Linking ontologies to one another and to the Cell Ontology with the COBrA ontology editor

Jonathan Bard & Stuart Aitken

Biomedical Science & Informatics

University of Edinburgh, UK

Context

♦ Our interest is in linking anatomy ontologies to use information from one species to help understand another (the XSPAN project)

♦ Links are to be based on homology, analogy and cell-type relationships

♦ When database curators met to produce an ontological framework for mutant phenotypes, it was clear that they needed an Ontology of Cell Types

♦ Michael Ashburner, David States and I then drafted an Ontology of Cell Types

♦ Cell types are a good way to link organisms (e.g. all have striated muscles, nerves, etc!) for XSPAN

♦ Archiving such links is however very complicated - Excel is not enough!

♦ Stuart Aitkin (ontologist/logician) + Roman Korf produced the COBrA tool

What I will do

♦ Summarise the specification for the Cell Ontology

♦ Use DAG-edit to show the Cell Ontology (a version)

♦ Making the β version

♦ Summarise the initial specification for COBrA

♦ Add the wish-list specification for COBrA

♦ Illustrate how COBrA works

♦ Highlight some technical problems in formats for ontologies that need clearing up

Acknowledgements

The Cell Ontology

- Michael Ashburner (curator)

- David States

- Jonathan Bard

  Others, particularly from the plant

  domain are feeding in cell data

  - and we want lots more!

COBrA

- Stuart Aitken

- Roman Korf

- (Jonathan Bard)

Cell Ontology Specification

♦ Cover all living species - bacteria, fungi, plants, vertebrates, invertebrates

♦ Facilitate interoperability across species databases by providing a cell type ID that was (as far as possible) species independent

  o This meant trying to set phylum/species-specific cell types as deep nodes

♦ Include all cell types

♦ Archive these cell types in all the obvious ways that a user might search on

  o Morphology

  o Function

  o Lineage etc   +

  o Organism

♦ As any cell type would thus have multiple parents, we required a DAG with rules

  o Is-a (subClass of)

  o Gives rise to

The current α version of the cell Ontology

- ◆ Meets the planned criteria

- ◆ Includes >650 items with CL IDs

- ◆ Has terms and synonyms from the eVOC cell ontology

- ◆ Is written in DAG-edit format Posted on OBO

- ◆ Can be downloaded from the OBO site

α -> β => feedback, additions, criticisms

The cell ontology in DAG-edit

The mapping problem for XSPAN

◆ The intention is to establish links between "equivalent" tissues in Drosophila, C elegans, zebrafish and mouse (= human) on the basis of

  o Homology (evolution or lineage)

  o Analogy (= function)

  o Cell type (common IDs)

  o Any other reason (common genetic networks?)

◆ One way to store links is to keep all this data in a spread sheet

      - impractical as it has to be computer readable for genetic access

◆ A better way is to store links between pairs of organisms in an ontology

  o This needs a purpose-built editing/mapping tool ……………………………..

Specifying the COBrA tool

♦ Assume appropriate developmental anatomies (another story) in ontology form

♦ We **need** a tool that will

    o Read 2 anatomy (or any other) ontologies   √

    o Incorporate the Cell Ontology   √

    o Make links across the ontologies   √

    o Define new relationships   √

    o Store the links in machine-readable form (e.g. an ontology)   √

We would also like the tool to be able to

o   Read & write all ontology formats (OWL, RDFS, GO, XML, DAG-edit)   √

o   Support ontology editing (create, drag-drop, merge etc)   √

o   Allow searching   √

o   Show the type of relationship connecting nodes and leaves   √

o   Permit external validation of ontologies   √

o   Display DAG relationships   √ish!

COBrA in action …..

COBrA is downloadable from

http://www.xspan.org/applications/cobra/index.html

and accessible from the GO tools page

Some oddities that emerged while making COBrA

- **COBrA software**

  - reads all standard ontology formats (XML, OWL, RDFS, GO, DAG-edit)

  - translates these ontologies internally into a format that is a superset of all

  - Writes new ontologies in any standard format that is specified

- COBrA software was designed to meet all external standards (XML, OWL, RDFS)

- Stuart Aitken found 3 inconsistencies/problems in translating formats

- To ensure future interoperability, these problems need solving

The 4 problems

1. **There are ID format inconsistencies**

♦ It turns out that GO & DAG-edit use the ID for the localname of an ontology term, but these always include **:** e.g. cell (e.g. GO**:**0005623)

♦ XML and RDFS standards forbid **:** in allowable names (Qnames)

♦ Existing software libraries such as Jena require XML Qnames for concepts. OWL-based tools for GO will need to use these existing resources

The solution is to replace **:** with **.** as in GO.0005623

2. **Access to OBO ontologies is not automatic**
Validation using existing ontology tools requires files to be at the stated **location** so that they can be downloaded. The OBO site provides a link rather than the ontology

The solution is to provide full namespaces at OBO

3. Logical relationships within GO

♦ The meaning of the **is_a** relation is not precisely defined

    o SA proposes the use of **rdfs:subClassOf** for **is_a.**

♦ The meaning of the **part_of** relations is ill-defined and not always transitive (**T**)

    – consider anatomical relationships

    1. The coronoid notch **is part of** the ulna bone (**a bit of**)   **T →**

    2. The bone marrow **is part of** the humerus (**contained within**)   **non-T**

    3. The collar bone **is part of** the skeleton (**member of the group**)   **T ←**

    4. Neural crest **is part of** head mesenchyme (**component of mixture**)   **T →**

♦ SA suggests that we should revise the GO class restrictions to model **part_of**

    o SA proposes as an interim solution using an **OWL-Full ontology** for **part_of**

4. GO/DAG-edit format limitation

♦ Semantic web tools (for reasoning etc) do not necessarily read GO and DAG-edit formats

♦ OBO ontologies will need to be translated into an OWL format so that they can be handled by such tools (the COBrA tool can help here)

♦ Discussion about standardisations will be required and a concensus agreed

Conclusions

♦ The cell ontology is available for use from the OBO site

 - please send criticisms, suggestions etc to Michael Ashburner

♦ The COBrA tool is now available from the GO tools site for

    o Editing ontologies

    o Linking ontologies

    o Ontology format translations

♦ Comments on the format problems raised by Stuart Aitken

should be addressed to stuart@aiai.ed.ac.uk