LISBOA
UNIVERSIDADE DE LISBOA

LASIGE
Large-Scale Informatics Systems Laboratory

ISWC 2014
Riva del Garda - Trentino, Italy

UNIVERSITY OF OXFORD

Towards annotating potential incoherences in BioPortal mappings

Daniel Faria, Ernesto Jiménez-Ruiz, Catia Pesquita, Emanuel Santos, Francisco M. Couto

October 23rd, 2014

SOMER
http://somer.fc.ul.pt/

Optique

Outline

- Introduction
  - BioPortal
  - Close Match vs. Equivalence
  - Mapping Repair
- Evaluation
  - Dataset
  - Automatic Repair Evaluation
  - Manual Analysis
- Conclusions
- Future Work

Introduction

BioPortal

- 373 ontologies
- 13 million mappings between them
- Mappings represented as a 4-tuple ‹e_1; e_2; *Rel*; *Ann*›
- Relationships can be:
  - *skos:exactMatch* (same URI)
  - *skos:closeMatch* (used for LOOM and UMLS mappings)
  - *skos:relatedMatch* (used for OBO cross-references)
  - *skos:narrowMatch* / *skos:broadMatch*

Introduction

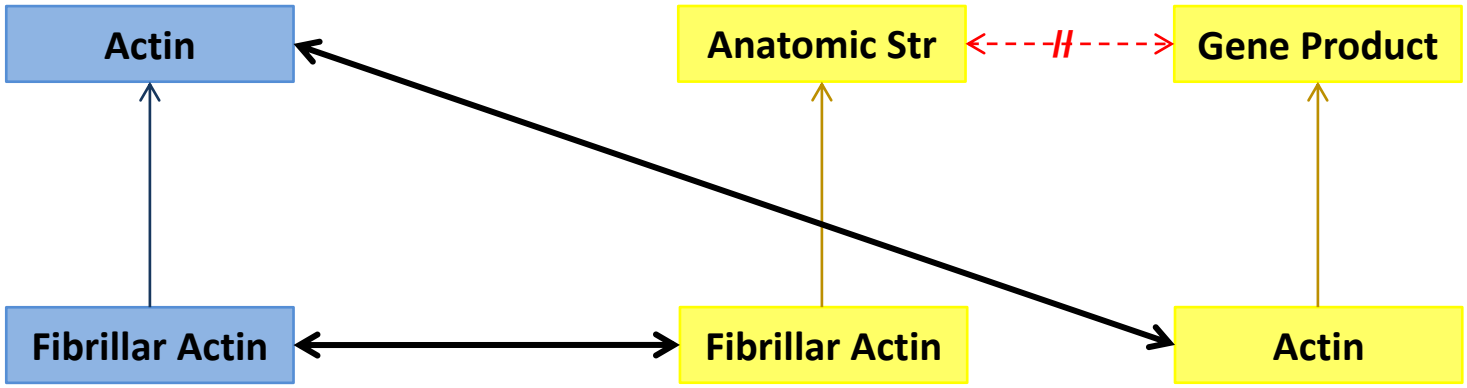
Close Match vs. Equivalence

- *skos:closeMatch* lacks a clear semantic meaning:
  - BioPortal need not be concerned with exact mapping semantics or their logical entailments
- *owl:equivalentClass* is a stronger assertion, but:
  - Without it, mappings are of limited use
- Most users will implicitly or explicitly assume equivalence
- Providing information about the logical entailments under this assumption will be helpful for these users

Introduction

Mapping Repair

- Integrating two ontologies through an alignment may lead to unsatisfiable classes (even if everything is “correct”, though generally it is a sign that something is wrong)



Introduction

Mapping Repair

- Incoherent alignment: set of mappings that causes unsatisfiable classes
- Conflict set: minimal set of mappings responsible for one or more unsatisfiabilities; removing one mapping will solve them
- Mapping repair: remove mappings from an incoherent alignment so as to increase its coherence; minimize the impact of the procedure on the alignment

Introduction

Mapping Repair

- Repair algorithms can be classified with regard to:
  - Completeness of analysis (complete vs. approximate repair)
  - Optimality of removal (optimal, global heuristic, or local heuristic)
- Complete analysis and optimal removal do not scale well for large ontologies with many unsatisfiabilities

Introduction

Mapping Repair

- AML-Repair: approximate, global heuristic (computes all conflict sets of mappings, then resolves them)
- LogMap-Repair: approximate, local heuristic (resolves conflict sets individually for each unsatisfiable class)
- Nearly equivalent with respect to completeness
- LogMap-Repair more efficient
- AML-Repair closer to optimality

Evaluation

Dataset

- 19 ontology pairs from BioPortal, such that:
  - Each pair had >500 mappings
  - At least one ontology had disjointness clauses
  - Both ontologies were biomedical
- 11 pairs had incoherent alignments

Evaluation

Automatic Repair Evaluation

- Completeness:
  - AML-Repair and LogMap-Repair resolve most unsatisfiable classes, produce coherent alignments in many cases
  - AML-Repair less complete than LogMap-Repair, but has been improved
- Optimality:
  - Repairs not aggressive (at most 6-7% of input mappings removed, in task where 61% involved in conflicts)

Evaluation

Automatic Repair Evaluation

- Relevance:
  - 11 of 19 BioPortal alignments incoherent
  - On average 22% of BioPortal mappings from incoherent alignments involved in conflict sets
  - Some key mappings are involved in multiple unsatifiabilities
- Efficiency:
  - Both systems tackled largest tasks in a few minutes

Evaluation

Manual Analysis

- Erroneous mappings:
  - Involve obsolete / deprecated / retired classes
  - Non-related entities (back [body part] vs. back [direction])
  - Related entities but not closely (a gene vs. the protein it encodes)
- Correct mappings but in conflict with erroneous ones
- Errors account for over 60% of removed mappings

Conclusions

- Identifying incoherence-causing BioPortal mappings is:
  - Critical for applications that require logical coherence
  - Helpful for finding potentially erroneous mappings

- Approximate mapping repair systems are:
  - Capable of tackling large datasets such as BioPortal
  - Able to identify most logical conflicts

Future Work

- Goals:
  - Analyze all BioPortal mappings
  - Extend BioPortal by storing and enabling querying of annotations about mappings involved in logical conflicts
  - Develop lightweight mapping analyzer, to analyze new mappings upon submission (preclude errors, remain up to date)
- Challenges:
  - Analyze new mappings on-the-fly, particularly involving huge ontologies
  - Tackle complex mapping networks between three or more ontologies?

Acknowledgments & Notices

We are grateful to Ray Fergerson for his ready interest and assistance in our goals to extend BioPortal.

We would also like to thank Bernardo Cuenca Grau, Ian Horrocks and Isabel F. Cruz for their invaluable help in the development of LogMap and AML.

This work was supported by the EU FP7 IP project Optique (no. 318338) and the EPSRC project Score!, and by the Portuguese FCT through the SOMER project (PTDC/EIA-EIA/119119/2010) and the LASIGE Strategic Project (PEst-OE/EEI/ UI0408/2014).

LogMap is open source and available at https://code.google.com/p/logmap-matcher/

AML is open source and available at https://github.com/AgreementMakerLight

There is a PhD scholarship opportunity in our research group: http://goo.gl/KrYJpS

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA E DO ENSINO SUPERIOR

SOMER
http://somer.fc.ul.pt/

Optique