BioUML – open source extensible workbench for systems biology
http://www.biouml.org

Biosoft.Ru/DevelopmentOnTheEdge.com

Laboratory of Bioinformatics, Design Technological Institute of Digital Techniques

Novosibirsk, Russia.
Main BioUML concepts and ideas

- Visual modeling
- Plug-in based architecture (Eclipse platform runtime from IBM).
- Meta model – problem domain neutral level of abstraction that describes system as compartmentalized graph.
- Diagram type concept – formally defines graphical notation and provides its incorporation into BioUML workbench.
- Module concept - allows developer to incorporate databases on biological pathways into BioUML framework taking into account database peculiarities.
Visual Modeling

The problem of modeling and simulating of complex systems can be significantly simplified for customers by using computer systems providing visual modeling.

These visual depictions offer alternative syntax to completely and formally specify models.

A number of visual syntaxes were developed and implemented for formal description of biological systems, but there is no common standard.
Some graphical notations for biological pathways


Database specific notations:
- KEGG/Metabolic pathways; GeneNet system; TRANSPATH
- …
Plug-in based architecture

A **plug-in** is the smallest unit of BioUML workbench function that can be developed and delivered separately into BioUML workbench. A plug-in is described in an XML manifest file, called plugin.xml. The parsed contents of plug-in manifest files are made available programmatically through a **plug-in registry** API provided by Eclipse runtime.

- **extension points** are well-defined function points in the system where other plug-ins can contribute functionality.

- **extension** is a specific contribution to an extension point. Plug-ins can define their own extension points, so that other plug-ins can integrate tightly with them.
BioUML meta model

The core of BioUML workbench is meta model.

Unlike UML meta mode BioUML meta model is problem domain neutral and provides an abstract layer for comprehensive formal description of wide range of biological and other complex systems.

Content of databases on biological pathways or SBML models are expressed in terms of meta model and then can be used by other workbench plug-ins.
Example of formalized description of system from two chemical reactions

System structure is described as a graph.

Mathematical model of the system:

\[
\begin{pmatrix}
\end{pmatrix}
\]

Description of system components in the database:

<table>
<thead>
<tr>
<th>ID</th>
<th>A</th>
<th>CC . . . .</th>
<th>R1</th>
<th>A-&gt;B</th>
<th>B</th>
<th>CC . . . .</th>
<th>R2</th>
<th>B-&gt;C</th>
<th>C</th>
<th>CC . . . .</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Main BioUML meta-model classes to describe system structure as compartmentalized graph
Mapping of SBML models into BioUML meta model

SBML

<table>
<thead>
<tr>
<th>Model</th>
<th>Compartment</th>
<th>Species</th>
<th>Reaction</th>
<th>SpeciesReference</th>
<th>FunctionDefinition</th>
<th>Rule</th>
<th>Event</th>
</tr>
</thead>
</table>

BioUML meta-model

- **DATA**
  - Description of system components
- **graph**
  - System structure is described as a graph
- **Math**
  - Mathematical model of the system

- **RDF notes**
  - math
  - variable
  - volume
  - initialAmount
  - trigger, delay
  - eventAssignments
Diagram type concept

Diagram type defines:

- what system components can be shown in the diagram;
- diagram view builder - it is used to generate view for each diagram element taking into account problem domain peculiarities;
- semantic controller - provides semantic integrity of the diagram during its editing;
- filters – hide or highlight diagram elements according to some selection criteria.
Module concept

The module concept allows to developer define new diagram types and incorporate other databases on biological pathways into BioUML framework.

The module defines mapping of database content into diagram elements and diagram types that can be used with the database. Module also provides query engine that can be used by BioUML workbench to find interacting components of the system.
Graph search engine
Modules

- standard BioUML module for biological pathways;
- module for models in SBML format;
- module for models in CellML format (only biochemical models);
- GeneNet database module;
- module for KEGG/Pathways database (draft);
- module for TRANSPATH database (draft);
- GeneOntology (under development).
Standard BioUML module for biological pathways

The module defines most common biological data types (gene, protein, RNA, substance, reaction, etc.), they mapping into simple text database and three diagram types for description of biological pathways on several semantic levels:

1. Semantic network (ontology) - this diagram type is used to describe semantic relationships between system components, system states, and related problem domain concepts.

2. Pathway diagram type is used for formalized description of biological pathway structure. This diagram type uses GeneNet graphical notation.

3. Pathway simulation diagram type is extension of pathway structure diagram, where variables are associated with graph nodes and differential equations with graph edges. This allows to BioUML workbench automatically generate mathematical model of the system and simulate its dynamics.
Conceptual model of heavy metals influence on pathogenesis of chronic lung diseases
Minimal Mitotic Oscillator

Citation

Description
This is a widely cited minimal (3-variable) model for a mitotic oscillator. The three variables represent Cyclin (C), inactive cdc-2 Kinase (M) and an active cdc-2 Kinase (X).
http://www.pnas.org/cgi/content/abstract/88/20/9107

Rate constant Reaction
X1 = 0.1 (half life) Comp(M) → M
Suggested approach can be applied for modeling biological systems using:

- ✔ Systems of ordinary differential equations
- ☐ Systems of algebra-differential equations
- ✔ State and transition diagrams
- ✔ Hybrid models
- ☐ Boolean and logical networks
- ☐ Petri nets
- ☐ Markov chains
- ☐ Stochastic models
- ☐ …

Limitations (to be overcame)

- ☐ Spatial models
- ☐ PDE
- ☐ …
Biological databases

Data search and retrieving

Formal description of structure of biological system

Visual modeling

Automated code generation for model simulation of model behavior

MATLAB code

... code

Java code

Simulating using MATLAB. JMatLink allows to BioUML workbench to start MATLAB and retrieve simulations results

Java simulation plug-in. Contains ODE solvers ported from odeToJava and methods for hybrid models support.
Simulation result

ID: res 1
SN: ode23
FROM: 0.0
TO: 100.0
Cyclonet database

Using BioUML workbench we are developing Cyclonet database – a database on cell cycle regulation in eukaryotes.

Using BeanExplorer Enterprise Edition database content is available through the Internet.
http://cyclonet.biouml.org

The database contains information about cell cycle specific genes, proteins, protein complexes and their interactions, diagrams of cell cycle regulation for vertebrates, models of cell cycle and results of their analyses, microarray data, literature references and other related resources.
Cyclonet system architecture

Client side:  
- Web browser

Server side:  
- Servlet container: Tomcat
- BeanExplorer Enterprise Edition
- JDBC
- Cyclonet module
- BioUML workbench
- GeneOntology MySQL database
- Cyclonet MySQL database
Ontology diagram type

... the description should be here ...
Biopath (codename) database

http://biopath.biouml.org

Purpose of this project is to create a formal description of pathogeneses of chronic respiratory diseases and optimization of their treatment.

The work is being done jointly with the Institute of Medical Problems of North (Krasnoyarsk) and the Institute of Pulmonology (Moscow).
Current works

- **Graph search engine**
  - Graph search engine completion and testing
  - Adaptation of new graph layout library
- **Complete support of SBML level 2**
  - Time delay
  - Algebraic rules
  - Units
  - SBML semantic tests
- **Formula editor**
- **State and transition diagrams.**
Improved version of force directed graph layout
(takes into account nodes size)

(lay-out time ~50ms, 300 iterations)
justAtest(x_1) = \begin{cases} 
\sin(-x_1) & \text{if } x < 0 \\
0 & \text{if } x == 0 \\
\frac{y_1 + y_2}{1 + e^{x + 5}} & \text{otherwise}
\end{cases}

function justAtest(x_1) =
    if(x<0) sin(-x_1);
    if(x==0) 0/(1+(y_1+y_2)/2);
    otherwise x_1+e^x(x+5)

AST tree dump:
Start
   Function declaration: justAtest
      Variable: "x_1"
      Piecewise: 3
         Piece
            Function "<"
               Variable: "x"
               Constant: 0
            Function "sin"
Further works

- Library of predefined kinetic laws
- Stochastic modeling
- New diagram type for gene regulatory regions
- …
- BioUML can be a test platform where we can try to implement proposals for SBML level 3, so people will get a possibility to evaluate them in practice.
- We will be glad to help third party developers to integrate their software with BioUML workbench.
- We are open for any collaboration, joint projects, etc.
Availability

BioUML workbench (including source code) is freely available at http://www.biouml.org

There is special forum dedicated to BioUML workbench where you can post your questions and suggestions: http://groups.yahoo.com/group/biouml/

Cyclonet database  – http://cyclonet.biouml.org
Biopath database  – http://biopath.biouml.org
Acknowledgments

Part of this work was partially supported by following grants: Volkswagen-Stiftung (I/75941), INTAS Nr. 03-51-5218 and RFBR Nr. 04-04-49826-a.

Author is grateful to for useful comments, discussions and technical support

Alexander Kel
Sergey Zhatchenko

Software developers
Mikhail Puzanov      Igor Tyazhev      Ruslan Sharipov
Vasiliy Hudyakov     Vlad Zhvaleev    Elena
Cheremushkina       Alexandr Koshukov  Oleg Onegov
Artem Shaidukov

Annotators