Metarel: An Ontology to Support the Inferencing of Semantic Web Relations within Biomedical Ontologies

Ward Blondé

Dept. of Applied Math., Biometrics and Process Control

University of Ghent, Belgium

ward.blonde@ugent.be

Contents

1. Introduction

2. Approach and architecture

3. Relational closures

4. Results

5. Conclusions

Introduction

- The ever increasing biological knowledge needs automated integration

- Ontologies provide a good scaffold for such integration:

    – High expressivity for detailed knowledge

    – Extensional knowledge can be inferred from a maintained core (the intensional knowledge)

    ➔ The Semantic Web can help us out

Introduction

- Biomedical ontologies contain only the intensional knowledge → the scaffold
- The extensional knowledge can be inferred from this scaffold by computational reasoning
- The rules that say which inferences can be made are mostly provided by the semantics of the relations in the ontology
  → We automated inferences starting from Metarel, an ontology that captures the semantics of relations

Next section

1. Introduction

2. Approach and architecture

3. Relational closures

4. Results

5. Conclusions

Reasoning with METAREL

- Metarel was created to allow relational inferencing in ontologies

- It is a meta-ontology for relations that exists in OBOF and RDF, compatible with RO

- Relational closures like reflexivity, transitivity, compositions, etc., can be created by using SPARUL, the update language for RDF

- An application in the RDF store BioGateway on the Gene Ontology (GO) and the GO Annotations (GOA) created 200 million inferences

BioGateway

Erick Antezana, Ward Blondé, Mikel Egana, Alistair Rutherford, Robert Stevens, Bernard De Baets, Vladimir Mironov and Martin Kuiper. BioGateway: a Semantic Systems Biology tool for the life sciences, BMC, 2009

Application architecture
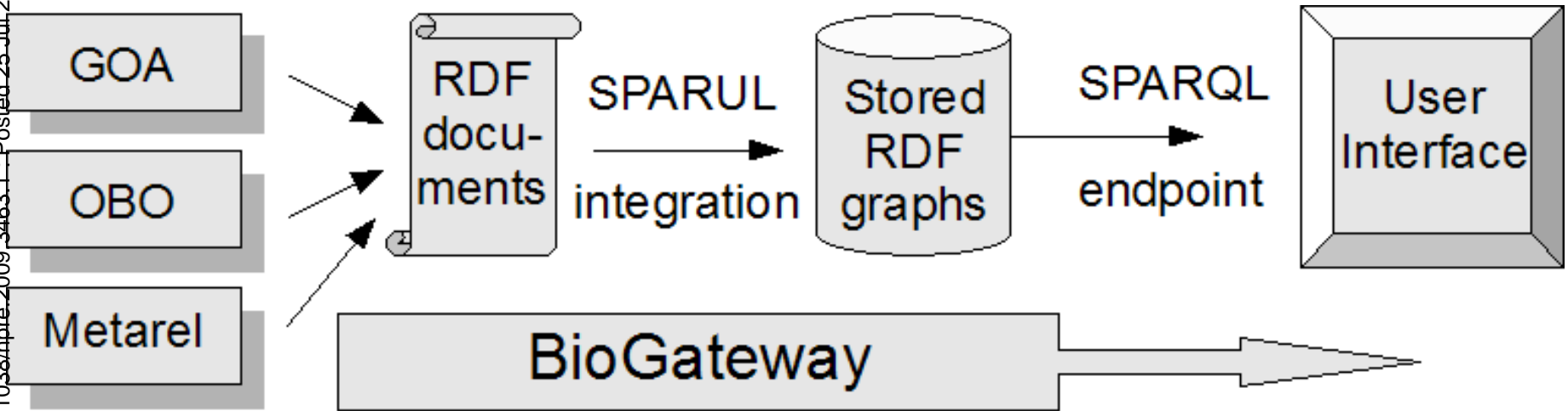
1. Biomedical ontologies and Metarel are translated to RDF
2. Integration and relational inferencing happens during the loading into the Virtuoso triple store with SPARUL
3. SPARQL querying can be done with the aid of a user interface
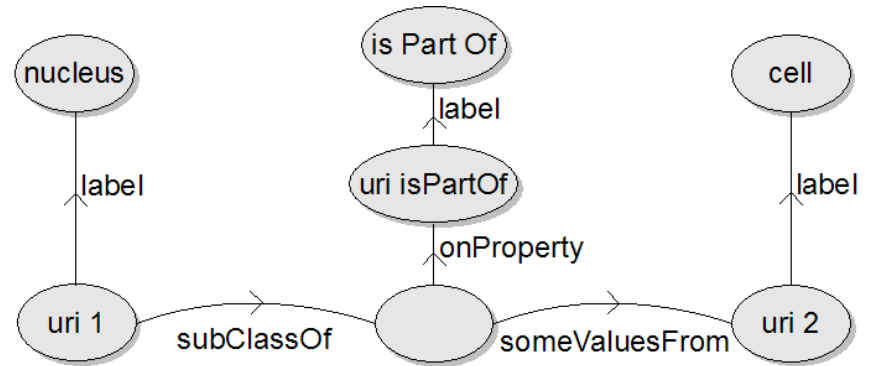
Metarel architecture

Two hierarchies for relations:
1. Is related to: Classes in this hierarchy can be written with a verb in the third person singular, like 'is a', 'is part of', 'is located in', …
2. Relation type: Classes in this hierarchy are meta-classes. Classes in hierarchy 1 are instances in this hierarchy.
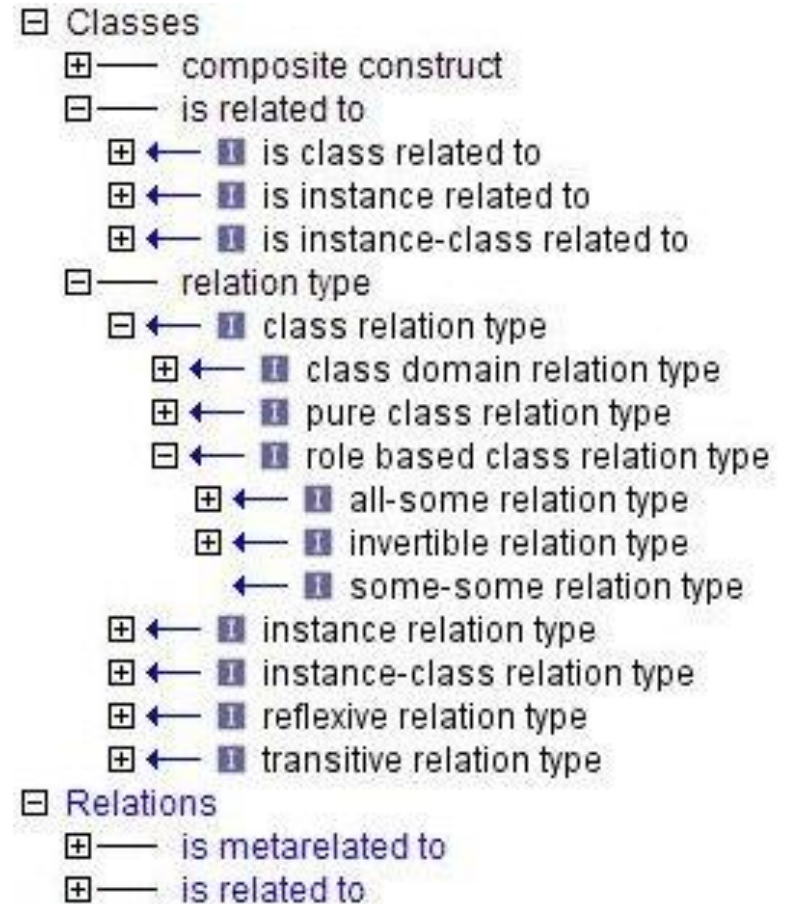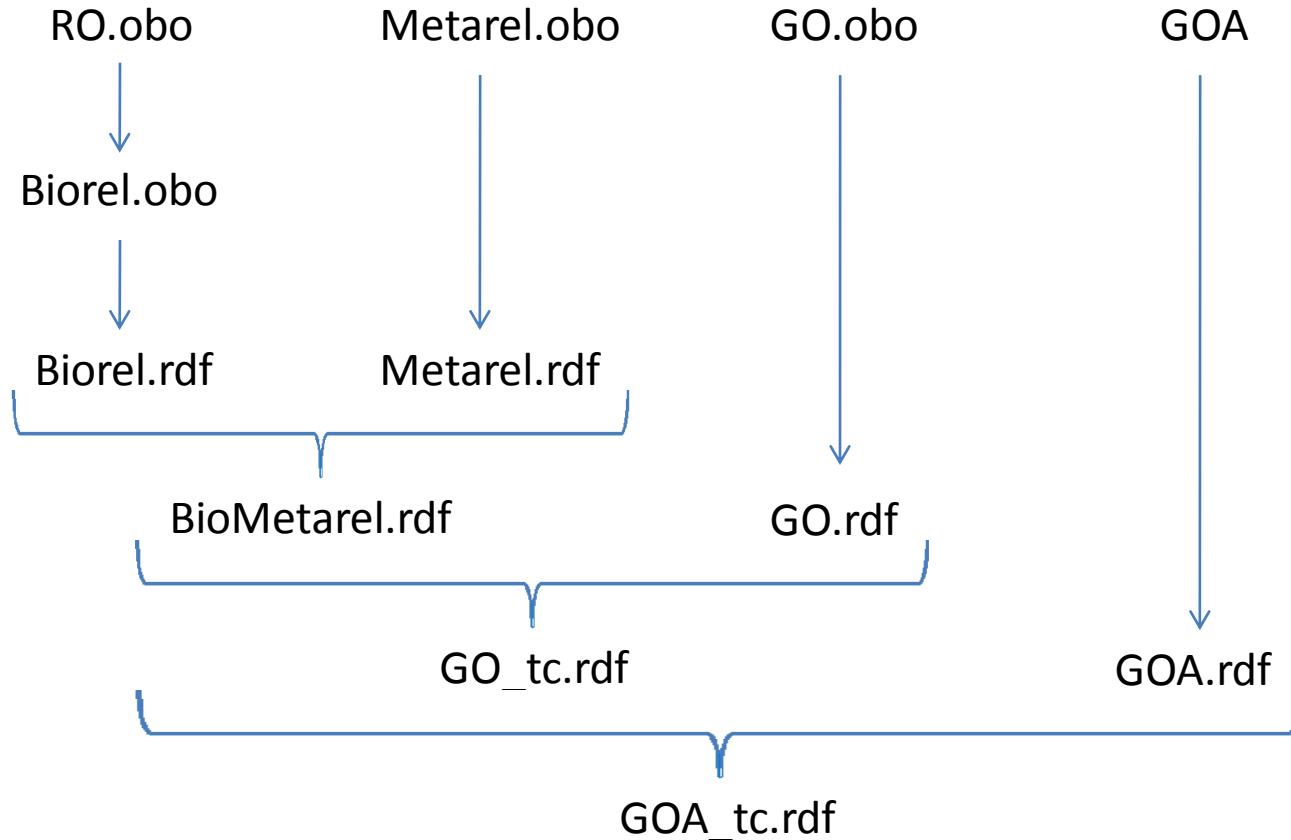
The ease of OBOF in RDF
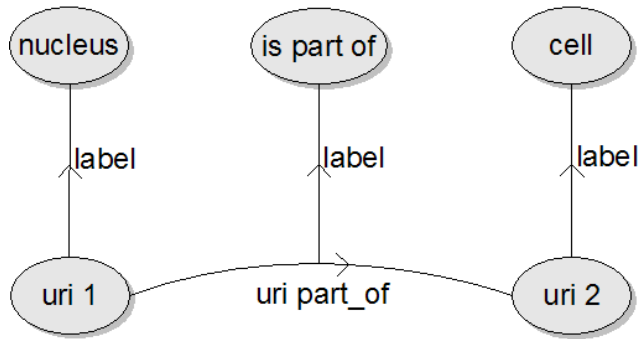
Direct link in OBOF/RDF

Indirect link in OWL/RDF

OBOF provides an all-some semantics for relations between classes, modeled with a single arrow.

METAREL integration with SPARUL

_tc: RDF graphs with total relational closures

A SPARUL closure query

```
BASE   <http://www.semantic-systems-biology.org/>
PREFIX ssb:<http://www.semantic-systems-biology.org/SSB#>
INSERT INTO GRAPH <25.H_sapiens_tc> {
 ?class1 ?resulting_relation ?class3.
}
WHERE {
 GRAPH <25.H_sapiens_tc> {
  ?class1 ?first_relation ?class2.
 }
 GRAPH <gene_ontology_edit_tc> {
  ?class2 ?second_relation ?class3.
 }
 GRAPH <biometarel> {
  ?composite ssb:first_relation ?first_relation.
  ?composite ssb:second_relation ?second_relation.
  ?composite ssb:resulting_relation ?resulting_relation.
 }
}
```
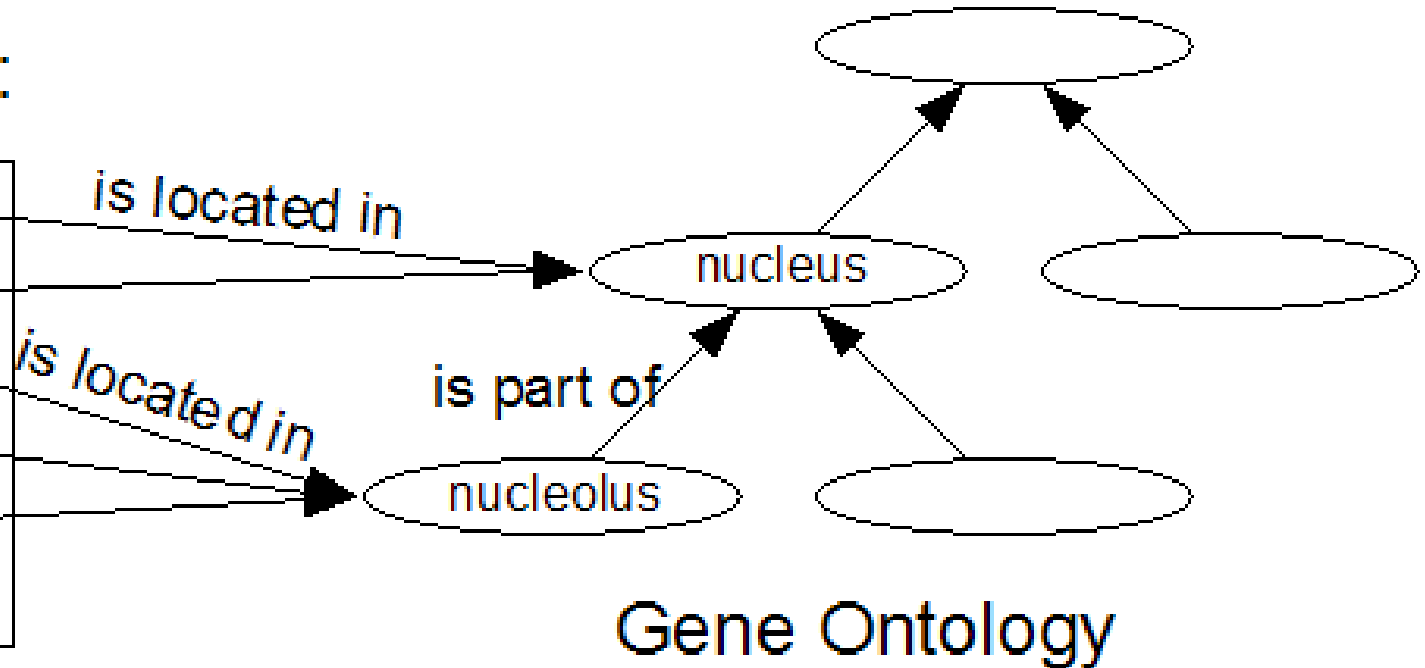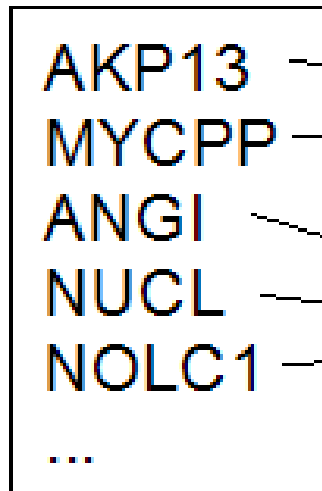
With this query, any pattern of composite relations is automatically retrieved from BioMetarel and updated over GO Annotations.

Next section

1. Introduction

2. Approach and architecture

3. Relational closures

4. Results

5. Conclusions

Problem statement

Proteins:

AKP13 — is located in → nucleus

MYCPP

ANGI — is located in

NUCL

NOLC1 → nucleolus

...

is part of

Gene Ontology
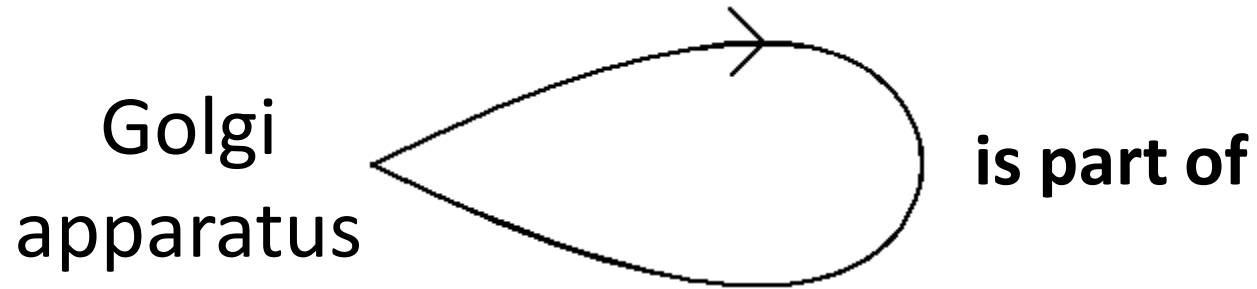
We want to get all the proteins that are located in the nucleus (with an easy SPARQL query).

Relational closures

- Reflexivity:

Golgi apparatus   **is part of**

Relational closures

- Transitivity:

is part of   is part of

nucleolus   nucleus

nuclear lumen

is part of
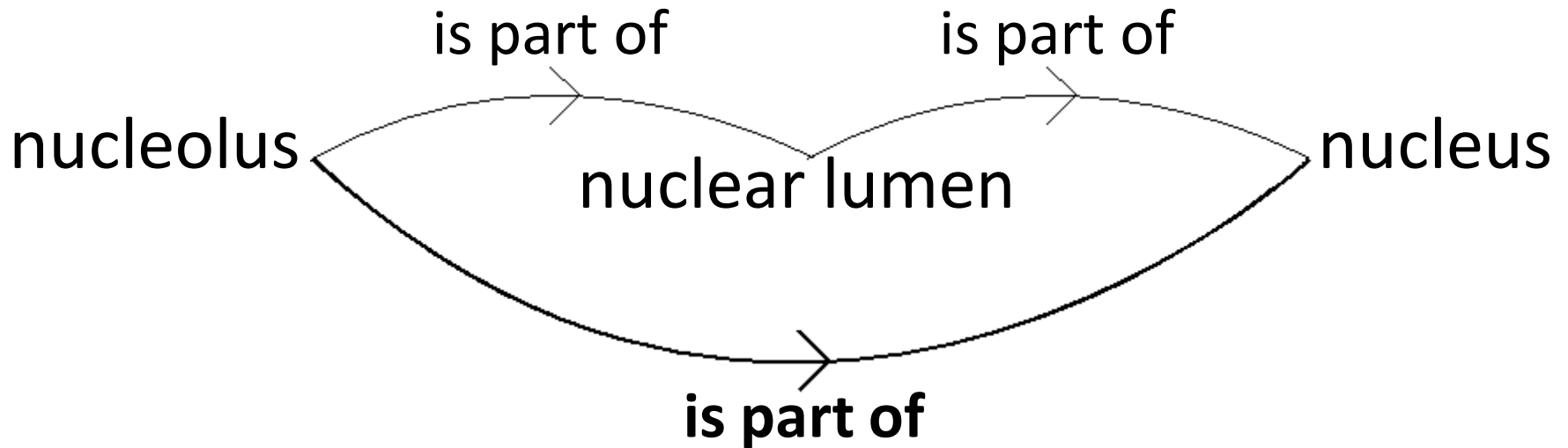
Relational closures

- Priority over is_a:

BRAF1_
HUMAN

has function

diaglycerol binding

is a

lipid
binding

has function
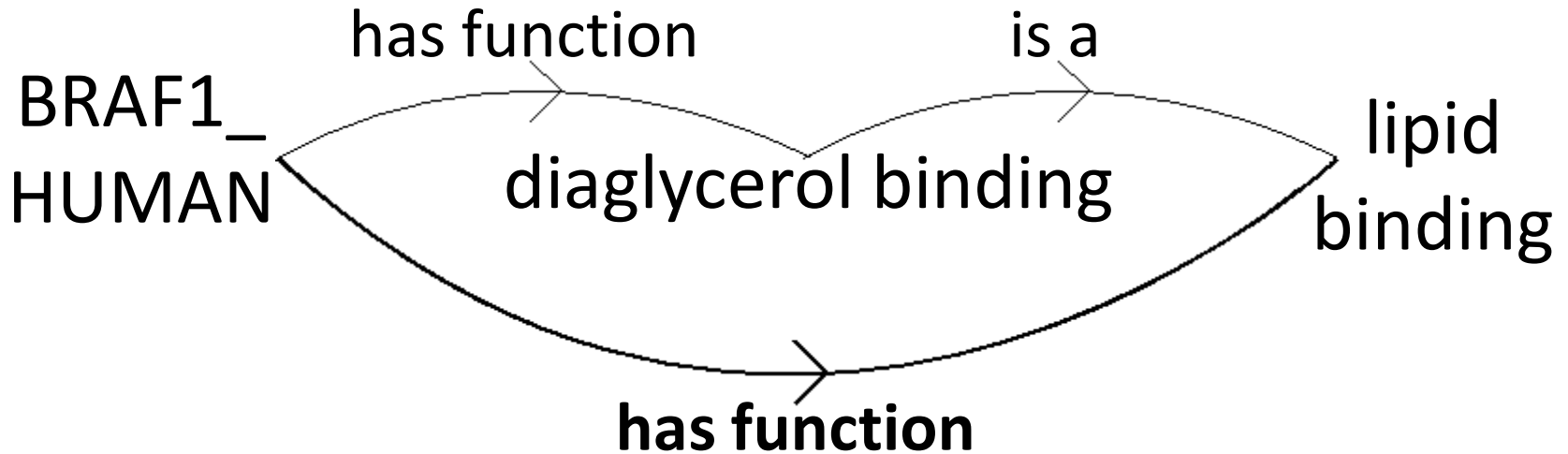
Relational closures

- Superrelations:

negatively regulates

AKIP_
HUMAN   mitosis

regulates
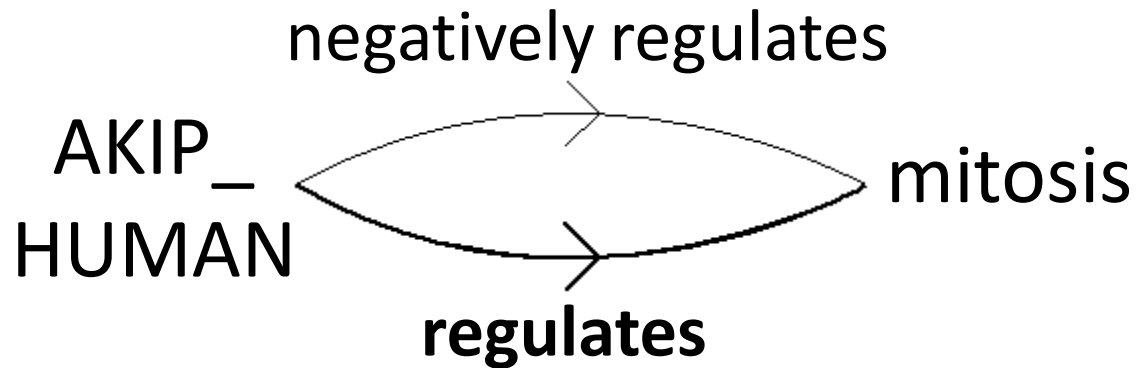
Relational closures

- Compositions:
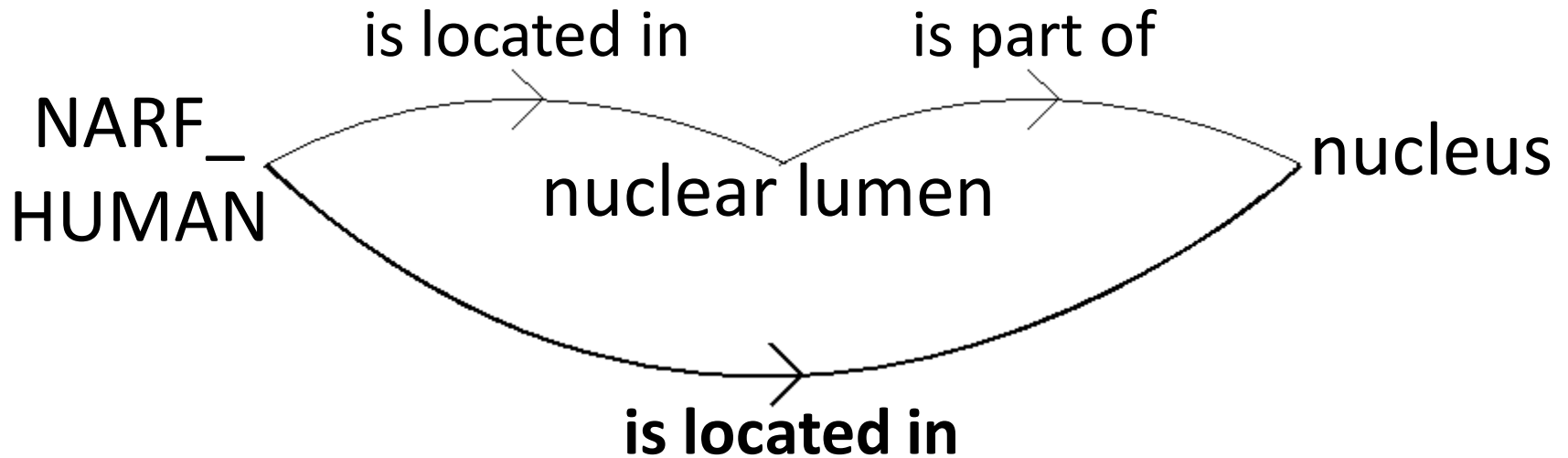
Relational closures

- Inverses? No!

is part of

uterus   abdomen

has part

OK:

is part of

nuclear lumen   nucleus

has part
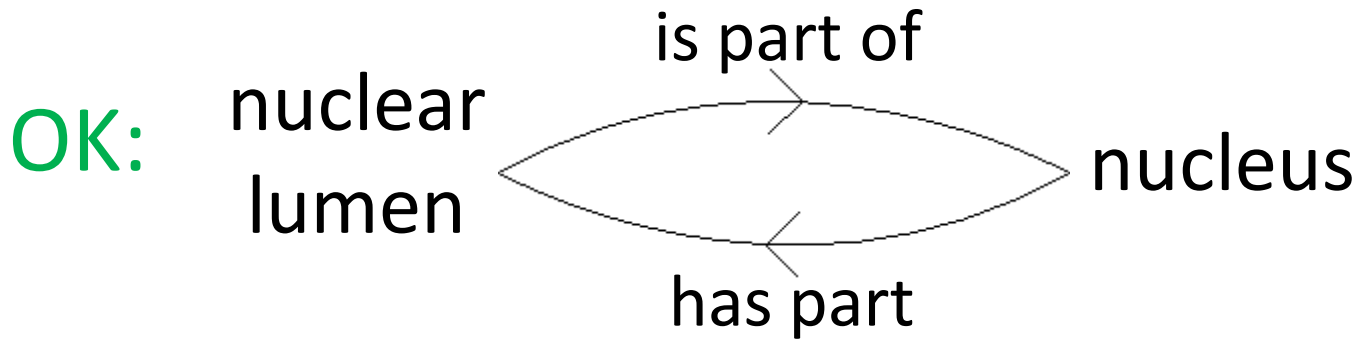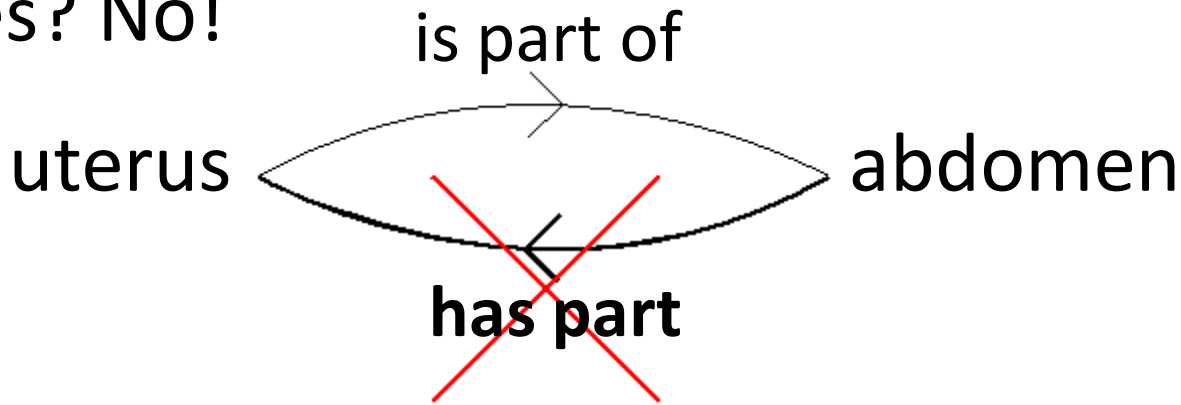
OBO ontologies need inverse relations where automated systems cannot infer them!

Next section

1. Introduction

2. Approach and architecture

3. Relational closures

4. Results

5. Conclusions

Results

- Millions of new relations were inferred:
  GO: from 0.57 m triples to 1.2 m triples
  GOA (human): from 3.3 m to 4.0 m triples
  GOA (all): about 200 million triples

- SPARQL queries on the closure graphs give *all* the expected results in a short time

- Complex queries can easily be addressed in this architecture

A complex query

Which proteins, in which organisms
- have function ion channel activity
- are located in the lysosome
- participate in ion transport   ??

9 proteins are found for direct annotations

11 proteins are found in closure graphs: subclasses in the hierarchy are also considered now

Simple queries get a simple syntax

```
BASE  <http://www.semantic-systems-biology.org/>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX ssb:<http://www.semantic-systems-biology.org/SSB#>
SELECT distinct ?protein
WHERE {
 {
  GRAPH <25.H_sapiens_tc> {
  ?protein ssb:located_in ssb:GO_0005634.
  }
 }
}
```

Query with total closures for GOA.

The same* query with only transitive closures in GO. (* But didn't we forget something?)

```
BASE  <http://www.semantic-systems-biology.org/>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX ssb:<http://www.semantic-systems-biology.org/SSB#>
SELECT distinct ?protein
WHERE {
 {
 GRAPH <25.H_sapiens> {
  ?protein ssb:located_in ssb:GO_0005634.
 }
 } UNION {
 GRAPH <25.H_sapiens> {
  ?protein ssb:located_in ?sublocation.
 }
 GRAPH <gene_ontology_edit_tc> {
  ?sublocation ssb:is_a ssb:GO_0005634.
 }
 } UNION {
 GRAPH <25.H_sapiens> {
  ?protein ssb:located_in ?sublocation2.
 }
 GRAPH <gene_ontology_edit_tc> {
  ?sublocation2 ssb:part_of ssb:GO_0005634.
 }
 } UNION {
 GRAPH <25.H_sapiens> {
  ?protein ssb:located_in ?sublocation3.
 }
 GRAPH <gene_ontology_edit_tc> {
  ?sublocation3 ssb:part_of ?sublocation4.
  ?sublocation4 ssb:is_a    ssb:GO_0005634.
 }
 } UNION {
 GRAPH <25.H_sapiens> {
  ?protein ssb:located_in ?sublocation5.
 }
 GRAPH <gene_ontology_edit_tc> {
  ?sublocation5 ssb:is_a ?sublocation6.
  ?sublocation6 ssb:part_of ssb:GO_0005634.
 }
 }
}
```
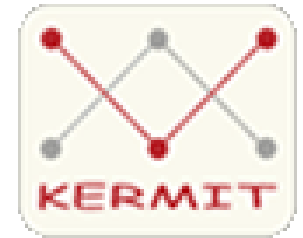
Last section

1. Introduction

2. Approach and architecture

3. Relational closures

4. Results

5. Conclusions

Conclusions

- RDF-based ontologies benefit from pre-inferenced relational closures for performant SPARQL querying

- SPARUL can infer millions of RDF triples with a small set of closure queries

- Metarel can be used to initiate such closures by providing all the necessary semantics for relations

Acknowledgements

- Erick Antezana (UGent, BE)

- Nirmala Seethappan (NTNU, NO)

- Bjørn Lindi (NTNU, NO)

- Bernard De Baets (UGent, BE)

- Vladimir Mironov (NTNU, NO)

- Martin Kuiper (NTNU, NO)

http://www.semantic-systems-biology.org/metarel

Thanks for your attention!