

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

Ernesto Jiménez-Ruiz¹ Bernardo Cuenca Grau²
Ulrike Sattler Thomas Schneider Rafael Berlanga¹

¹Computer Languages and Systems, Universitat Jaume I, Spain

²Computing Laboratory, University of Oxford, UK

IMG Seminar, 14th February 2008

Our approach in a nutshell

- Logic-based methodology for the re-use of ontologies
- **Safe** use of imported symbols
- **Economic** import of the relevant parts of external ontologies
- Tool support — Protégé plugin
- Work in progress!

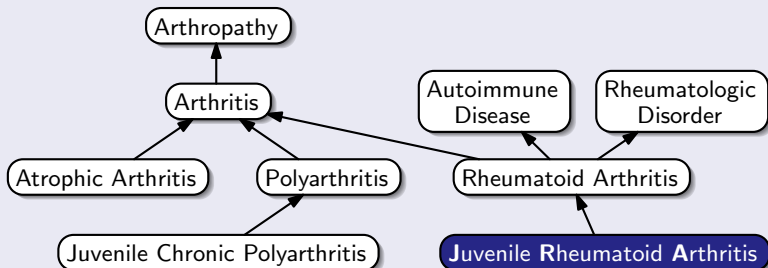
And now ...

- 1 Why ontology re-use?
- 2 A safe and economic methodology
- 3 Tool support and experiments
- 4 Conclusion and perspectives

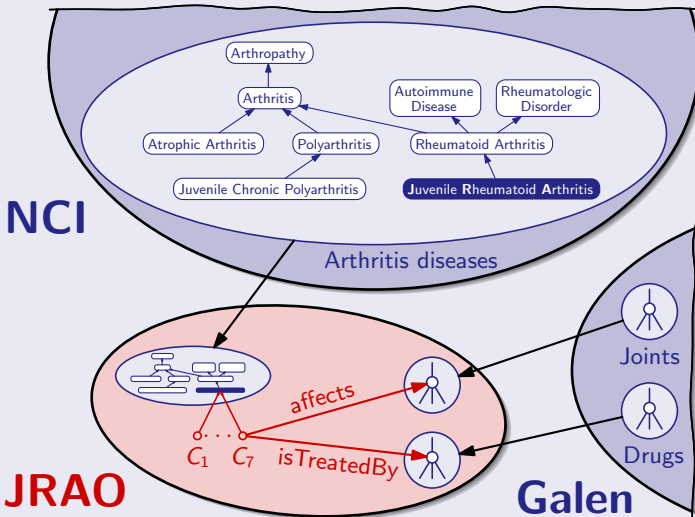
A re-use scenario: the Health-e-Child Project

- Build an ontology **JRAO** that describes JRA:
Juvenile Rheumatoid Arthritis
- Describe JRA subkinds by
 - Joints affected
 - Occurrence of concomitant symptoms — e.g., fever
 - Treatment with certain drugs
- Re-use information provided by biomedical ontologies
 - **NCI** — diseases, drugs, proteins etc.
 - **Galen** — human anatomy

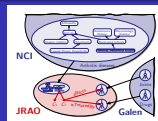
JRA and related diseases in NCI



Building JRAO



A case for safe and economic re-use



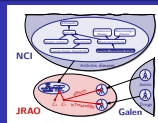
Reasons for re-use

- 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.

The two main import guarantees



Safety

Importing terms doesn't change their meaning.

JRAO \cup **NCI** \models JRA \sqsubseteq GeneticDisorder
iff **NCI** \models JRA \sqsubseteq GeneticDisorder.

Directed inwards: How do we re-use terms **in our local ontology**?

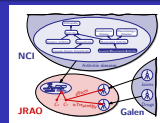
Module Coverage

Import all relevant parts of external ontologies.

JRAO \cup **NCI** \models JRA \sqsubseteq GeneticDisorder
iff **JRAO** \cup **NCI-module** \models JRA \sqsubseteq GeneticDisorder.

Directed outwards: How many external terms do we re-use?

The two main import guarantees



Safety

Importing terms doesn't change their meaning.

JRAO \cup **NCI** \models JRA \sqsubseteq GeneticDisorder
 iff **NCI** \models JRA \sqsubseteq GeneticDisorder.

Directed inwards: How do we re-use terms **in our local ontology**?

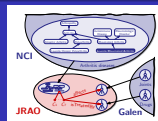
Module Coverage

Import all relevant parts of external ontologies.

JRAO \cup **NCI** \models JRA \sqsubseteq GeneticDisorder
 iff **JRAO** \cup **NCI-module** \models JRA \sqsubseteq GeneticDisorder.

Directed outwards: How many **external** terms do we re-use?

The third import guarantee



Module Independence

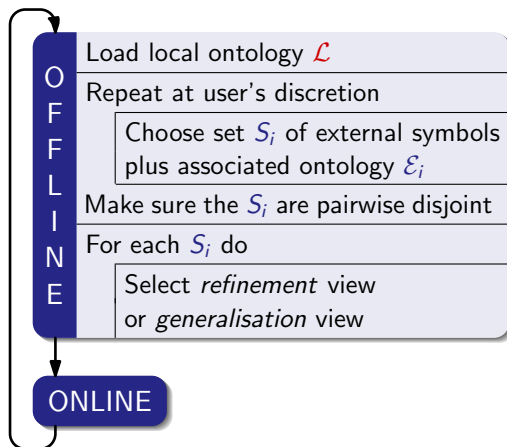
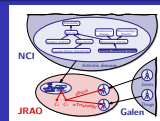
The order of imports doesn't matter.

If it is safe to import an **NCI** module,
then this is still the case after importing a **Galen** module.

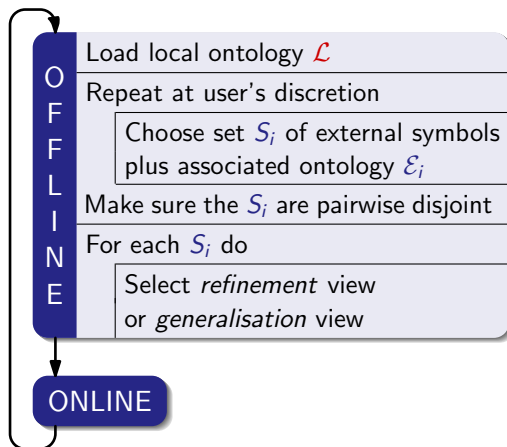
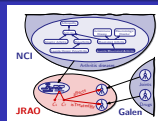
And now ...

- 1 Why ontology re-use?
- 2 A safe and economic methodology**
- 3 Tool support and experiments
- 4 Conclusion and perspectives

A working cycle: The Offline Phase

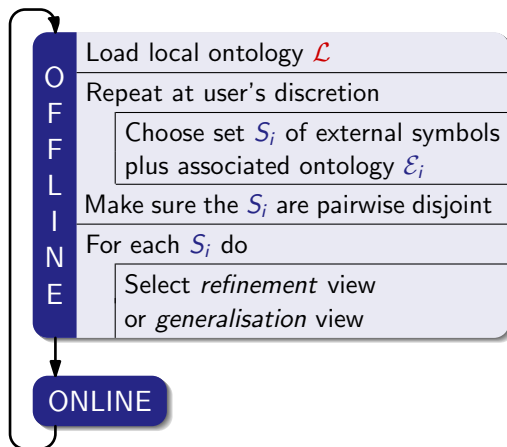
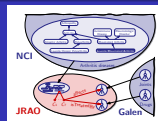


A working cycle: The Offline Phase



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

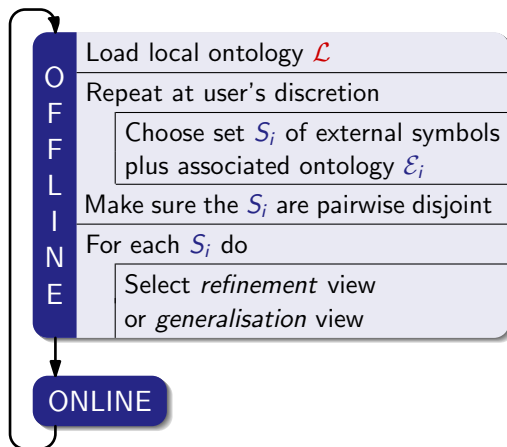
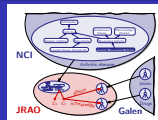
A working cycle: The Offline Phase



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

$$S_2 = \{KneeJoint, Fever\} \\ \mathcal{E}_2 = \mathbf{Galen}$$

A working cycle: The Offline Phase



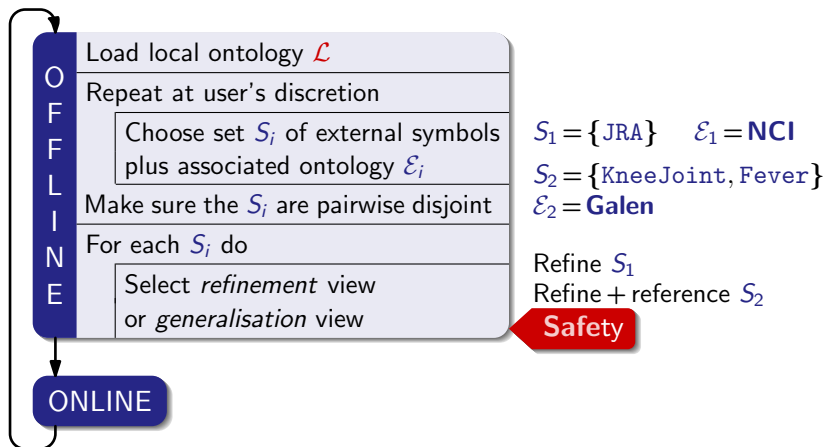
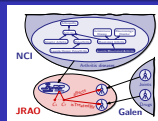
$S_1 = \{\text{JRA}\} \quad \mathcal{E}_1 = \text{NCI}$

$S_2 = \{\text{KneeJoint}, \text{Fever}\}$
 $\mathcal{E}_2 = \text{Galen}$

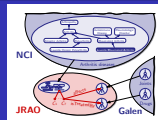
Refine S_1

Refine + reference S_2

A working cycle: The Offline Phase



Formalising the Safety Guarantee



Safety

Importing terms doesn't change their meaning.

Definition (Safety)

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

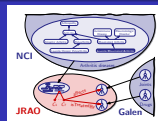
For every \mathcal{E}'_i with $\text{Sig}(\mathcal{L}) \cap \text{Sig}(\mathcal{E}'_i) \subseteq S_i$,

for all axioms α with $\text{Sig}(\alpha) \subseteq S_i$, $\mathcal{L} \cup \mathcal{E}'_i \models \alpha$ iff $\mathcal{E}'_i \models \alpha$.

Example

$\text{JRAO} \cup \text{NCI} \models \text{JRA} \sqsubseteq \text{GeneticDisorder}$
iff $\text{NCI} \models \text{JRA} \sqsubseteq \text{GeneticDisorder}.$

Providing safety



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

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

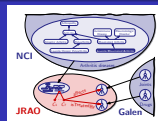
Locality ...

- is a syntactic approximation of conservativity.
- can be decided efficiently; conservativity often can't.
- comes in two “flavours” for refinement/generalisation.
- is sufficient, but not necessary, for safety.

If non-local axioms are found, the user may want to repair \mathcal{L} .

Example: $C_7 \sqsubseteq \text{JRA}$ ✓ $\text{GeneticDisorder} \sqsubseteq C_7$ ✓
 $\text{JRA} \sqsubseteq \text{GeneticDisorder}$ ✗

Providing safety



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

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

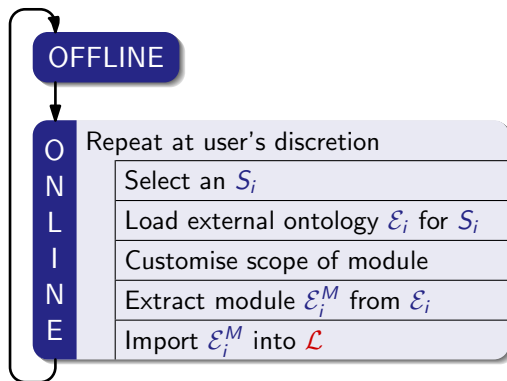
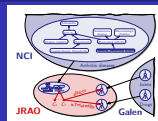
Locality ...

- is a syntactic approximation of conservativity.
- can be decided efficiently; conservativity often can't.
- comes in two “flavours” for refinement/generalisation.
- is sufficient, but not necessary, for safety.

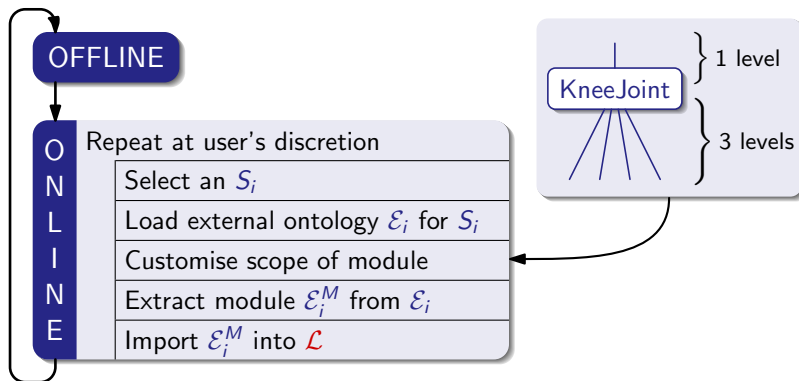
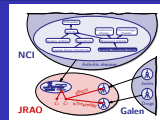
If non-local axioms are found, the user may want to repair \mathcal{L} .

Example: $C_7 \sqsubseteq \text{JRA} \checkmark$ $\text{GeneticDisorder} \sqsubseteq C_7 \checkmark$
 $\text{JRA} \sqsubseteq \text{GeneticDisorder} \times$

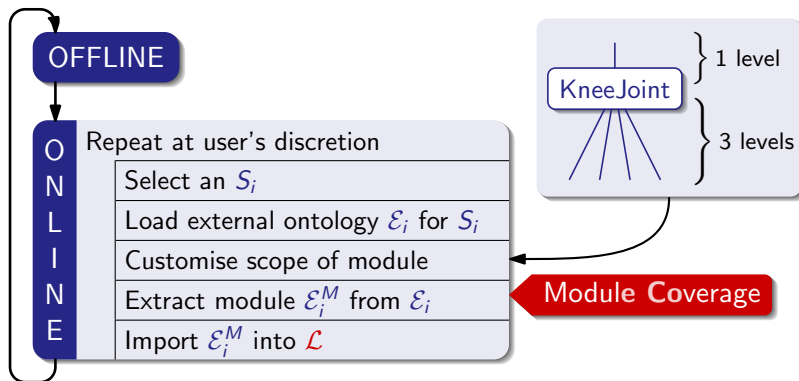
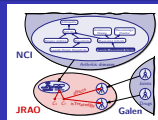
The Online Phase



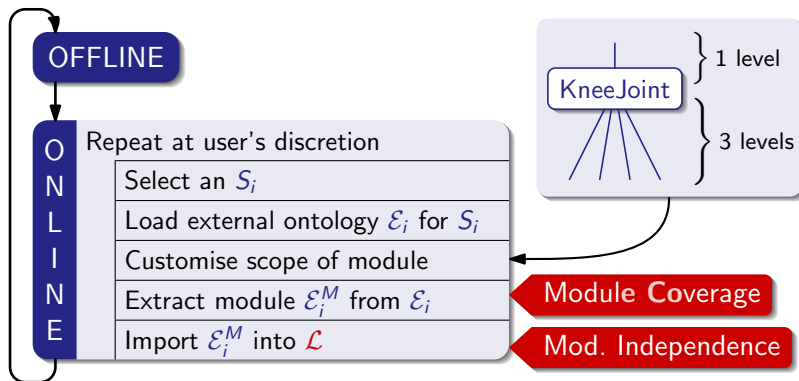
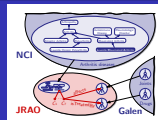
The Online Phase



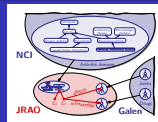
The Online Phase



The Online Phase



Formalising the Module Coverage Guarantee



Module coverage

Import all relevant parts of external ontologies.

Definition (Module coverage)

Let $\mathcal{E}_i^M \subseteq \mathcal{E}_i$ such that $S_i \subseteq \text{Sig}(\mathcal{E}_i^M)$.

\mathcal{E}_i^M guarantees coverage of S_i if:

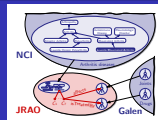
For every \mathcal{L}' with $\text{Sig}(\mathcal{L}') \cap \text{Sig}(\mathcal{E}_i) \subseteq S_i$,

for all ax. α with $\text{Sig}(\alpha) \subseteq S_i$, $\mathcal{L}' \cup \mathcal{E}_i \models \alpha$ iff $\mathcal{L}' \cup \mathcal{E}_i^M \models \alpha$.

Example

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

Formalising the Module Coverage Guarantee



Module coverage

Import all relevant parts of external ontologies.

Definition (Module coverage)

Let $\mathcal{E}_i^M \subseteq \mathcal{E}_i$ such that $S_i \subseteq \text{Sig}(\mathcal{E}_i^M)$.

\mathcal{E}_i^M guarantees coverage of S_i if:

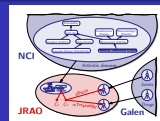
For every \mathcal{L}' with $\text{Sig}(\mathcal{L}') \cap \text{Sig}(\mathcal{E}_i) \subseteq S_i$,

for all ax. α with $\text{Sig}(\alpha) \subseteq S_i$, $\mathcal{L}' \cup \mathcal{E}_i \models \alpha$ iff $\mathcal{L}' \cup \mathcal{E}_i^M \models \alpha$.

Example

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

Providing the Module Coverage Guarantee



- Coverage is provided by 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 and experiments**
- 4 Conclusion and perspectives

Safe Protégé Manager: a plugin

See demo ...

“Synthetic” Experiments

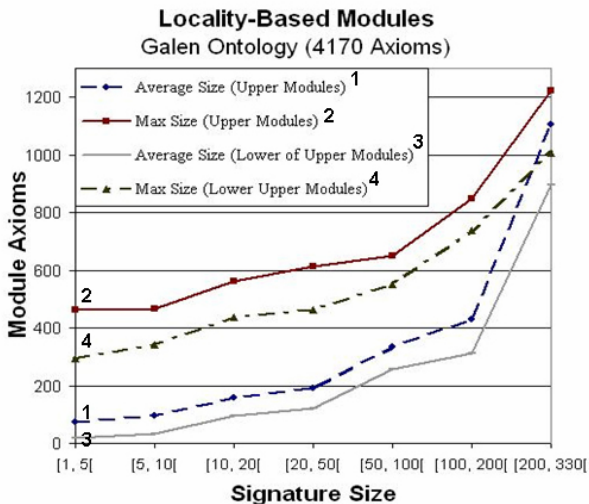
Setting

- Took arbitrary class name from **Galen** or **NCI**
- Added 0 . . . 3 levels of super/subclasses
- ↪ Signature of size 1 . . . 330
- Computed modules UM and LUM for each such signature

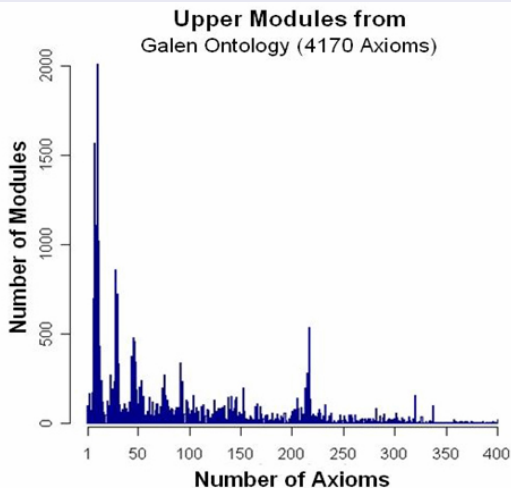
Results

- 99 % of **Galen** UM contain $< 10\%$ of **Galen**'s axioms
- 99 % of **Galen** LUM contain $< 5\%$ of **Galen**'s axioms
- similar findings for **NCI**

Statistics



Statistics



“Real-life” Experiments

Setting

- Health-e-Child context: JRA + Cardiomyopathies
- Manually selected + expanded signatures from **Galen** and **NCI**
- Computed LUM

Results

Disease	Ext. Ont.	# Sig.	# axioms	# classes	# properties
JRA	Galen	11	105	96	20
JRA	Galen	76	736	427	119
CMP	Galen	72	620	363	99
JRA	NCI	18	488	18	
JRA	NCI	144	5057	312	14
CMP	NCI	124	4751	321	11

Comparing Experiments

Setting

- SNOMED (health care; restricted language; 350,000 axioms)
- Initial signatures: terms from intensive care unit
- Computed LUM; conservativity-based modules (Wolter et al.); Seidenberg/Rector segments; PROMPT segments (Noy, Musen)

Results

# Sig.	# axioms segment : SNOMED		
	Wol++	Sei/Rec	LUM
≈ 4,000	2 %	2 %	4 %
≈ 16,000	7 %	7 %	10 %
≈ 24,000		10 %	15 %
time	4–5 s	< 1 s	4–7 s

And now ...

- 1 Why ontology re-use?
- 2 A safe and economic methodology
- 3 Tool support and experiments
- 4 Conclusion and perspectives**

Conclusion

- Logic-based approach to re-use of ontologies that is ...
 - **safe** — importing terms doesn't change their meaning
 - **economic** — the relevant parts are imported
- Re-use methodology
- Tool support
- Work in progress!

Perspectives

- Extend module scope customisation: “shopping for symbols”
 - Browse external ontology
 - Pick symbols
 - At each stage, view resulting module
 - “Check out” module
- Optimise module extraction
- Perform user study and improve interface
- Import “by reference” as opposed to “by value”
- Multi-user scenario

We want you...

- ... to test our plugin and give us feedback
 - `protege.stanford.edu`
 - `krono.act.uji.es/people/Ernesto/safety-ontology-reuse`
- ...r favourite ontologies and real-life signatures!

We want you...

- ... to test our plugin and give us feedback
 - `protege.stanford.edu`
 - `krono.act.uji.es/people/Ernesto/safety-ontology-reuse`
- ...r favourite ontologies and real-life signatures!

Thank you!

More links

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