ARCHITECTURE OF SOFTWARE TOOLKIT FOR STORING AND OPERATING WITH BIOSYSTEMS MODELS

Miginsky D.S.*,1,2, Suslov V.V.1, Rasskazov D.A.1, Podkolodny N.L.1, Kolchanov N.A.1
1 Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; 2 Novosibirsk State University, Novosibirsk, 630090, Russia
* Corresponding author: e-mail: shadow@bionet.nsc.ru

Key words: ecosystems, gene networks, visual modeling, databases

SUMMARY

Motivation: Each biosystem – from cell-level to biosphere – could be viewed as a self-regulating hierarchical system of elements interconnected into the network by the regulatory circuits. On different stages of biosystem’s evolution such a network could contain various number of elements and has various structure of interconnections that is the consequence of serious behaviour variations. Regarding to the biosystem’s type there are different types of elements and interconnections forming the domain of current biosystem (Zherikhin, 1994; Suslov et al., 2006). For instance the cell models contains microbodies, proteins and chemical substances as elements while the ecosystems includes concrete organisms, other biotic and abiotic factors, food (Luczkovich et al., 2003) and also completely different interaction types (De Jong-Brink, 1995; Schultze, Kondorosi, 1998; Adamo, 2002). The software capable to operate with different network types first of all need to be flexible enough and have an extensible domain. More precisely, each network type is described by separate domain. For each one user need to define the appropriate element types – and this must be the key feature of the software to develop.

Results: Currently the database capable to store different types of network is prototyped. The first two network types to be stored are gene and ecological networks. The database is capable to represent them, describe and associate with others stages of biosystem’s evolution in context of single model and also decompose such networks.

INTRODUCTION

The value of data on any biological system is greatly increasing if formalized and represented in structured computer database. Such a representation provides wide specter of capabilities to analyze it, share between many experts all over the world and integrate with other information resources of the same domain. Currently most databases on this field have different disadvantages: (i) have very complex multilayered structure with inconvenient navigation or (ii) their structure too inflexible to extend the domain (set of concepts) or (iii) not structured and formalized enough, e.g. represented with text unstructured files (Sergeyev et al., 2006). For small-scale models of tissues, organs and also gene and protein networks there are also some schemas (iv) containing elementary objects and interconnections with or without spatial location information – a sort of object-oriented models (Ananko et al., 2005). Domains of such biosystems are more clearly understandable and could be described as a finite set of well-defined concepts and interconnections between them. The problem to solve is to extrapolate such an object-oriented (OO) approach to large-scale systems such as ecosystems.
In this work we tried to use the experience of operating with cell-level models – gene networks (Ananko et al., 2005). The software operating with networks of any level have to operate with different domains. Thus there must be the capability for user to define new domains by himself. Next he can build his own networks based on concepts described in chosen domain. If necessary this database could be integrated or supplied with information from other databases with any of i-iv structures.

Also there must be relatively simple and effective software capable to edit such networks. The most natural and effective solution is a sort of visual modeling software tool using vector 2D-graphics to represent networks as marked graphs with defined layout. This solution also gives the capabilities to decompose the network. For example user could hide some elements of network or visualize some functional aspects only while hiding others.

METHODS AND ALGORITHMS

Regarding to the requirements described earlier the software toolkit need to consist of two main subsystems – database and visual editor. The most important is the first subsystem because it defines the most part of business-logic i.e. all concepts operated by the software and all operations with them.

It is better to think about this database as high-level specific domain-oriented DBMS (databases management software) with its own meta-model (Fig. 1). User could define his own “database structure” based on this meta-model. By the database-structure here we mean the specific Domain that contains a set of Terms (it could be think as OO class) each one representing specific concept from the domain. Each term has a name (for gene networks it could be “gene” or “protein” for instance) should be defined as a set Properties characterizing it. For instance term “gene” characterized by its “name”, list of alternative names (“synonyms”), link to “species” where it presented and link to “protein” it produces. Here each quoted name is a name of concrete property. Properties could be of different types and also could be either scalar or pointing (links). Links are very important because represents the relationships (associative first of all) between different concepts. Each link property is characterized by a set of terms to point from the same domain. Here is the first level of meta-model.

Figure 1. Database meta-model.
The second and third levels define the concrete *Objects* as instances of terms. All objects are grouped into *Models* – a sort of logical partitions treated to concrete domain. For user the model is the representation of concrete biological system. Each object is contains values of properties defined by its term.

The most part of information on each individual object is described on the second level. It could be viewed as a sort of dictionary containing elements for networks. Almost all properties gain the values on this level.

The third layer has almost the same structure as the second but assigned for representation of networks. The entire model with its objects of the 2nd and 3rd levels could be viewed as a network representing concrete biological system. Third level contains *Schemas* representing some parts or functional aspects of the network. Its combination describes the entire network. For smaller network there could be the only one schema but it is hardly possible to find so simple biosystem that could be described in that way. So schemas are one of the ways to decompose the model into smaller parts. There are two other mechanisms of decomposing – *Layers* and compartments. Layer is a reusable part of the schema – more than one schema could contain the same layer. Layers are also necessary for functional decomposition so each schema contains the set of depended ones. Viewing the schema user can show or hide some of them. Creating or editing one he could reuse already created layers.

Compartments unlike layers and schemas provide spatial decomposition. Any object could be a compartment if its term is marked as compartment. For instance the “cell nucleus” term could be the compartment in gene networks. In meta-model the compartment is defined as schema level object able to contain others. The hierarchy of compartments could be of any nesting level. Viewing the schema user could show or hide payload of any compartment and define the necessary detail level by this way.

**RESULTS AND DISCUSSION**

Currently the database has been prototyped. Implementation of this model is based on the general database integration technology (Miginsky *et al.*, 2006). This meta-models are pretty close in both works regardless to different purposes but this one has one additional level. This problem could be solved by mapping the domain into two separate but correlated domains from that one. For testing purposes we have imported some schemas from GeneNet (Ananko *et al.*, 2005) into this database. The functional prototype of the editor is implemented in GeneNet. Currently new version of the editor oriented for described meta-model is being developed.

The first two areas of application of this software are gene and ecological networks. For gene networks it counts as next generation of GeneNet system. In comparison with it this one provides more flexible mechanisms of operating with this networks by new methods of decomposition and plug-in architecture capable to integrate modules for analysis.

**ACKNOWLEDGEMENTS**

Authors are grateful to prof. M.G. Sergeev and dr. N.I. Yurlov for assistance in results discussion. The work is supported by the following grants: RFFI 03-04-48506, integration projects SB RAS 34, project “Computer Modeling and Experimental Design of Gene Networks”, RAS program on physicochemical biology, RAS presidium program “Biosphere Origin and Evolution” and innovation project of Federal Agency of Science and Innovations IT-CP.5/001 “Development of software for computer modeling and design in postgenomic system biology (system biology in silico)”. 
REFERENCES


