Safe and Economic Re-Use of Ontologies: A Logic-Based Methodology and Tool Support

Ernesto Jiménez-Ruiz[1] Bernardo Cuenca Grau[2]
Ulrike Sattler[3] *Thomas Schneider*[3] Rafael Berlanga[1]

[1]Computer Languages and Systems, Universitat Jaume I, Spain

[2]Computing Laboratory, University of Oxford, UK

[3]Computer Science, University of Manchester, UK

DL, 13–16 May 2008

Our approach in a nutshell

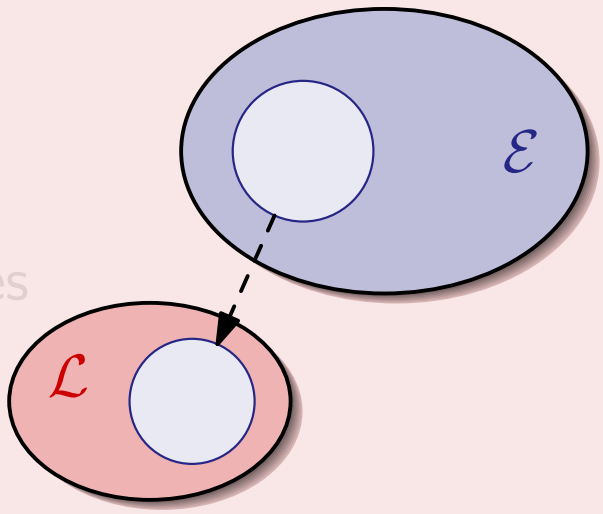
Logic-based methodology for the re-use of ontologies

Safe use of imported symbols

1. Don't change their meaning! ✓

Economic import of the external ontologies

2. Import only the relevant parts . . .
3. . . . without loss of information! ✓



- Tool support — Protégé plugin
- Work in progress!

Our approach in a nutshell

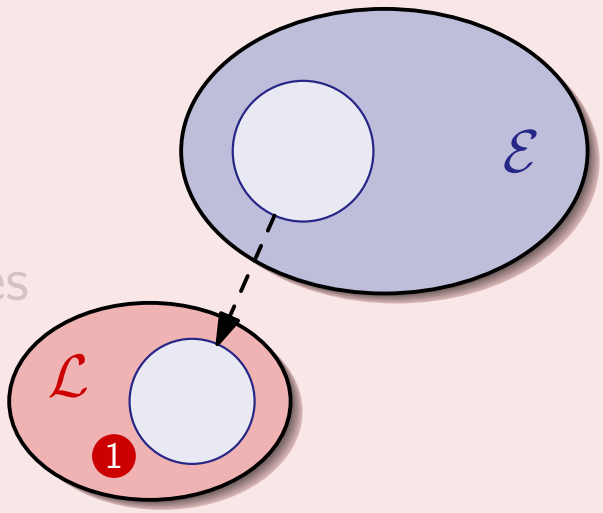
Logic-based methodology for the re-use of ontologies

Safe use of imported symbols

1. Don't change their meaning! ✓

Economic import of the external ontologies

2. Import only the relevant parts . . .
3. . . . without loss of information! ✓



- Tool support — Protégé plugin
- Work in progress!

Our approach in a nutshell

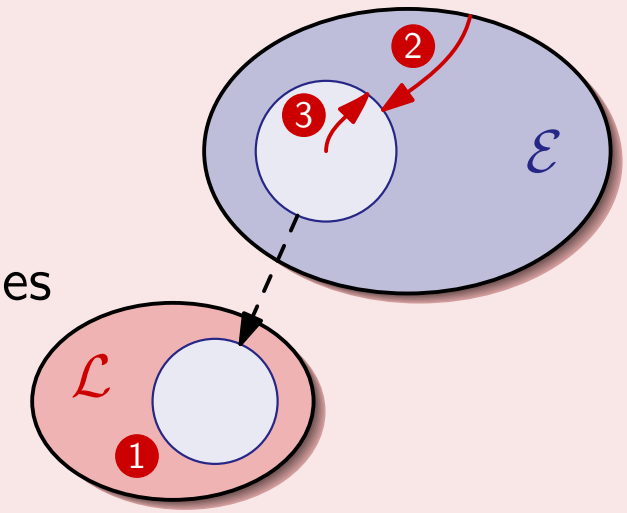
Logic-based methodology for the re-use of ontologies

Safe use of imported symbols

1. Don't change their meaning! ✓

Economic import of the external ontologies

2. Import only the relevant parts . . .
3. . . . without loss of information! ✓



- Tool support — Protégé plugin
- Work in progress!

Our approach in a nutshell

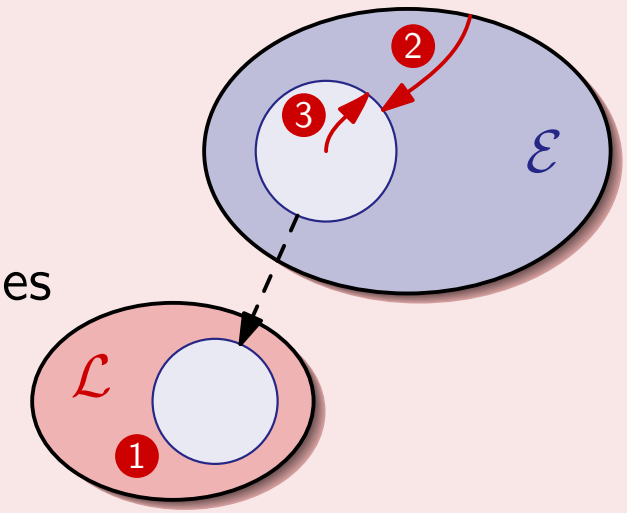
Logic-based methodology for the re-use of ontologies

Safe use of imported symbols

1. Don't change their meaning! ✓

Economic import of the external ontologies

2. Import only the relevant parts ...
3. ... without loss of information! ✓

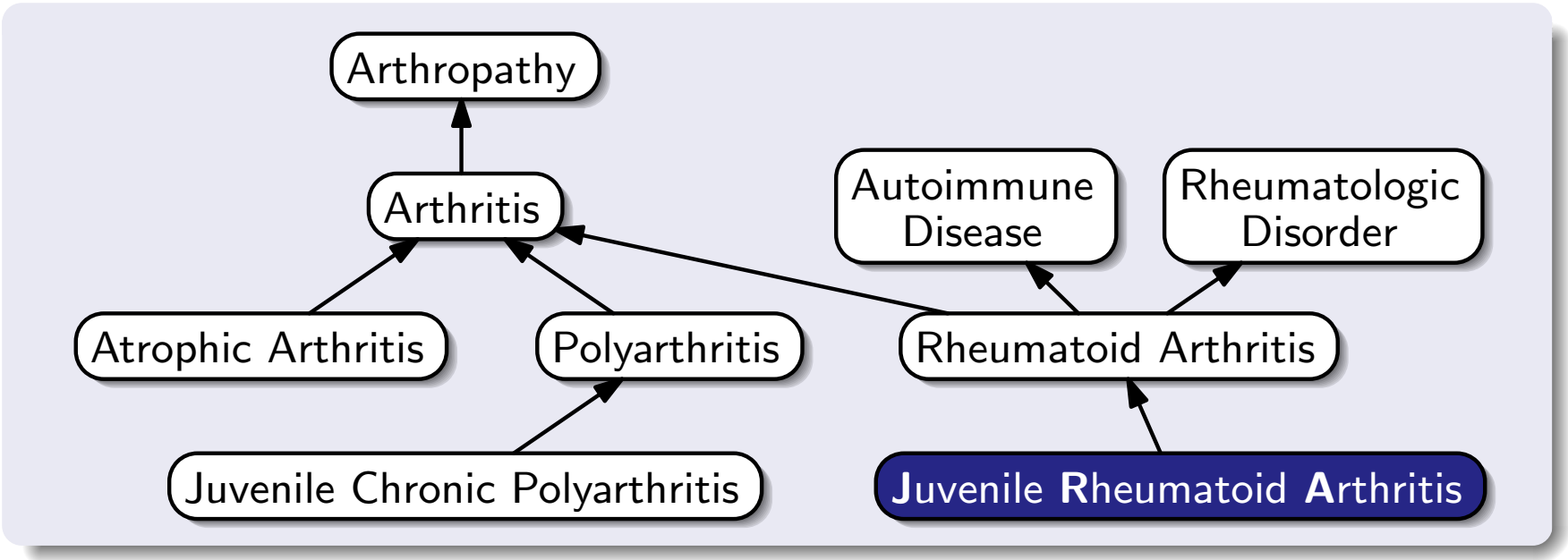


- Tool support — Protégé plugin
- Work in progress!

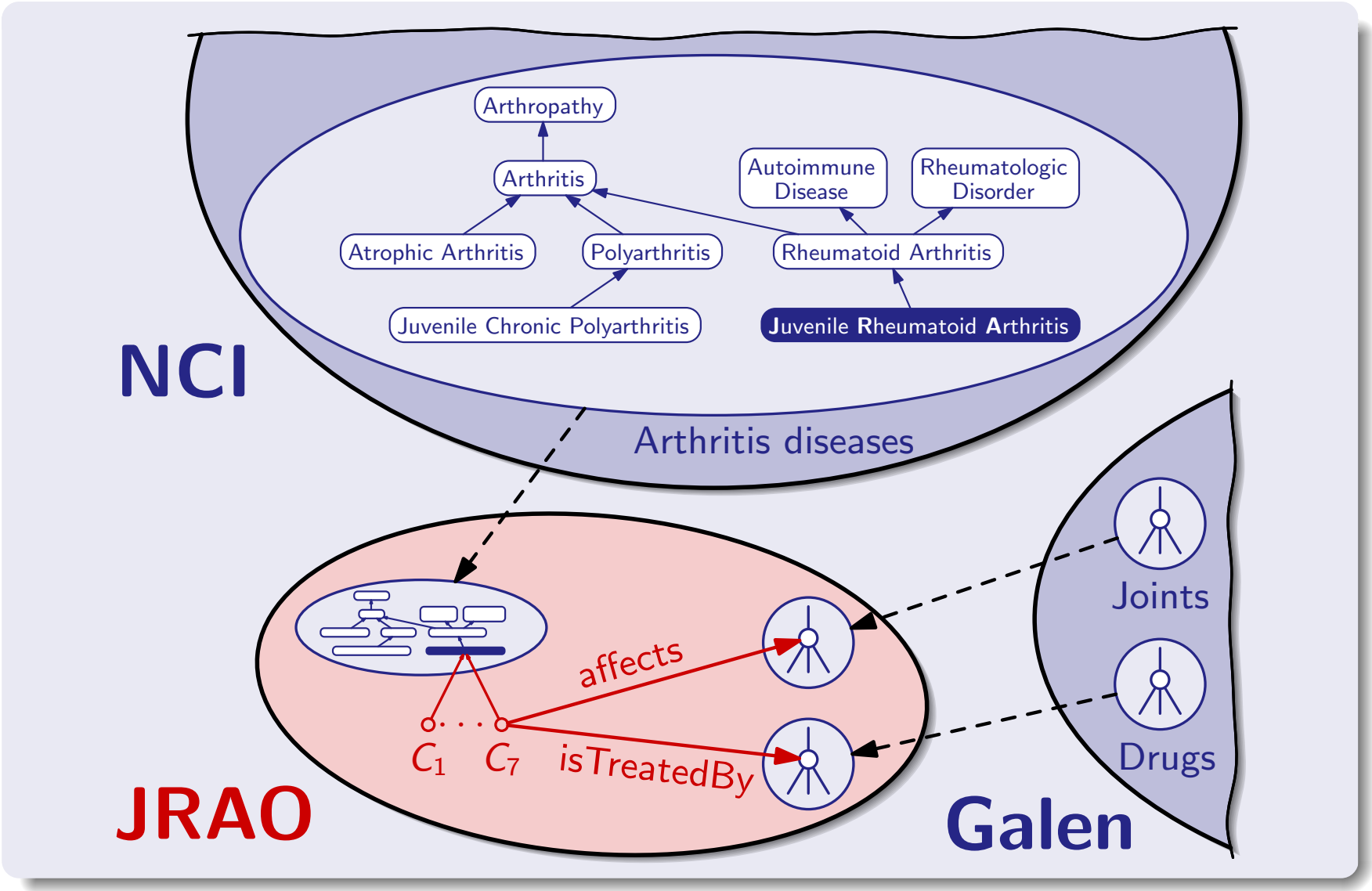
And now . . .

1. Why ontology re-use?
2. A safe and economic methodology
3. Tool support
4. Perspectives

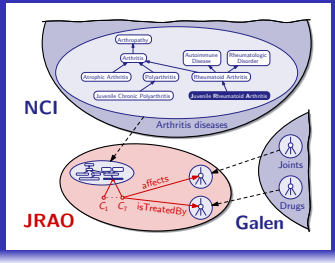
A re-use scenario: the *Health-e-Child* project



A re-use scenario: the *Health-e-Child* project



A case for safe and economic re-use



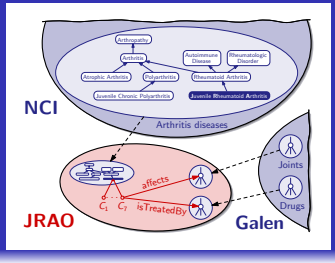
Reasons for re-use

- Saves time for re-writing
- Provides access to well-established knowledge
- Doesn't require expertise in drugs, proteins, anatomy etc.

Guarantees to provide

- **[safe]** Importing terms doesn't change their meaning.
- **[eco]** Import all relevant parts of external ontologies.
- [aux] The order of imports doesn't matter.

And now . . .



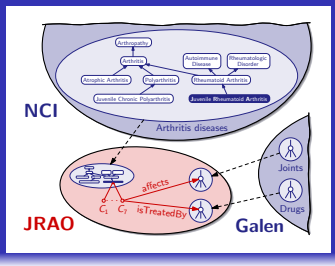
1 Why ontology re-use?

2 A safe and economic methodology

3 Tool support

4 Perspectives

A working cycle: the offline phase

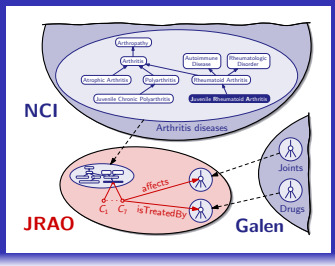


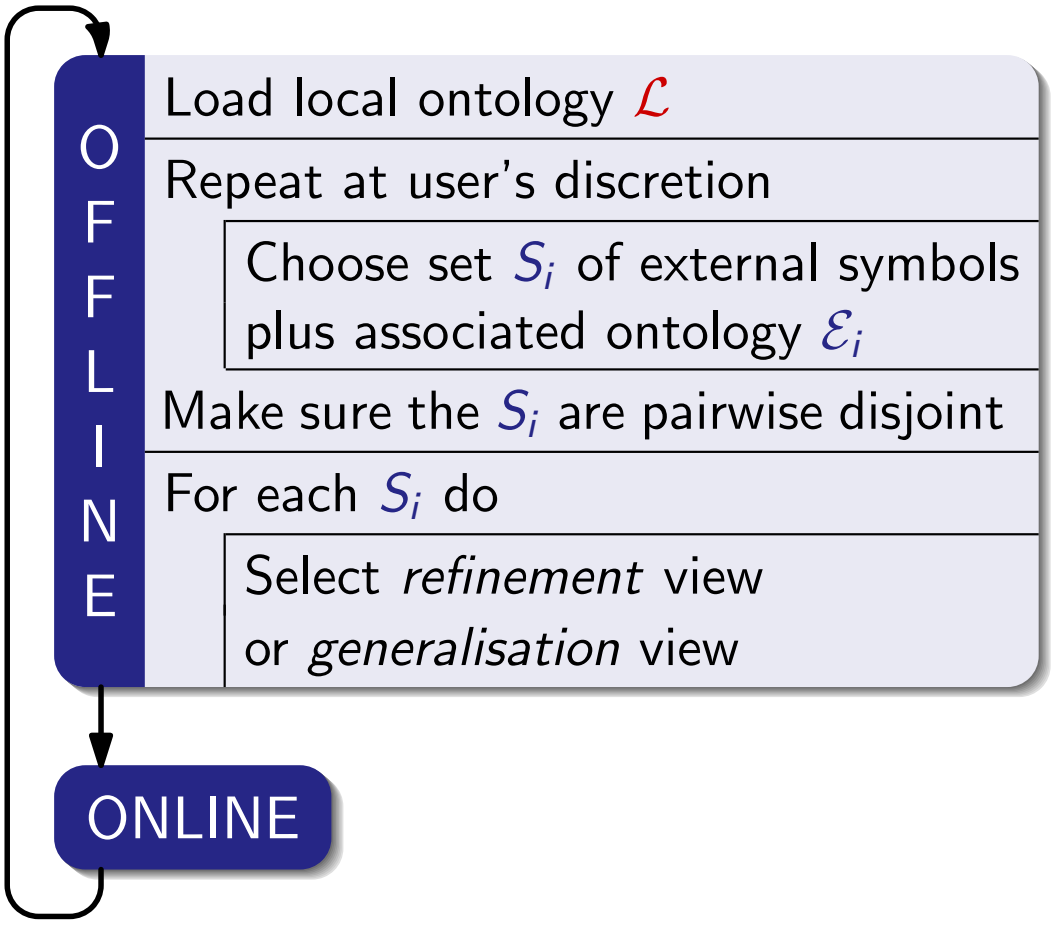
OFFLINE

- Load local ontology $\mathcal{L}$
- Repeat at user's discretion
  - Choose set S_i of external symbols plus associated ontology $\mathcal{E}_i$
- Make sure the S_i are pairwise disjoint
- For each S_i do
  - Select *refinement* view or *generalisation* view

ONLINE

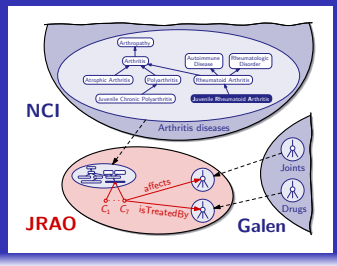
A working cycle: the offline phase





$S_1 = \{\texttt{JRA}\}$ $\mathcal{E}_1 = \textbf{NCI}$

A working cycle: the offline phase



OFFLINE

- Load local ontology $\mathcal{L}$
- Repeat at user's discretion
  - Choose set S_i of external symbols plus associated ontology $\mathcal{E}_i$
- Make sure the S_i are pairwise disjoint
- For each S_i do
  - Select *refinement* view or *generalisation* view

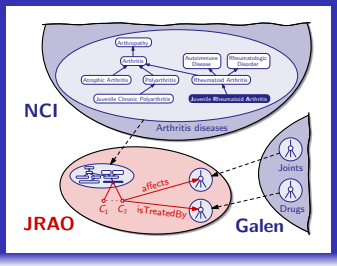
ONLINE

$S_1 = \{\texttt{JRA}\}$ $\quad \mathcal{E}_1 = \textbf{NCI}$

$S_2 = \{\texttt{KneeJoint}, \texttt{Fever}\}$

$\mathcal{E}_2 = \textbf{Galen}$

A working cycle: the offline phase



OFFLINE

- Load local ontology $\mathcal{L}$
- Repeat at user's discretion
  - Choose set S_i of external symbols plus associated ontology $\mathcal{E}_i$
- Make sure the S_i are pairwise disjoint
- For each S_i do
  - Select *refinement* view or *generalisation* view

ONLINE

$S_1 = \{\texttt{JRA}\}$ $\quad \mathcal{E}_1 = \textbf{NCI}$

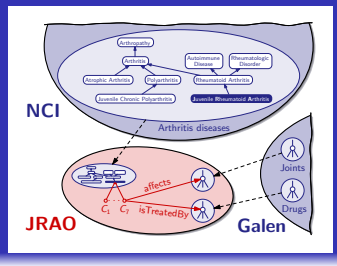
$S_2 = \{\texttt{KneeJoint}, \texttt{Fever}\}$

$\mathcal{E}_2 = \textbf{Galen}$

Refine S_1

Refine + reference S_2

A working cycle: the offline phase



OFFLINE

- Load local ontology $\mathcal{L}$
- Repeat at user's discretion
  - Choose set S_i of external symbols plus associated ontology $\mathcal{E}_i$
- Make sure the S_i are pairwise disjoint
- For each S_i do
  - Select *refinement* view or *generalisation* view

ONLINE

$S_1 = \{\texttt{JRA}\}$ $\quad \mathcal{E}_1 = \textbf{NCI}$

$S_2 = \{\texttt{KneeJoint}, \texttt{Fever}\}$

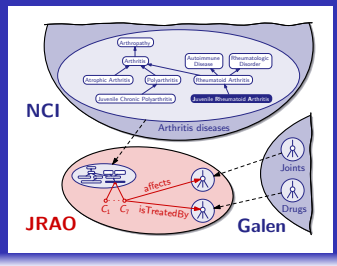
$\mathcal{E}_2 = \textbf{Galen}$

Refine S_1

Refine + reference S_2

Safety

Formalising the *Safety Guarantee*



Safety

Importing terms doesn't change their meaning.

Example

$$\mathbf{JRAO} \cup \mathbf{NCI} \models \mathtt{JRA} \sqsubseteq \mathtt{GeneticDisorder}$$

iff

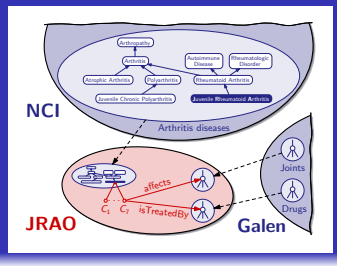
$$\mathbf{NCI} \models \mathtt{JRA} \sqsubseteq \mathtt{GeneticDisorder}.$$

Definition (Safety)

$\mathcal{L}$ guarantees safety if for every $i = 1, \ldots, n$:

For every $\mathcal{E}'_i$ with $\mathsf{Sig}(\mathcal{L}) \cap \mathsf{Sig}(\mathcal{E}'_i) \subseteq S_i$,
$\mathcal{L} \cup \mathcal{E}'_i$ is a conservative extension of $\mathcal{E}'_i$.

Formalising the *Safety Guarantee*



Safety

Importing terms doesn't change their meaning.

Example

$$\mathbf{JRAO} \cup \mathbf{NCI} \models \mathtt{JRA} \sqsubseteq \mathtt{GeneticDisorder}$$

iff

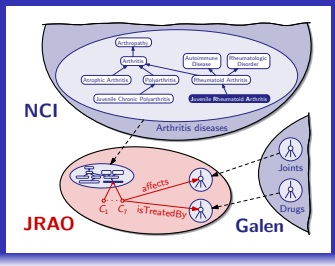
$$\mathbf{NCI} \models \mathtt{JRA} \sqsubseteq \mathtt{GeneticDisorder}.$$

Definition (Safety)

$\mathcal{L}$ guarantees safety if for every $i = 1, \ldots, n$:

For every $\mathcal{E}'_i$ with $\mathrm{Sig}(\mathcal{L}) \cap \mathrm{Sig}(\mathcal{E}'_i) \subseteq S_i$,
$\mathcal{L} \cup \mathcal{E}'_i$ is a conservative extension of $\mathcal{E}'_i$.

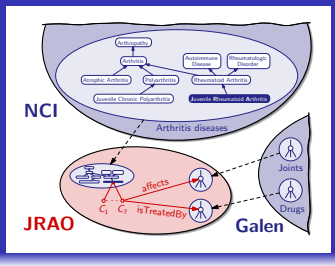
Approximating conservativity



$\mathcal{L} \cup \mathcal{E}_i'$ is a *deductive conservative extension* of $\mathcal{E}_i'$

$\forall \alpha$ with $\mathsf{Sig}(\alpha) \subseteq \mathsf{Sig}(\mathcal{E}_i')$, $\quad \mathcal{L} \cup \mathcal{E}_i' \models \alpha$ iff $\mathcal{E}_i' \models \alpha$.

Approximating conservativity



$\mathcal{L} \cup \mathcal{E}'_i$ is a *deductive conservative extension* of $\mathcal{E}'_i$

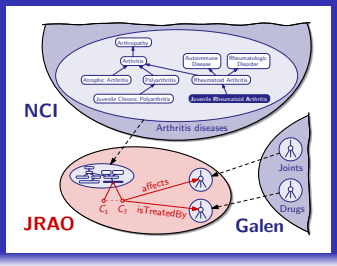
$\forall \alpha$ with $\mathsf{Sig}(\alpha) \subseteq \mathsf{Sig}(\mathcal{E}'_i)$, $\quad \mathcal{L} \cup \mathcal{E}'_i \models \alpha$ iff $\mathcal{E}'_i \models \alpha$.

$\Uparrow$

$\mathcal{L} \cup \mathcal{E}'_i$ is a *model-theoretic conservative extension* of $\mathcal{E}'_i$

$\forall \mathcal{I}$ for $\mathcal{E}'_i \quad \exists \mathcal{J}$ for $\mathcal{L} \cup \mathcal{E}'_i \quad : \quad \mathcal{I}|_{\mathsf{Sig}(\mathcal{E}'_i)} = \mathcal{J}|_{\mathsf{Sig}(\mathcal{E}'_i)}$.

Approximating conservativity



$\mathcal{L} \cup \mathcal{E}_i'$ is a *deductive conservative extension* of $\mathcal{E}_i'$

$\forall \alpha$ with $\mathsf{Sig}(\alpha) \subseteq \mathsf{Sig}(\mathcal{E}_i')$, $\quad \mathcal{L} \cup \mathcal{E}_i' \models \alpha$ iff $\mathcal{E}_i' \models \alpha$.

$\Uparrow$

$\mathcal{L} \cup \mathcal{E}_i'$ is a *model-theoretic conservative extension* of $\mathcal{E}_i'$

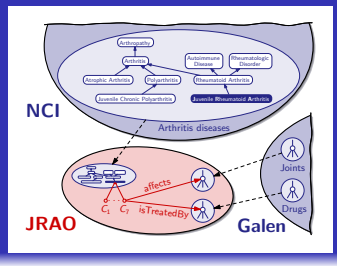
$\forall \mathcal{I}$ for $\mathcal{E}_i'$ $\quad \exists \mathcal{J}$ for $\mathcal{L} \cup \mathcal{E}_i'$ $\quad : \quad \mathcal{I}|_{\mathsf{Sig}(\mathcal{E}_i')} = \mathcal{J}|_{\mathsf{Sig}(\mathcal{E}_i')}$.

$\Uparrow$

$\mathcal{L}$ is *semantically $\bot$-local* w.r.t. $\mathcal{L} \cup \mathcal{E}_i'$

$\forall \mathcal{I}$ for $\mathcal{E}_i'$ $\quad \exists \mathcal{J}$ for $\mathcal{L} \cup \mathcal{E}_i'$ $\quad : \quad \mathcal{I}|_{\mathsf{Sig}(\mathcal{E}_i')} = \mathcal{J}|_{\mathsf{Sig}(\mathcal{E}_i')}$
such that $\quad X^{\mathcal{J}} = \emptyset \quad$ for each $\quad X \in \mathsf{Sig}(\mathcal{L}) \setminus S_i$.

Approximating conservativity

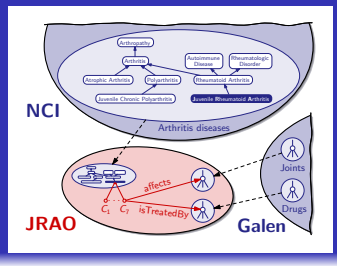


$\mathcal{L}$ is *semantically* $\bot$-*local* w.r.t. $\mathcal{L} \cup \mathcal{E}'_i$

$\forall \mathcal{I}$ for $\mathcal{E}'_i$ $\quad \exists \mathcal{J}$ for $\mathcal{L} \cup \mathcal{E}'_i$ $\quad : \quad \mathcal{I}|_{\mathsf{Sig}(\mathcal{E}'_i)} = \mathcal{J}|_{\mathsf{Sig}(\mathcal{E}'_i)}$

such that $\quad X^{\mathcal{J}} = \emptyset \quad$ for each $\quad X \in \mathsf{Sig}(\mathcal{L}) \setminus S_i$.

Approximating conservativity



$\mathcal{L}$ is *semantically $\bot$-local* w.r.t. $\mathcal{L} \cup \mathcal{E}'_i$

$\forall \mathcal{I}$ for $\mathcal{E}'_i$ $\quad \exists \mathcal{J}$ for $\mathcal{L} \cup \mathcal{E}'_i$ $\quad : \quad \mathcal{I}|_{\mathsf{Sig}(\mathcal{E}'_i)} = \mathcal{J}|_{\mathsf{Sig}(\mathcal{E}'_i)}$

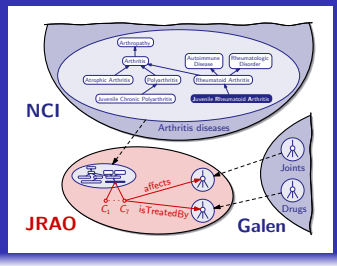
such that $\quad X^{\mathcal{J}} = \emptyset \quad$ for each $\quad X \in \mathsf{Sig}(\mathcal{L}) \setminus S_i$. $\quad \ast$

$\Uparrow$

$\mathcal{L}$ is *syntactically $\bot$-local* w.r.t. S_i

- all GCIs in $\mathcal{L}$ are of the form $C_\bot \sqsubseteq C$ or $C \sqsubseteq C_\top$ where $C_\bot^{\mathcal{I}} = \emptyset \quad$ and $\quad C_\top^{\mathcal{I}} = \Delta^{\mathcal{I}} \quad$ follow from $\ast$
- similar conditions for RIs and Trans(R) statements

Providing safety



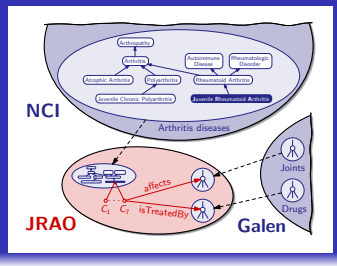
Examples

- $C_7 \sqsubseteq \texttt{JRA}$ ✓ $\bot$-local
- $\texttt{GeneticDisorder} \sqsubseteq C_7$ ✓ $\top$-local
- $\texttt{JRA} \sqsubseteq \texttt{GeneticDisorder}$ ✘ non-local

Theorem [Cuenca Grau, Horrocks, Kazakov, Sattler 2007]

If $\mathcal{L}$ is syntactically local w.r.t. each S_i, then $\mathcal{L}$ guarantees safety.

The online phase



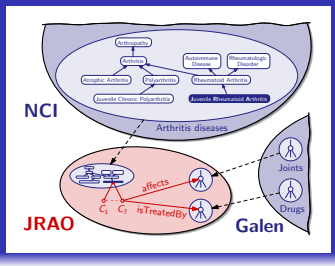
OFFLINE

ONLINE

Repeat at user's discretion

- Select an S_i
- Load external ontology $\mathcal{E}_i$ for S_i
- Customise scope of module
- Extract module $\mathcal{E}_i^M$ from $\mathcal{E}_i$
- Import $\mathcal{E}_i^M$ into $\mathcal{L}$

The online phase

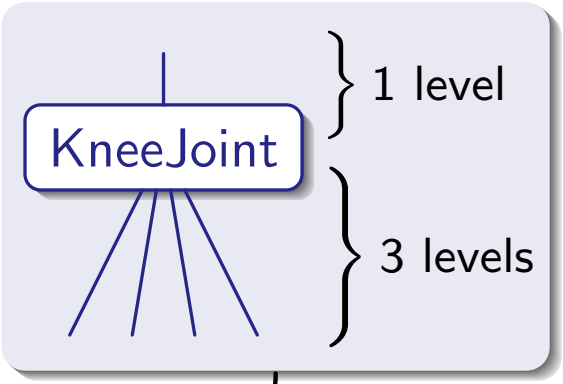


OFFLINE

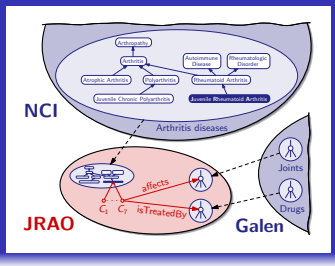
ONLINE

Repeat at user's discretion

- Select an S_i
- Load external ontology $\mathcal{E}_i$ for S_i
- Customise scope of module
- Extract module $\mathcal{E}_i^M$ from $\mathcal{E}_i$
- Import $\mathcal{E}_i^M$ into $\mathcal{L}$



The online phase

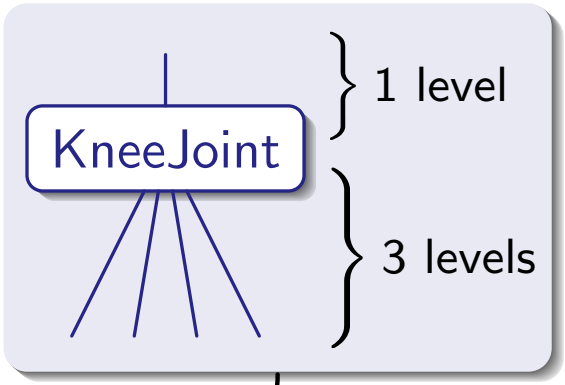


OFFLINE

ONLINE

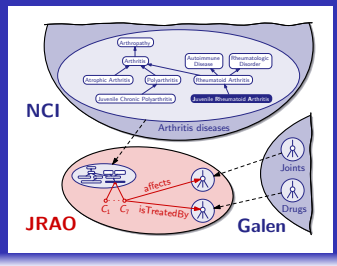
Repeat at user's discretion

- Select an S_i
- Load external ontology $\mathcal{E}_i$ for S_i
- Customise scope of module
- Extract module $\mathcal{E}_i^M$ from $\mathcal{E}_i$
- Import $\mathcal{E}_i^M$ into $\mathcal{L}$



Module Coverage

Formalising the *Module Coverage Guarantee*



Module coverage

Import all relevant parts of external ontologies.

Example

$$\mathbf{JRAO} \cup \mathbf{NCI} \models \texttt{JRA} \sqsubseteq \texttt{GeneticDisorder}$$

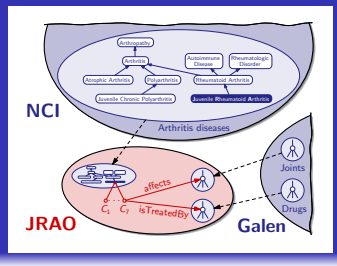
iff $\mathbf{JRAO} \cup \mathbf{NCI}\text{-module} \models \texttt{JRA} \sqsubseteq \texttt{GeneticDisorder}.$

Definition (Module coverage)

Let $\mathcal{E}_i^M \subseteq \mathcal{E}_i$ with $S_i \subseteq \mathrm{Sig}(\mathcal{E}_i^M)$. $\mathcal{E}_i^M$ guarantees coverage of S_i if:

For every $\mathcal{L}'$ with $\mathrm{Sig}(\mathcal{L}') \cap \mathrm{Sig}(\mathcal{E}_i) \subseteq S_i$,
$\mathcal{L}' \cup \mathcal{E}_i$ is a conservative extension of $\mathcal{L}' \cup \mathcal{E}_i^M$.

Formalising the *Module Coverage Guarantee*



Module coverage

Import all relevant parts of external ontologies.

Example

$$\mathbf{JRAO} \cup \mathbf{NCI} \models \texttt{JRA} \sqsubseteq \texttt{GeneticDisorder}$$

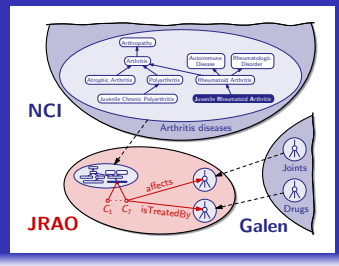
iff $\mathbf{JRAO} \cup \mathbf{NCI}\text{-module} \models \texttt{JRA} \sqsubseteq \texttt{GeneticDisorder}$.

Definition (Module coverage)

Let $\mathcal{E}_i^M \subseteq \mathcal{E}_i$ with $S_i \subseteq \mathrm{Sig}(\mathcal{E}_i^M)$. $\mathcal{E}_i^M$ guarantees coverage of S_i if:

For every $\mathcal{L}'$ with $\mathrm{Sig}(\mathcal{L}') \cap \mathrm{Sig}(\mathcal{E}_i) \subseteq S_i$,
$\mathcal{L}' \cup \mathcal{E}_i$ is a conservative extension of $\mathcal{L}' \cup \mathcal{E}_i^M$.

Providing coverage



- Coverage is again provided using locality.
- Locality-based modules =
  syntactic approximations of conservativity-based modules
  - in general not minimal
  - efficiently computable

And now . . .

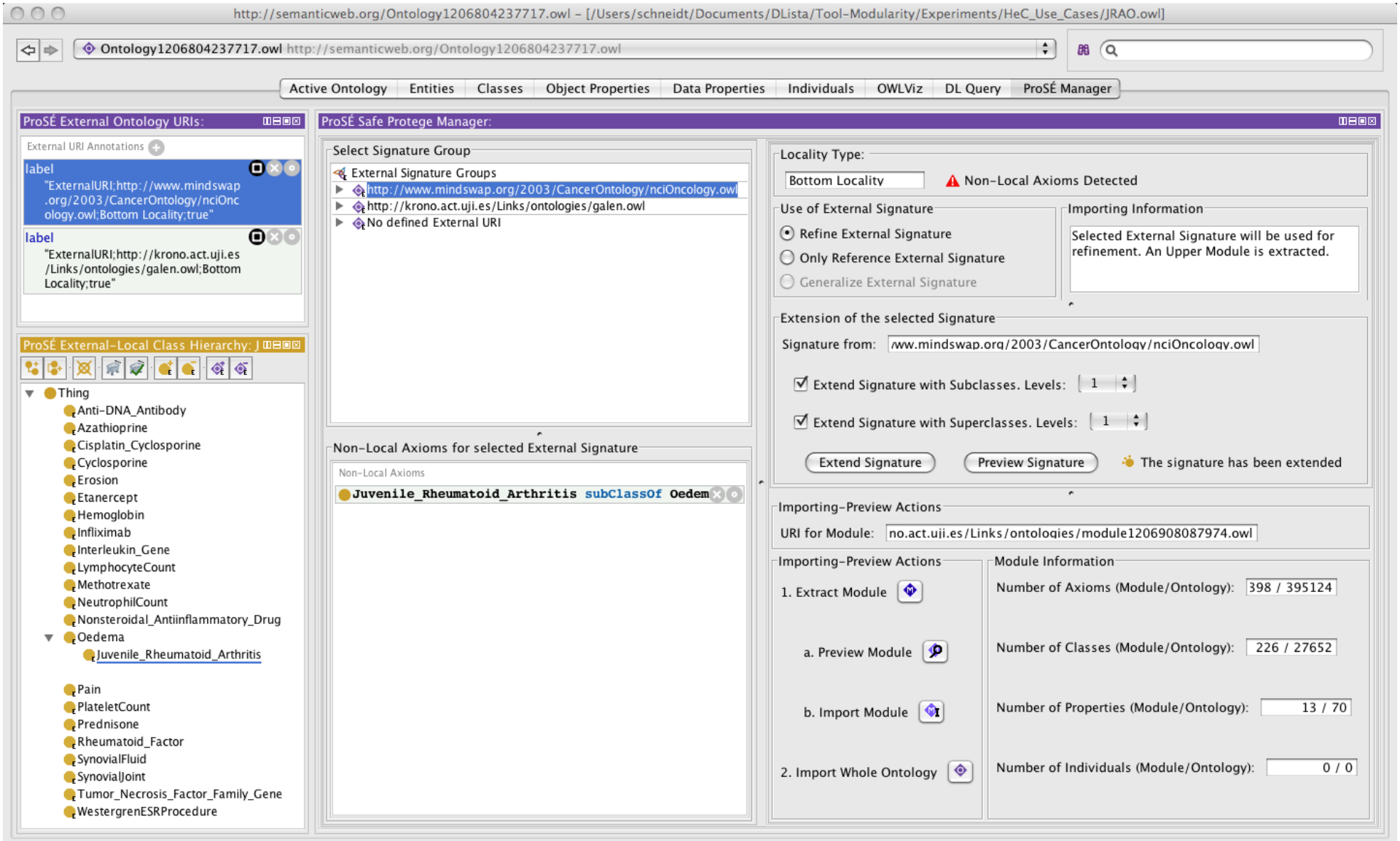
1 Why ontology re-use?

2 A safe and economic methodology

3 Tool support

4 Perspectives

A Protégé Plugin for the reuse of Ontologies: Safe and Économique



And now . . .

1. Why ontology re-use?
2. A safe and economic methodology
3. Tool support
4. Perspectives

“Shopping for symbols”

Extend module scope customisation:

- Browse external ontology and pick symbols
- At each stage, view resulting module
- “Check out” module

⇝ Treemaps?

Other plans

- Optimise module extraction
- Import "by reference" as opposed to "by value"
- Multi-user scenario
- Module extraction service at `owl.cs.manchester.ac.uk`
- Modularity tool tutorial at ISWC 2008
- Perform user study and improve interface

Other plans

- Optimise module extraction
- Import "by reference" as opposed to "by value"
- Multi-user scenario
- Module extraction service at `owl.cs.manchester.ac.uk`
- Modularity tool tutorial at ISWC 2008
- Perform user study and improve interface

Other plans

- Optimise module extraction
- Import "by reference" as opposed to "by value"
- Multi-user scenario
- Module extraction service at `owl.cs.manchester.ac.uk`
- Modularity tool tutorial at ISWC 2008
- Perform user study and improve interface

Invitation

We want you...

- ... to work with us on incorporating our services into your workflows!
- ...r favourite ontologies and real-life signatures!

Contact

schneider@cs.man.ac.uk

Invitation

We want you...

- ... to work with us on incorporating our services into your workflows!
- ...r favourite ontologies and real-life signatures!

Contact

schneider@cs.man.ac.uk

Thank you!

"Synthetic" experiments

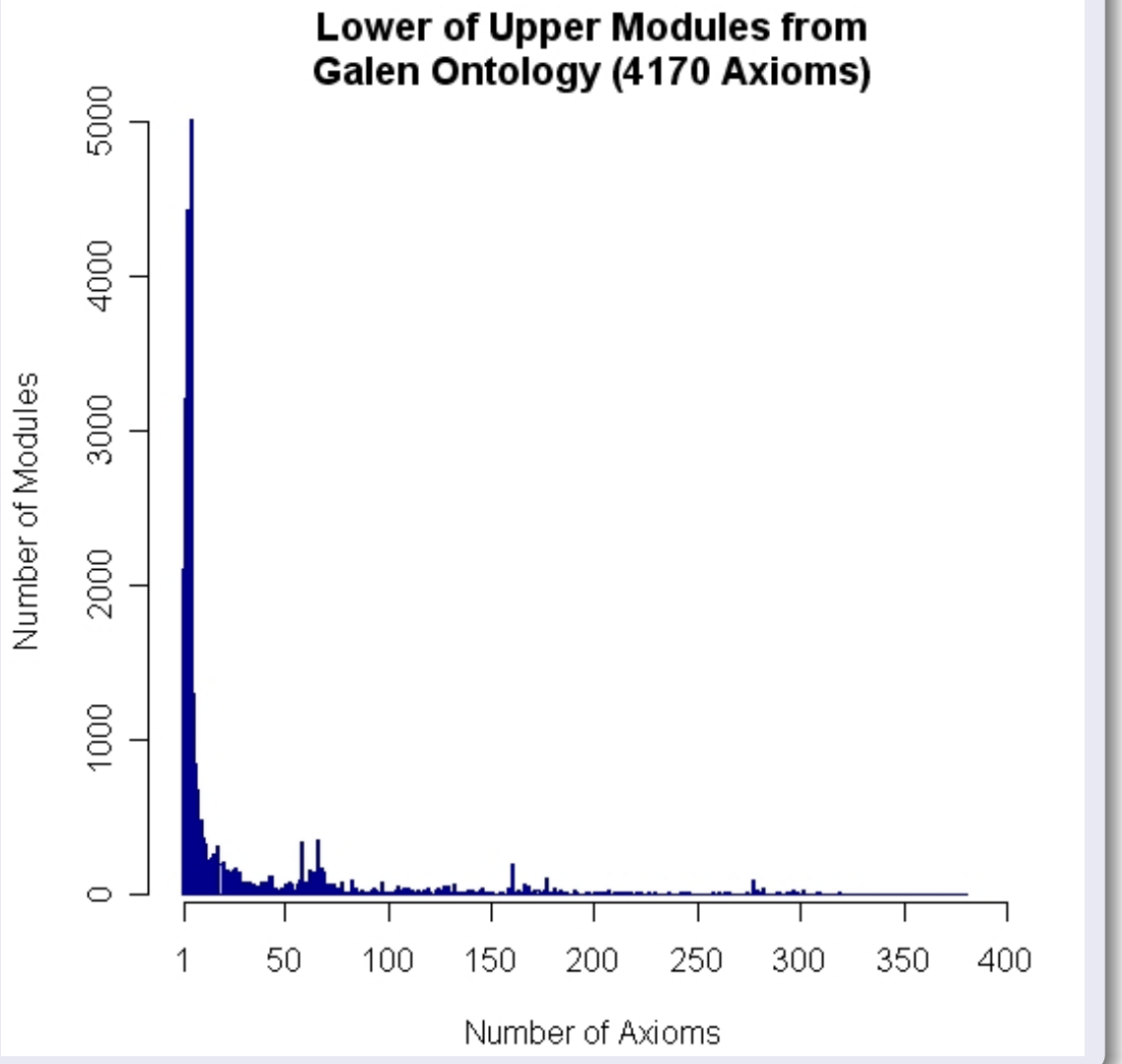
Setting

- Randomly generated signatures of size $1 \ldots 330$
- Computed *Lower of Upper Module (LUM)* for each such signature

Results

- 99 % of **Galen** LUMs contain < 5 % of **Galen**'s axioms
- similar findings for **NCI**

Statistics


Lower of Upper Modules from
Galen Ontology (4170 Axioms)
Number of Modules
0
1000
2000
3000
4000
5000
1
50
100
150
200
250
300
350
400
Number of Axioms

"Real-life" experiments

Setting

LUMs for manually selected signatures from **Galen** and **NCI**
(*Health-e-Child* context: JRA + Cardiomyopathies)

Results

Ext. Ont.	# Sig.	# axioms	
Galen	11	105	(2.5 %)
Galen	72	620	(14.9 %)
Galen	76	736	(17.6 %)
NCI	18	488	(0.1 %)
NCI	124	4751	(1.2 %)
NCI	144	5057	(1.3 %)

Comparing experiments

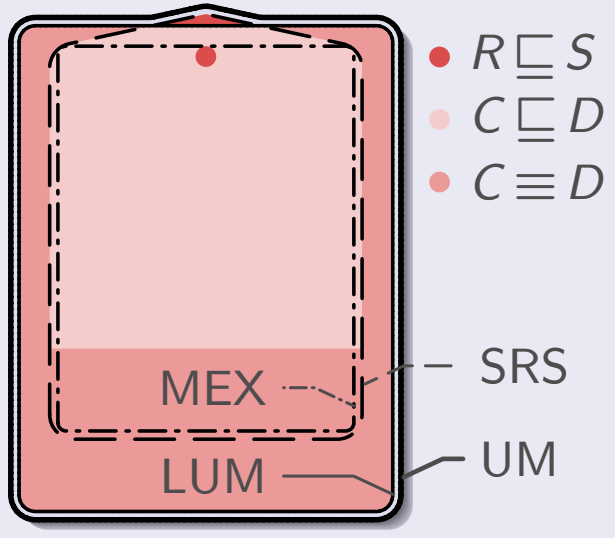
Setting

- SNOMED (health care; restricted language; 350,000 axioms)
- Initial signatures: terms from intensive care unit
- Compared UM, LUM to MEX (conservativity-based modules) and SRS (Seidenberg/Rector segments)

Results

# Sig.	# axioms in % MEX	SRS	(L)UM
4,000	2	2	4
16,000	7	7	10
24,000	10	10	15
time	4–5 s	1 s	4–7 s

$\rightsquigarrow$



More links

Protégé and ProSÉ

- protege.stanford.edu
- krono.act.uji.es/people/Ernesto/safety-ontology-reuse

Health-e-Child

- www.health-e-child.org

NCI and Galen

- nciterms.nci.nih.gov/NCIBrowser/Dictionary.do
- ftp1.nci.nih.gov/pub/cacore/EVS/NCIThesaurus
- www.co-ode.org/galen