Systems biology

libAnnotationSBML: a library for exploiting SBML annotations

Neil Swainston1,∗ and Pedro Mendes1,2

1Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre, University of Manchester, Manchester M1 7DN, UK and 2Virginia Bioinformatics Institute, Virginia Tech, Washington St. 0477, Blacksburg, VA 24061, USA

ABSTRACT

Summary: The Systems Biology Markup Language (SBML) is an established community XML format for the markup of biochemical models. With the introduction of SBML level 2 version 3, specific model entities, such as species or reactions, can now be annotated using ontological terms. These annotations, which are encoded using the resource description framework (RDF), provide the facility to specify definite terms to individual components, allowing software to unambiguously identify such components and thus link the models to existing data resources.

libSBML is an application programming interface library for the manipulation of SBML files. While libSBML provides the facilities for reading and writing such annotations from and to models, it is beyond the scope of libSBML to provide interpretation of these terms. The libAnnotationSBML library introduced here acts as a layer on top of libSBML linking SBML annotations to the web services that describe these ontological terms. Two applications that use this library are: SbmlSpeciesNameExtractor finds name synonyms of SBML model entities, and SbmlReactionBalancer checks SBML files to determine whether specified reactions are elementally balanced.

Availability: http://mcisb.sourceforge.net/

Contact: neil.swainston@manchester.ac.uk

1 INTRODUCTION

The minimum information requested in the annotation of SBML models (Hucka et al., 2003) defines guidelines for annotation of biochemical models. The annotation of models with the MIRIAM standard provides a number of significant advantages in the development of computational tools and applications that can reason over them (Kell and Mendes, 2008).

An example is the task of comparing or merging two biochemical models. Before the introduction of MIRIAM, individual components of SBML models (Hucka et al., 2003) were identified solely by free-text, human-readable, name attributes, often resulting in equivalent components being named differently in different models. As naming conventions are non-standard, it is impossible to definitively match these components computationally, and the process of model merging then requires human input to resolve ambiguities. Providing MIRIAM-compliant annotations allows a component to be unambiguously identified by associating it with one or more terms from publicly available databases such as ChEBI (Degtyarenko et al., 2008) or KEGG (Kanehisa et al., 2000) (Fig. 1).

2 FEATURES

The species elements in Figure 1 are both annotated with MIRIAM-compliant terms. libSBML (Bornstein et al., 2008) provides the facility for reading a given SBML element’s annotation and hence could be used to determine that species1 and species2 are annotated with ChEBI term CHEBI:4167 and KEGG Compound C00031, respectively. From this, it may be concluded that the compounds represented by these species are different. However, manual inspection of the database references in ChEBI and KEGG show that both species are annotated with references that share the same chemical structure, and hence are equivalent.

Performing such a comparison computationally is beyond the scope of libSBML. To do so, the annotations must be ‘dereferenced’ by querying the two databases via their web service interfaces. This task is complicated particularly because each of the web services has non-standard interfaces.

The libAnnotationSBML library creates a unified framework for supporting MIRIAM-compliant annotations by wrapping these divergent web services into a Java API, allowing each web service to be queried in a consistent manner. The library itself can act as a layer on top of the libSBML API.

The library is built dynamically by querying the MIRIAM web service (Laibe and Le Novère, 2007), which provides a collection of data types that are recommended for use in model annotation. The web service provides details of each of these data types

∗To whom correspondence should be addressed.
including names, URNs and physical URLs to resources. From this, a collection of Ontology objects are instantiated, one for each data type supported specified in MIRIAM.

Individual OntologyTerms objects are built up from an Ontology object and a unique identifier. Once instantiated, the OntologyTerm provides a number of methods, specified in Figure 2. The implementation of these methods is performed by mapping the calls to an appropriate call to the data type’s web service, where such a web service exists.

The OntologyTerm class can be extended to provide methods specific to the SBML element that is being described. For example, a metabolite species element annotated with a ChEBI term will return a ChEBITerm object, providing a method for accessing the chemical formula of the metabolite. Similarly, a KEGG Reaction annotation will return a KeggReactionTerm object, providing methods for accessing reactants and products, each returned as OntologyTerms.

Applying libAnnotationSBML to the SBML in Figure 1 will associate an OntologyTerm with each of the species. Calling getTerm() on these ChEBI and KEGG OntologyTerm objects returns ‘β-glucopyranose’ and ‘β-glucose’, respectively. Considering the initial problem of comparing SBML components, this provides an example of why names cannot be used reliably to perform this task. A more reliable approach is to exploit the fact that many data resources cross-reference one another. For example, entries in the ChEBI database can provide details of the equivalent term in KEGG, and vice versa. The OntologyTerm class supports this by implementing a getXrefs() method which returns cross references themselves as OntologyTerms, along with a predicate, defined in libSBML, that indicates the relationship between them. When an OntologyTerm references an equivalent entity in a different database, the predicate libsbmlConstants.BQB_ENCODES is returned. In the case of a genomic database entry cross referencing an entry in a proteomic database, libsbmlConstants.BQB_ENCODES is used. Utilizing this method, it can be determined computationally that the ChEBI and KEGG terms cross-reference one another, and hence species1 and species2 can be unambiguously determined to represent equivalent entities.

The libAnnotationSBML library facilitates the rapid development of tools to manipulate SBML annotation terms. The library can be used to add annotation to unannotated SBML models, using a similar approach to semanticSBML (Schulz et al., 2006). libAnnotationSBML can annotate both metabolites and proteins, exploiting the search facility that exists in both the ChEBI and UniProt web services (The UniProt Consortium, 2008). The focus of libAnnotationSBML is to develop tools to manipulate already annotated models. An example of such a tool is the SbmlSynonymExtractor, which takes annotated SBML as input, and returns a mapping of all species terms to their name synonyms, harvested from ChEBI, KEGG or UniProt. Another tool, the SbmlReactionBalancer, determines whether the reactions specified within an SBML file are elementally balanced by querying the ChEBI web service to retrieve chemical formulae of reaction participants.

libAnnotationSBML was used extensively in the development of a genome-scale model of yeast metabolism, the first model of this scale in which all compartments, metabolites, enzymes and complexes are unambiguously defined using MIRIAM-compliant annotations (Herrgård et al., 2008).

3 IMPLEMENTATION AND DISTRIBUTION

The API is written in Java 1.5 and is dependent on libSBML v3. It is supported in Linux, Windows and MacOS X and is distributed as source code and associated build files under the open source Academic Free Licence v3.0 from http://mcisb.sf.net/ along with other tools described in this manuscript.

ACKNOWLEDGEMENTS

The authors thank the BBSRC and EPSRC for financial support of the Manchester Centre for Integrative Systems Biology, of which this work was an integral part.

Conflict of Interest: none declared.

REFERENCES