A Methodology to Migrate the Gene Ontology to a Description Logic Environment using DAML+OIL

Chris Wroe, Robert Stevens, Carole Goble
University of Manchester, UK

Michael Ashburner
EBI, Hinxton, UK

http://gong.man.ac.uk
Gene Ontology Next Generation Project (GONG)

• Demonstrate the utility of finer grained concept descriptions in DAML+OIL
• Develop methodologies and tools to support the process

http://gong.man.ac.uk
Translating theory into practice

- Gene Ontology provides a service to the model organism database community
- Description logic (DL) is a technology born out of computer science research
- DAML+OIL becoming a standard ontology interchange language underpinned by DL
- Adopted by W3C and will soon become Ontology Web Language (OWL)
GONG - proof of concept

- Maintaining an exhaustive is-a structure

http://gong.man.ac.uk
Example: heparin biosynthesis

[chemical] biosynthesis (GO:0009058)

[i] carbohydrate biosynthesis (GO:0016051)

[i] aminoglycan biosynthesis (GO:0006023)

[i] heparin biosynthesis (GO:0030210)

http://gong.man.ac.uk
Example: heparin biosynthesis

[chemical] biosynthesis (GO:0009058)

[i] carbohydrate biosynthesis (GO:0016051)

Axis 1:

Chemicals

[i] aminoglycan biosynthesis (GO:0006023)

[i] heparin biosynthesis (GO:0030210)

Axis 2:

Process

[i] heparin metabolism (GO:0030202)

[i] heparin biosynthesis (GO:0030210)

http://gong.man.ac.uk
Example: heparin biosynthesis

[chemical] biosynthesis (GO:0009058)
[i] carbohydrate biosynthesis (GO:0016051)

Axis 1:
[i] aminoglycan biosynthesis (GO:0006023)
[i] glycosaminoglycan biosynthesis (GO:0006024)
[i] heparin biosynthesis (GO:0030210)

Axis 2:
[i] heparin metabolism (GO:0030202)
[i] heparin biosynthesis (GO:0030210)

http://gong.man.ac.uk
Is this important?

- Missing is-a not noticed by users

- BUT... improves fidelity of DB record retrieval.
  
  - Asking for gene products involved in 'glycosaminoglycan biosynthesis' will lead to an additional result:

    O94923 SPTr ISS - D-glucuronyl C5-epimerase (Fragment)

http://gong.man.ac.uk
How can DAML+OIL support the task?

- Step 0. Translate to DAML+OIL syntax
  - Provided by OilEd (ontology editing tool)
  http://oiled.man.ac.uk

<table>
<thead>
<tr>
<th>GO RDF</th>
<th>DAML+OIL</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="">go:term</a></td>
<td><a href="">daml:Class</a></td>
</tr>
<tr>
<td><a href="">go:isa</a></td>
<td><a href="">daml:subClassOf</a><a href="">daml:Class</a></td>
</tr>
<tr>
<td><a href="">go:part-of</a></td>
<td><a href="">daml:subClassOf</a><a href="">daml:Restriction</a></td>
</tr>
<tr>
<td></td>
<td><a href="">daml:onProperty</a>&lt;daml:ObjectProperty</td>
</tr>
<tr>
<td></td>
<td>rdf:resource=&quot;go:part-of&quot;/&gt;</td>
</tr>
<tr>
<td></td>
<td><a href="">daml:hasClass</a><a href="">daml:Class</a></td>
</tr>
</tbody>
</table>

http://oiled.man.ac.uk

http://gong.man.ac.uk
DAML+OIL definitions for metabolism concepts

• heparin biosynthesis
  – class heparin biosynthesis defined
  subClassOf biosynthesis
  restriction onProperty acts_on hasClass heparin
  (acts_on is unique)
  – Paraphrase: biosynthesis which acts solely on heparin

• glycosaminoglycan biosynthesis
  – class glycosaminoglycan biosynthesis defined
  subClassOf biosynthesis
  restriction onProperty acts_on hasClass glycosaminoglycan

http://gong.man.ac.uk
A chemical ontology

- Mapped chemical concepts to MeSH (using UMLS tools/API e.g. Norm).
- Created a DAML+OIL ontology from a subset of the MeSH chemical taxonomy
- Provides the following information:

  carbohydrates
    [i] polysaccharides
      [i] glycosaminoglycans
      [i] heparin

http://gong.man.ac.uk
Paraphrased reasoning process

- **heparin biosynthesis**
  - class heparin biosynthesis defined
    subclassOf biosynthesis
    restriction onProperty acts_on hasClass heparin

- **glycosaminoglycan biosynthesis**
  - class glycosaminoglycan biosynthesis defined
    subclassOf biosynthesis
    restriction onProperty acts_on hasClass glycosaminoglycan
Inferring a new is-a link

• heparin biosynthesis
  - class heparin biosynthesis defined
    subClassOf biosynthesis
    restriction onProperty acts_on hasClass heparin

• glycosaminoglycan biosynthesis
  - class glycosaminoglycan biosynthesis defined
    subClassOf biosynthesis
    restriction onProperty acts_on hasClass glycosaminoglycan
Output

- OilEd API reports additional inferred is-a relationships.
  E.g. *heparin biosynthesis* has new is-a parent *glycosaminoglycan biosynthesis*
- Report sent to GO editorial team for comment.
- They makes changes to GO if appropriate and sends back queries

http://gong.man.ac.uk
Results

• Carbohydrate metabolism ~250 concepts
  - 22 additional is-a links 17 of which now in GO
• Amino acid metabolism ~ 250 concepts
  - Further 17 additional is-a links now in GO
• GO team will be reviewing results for metabolism as a whole once we have the tools to support the process
• Useful results come from even a partial coverage

http://gong.man.ac.uk
Build a practical environment

• Tools needed for:
  - Creating DAML+OIL definitions
  - Tracking changes
  - Reporting reasoning results
  - Viewing definitions

http://gong.man.ac.uk
Multi-step workflow

- Current GO working copy
  - Automatic translation
    - GO DAML
      - Automatic subsetting
        - GO subset
          - Automatic dissection
            - Dissected GO subset
              - included in
              - Manual changes to definitions
  - UMLS
    - MeSH subset + flat list of unclassified chemicals
  - External changes to taxonomy
  - GO editorial team
  - Changes to taxonomy

- Database of new subsumptions and inconsistencies
- Combined ontology
- Changes
- Classification
- Classified ontology

http://gong.man.ac.uk
Reporting tools

- carbohydrate metabolism
  + polysaccharide metabolism
  + disaccharide metabolism classification change
  + proteoglycan metabolism classification change
  + oligosaccharide metabolism classification change
  + glucuronic acid metabolism
- aminoglycan metabolism
  + aminoglycan biosynthesis
    - glycosaminoglycan metabolism classification change
      + hyaluronan metabolism
      + chondroitin sulfate metabolism
      + dermatan sulfate metabolism
      glycosaminoglycan catabolism classification change
- heparin metabolism
  heparin biosynthesis classification change
  heparin catabolism classification change
- glycosaminoglycan biosynthesis classification change
  keratan sulfate biosynthesis
  + aminoglycan catabolism
  + organic acid metabolism
  + peptidoglycan metabolism classification change
  + alditol metabolism classification change
  + aldonic acid metabolism
  + aminosugar metabolism
  mannan metabolism classification change

http://gong.man.ac.uk
Using Amigo to view definitions

http://gong.man.ac.uk
Conclusion & future plans

• Description logic approach is useful
• Beginning to build a practical environment
• Extend quantity and detail of DAML+OIL definitions
  - ~2000 complete but un-validated metabolism definitions
  - ~2000 partial enzyme definitions
• Continue prototyping and developing tools to support the process.

http://gong.man.ac.uk
Acknowledgments

• Jane Lomax and Midori Harris of the GO editorial team for help and advice and responding to the suggested changes
• UMLS and MeSH which provided valuable resources for chemical information
• Sean Bechhofer for development on OilEd

• Project funded as a subcontract of the DARPA DAML programme

http://gong.man.ac.uk