STON translator: applying graph database to standard knowledge representation in Systems Biology

Vasundra Touré, Alexander Mazein, Irina Roznovsky, Mansoor Saqi, Johann Pellé and Charles Auffray

1 European Institute for Systems Biology and Medicine (EISBM), CNRS-ENS-UCBL, 50 Avenue Tony Garnier, 69007 Lyon, France; *IMI-eTRIKS consortium
2 Université Paris-Sud, UFR Sciences Bât. 301, 91405 Orsay cedex, France

Motivation
In Systems Biology, complex heterogeneous networks typically represent biological pathways. Graph databases are a natural way of representing biological networks, and can offer some advantages over relational databases in efficient querying of the network. A tool for importing biological pathways from the popular SBGN format and representing them in a graph database format is described.

STON: SBGN TO Neo4j translator
STON has been developed to allow import and translation of metabolic, signalling and gene regulatory pathways presented in SBGN Process Description and SBGN Activity Flow languages to a graph-oriented format compatible with Neo4j.

1) Translation rules
Rules were needed to be specified in order to develop STON application: syntactic and semantic translation of SBGN objects into Neo4j. For the most of the SBGN objects, translation was straightforward: node glyphs from SBGN will be nodes in Neo4j and arcs from SBGN will be relationships in Neo4j. Properties store details about nodes and relationships.

Particular cases:
• Complexes are represented as nodes with biochemical entities related to them with the relationship “belongs_to”.
• A property (“reversible”) has been created for process nodes to determine the reversibility of a reaction (true/false value).

2) Development
STON is a Java-based framework using two sets of external Java libraries: namely LibSBGN (support SBGN maps) and Neo4j libraries (allow Neo4j graph development). In addition, a graphical user interface has been developed in order to provide a better access to STON’s functionalities. The program has been tested using several examples available from LibSBGN and from SBGN bricks SourceForge (http://sbgnbricks.sourceforge.net/).

3) Usage of STON framework

Conclusion and perspectives
Applying graph database technologies to biological network data management provide new advanced opportunities of study in Systems Biology.

Possible applications include:
1) combining multiple pathways for developing even more complex networks;
2) combining different levels of granularity (PD and AF).

Tools of interest
Systems Biology Graphical Notation
• Standard for representation of biological networks.
• Set of complementary languages.

Neo4j Graph Database
• «Property graph» database, NoSQL.
• CYPHER query easy-to-use language.

References
Available on: http://sourceforge.net/projects/ston/

Acknowledgements
Authors are grateful for financial support from the eTRIKS and U-BIOPRED consortium (IMI 115446 and 115010).

Contact: vtoure@eisbm.org