

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

Ernesto Jiménez-Ruiz¹ Bernardo Cuenca Grau²
Ulrike Sattler³ *Thomas Schneider*³ Rafael Berlanga¹

¹Computer Languages and Systems, Universitat Jaume I, Spain

²Computing Laboratory, University of Oxford, UK

³Computer Science, University of Manchester, UK

Ontogenesis, 25 April 2008

Our approach in a nutshell

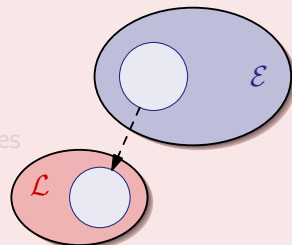
Logic-based methodology for the re-use of ontologies

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 ProSÉ
- Work in progress!

Our approach in a nutshell

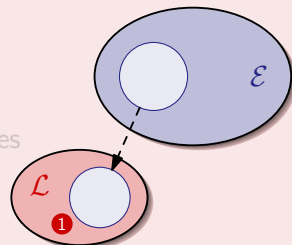
Logic-based methodology for the re-use of ontologies

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 ProSÉ
- Work in progress!

Our approach in a nutshell

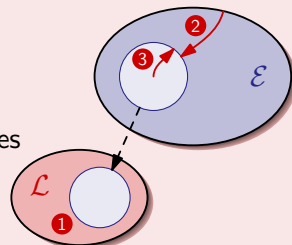
Logic-based methodology for the re-use of ontologies

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 ProSÉ
- Work in progress!

Our approach in a nutshell

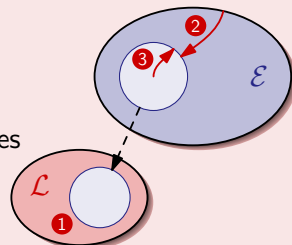
Logic-based methodology for the re-use of ontologies

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! ✓

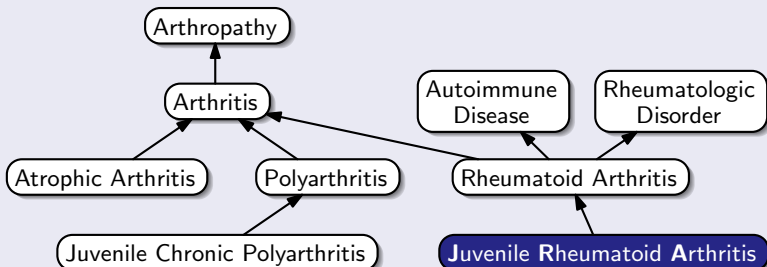


- Tool support — Protégé plugin ProSÉ
- Work in progress!

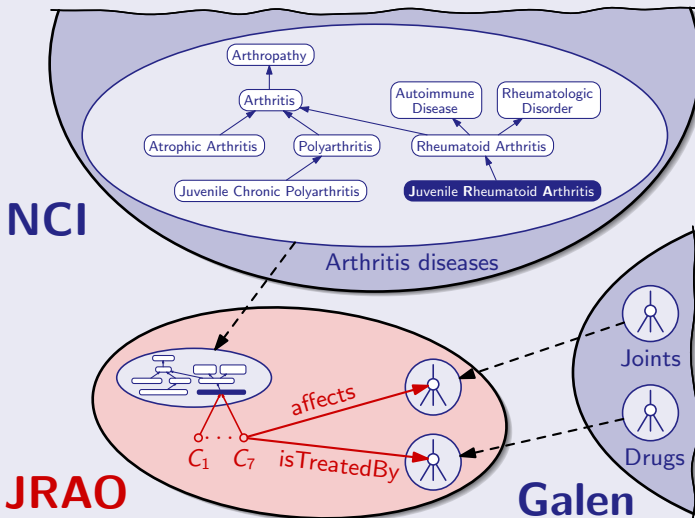
And now ...

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

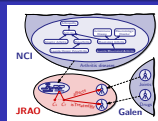
Reasons for re-use

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

Guarantees to provide

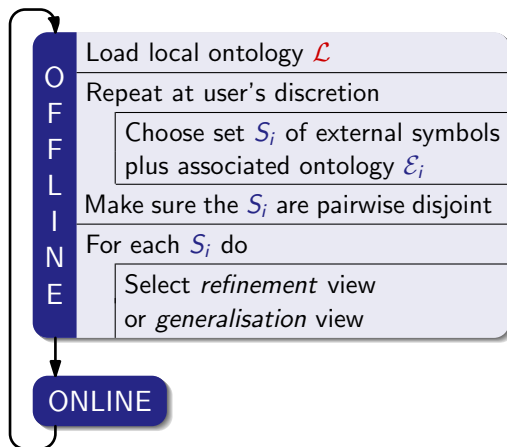
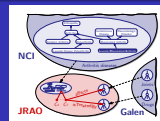
- **[safe]** Importing terms doesn't change their meaning.
- **[eco]** Import all relevant parts of external ontologies.
- **[aux]** The order of imports doesn't matter.

And now ...

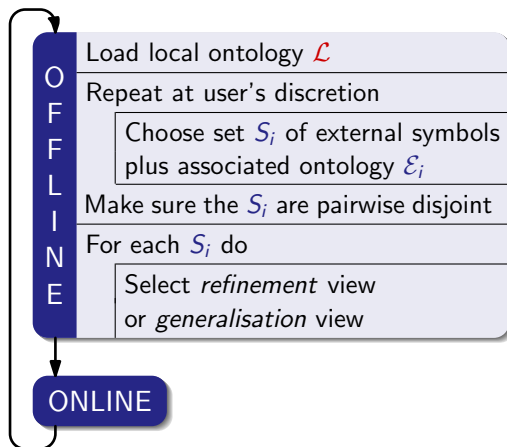
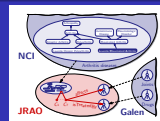


- 1 Why ontology re-use?
- 2 A safe and economic methodology
- 3 Tool support and Experiments
- 4 Perspectives

A working cycle: the offline phase

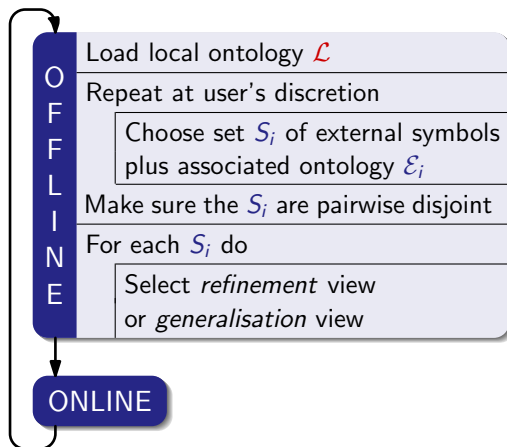
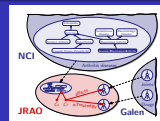


A working cycle: the offline phase



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

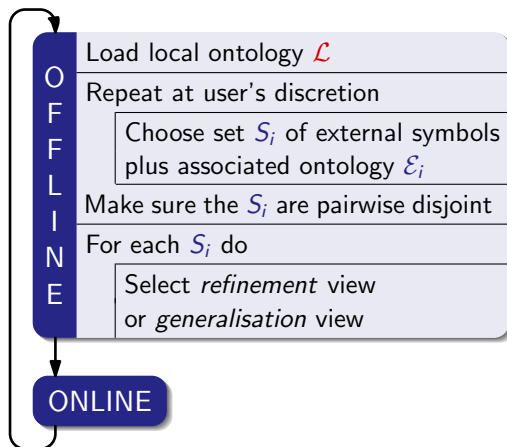
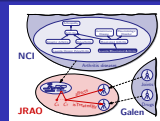
A working cycle: the offline phase



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

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

A working cycle: the offline phase



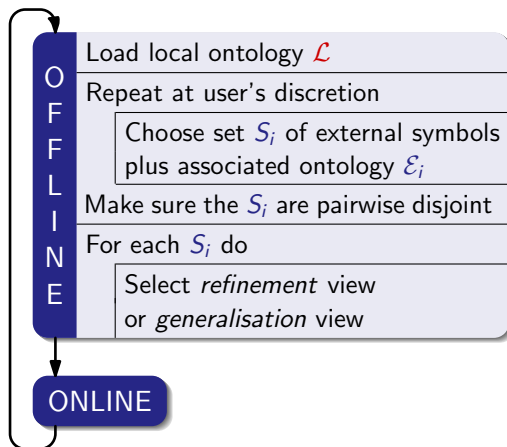
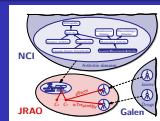
$$S_1 = \{JRA\} \quad \mathcal{E}_1 = \mathbf{NCI}$$

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

Refine S_1

Refine + reference S_2

A working cycle: the offline phase



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

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

Refine S_1

Refine + reference S_2

Safety

The diagram shows three interconnected ontologies. At the top is the NCI ontology, which includes a hierarchy of terms like 'Disease', 'Organ', and 'Anatomical structure'. Below it is the JRAO ontology, which includes terms like 'Disease', 'Organ', and 'Anatomical structure'. To the right is the Galen ontology, which includes terms like 'Disease', 'Organ', and 'Anatomical structure'. Arrows indicate relationships between these ontologies, with labels such as 'Artificial diseases', 'Disease', and 'Therapy'.

Safety

Importing terms doesn't change their meaning.

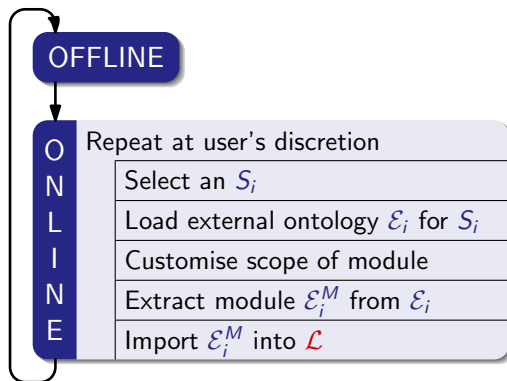
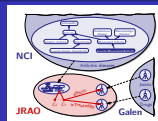
Example

$$\begin{array}{l} \text{JRAO} \cup \text{NCI} \models \text{JRA} \sqsubseteq \text{GeneticDisorder} \\ \text{iff} \quad \text{NCI} \models \text{JRA} \sqsubseteq \text{GeneticDisorder}. \end{array}$$

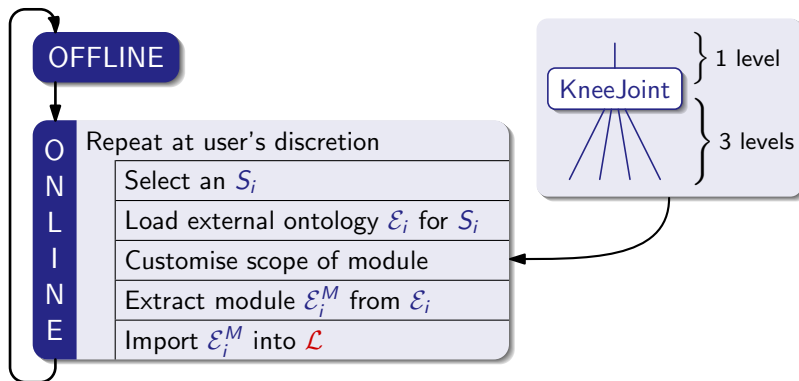
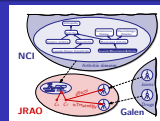
Safety is approximated by locality.

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

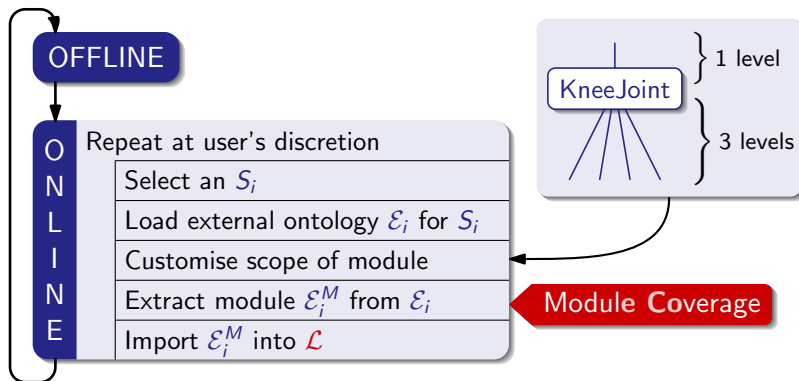
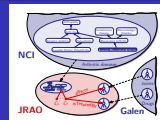
The online phase



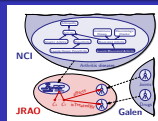
The online phase



The online phase



Formalising the *Module Coverage Guarantee*



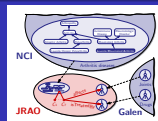
Module coverage

Import all relevant parts of external ontologies.

Example

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

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 and Experiments**
- 4 Perspectives

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

http://semanticweb.org/Ontology1206804237717.owl - [Users/schneidt/Documents/DLista/2008/prose/Experiments/HeC_Use_Cases/JRAO.owl]

Ontology1206804237717.owl http://semanticweb.org/Ontology1206804237717.owl

Active Ontology Entities Classes Object Properties Data Properties Individuals OWL Viz DL Query ProSE Manager

ProSE External-Local Class Hierarchy: Oedema

- Thing
 - Juvenile_Rheumatoid_Arthritis
 - Oedema
 - Anti-DNA_Antibody
 - Azathioprine
 - Cisplatin_Cyclosporine
 - Cyclosporine
 - Erosion
 - Etanercept
 - Hemoglobin
 - Infliximab
 - Interleukin_Gene
 - LymphocyteCount
 - Methotrexate
 - NeutrophilCount
 - Nonsteroidal_Antinflammatory_Drug
 - Pain
 - PlateletCount
 - Prednisone
 - Rheumatoid_Factor
 - SynovialFluid
 - SynovialJoint
 - Tumor_Necrosis_Factor_Family_Gene
 - WestergrenESRProcedure

ProSE Safe Protege Manager

Select Signature Group Create/Modify External URIs

External Signature Groups

- http://www.mindswap.org/2003/CancerOntology/nciOntology
- http://krono.act.uji.es/Links/ontologies/galen.owl
- No defined External URI

Locality Type:

Bottom Locality ▲ Non-Local Axioms Detected

Non-Local Axioms for selected External Signature

Non-Local Axioms

Oedema subClassOf Juvenile_Rheumatoid_Arthritis

Extension of the selected Signature (Optional)

Signature from: http://krono.act.uji.es/Links/ontologies/galen.owl

☒ Extend Signature with Subclasses. Levels: 1

☒ Extend Signature with Superclasses. Levels: 1

Extend Signature Clear Extended Signature Preview Signature

URI for Module

URI for Module: ct.uji.es/ontologies/module1209059675139Fromgalen.owl

Module Coverage

☐ Locality-based Module

☒ Minimal Locality-based Module

Importing Information

Selected URI was defined to be used for refinement and/or reference. An Lower of Upper Module (LUM) for the

Extract Entity Annotation Axioms

Importing-Preview Actions

1. Extract Module

a. Preview Module

b. Import Module

b. Come Back

2. Import Whole Ontology

Extracted Module Information

Number of Axioms (Module/Ontology): 140 / 4170

Number of Classes (Module/Ontology): 98 / 2748

Number of Properties (Module/Ontology): 25 / 413

Number of Individuals (Module/Ontology): 0 / 0

“Synthetic” experiments

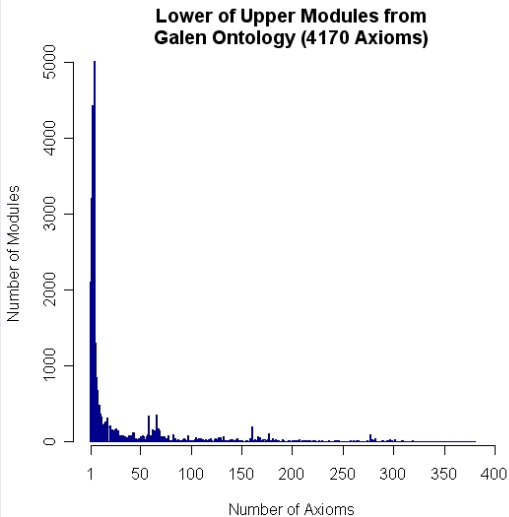
Setting

- Randomly generated signatures of size 1 . . . 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



“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 %)

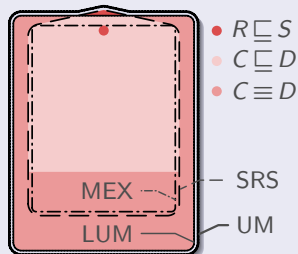
Comparing experiments

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



And now ...

- 1 Why ontology re-use?
- 2 A safe and economic methodology
- 3 Tool support and Experiments
- 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!

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