OREMP: Ontology Reasoning Engine for Molecular Pathways

Renato Umeton[1,2], Beracah Yankama[1], Giuseppe Nicosia[3], C. Forbes Dewey, Jr.[1]

[1]Massachusetts Institute of Technology, Cambridge, MA; [2]University of Calabria, Cosenza, Italy; [3]University of Catania, Catania, Italy.

http://cytosolve.mit.edu/oremp

ABSTRACT

The information about molecular processes is shared continuously in the form of runnable pathway collections, and biomedical ontologies provide a semantic context to the majority of those pathways. Recent advances in both fields pave the way for a scalable information integration based on aggregate knowledge repositories, but the lack of overall standard formats impedes this progress. Here we propose a strategy that integrates these resources by means of extended ontologies built on top of a common meta-format. Information sharing, integration and discovery are the primary features provided by the system; additionally, two current field applications of the system are reported.

INTRODUCTION

Input information resources:

- Quantitative pathways (models)
- Ontologies

Goal integration cycle based on shared knowledge-bases[1]

- ChEBI[2]
- KEGG[3]
- BioPortal[4]
- BioModels.net[5]
- CellML repository[6]

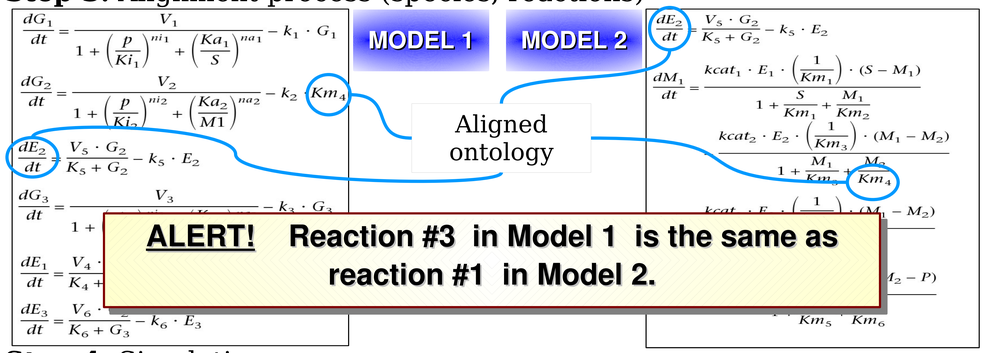
However, there is still a large chasm between today's functionality and the true ability to use the ontology data to inform molecular pathways.

Use Case 1: Model Alignment for Parallel Simulation with Cytosolve[7]

Step 1. Select or upload biochemical models

Step 2. Set initial concentrations

Step 3. Alignment process (species, reactions)



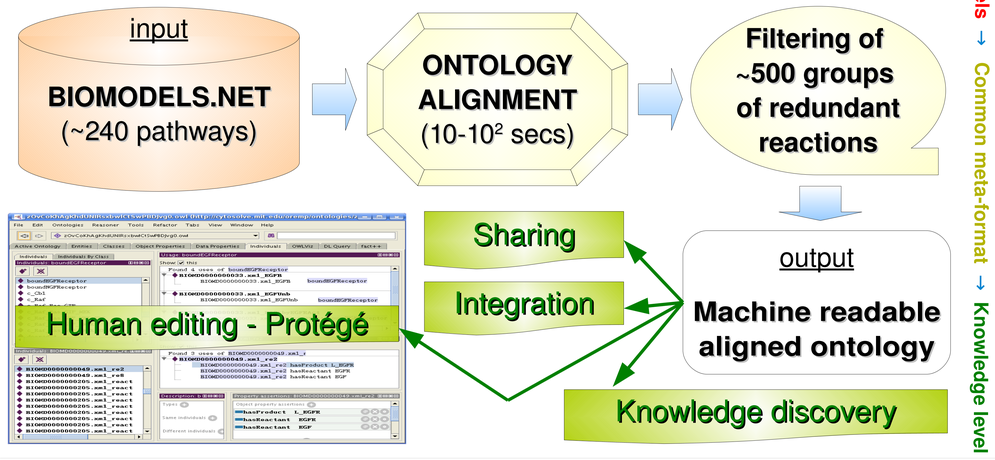
Step 4. Simulation

Step 5. Share the aligned models by means of an OWL file

Use Case 2: Aligned Ontologies

Feeding step 1. Access one or more database of models

Feeding step 2. Align the knowledge there discovered



PROCESSING FLOW

1. The *data access facility* collects information about multiple pathways and existing biological databases;

2. A *parser component* reads different file formats (i.e., XML, RDF, SBML, CellML, etc) and extracts relevant information;

3. The *core module* assembles the knowledge, parsed from different sources, into a coherent ontology (based on our meta-format); and

4. The *logic component* can annotate all of the species from a collection of reactions and do automated comparisons, identification of common species, and duplicate reactions

RESULTS

Sharing: Despite disparate initial data formats, the biochemical information described in each pathway is now homogeneously represented. This enables the direct reuse of components (such as species or reactions) coming from different sources.

Integration: Our system ensures a consistent merging of the resources, automatically aligning the species and showing the end-user possible duplications among reactions in the different pathways.

Knowledge discovery: Once the species alignment is done and duplicate reaction have been detected, a new step is taken: the set of "alternative circuits" is computed.

Human editing with Protégé: The information gathered is exported in OWL. With the OWL file we use the semantic tool, Protégé[9], to visually edit, compare, and finalize the biochemical information. With the OWL query interface, the user can now formulate "semantically-enabled" queries that were impractical when dealing with the previously heterogeneous, unaligned data repositories.

REFERENCES

1. A. Bauer-Mehren et al., Mol. Sys. Bio., **5**, 290, 2009.
2. K. Degtyarenko et al., Nucleic Acids Res., **36**, D344, 2008.
3. M. Kanehisa et al., Nucleic Acids Res., **36**, 484, 2008.
4. N. Noy et al., Nucleic Acids Res., **37**, W170, 2009.
5. N. Le Novere et al., Nucleic Acids Res., **34**, 689, 2006.
6. C. M. Lloyd et al., Bioinformatics, **24**, 2122, 2008.
7. S. Ayyadurai et al., Cell. Mol. Bioeng., *In press*, 2010.
8. N. Le Novere et al., Nat. Biotechnol., **23**, 1509, 2005.
9. N. Noy et al., IEEE Intell. Syst. Appl., **16**, 60, 2001.

ACKNOWLEDGEMENTS

This work is supported by the NIH and the Singapore-MIT Alliance