Towards a Top-Domain Ontology for Linking Biomedical Ontologies

Holger Stenzhorn[a,b]

Elena Beißwanger[c]

Stefan Schulz[a]

[a] Department of Medical Informatics, Freiburg University Hospital, Germany

[b] Institute for Formal Ontology and Medical Information Science (IFOMIS), Saarbrücken, Germany

[c] Jena University Language and Information Engineering (JULIE) Lab, Germany

Introduction

- Tremendous increase in collected data in life sciences in the last years – especially in biomedicine with its genomics research

- Obvious need to analyze and structure this gained data to extract new knowledge from it

- Our focus is on the scientific efforts to use ontologies to achieve this particular task

Problems with Current Ontologies

- Most existing biomedical ontologies are developed mostly independently – even if they have overlapping content

- Each ontology only embraces a distinct scenario with a mere partial view of the field

- An overarching resource has been missing to link and interface the independent ontologies

- This could help to use the single ontologies more efficiently and create synergetic effect

Goals of Our Work

- Creation of the top-domain ontology *BioTop* to achieve this interfacing and linking task

- Concentration upon a smaller selection of ontologies from the *Open Biomedical Ontologies* (*OBO*)…

  - e.g. *Gene Ontology* (*GO*), *Foundational Model of Anatomy* (*FMA*), *Chemical Enties of Biololgical Interest* (*ChEBI*)

- Fulfil the principles of *OBO* in order to join it

Basis of Our Work

- *BioTop* is based on *GENIA* which is a small ontology (45 classes) that has been…
  - developed in the biological natural language processing (BioNLP) community
  - is intended to be a formal model of cell signalling reactions in humans
  - is limited to highly general classes on biochemical substances and their location in organisms

Basis of Our Work

- *BioTop* we tried to fix problems found within *GENIA* such as…

  - Missing or unclear definitions and documentation

  - Taxonomy sometimes contradicting biological or ontological intuitions

- A complete analysis of the found drawback can be found in our AMIA 2006 paper

Ontology Types

- Top Ontology – contains only very small set of highl-level, general classes (e.g. „Continuant", „*Function*" or „*Object*")

- Top-Domain Ontology – holds essential core domain classes to interface top and domain ontologies (e.g. „*Organism*" or „*Cell*")

- Domain Ontology – has a multitude of low-level, domain-specific classes to describe a domain of interest (e.g. „Antisense RNA Transcription")

Ontology Types
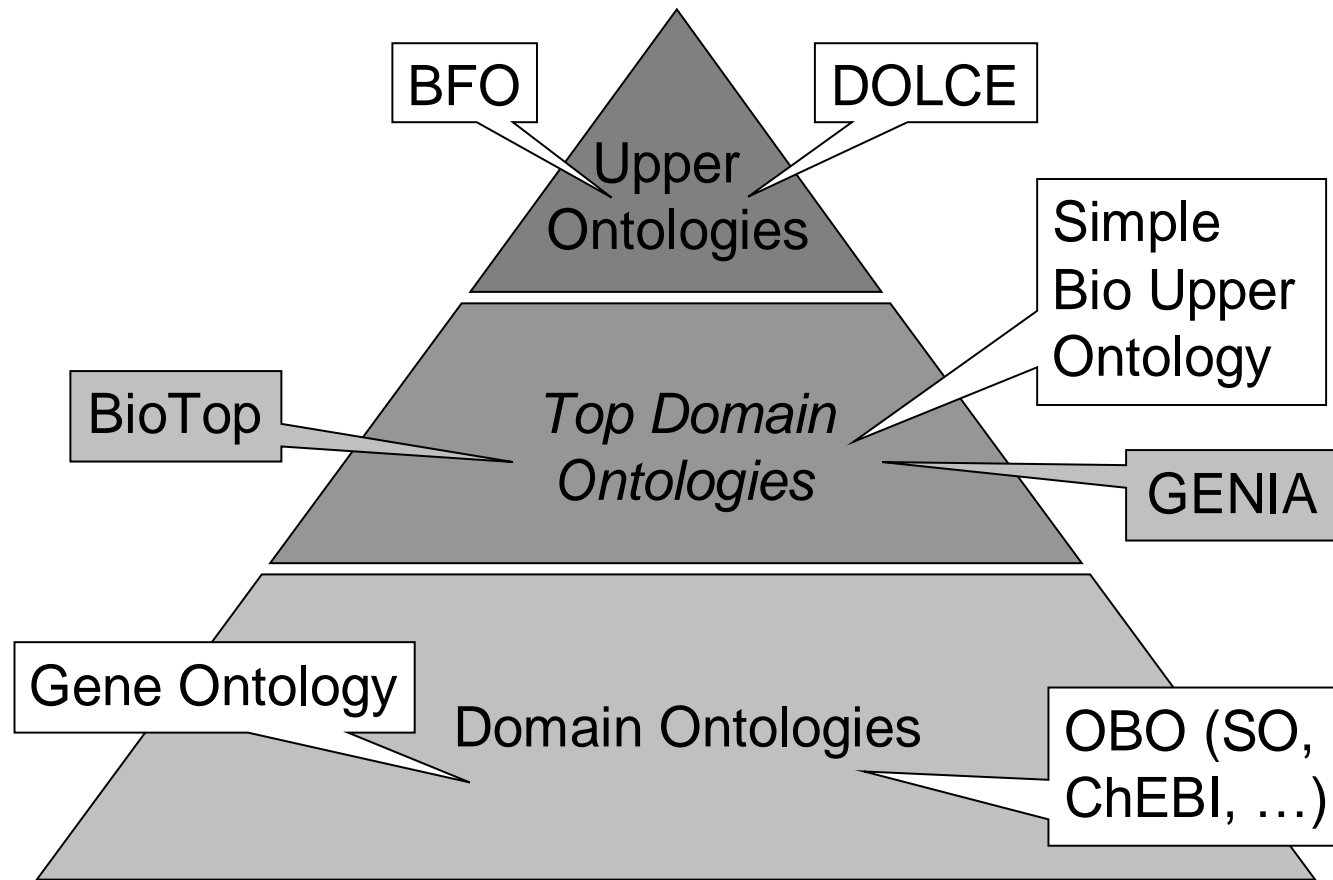
Some OBO Principles

- *BioTop* has a clearly defined topic of being a topic-domain ontology and contains only the classes to fulfil its task

- The development is a collaborative effort between several institutions involving other experts via e.g. face-to-face discusssions or mailing lists

- It is completely open-source and freely available to everybody from its website

Integration of Top Ontology

- Initial *BioTop* version was based on out own set of top-level classes based both on *BFO* and *DOLCE*

- When *BFO* became available in OWL-DL we imported this into *BioTop*

- Integration was straighforward for classes taken directly from *BFO*

- For classes taken from *DOLCE* some remodeling took place via introducing mediator classes

Some OBO Principles

- We include precise textual definitions to…
  - avoid ambiguities many biomedical terms possess
  - make the ontology understandable for humans

- It is implemented in a common language (OWL-DL) that allows the use of standard tools and applications

Example Links to Ontologies

Biological Process	Biological Process (GO)
Protein Function	Molecular Function (GO)
Cell Component	Cell Component (GO)
Cell	Cell (FMA)
Atom	Atoms (ChEBI)
Subatomic Particle	Elementary Particles (ChEBI)
Organic Compound	Organic Molecular Entities (ChEBI)
Tissue	Tissue (FMA)

Example Links to Ontologies
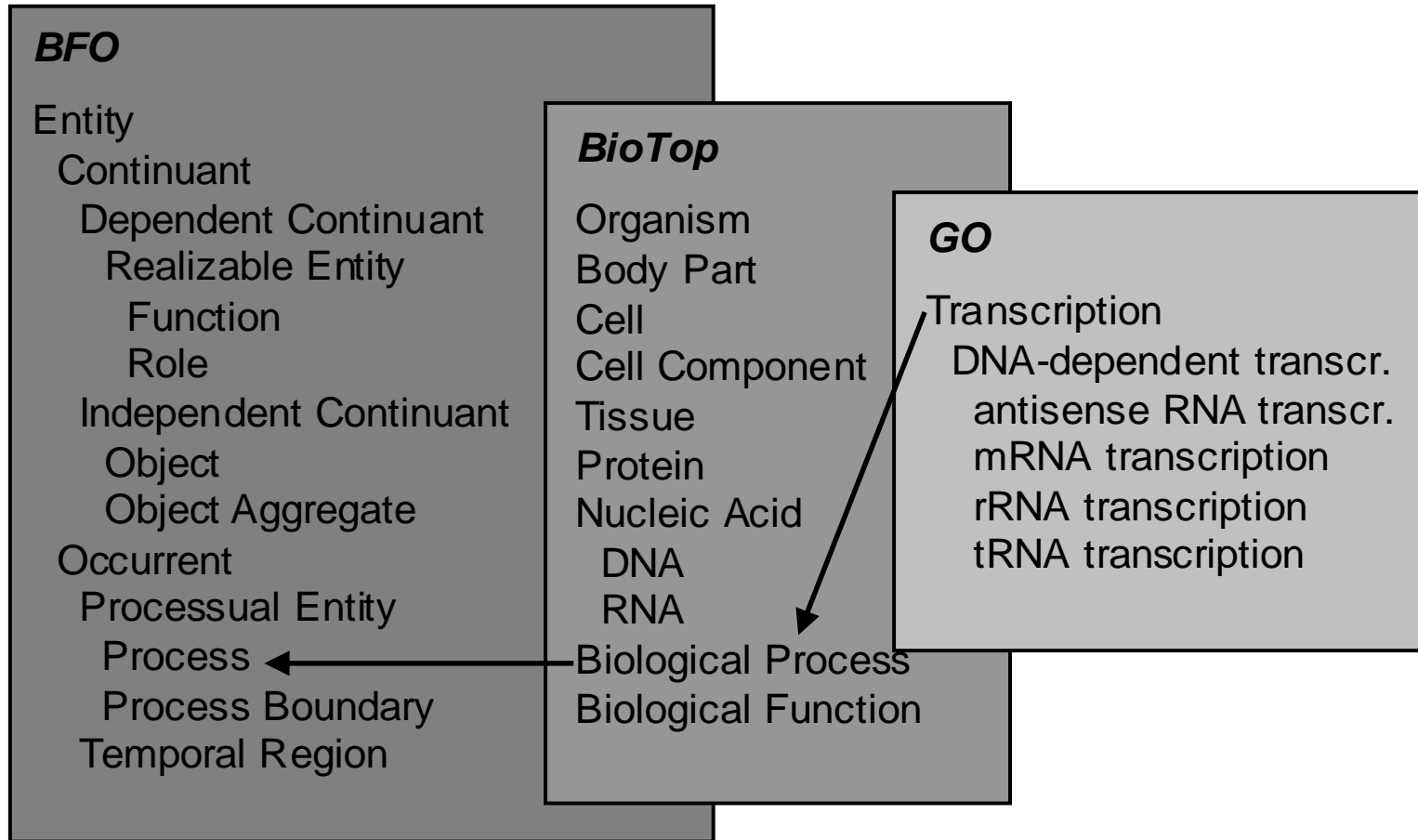
Usage Scenarios and Future

- *BioTop* is used in the EU project *BOOTStrep* for various NLP purposes
  - classes are used to semantically annotate named entities automatically identified in literature abstracts

- *BioTop* as basis for the developement of new ontologies in the framework of *OBO* and elsewhere

- Continuous amelioration and expansion of the existing interface classes to expressed needs by users

Related Work

- Currently two other top-domain ontology for biology/biomedicine are worked upon, i.e. *Simple Bio Upper Ontology* and *GFO-Bio*

- To have several top-domain ontologies is counterproductive to their actual goal

- Therefore we want to collaborate with the other two to achieve a unified ontology
  - work together with the author of the first ontology (Alan Rector) is already ongoing

Website

- All BioTop material (including the latest version of its OWL-DL implementation) is available from its website

  www.ifomis.org/biotop

The End