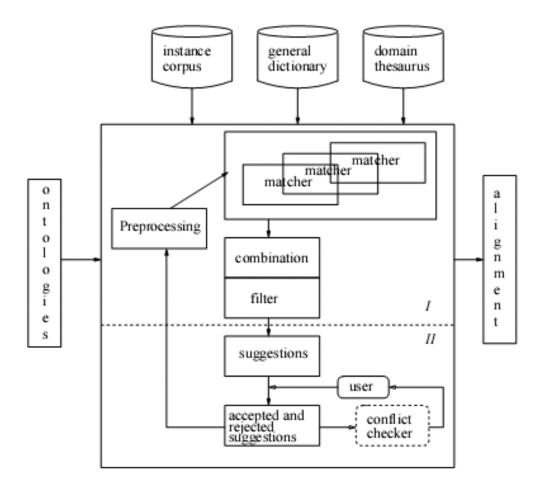
2006

### OAEI Anatomy Track 2007-2016\*

### Components

- Almost all systems implement preprocessing, matchers, combination, filtering components
- Debugging component and GUI rarely implemented
- Matching strategies:
  - Variety of string-based strategies
  - Most often string and structured-based strategies
- Use of background knowledge
  - Almost all systems use sources of background knowledge
- \* <u>https://doi.org/10.1186/s13326-017-0166-5</u>

### **An Alignment Framework**



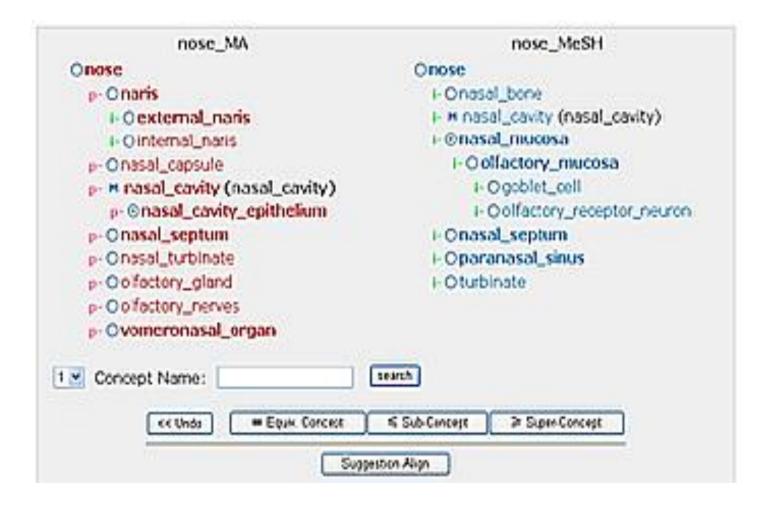
### Example: Alignment System SAMBO – matchers, combination, filter

			Align Concept in mouse and h	uman	
matchers	1.0NGram1.0TermBasic1.0TermWN1.0UMLSM1.0Naive Bayes	single threshold: double threshold:	0.6 • upper 0.6 lower 0.4 ©	weighted-sum combination    maximum-based combination	use preprocessed data
Start Com	utation Finish Computation	Interrupt Computation	interrupt at: 1000		

## **Example: Alignment System SAMBO** – suggestion mode

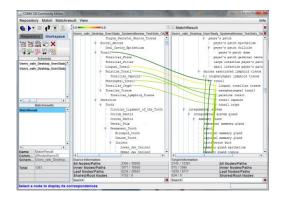
nose_MA	nose_MeSH		
nasal_cavity_epithelium definition: MA:0001324 synonym: nasal mucosa part-of: nasal_cavity	nasal_mucosa definition: MESH:A.04.531.520 synonym: nasal epithelium part-of:		
nasal_cavity_epithelium nasal_mucosa			
new name for the equivalent conce	pts:		
= Equiv. Concepts Sub-Concept	Super-Concept << Undo >> Skip to Next		

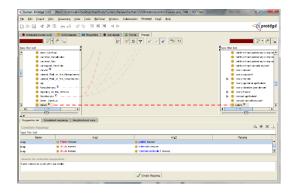
### Example: Alignment System SAMBO – manual mode



## **Ontology Alignment User Tasks**

Navigation & exploration **♦**Panning **♦**Zooming ♦Scrolling ♦Expanding *♦*Collapsing



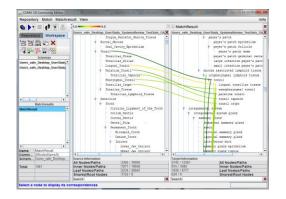


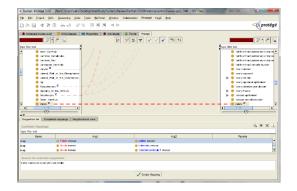
Aligning and Marging Biomedical Optiologies			
		concept Bresh	
	Mapping Can	lidate Details	
human		mouse	
Anatomic_Structure_System_or_Substance > 0 Body_Part	Î	mouse anatomy     organ system	í
<ul> <li>Organ_of_Special_Sense_Part</li> <li>Ear_Part</li> </ul>		© muscle © sensory organ	
⊷ ○ Internal_Ear     • ○ Semicircular_Canal     • ○ External_Canal		<ul> <li>skeletal system</li> <li>back</li> <li>abdomen/pelvis/perineum</li> </ul>	
O External_Canal     O Posterior_Semicircular_Canal     O Superior_Semicircular_Canal		<ul> <li>abdomen/pervis/perineum</li> <li>thorax</li> <li>bead</li> </ul>	
Antihelix     Additory Ossicle		© neck © lower back	
- O Malleus		upper back     abdomen	
⊢ <sup>©</sup> Stapes ⊢ <sup>©</sup> Common_Bony_Limb		© pelvis ⊙ chest	
O Processus_Cochleariformis     Spiral_Organ_of_Corti		© thoracle cavity © arm	
- O Organ_of_Corti - O Fenestra_Rotunda		O lower arm	

Printer's added to over	given and exactly breedy pr	The state the state has been done on a	inclusion faits versions ap	mand joint word d'ait stamptopper	interes erter bereiter bie ber
citre					antes .
and Shipteenter	Bonn a berr dartin the Bergtas	I som jass wel	**		is a gutta corpatione
2	Paker Barghao	Environ ( e.e. the effective	Fast Thip	Hitifigo phatas geal	cip nees orps
ratto + valla		1 mg		ab cold	
patieta conT	Joint of the Polish Souls-	Donal_Tegion_of_	Tes Tope		avail
Silver on participation	Acces Cause Jane Lacer Jane	Lener		and presmal mannel to	
	Consistential large Jury	Palviz_Bu	the Office Deternity Per		Autom
				feet pas affet dat bons	hirding jaks jel
preserve ashing mind	Period 21 June Metaasponsalmanal.	sist	111		
			Dawrity_Tat	wedgets and her an	an de jui
	Discriburgie Jahre Centrel and ad Join	4 Intel			nonalista rargiturpa
			Hadless long same	ap.fincles.pattere furthere	
colai ant	Selenal contraction of the lighted	Anatomic_Clusters_Sydem_	or_Gabdance		ate
M	entrances of the artifician Laren		finderst strikters		tanal terre cubo
e la mo	A NUMBER OF TAXABLE PARTY			tionis forment pit ad an a	
		Gasthon's Berg P		bind teb been	awaral
					Las MARRIAN CONTRACTOR
a photo sure	Synamucal	1.11	she alder being	ibone / seedi	
				Boresand digit hore	
		htereenfeat_De	× /		ing have - investored
Esagep cc	monotights have Arrightatte	warmebrall, Dolaran		hand been small been	11
		Verset Exten	utian arabone		house fag house
10.00	bang bans datianajam bars	Manual ad a latel, Davis v., P		he part with lawser Appen	leg bare d
forwal in					upper her
	2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	Savarse	System apper arts books a	in tere	
545.000	spatial assessments house	Statuta _ System_Pat		040703	id bern el gastass sami
auf stat 11	atasxetd boxa	Versibitan Nabilit Process		no să bore e	un effe al barr
chernid	//		Capity Serve		
	optopp + have seen anne	Metaovytic	Brianas Barr		
	intergranital bone	The state of the s			searching
soft palars	and particular books		Ber Ders		
			FAMP: Boar		has
soahiras	Efter ald_S	Wagelble Date			
contracts		angun Ventlan Sont 1	Tenu		/ 9
	Happy inter	mini Sanni Bom		Dera	staterys water
Frid	100	the man	sarn Long done Maintenn		Deally
stanced aniest	Secul Felder	In Berelfaine Bere	Ices of the Distorty	Pari Sera	
CREARCHES   chrt	and the	For Ftidate		Lower Edwards	V
and item	Terpost_Boss	Antello Rece Hars P	Retainers difficia	Calinyyan Calinyyan	broking. In
contral anti-christ			and the second sec		
con in	gingine-aljubit brierer Tabina	this gars		Hadaala' Ga	Te lan
e-reistk			Rear at the uppterfehred #	r burd Tamit Doos	chendrecombur
Taxas je	1	Flat Sour	Szarala		
(agghowhore	ater jabe st cl			lagezaid fine Cabels_Bon	

#### **Ontology Alignment User Tasks**

Align **♦**Validating ♦Evaluating **♦**Revising *♦*Debugging *♦*Sharing





Airgening 2nd Marging Riomedical Octobelogies		0
	concept Broth	
Mapping	Candidate Details	
human	mouse	
Anatomic_Structure_System_or_Substance	mouse anatomy	i i
- O Body_Part	🛛 organ system	
0 Organ_of_Special_Sense_Part	0 muscle	
⊨ <sup>©</sup> Ear_Part	sensory organ	
+ 🗢 Internal_Ear	skeletal system	
🖉 🔍 Semicircular_Canal	0 back	
+ 0 External_Canal	abdomen/pelvis/perineum	
Posterior_Semicircular_Canal	O thorax	
+ O Superior_Semicircular_Canal	() head	
i. C Antibelix	0 neck	
- O Auditory_Ossicle	O lower back	
Malleus	upper back	
+ 0 Incus	abdomen	
⊢ <sup>©</sup> Stapes	O pelvis	
Common_Bony_Limb	O chest	
+ O Processus_Cochleariformis	O thoracic cavity	
🖬 😑 Spiral_Organ_of_Corti	⊖ arm	
- Organ_of_Corti	© lower arm	
Fenestra_Rotunda	O upper arm	

Printer and interesting i have a set of each stationy	here out open pper appe	for stars property and	related in the version and	speak joint word d'ant squiptingper	beite bei eine an an ber ber
calarte					when
Total Shipticulture Bonne hart dastangage		lane jam wels			a ja e datta corpel jare
antaria antaria	a Barghain In	near (AA/16 KFs	Tast Thip	Hilligino phatas geal	icht hasferp
name walk		1 mg		at cut a	
polisis due ?	has beet-		Tes finger		avail
(II)		local_legiss_of_fo		and the second se	Unit plate or or left
Status and and a status of the	Take Acre			perina della d	ins of fact
Catteretes 2 Jarks		Palote_Buth	White Distanty Per		
			1/2 *	feet phe affert eigt bone	hirefint jala jel
pressy salay friend an out of Salat " Mathematic	and a large at a start of the		Awaite Tat		at de la
Sabider shell Jone By St	ba .		and the second sec	wedging and how as	
Discontent date take	whet being design for				INVITED IN THE PARTY
			Hadless long same	apdivelageallare faithers	
cost and setting constrain	Art all the lord hannes	relo_Chuckasa_System_or			ald
Verelail.cont			lieb long I	torpathene over	-lamat harn - ander
vitres Webstancetering Line as the	the Jurist		faulish erghers		1
- III III		thona Been Pa		timit fparit gbatarn	awter
		and any a		fallons hindlich bern	I BILL MARRIAN CARACTER
a photo sure					
	Spartweis- Jard		staulder bereiten	a poss	
	7	international from		Berrhand digit horr	ing have - investored
gaugegeet monotight have		Sebral, Colaras	dan arabasa	hand bong ment town there .	at bard
femoral base liamanipite to	No.	nut Liting	ian arrows		
terrightens o sare	-Me	tales about System, Per		he paravah hann Appe	rleg bare /
Rowal house	diam'r	Barbrie St			MOONT has
		Over fat	den apper arts banes /		
savous spatial transmission boost	Spenal Versibilat				id bern ei gattos sarsi
stored paratalities prayers	2218		Canada Parter	eesd fase	an eile al b an
	Metary		value seas		
annin #\$\$ \$777 Chata wear ann		then	Sciences Barr		
interpanetal bone		(Dal)			sequences
self palars		0.0	IT_DITE		
			Calling Room		has
support of the second	Efer eld_Bingelabs	Dans			
	Mailla Savenne VI	unitan Bont	Ferrus		harmer 1
Baset, inter			Lorg dere Meldens	C.C.F.	
	ternt religible Beret				fente
	atta atta	Corr. Philans	cos of the Deliverty	Part Bars	
comancellas   chut	and has	1 1 /	Sermal de	Come Induceda	In
100		Rate Hare PH	Bananna Hills als	Caligoram	h rolling.
conical web che at ginging al john					
	rferer, Tabinate den		NV.	Hudsala De	ter they
ay mphysike			tear at the uppersonal t	Tamat Done	chendrecembur
those print		2710 5	rapula	Trapezoid Sone Th reason Barn	
	and the second second	mat_loss Meat He		ta Cheer Los	

#### **Ontology Alignment**

- Ontology alignment
- Ontology alignment strategies
- Evaluation of ontology alignment strategies
- Ontology alignment challenges

#### **Ontology Alignment Evaluation Initiative (OAEI)**

- Since 2004
- Evaluation of systems
- Different tracks (2017)
  - anatomy, conference, large biomedical ontologies, disease and phenotype
  - multilingual: multifarm (9 languages)
  - □ process model
  - □ interactive
  - □ instance
  - □ link discovery for spatial data

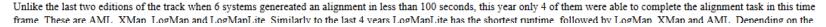
Evaluation

Ontology Alignment Evaluation Initiative - OAEI-2016 Campaign

#### **Results for OAEI 2016 - Anatomy track**

In the following, we analyze all participating systems that could generate an alignment. The listing comprises of 13 entries. As previous years some of the s different versions. LogMap participated with LogMap, LogMapBio and a lightweight version LogMapLite that uses only some core components. Similarly with two versions, DKP-AOM and DKP-AOM-Lite. There are a number of systems which participate in the anatomy track for the first time. These are Alii LYAM. While every year there are several new participants in this track, there are also systems participating for several years in a row. LogMap is a constai and XMap joined the track in 2013. DKP-AOM, Lily and CroMatcher participate for the second year in a row in this track. Lily participated in the track ba participated in 2013 but did not produce an alignment within the given time frame back then. For more details, we refer the reader to the papers presenting have 10 different systems (not counting different versions) which generated an alignment.

Matcher	Runtime	Size	Precision	F-Measure	Recall	Recall+	Coherent
合 - 탄	순 🖑	순 문	순 문	습 문	- 10 - 10 - 10 - 10 - 10 - 10 - 10 - 10	순 문	순 문
AML	47	1493	0.95	0.943	0.936	0.832	+
LPHOM	1601	1555	0.709	0.718	0.727	0.497	-
Lily	272	1382	0.87	0.83	0.794	0.515	-
DKP-AOM	379	207	0.99	0.238	0.135	0.0	+
ХМар	45	1413	0.929	0.896	0.865	0.647	+
DKP-AOM- Lite	372	207	0.99	0.238	0.135	0.0	+
FCA_Map	117	1361	0.932	0.882	0.837	0.578	-
Alin	306	510	0.996	0.501	0.335	0.0	+
LogMap	24	1397	0.918	0.88	0.846	0.593	+
LogMapBio	758	1531	0.888	0.892	0.896	0.728	+
LYAM	799	1539	0.863	0.869	0.876	0.682	-
LogMapLite	20	1147	0.962	0.828	0.728	0.288	-
CroMatcher	573	1442	0.949	0.925	0.902	0.773	-
StringEquiv	-	946	0.997	0.766	0.622	0.000	-



#### **Evaluation Measures**

Precision:

# correct mapping suggestions

# mapping suggestions

Recall:

# correct mapping suggestions

# correct mappings

• F-measure: combination of precision and recall

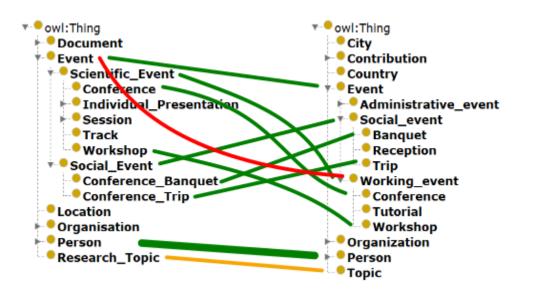
$$J(A,B)=rac{|A\cap B|}{|A\cup B|}$$

http://oaei.ontologymatching.org/2007/results/anatomy/

	Reference	AOAS	Sambo	ASMOV	RiMOM	LabelEq.
Reference	-	0.756	0.687	0.598	0.316	0.600
AOAS	0.756	-	0.738	0.679	0.308	0.706
Sambo	0.687	0.738	-	0.573	0.302	0.576
ASMOV	0.598	0.679	0.573	-	0.282	0.681
RiMOM	0.316	0.308	0.302	0.282	-	0.277
LabelEq.	<mark>0.600</mark>	0.706	0.576	0.681	0.277	-
Falcon-AO	0.578	0.657	0.562	0.620	0.272	0.814
ТахоМар	0.464	0.484	0.443	0.461	0.264	0.515
AgreementM.	0.423	0.435	0.401	0.408	0.252	0.444
Prior+	0.420	0.416	0.407	0.376	0.258	0.404
Lily	0.349	0.334	0.326	0.297	0.224	0.308
X-Som	0.205	0.239	0.206	0.232	0.089	0.333
DSSim	0.109	0.115	0.103	0.106	0.065	0.121

Precision & recall do not show:

- Often & rarely found mappings
- Correct & incorrect mappings
- (Dis)Agreement between tools
- Regions with the same or different number of mappings
- Fine-grained evaluation



#### In practice:

- Write custom scripts
  - □ Error-prone
  - $\Box$  New question = new script
  - □ ...
- Explore their results
   in (sortable) tables ...

## How to write queries about unexpected findings?

#### List all mappings computed by AML in the Anatomy track of the OAEI?

- 1 SELECT ma.tool, ma.year, cl1.label, cl2.label, ma.relation,
- 2 ma.measure, m.relation, m.trivial, m.reference
- 3 FROM oaei\_anatomy\_track.matchingsInAlignment ma
- 4 INNER JOIN conceptLabels cl1 on ma.source = cl1.name
- 5 INNER JOIN conceptLabels cl2 on ma.target = cl2.name 6 INNER JOIN matching m on m.source=ma.source and m.ta
  - INNER JOIN matching m on m.source=ma.source and m.target=ma.target
  - where tool ='AML';

tool	year	label	label	relation	measure	relation	trivial	reference
AML	2013	corpus callosum	Corpus_Callosum	=	0.883599997	=	1	1
AML	2013	cochlea	Cochlea	=	0.883599997	=	1	1
AML	2013	stomach lesser curvature	Lesser_Curvature	=	0.744212985	=	0	1
AML	2013	intercostales internus	Internal_Intercostal_Muscle	=	0.789600015	=	0	0
AML	2013	lateral ventricle ependyma	Lateral_Ventricle_Ependyma	=	0.883599997	=	1	1
AML	2013	mylohyoid	Mylohyoid	=	0.883599997	=	1	1
AML	2013	right lung alveolus	Right_Lung_Alveolus	=	0.971960008	=	1	1
AML	2013	rib 7	Rib_7	=	0.780200005	=	1	1
AML	2013	hand digit 4 phalanx	Hand_Digit_4_Phalanx	=	0.971960008	=	1	1
AML	2013	sphincter pupillae	Sphincter_Pupillae_Muscle	=	0.783222914	=	0	1
AML	2013	hand digit skin	Hand_Digit_Skin	=	0.883599997	=	1	1
AML	2013	urinary bladder wall	Bladder_Wall	=	0.653980553	=	0	1
AML	2013	iliacus	Iliacus	=	0.883599997	=	1	1
AML	2013	thyroid gland follicle	Thyroid_Gland_Follicle	=	0.883599997	=	1	1
AML	2013	metatarsal bone digit 1	Metatarsal_Bone_Digit_1	=	1.000000000	=	1	1
AML	2013	pectoralis minor	Pectoralis_Minor	=	0.971960008	=	1	1
AML	2013	retina layer	Retina_Layer	=	0.883599997	=	1	1
AML	2013	lacrimal sac	Lacrimal_Sac	=	0.883599997	=	1	1
лмі	2013	hand digit 5	Hand Dinit 5	-	0 883500007	-	1	1

#### **Evaluation Use Cases**

- Selecting, combining & fine tuning tools
- Matchers development
- Ontology alignment evolution
- Validating & debugging of ontology alignments and RA
- Collaborative ontology alignment

#### Alignment Cubes Towards Interactive Visual Exploration and Evaluation of Multiple Ontology Alignments

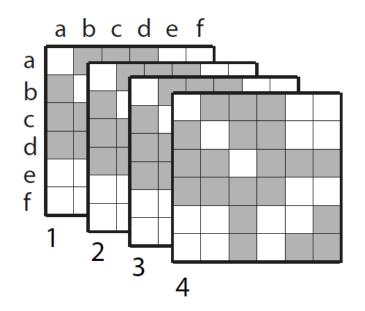
Valentina Ivanova, Benjamin Bach, Emmanuel Pietriga, Patrick Lambrix





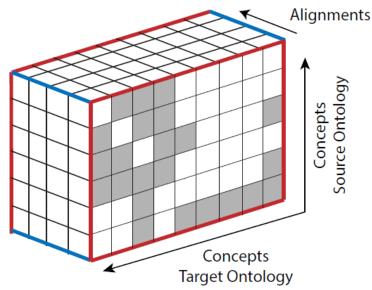
(naín\_

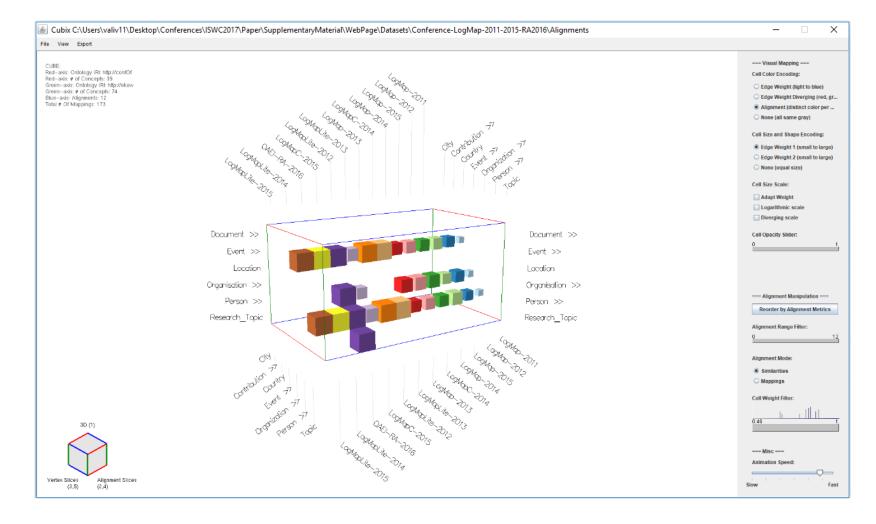
Submitted to ISWC 2017



An alignment is represented by a matrix

Stacking several matrices together



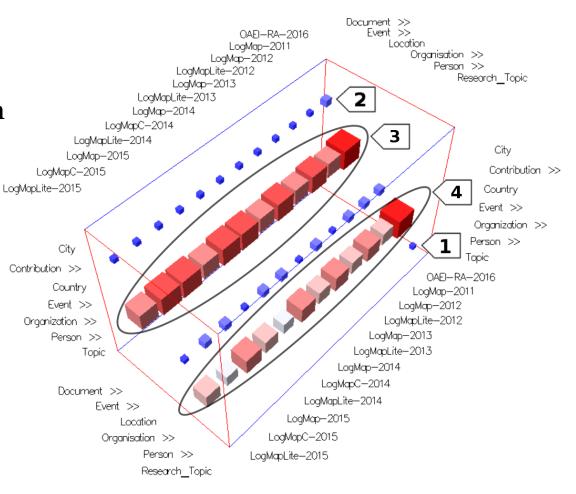


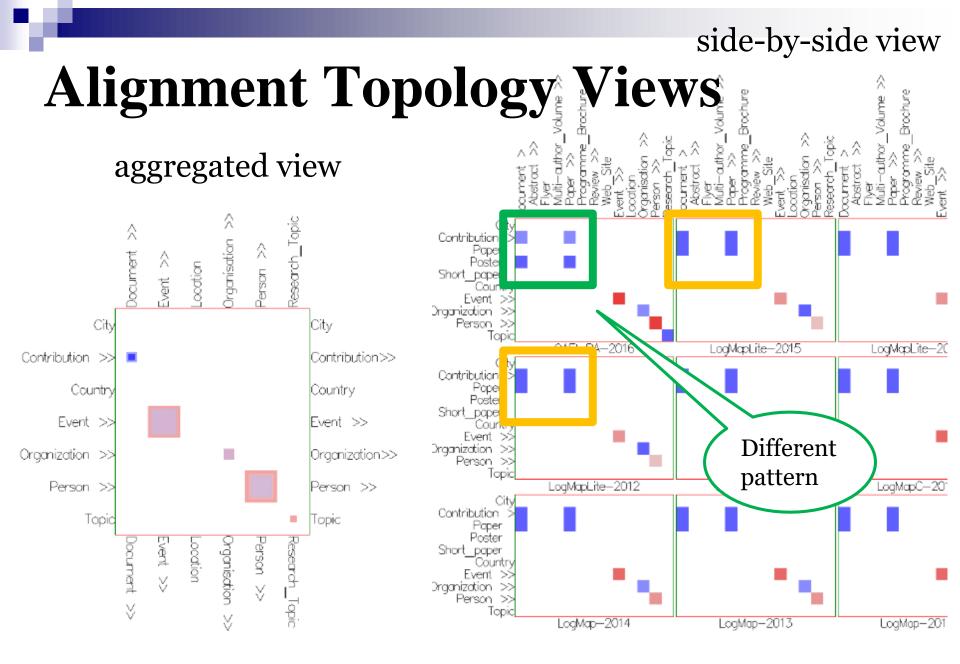
#### 3D view:

- drives *initial explorat*ion
- supports identifying regions of interest
- can possibly yield some high-level insights

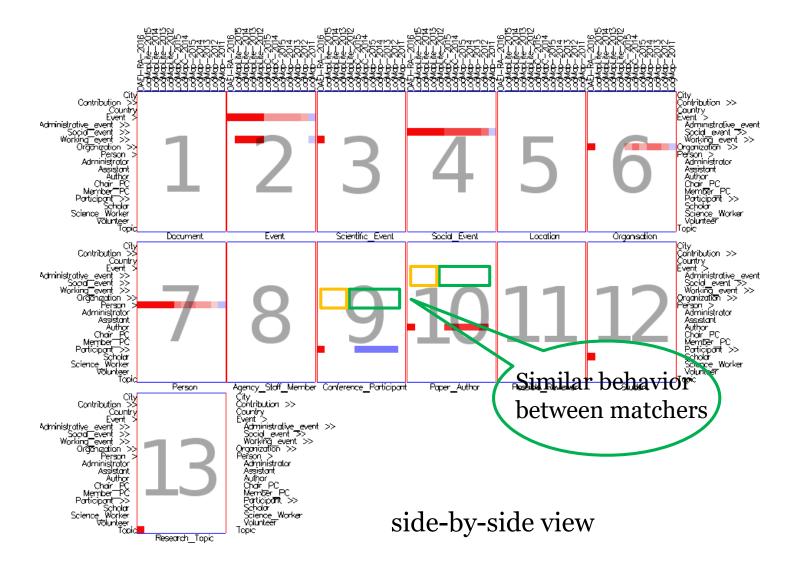
Paired aggregated & detailed side-by-side views:

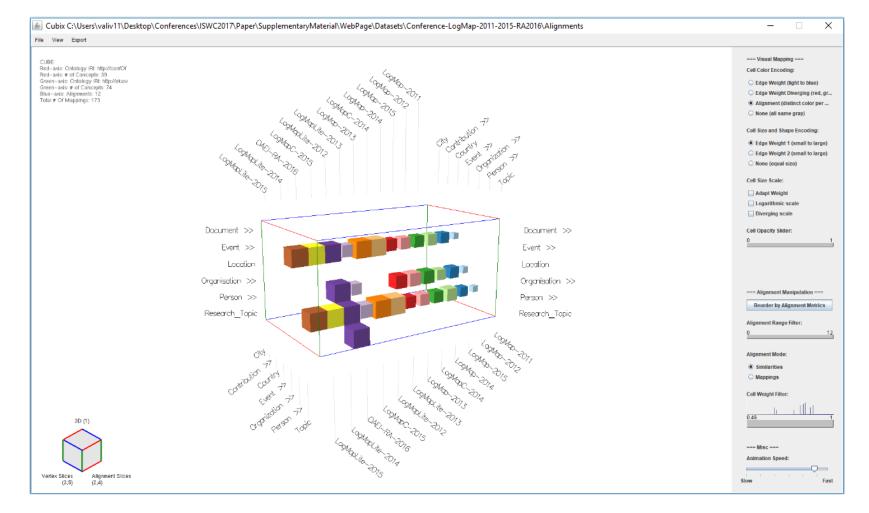
- Alignment topology
- Concepts network





#### **Concepts Network Views**





#### http://www.ida.liu.se/~patla00/research/AlignmentCubes/

#### **Ontology Alignment**

- Ontology alignment
- Ontology alignment strategies
- Evaluation of ontology alignment strategies
- Ontology alignment challenges

#### Algorithmic approaches

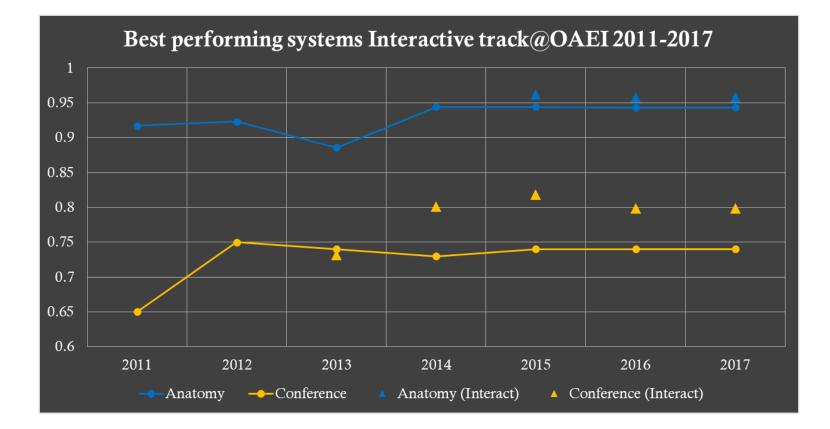
- Large-scale matching evaluation
- Efficiency of matching techniques
- Matching with background knowledge
- Matcher selection, combination and tuning
- User-engaging approaches
  - User involvement
  - Explanation of matching results
  - Social and collaborative matching
  - Alignment management: infrastructure and support

- Large-scale matching evaluation
- Efficiency of matching techniques
  - parallellization
  - distribution of computation
  - approximation of matching results (not complete)
  - modularization of ontologies
  - optimization of matching methods

Matching with background knowledge

- partial alignments
- reuse of previous matches
- □ use of domain-specific corpora
- use of domain-specific ontologies
- Matcher selection, combination and tuning
   recommendation of algorithms and settings

#### **Interactive Ontology Alignment**



- User involvement
   visualization
   user feedback
- Explanation of matching results
- Social and collaborative matching
- Alignment management: infrastructure and support

#### **Further Reading**

Starting points for further studies

#### **Further Reading Ontology Alignment**

- <u>http://www.ontologymatching.org</u>
   (plenty of references to articles and systems)
- Ontology alignment evaluation initiative: <u>http://oaei.ontologymatching.org</u> (home page of the initiative)
- Euzenat, Shvaiko, *Ontology Matching*, Springer, 2007.

## Further Reading OAEI – Experience Papers

- Dragisic Z, Ivanova V, Li H, Lambrix P, Experiences from the Anatomy track in the Ontology Alignment Evaluation Initiative. J Biomedical Semantics 8:56, 2017.
- Shvaiko P, Euzenat J., Ontology matching: state of the art and future challenges. IEEE Trans Knowl Data Eng. 25(1):158–76, 2013.
- Euzenat J, Meilicke C, Stuckenschmidt H, Shvaiko P, Trojahn C, Ontology alignment evaluation initiative: Six years of experience. J Data Semant. XV:158–92, 2011.
- Cheatham M, Hitzler P, String similarity metrics for ontology alignment. In: The Semantic Web - ISWC 2013, 12th International Semantic Web Conference, LNCS 8219, 294–309, 2013.

#### **Further Reading Ontology Alignment**

Systems at LiU / IDA / ADIT

 Lambrix, Tan, SAMBO – a system for aligning and merging biomedical ontologies, *Journal of Web Semantics*, 4(3):196-206, 2006.

(description of the SAMBO tool and overview of evaluations of different matchers)

Lambrix, Tan, A tool for evaluating ontology alignment strategies, *Journal on Data Semantics*, VIII:182-202, 2007.
 (description of the KitAMO tool for evaluating matchers)

- Lambrix P, Kaliyaperumal R, <u>A Session-based Ontology Alignment Approach</u> <u>enabling User Involvement</u>, *Semantic Web Journal* 8(2):225-251, 2017.
- Ivanova V, Bach B, Pietriga E, Lambrix P, <u>Alignment Cubes: Towards Interactive</u> <u>Visual Exploration and Evaluation of Multiple Ontology Alignments</u>, 16th International Semantic Web Conference, 400-417, 2017.

#### Further Reading Ontology Alignment

 Chen, Tan, Lambrix, Structure-based filtering for ontology alignment, *IEEE WETICE workshop on semantic technologies in collaborative applications*, 364-369, 2006.

(double threshold filtering technique)

- Tan, Lambrix, A method for recommending ontology alignment strategies, *International Semantic Web Conference*, 494-507, 2007.
   Ehrig, Staab, Sure, Bootstrapping ontology alignment methods with APFEL, *International Semantic Web Conference*, 186-200, 2005.
   Mochol, Jentzsch, Euzenat, Applying an analytic method for matching approach selection, *International Workshop on Ontology Matching*, 2006.
   (recommendation of alignment strategies)
- Lambrix, Liu, Using partial reference alignments to align ontologies, *European* Semantic Web Conference, 188-202, 2009.
   (use of partial alignments in ontology alignment)

# Ontology Debugging and Completion

#### **Defects in Ontologies**

- Syntactic defects
  - □ E.g. wrong tags or incorrect format
- Semantic defects
  - E.g. unsatisfiable concepts, incoherent and inconsistent ontologies
- Modeling defects
  - □ E.g. wrong or missing relations

#### **Example: Incoherent Ontology**

#### Example: DICE ontology

 Brain ⊑ CentralNervousSystem п BodyPart п ∃systempart.NervousSystem п ∃ region.HeadAndNeck п ∀region.HeadAndNeck

A brain is a central nervous system and a body part which has a system part that is a nervous system and that is in the head and neck region.

■ CentralNervousSystem ⊑ NervousSystem

A central nervous system is a nervous system.

■ BodyPart ⊑¬NervousSystem

Nothing can be at the same time a body part and a nervous system.

#### Slide from G. Qi

### **Example: Inconsistent Ontology**

#### Example from Foaf:

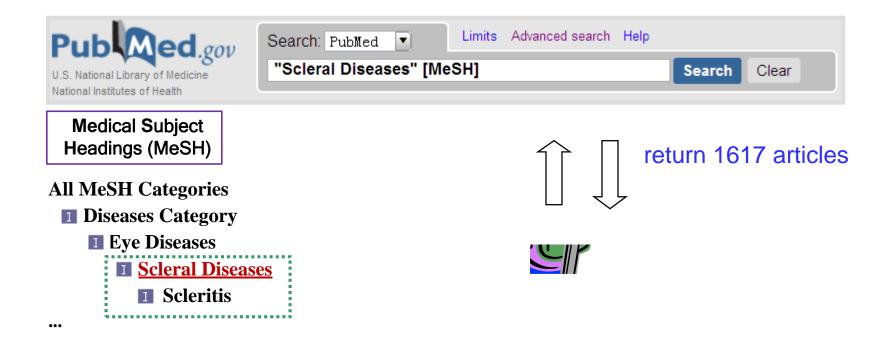
- Person(timbl)
- Homepage(timbl, <u>http://w3.org/</u>)
- Homepage(w3c, <u>http://w3.org/</u>)
- Organization(w3c)
- InverseFunctionalProperty(Homepage)
- DisjointWith(Organization, Person)
- Example from OpenCyc:
  - ArtifactualFeatureType(PopulatedPlace)
  - ExistingStuffType(PopulatedPlace)
  - DisjointWith(ExistingObjectType,ExistingStuffType)
  - ArtifactualFeatureType 
     ExistingObjectType

## **Example: Missing is-a Relations**

- In 2008 Ontology Alignment Evaluation Initiative (OAEI) Anatomy track, task 4
  - Ontology MA : Adult Mouse Anatomy Dictionary (2744 concepts)
  - Ontology NCI-A : NCI Thesaurus anatomy (3304 concepts)
  - □ 988 mappings between MA and NCI-A
    - 121 missing is-a relations in MA
    - 83 missing is-a relations in NCI-A

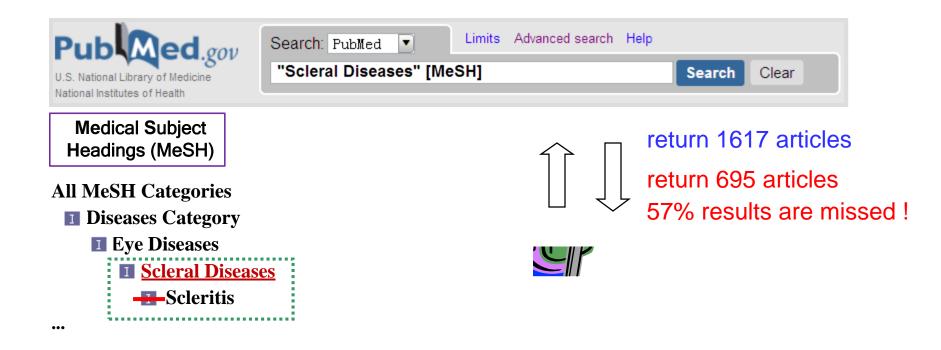
## **Influence of Missing Structure**

Ontology-based querying.



## **Influence of Missing Structure**

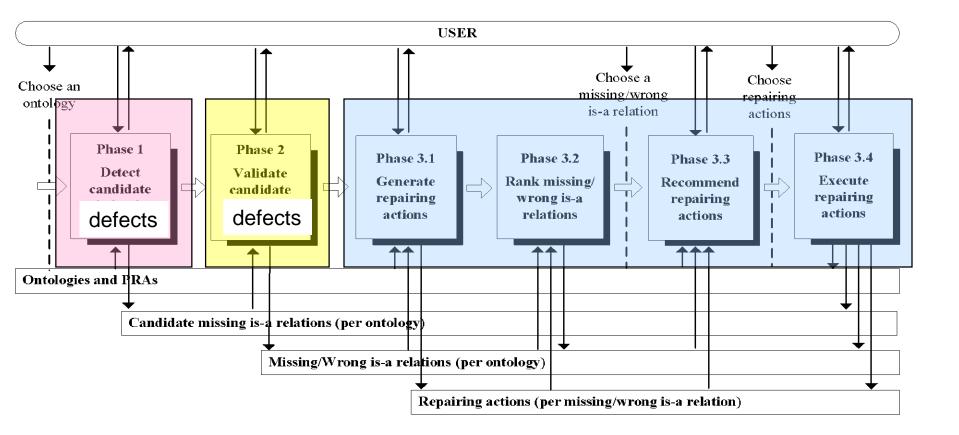
#### Incomplete results from ontology-based queries



# Defects in Ontologies and Ontology Networks

- Ontologies and ontology networks with defects, although often useful, also lead to problems when used in semantically-enabled applications.
- → Wrong conclusions may be derived or valid conclusions may be missed.

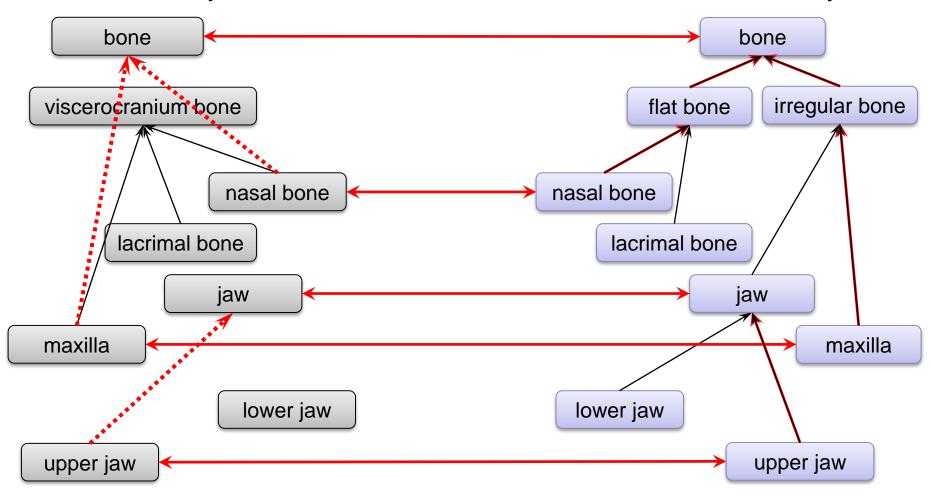
#### **Overview of Debugging Approach**



# **Example: Missing Structure**

mouse anatomy

NCI anatomy

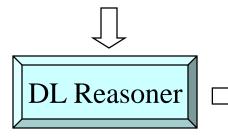


# Debugging semantic defects

# **Example: Incoherent Ontology**

Consider the following TBox  $T^*$ , where A, B and C are primitive and  $A_1, \ldots, A_7$  defined concept names:

 $\begin{array}{ll} ax_1:A_1 \stackrel{.}{\sqsubseteq} \neg A \sqcap A_2 \sqcap A_3 & ax_2:A_2 \stackrel{.}{\sqsubseteq} A \sqcap A_4 \\ ax_3:A_3 \stackrel{.}{\sqsubseteq} A_4 \sqcap A_5 & ax_4:A_4 \stackrel{.}{\boxminus} \forall s.B \sqcap C \\ ax_5:A_5 \stackrel{.}{\sqsubseteq} \exists s. \neg B & ax_6:A_6 \stackrel{.}{\sqsubseteq} A_1 \sqcup \exists r.(A_3 \sqcap \neg C \sqcap A_4) \\ ax_7:A_7 \stackrel{.}{\sqsubseteq} A_4 \sqcap \exists s. \neg B \end{array}$ 



The ontology is incoherent!

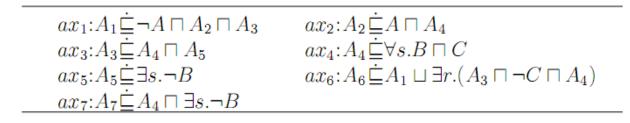
The set of unsatisfiable concepts are :  $\{A_1, A_3, A_6, A_7\}$ .



What are the root causes of these defects?

# **Explain the Semantic Defects**

• We need to identify the sets of axioms which are necessary for causing the logic contradictions.



• For example, for the unsatisfiable concept " $A_1$ ", there are two sets of axioms.

$$ax_1:A_1 \sqsubseteq \neg A \sqcap A_2 \sqcap A_3 ax_2:A_2 \sqsubseteq A \sqcap A_4$$

$$ax_{1}:A_{1} \sqsubseteq \neg A \sqcap A_{2} \sqcap A_{3}$$
$$ax_{3}:A_{3} \sqsubseteq A_{4} \sqcap A_{5}$$
$$ax_{4}:A_{4} \sqsubseteq \forall s.B \sqcap C$$
$$ax_{5}:A_{5} \sqsubseteq \exists s. \neg B$$

#### Minimal Unsatisfiability Preserving Sub-TBoxes (MUPS)

**Definition 1** Let A be a concept which is unsatisfiable in a TBox  $\mathcal{T}$ . A set  $\mathcal{T}' \subseteq \mathcal{T}$  is a *minimal unsatisfiability-preserving sub-TBox (MUPS)* of  $\mathcal{T}$  if

- A is unsatisfiable in  $\mathcal{T}'$ , and
- A is satisfiable in every sub-TBox  $\mathcal{T}'' \subset \mathcal{T}'$ .

We will abbreviate the set of MUPS of  $\mathcal{T}$  and A by  $mups(\mathcal{T}, A)$ .  $mups(\mathcal{T}^*, A_1) = \{\{ax_1, ax_2\}, \{ax_1, ax_3, ax_4, ax_5\}\}$ 

• The MUPS of an unsatisfiable concept imply the solutions for repairing.

 $\rightarrow$  Remove at least one axiom from each axiom set in the MUPS

# Example

$$mups(\mathcal{T}^*, A_1) = \{\{ax_1, ax_2\}, \{ax_1, ax_3, ax_4, ax_5\}\}$$
$$mups(\mathcal{T}^*, A_3) = \{\{ax_3, ax_4, ax_5\}\}$$
$$mups(\mathcal{T}^*, A_6) = \{\{ax_1, ax_2, ax_4, ax_6\}, \\ \{ax_1, ax_3, ax_4, ax_5, ax_6\}\}$$
$$mups(\mathcal{T}^*, A_7) = \{\{ax_4, ax_7\}\}$$

• Possible ways of repairing all the unsatisfiable concepts in the ontology:

$$\{ax_1, ax_3, ax_4\}$$



How to represent all these possibilities?

#### Minimal Incoherence Preserving Sub-TBox (MIPS)

**Definition 2** Let  $\mathcal{T}$  be an incoherent TBox. A TBox  $\mathcal{T}' \subseteq \mathcal{T}$  is a minimal incoherencepreserving sub-TBox (MIPS) of  $\mathcal{T}$  if

- $\mathcal{T}'$  is incoherent, and
- every sub-TBox  $\mathcal{T}'' \subset \mathcal{T}'$  is coherent.

$$\begin{split} mups(\mathcal{T}^*, A_1) &= \{ \{ax_1, ax_2\}, \{ax_1, ax_3, ax_4, ax_5\} \} \\ mups(\mathcal{T}^*, A_3) &= \{ \{ax_3, ax_4, ax_5\} \} \\ mups(\mathcal{T}^*, A_6) &= \{ \{ax_1, ax_2, ax_4, ax_6\}, \\ \{ax_1, ax_3, ax_4, ax_5, ax_6\} \} \\ mups(\mathcal{T}^*, A_7) &= \{ \{ax_4, ax_7\} \} \end{split}$$

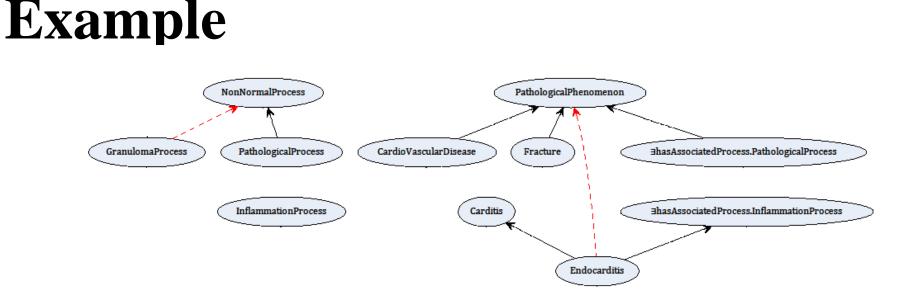
We will abbreviate the set of MIPS of  $\mathcal{T}$  by  $mips(\mathcal{T})$ . For  $\mathcal{T}^*$  we get three MIPS:

$$mips(\mathcal{T}^*) = \{ \{ax_1, ax_2\}, \{ax_3, ax_4, ax_5\}, \{ax_4, ax_7\} \}$$

A possible repairing is  $\{ax_i\} \cup \{ax_j\} \cup \{ax_k\}$ , where

- $ax_i \in \{ax_1, ax_2\}$
- $ax_j \in \{ax_3, ax_4, ax_5\}$
- $ax_k \in \{ax_4, ax_7\}$

# Completing the is-a structure of ontologies



#### **Repairing actions:**

{Endocarditis  $\doteq$  PathologicalPhenomenon, GranulomaProcess  $\doteq$  NonNormalProcess}

{Carditis  $\doteq$  CardioVascularDisease, GranulomaProcess  $\doteq$  PathologicalProcess}

{Carditis  $\sqsubseteq$  Fracture, GranulomaProcess  $\doteq$  NonNormalProcess}

# **Description Logic EL**

#### Concepts

Atomic concept	Α
Universal concept	T
Intersection of concepts	СпD
Existential restriction	∃r.C

#### Terminological axioms: equivalence and subsumption

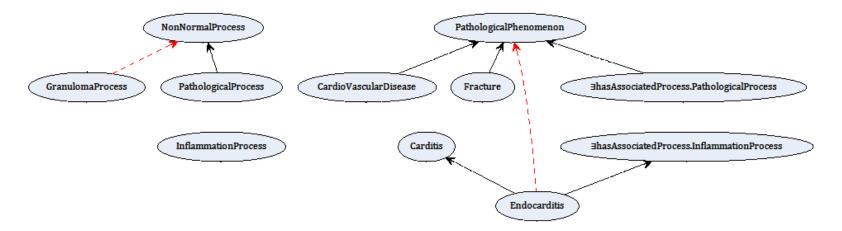
# **Generalized Tbox Abduction Problem – GTAP(T,C,Or,M)**

- Given
  - □**T** a Tbox in EL
  - □C- a set of atomic concepts in T
  - $\Box M = \{Ai \subseteq Bi\}_{i=1..n} and \forall i:1..n: Ai, Bi \in \boldsymbol{C}$
  - $\Box \text{ Or: } \{\text{Ci} \subseteq \text{Di} \mid \text{Ci, } \text{Di} \in \textbf{C}\} \rightarrow \{\text{true, false}\}$

Find

$$\label{eq:second} \begin{split} \square \ S &= \{ E_i \subseteq F_i \}_{i=1..k} \, \text{such that} \\ \forall \ i:1..k: \, E_i, \, F_i \in \boldsymbol{C} \, \, \text{and} \, \, Or(E_i \subseteq F_i) = true \\ \text{and} \, T \, U \, S \, \text{is consistent and} \, T \, U \, S \mid = M \end{split}$$

# **Example: GTAP**



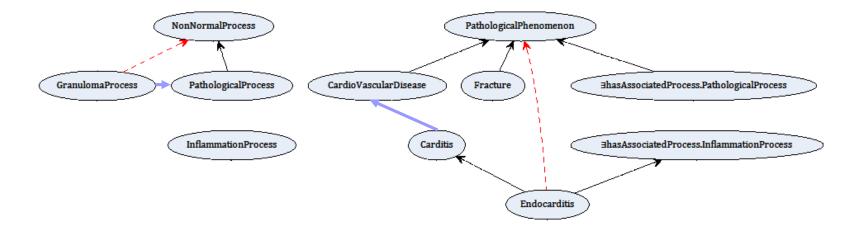
 $C = \{$  GranulomaProcess, CardioVascularDisease, PathologicalPhenomenon, Fracture, Endocarditis, Carditis, InflammationProcess, PathologicalProcess, NonNormalProcess $\}$ 

 $T = \{ \text{ GranulomaProcess} \sqsubseteq \top, \text{ hasAssociatedProcess} \trianglerighteq \top \times \top, \\ \text{CardioVascularDisease} \trianglerighteq \text{ PathologicalPhenomenon, Fracture} \trianglerighteq \text{ PathologicalPhenomenon,} \\ \exists \text{hasAssociatedProcess.PathologicalProcess} \trianglerighteq \text{ PathologicalPhenomenon,} \\ \text{Endocarditis} \trianglerighteq \text{ Carditis, Endocarditis} \trianglerighteq \exists \text{hasAssociatedProcess.InflammationProcess,} \\ \text{PathologicalProcess} \end{Bmatrix} \\ \end{cases}$ 

 $M = \{$  Endocarditis  $\stackrel{.}{\sqsubseteq}$  PathologicalPhenomenon, GranulomaProcess  $\stackrel{.}{\sqsubseteq}$  NonNormalProcess  $\}$ 

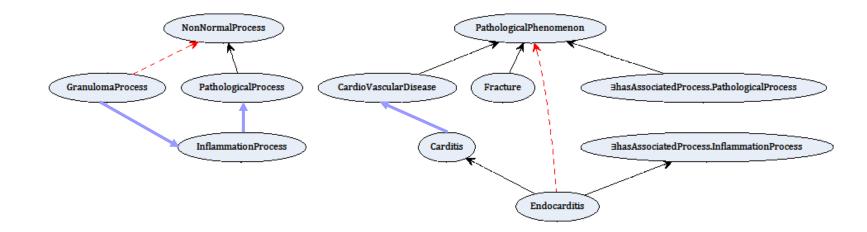
## **Preference Criteria**

#### There can be many solutions for GTAP



# **Preference Criteria**

#### There can be many solutions for GTAP



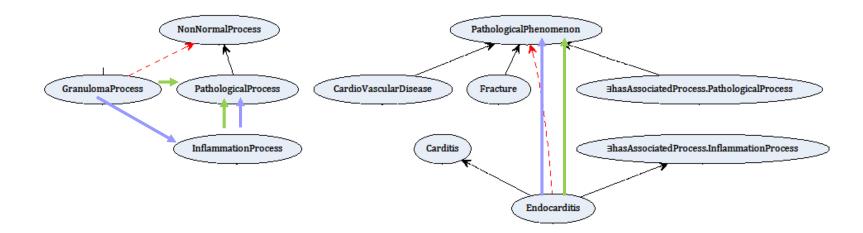
Not all are equally interesting.

### **More Informative**

- Let S and S' be two solutions to GTAP(T,C,Or,M). Then,
- S is more informative than S' iff  $\mathbf{T} \cup S \models S'$  but not  $\mathbf{T} \cup S' \models S$
- S is equally informative as S' iff  $\mathbf{T} \cup S \models S'$  and  $\mathbf{T} \cup S' \models S$

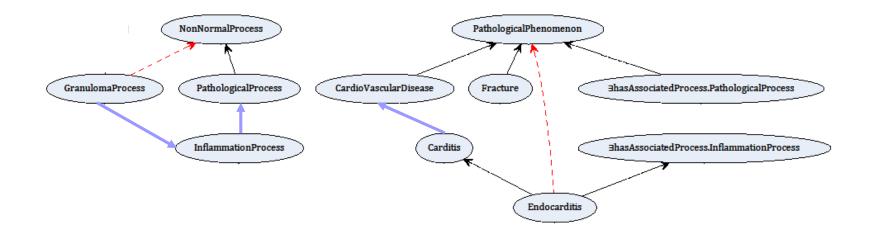
# **More Informative**

#### Blue' solution is more informative than 'green' solution



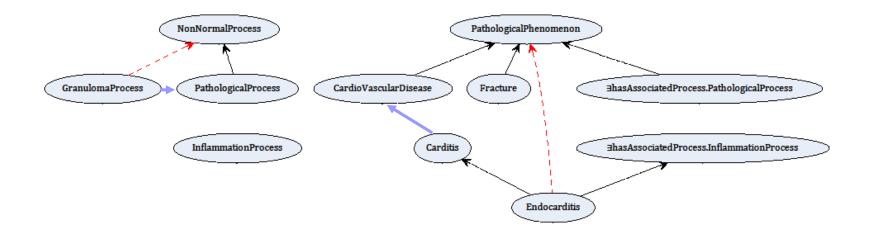
# **Semantic Maximality**

A solution S to GTAP(T,C,Or,M) is semantically maximal iff there is no solution S' which is more informative than S.



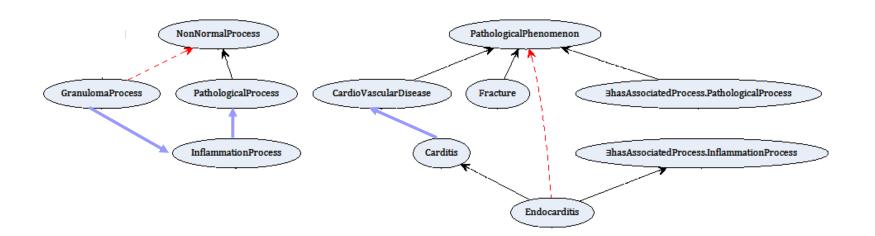
# **Subset Minimality**

A solution S to GTAP(T,C,Or,M) is subset minimal iff there is no proper subset S' of S that is a solution.



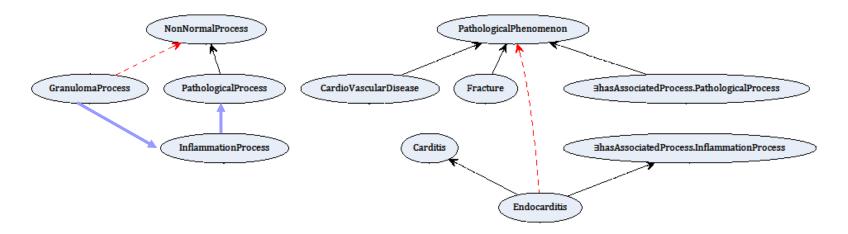
# **Combining with Priority for Semantic Maximality**

A solution S to GTAP(T,C,Or,M) is maxmin optimal iff S is semantically maximal and there is no other semantically maximal solution that is a proper subset of S.



#### **Combining with Priority for Subset Minimality**

A solution S to GTAP(T,C,Or,M) is minmax optimal iff S is subset minimal and there is no other subset minimal solution that is more informative than S.



#### **Combining with Equal Preferences**

- A solution S to GTAP(T,C,Or,M) is skyline optimal iff there is no other solution that is a proper subset of S and that is equally informative than S.
  - All subset minimal, minmax optimal and maxmin optimal solutions are also skyline optimal solutions.
  - Semantically maximal solutions may or may not be skyline optimal.

# **Preference Criteria - Conclusions**

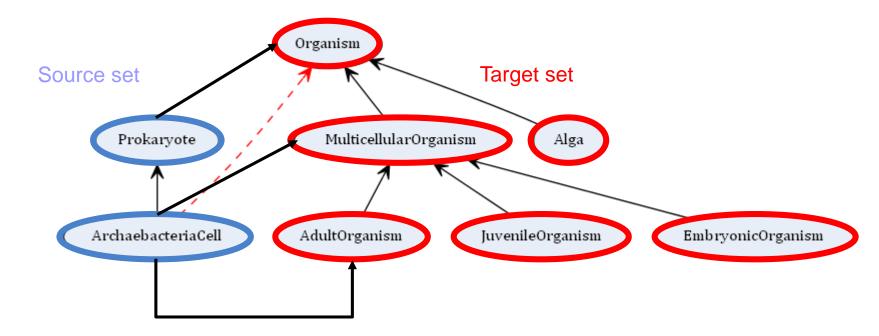
- In practice it is not clear how to generate maxmin or semantically maximal solutions (the preferred solutions)
- Skyline optimal solutions are the next best thing and are easy to generate

# Approach

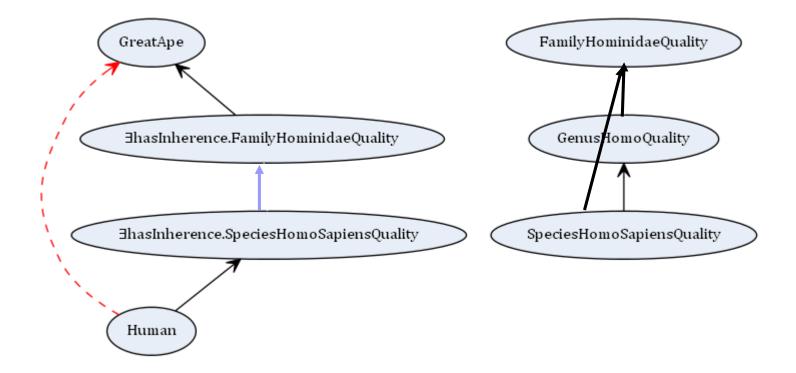
#### Input

- □ Normalized EL TBox
- Set of missing is-a relations (correct according to the domain)
- Output a skyline-optimal solution to GTAP
- Iteration of three main steps:
  - Creating solutions for individual missing is-a relations
  - Combining individual solutions
  - Trying to improve the result by finding a solution which introduces additional new knowledge (more informative)

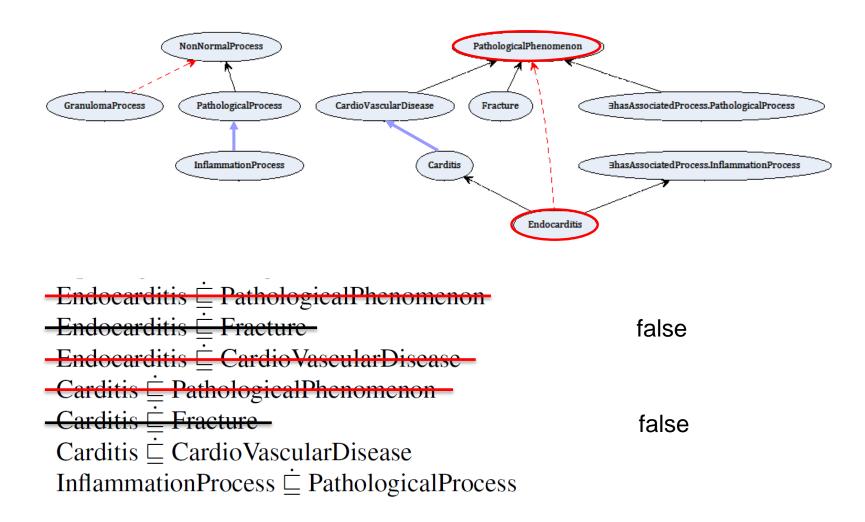
# **Intuition 1**



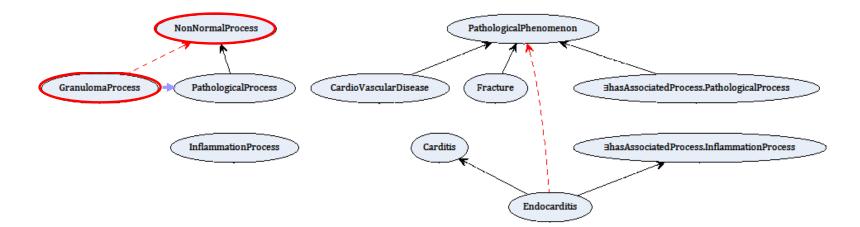
### **Intuitions 2/3**



#### **Example: Repairing Single is-a Relation**



#### **Example: Repairing Single is-a Relation**



GranulomaProcess E NonNormalProcess GranulomaProcess E PathologicalProcess

#### **Algorithm: Repairing Multiple is-a Relations**

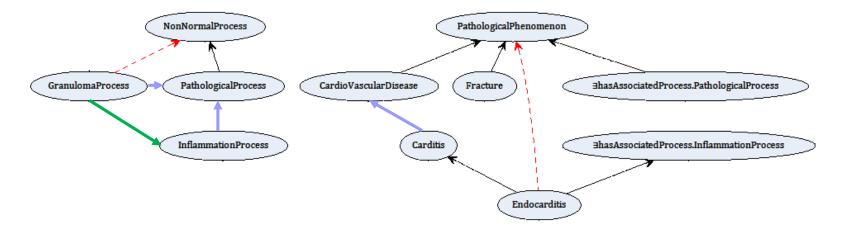
- Combine solutions for individual missing is-a relations
- Remove redundant relations while keeping the same level of informativness
- Resulting solution is a skyline optimal solution

{InflammationProcess  $\sqsubseteq$  PathologicalProcess, Carditis  $\doteq$  CardioVascularDisease, GranulomaProcess  $\doteq$  PathologicalProcess}

# **Algorithm: Improving Solution**

- Solution S from previous step may contain relations which are not derivable from the ontology.
- These can be seen as new missing is-a relations.
- We can solve a new GTAP problem: GTAP(T U S, C, Or, S)

# **Example: Improving Solutions**



#### GranulomaProcess $\doteq$ InflammationProcess

{InflammationProcess  $\sqsubseteq$  PathologicalProcess, Carditis  $\doteq$  CardioVascularDisease, GranulomaProcess  $\doteq$  InflammationProcess}

# **Algorithm Properties**

#### Sound

Skyline optimal solutions

# Experiments

Two use-cases

Case 1: given missing is-a relations
 AMA and a fragment of NCI-A ontology – OAEI 2013

- AMA (2744 concepts) 94 missing is-a relations
   → 3 iterations, 101 in repairing (47 additional new knowledge)
- NCI-A (3304 concepts) 58 missing is-a relations
   → 3 iterations, 54 in repairing (10 additional new knowledge)
- Case 2: no given missing is-a relations Modified BioTop ontology
  - Biotop (280 concepts, 42 object properties) randomly choose is-a relations and remove them: 47 'missing' → 4 iterations, 41 in repairing (40 additional new knowledge)

# **Further Reading**

Starting points for further studies

# **Further Reading Ontology Debugging**

http://www.ida.liu.se/~patla/DOOM/

Semantic defects

- Schlobach S, Cornet R. Non-Standard Reasoning Services for the Debugging of Description Logic Terminologies. 18th International Joint Conference on Artificial Intelligence - IJCAI03, 355-362, 2003.
- Schlobach S. <u>Debugging and Semantic Clarification by Pinpointing</u>. 2nd European Semantic Web Conference - ESWC05, LNCS 3532, 226-240, 2005.

# **Further Reading Ontology Completion**

Completing ontologies

- Fang Wei-Kleiner, Zlatan Dragisic, Patrick Lambrix. <u>Abduction Framework</u> for Repairing Incomplete EL Ontologies: Complexity Results and <u>Algorithms</u>. 28th AAAI Conference on Artificial Intelligence - AAAI 2014, 1120-1127, 2014.
- Lambrix P, Ivanova V, <u>A unified approach for debugging is-a structure and</u> <u>mappings in networked taxonomies</u>, *Journal of Biomedical Semantics* 4:10, 2013.
- Lambrix P, Liu Q, <u>Debugging the missing is-a structure within taxonomies</u> <u>networked by partial reference alignments</u>, *Data & Knowledge Engineering* 86:179-205, 2013.

# Further Reading Systems

#### RepOSE

http://www.ida.liu.se/~patla00/research/RepOSE/

video:

http://videolectures.net/semantic\_ivanova\_taxonomies/

OOPS! <u>http://oops.linkeddata.es/</u>