

На правах рукописи



УДК 004.652.5; 004.655; 519.179

МИГИНСКИЙ Денис Сергеевич

**МЕТОДЫ И СРЕДСТВА РЕКОНСТРУКЦИИ СЕТЕВЫХ
МОДЕЛЕЙ СЛОЖНЫХ БИОЛОГИЧЕСКИХ СИСТЕМ**

Специальность 05.13.11. – математическое и программное
обеспечение вычислительных машин, комплексов и
компьютерных сетей

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата физико-математических наук

Новосибирск 2009

Работа выполнена в Учреждении Российской академии наук
Институте цитологии и генетики Сибирского отделения РАН

Научные руководители: Колчанов Николай Александрович
академик РАН, доктор
биологических наук

Лаврентьев Михаил Михайлович
доктор физико-математических
наук

Официальные оппоненты: Витяев Евгений Евгеньевич
доктор физико-математических наук

Скопин Игорь Николаевич
кандидат физико-математических
наук

Ведущая организация: Московский государственный
технический университет имени
Н.Э. Баумана

Защита состоится 21 декабря 2009 г. в 16 ч. 00 мин. на заседании
диссертационного совета ДМ003.032.01 при Институте систем
информатики им. А.П. Ершова Сибирского отделения РАН по
адресу:
630090, Новосибирск, пр. Акад. Лаврентьева, 6.

С диссертацией можно ознакомиться в читальном зале библиотеки
ИСИ СО РАН (г. Новосибирск, пр. Акад. Лаврентьева, 6)

Автореферат разослан 20 ноября 2009 г.

Ученый секретарь
диссертационного совета

к.ф. – м.н.



Ф.А. Мурзин

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы

Под сетевыми моделями будем понимать ориентированные графы, отражающие структурные и функциональные характеристики моделируемой системы и представляющие их в терминах вершин и дуг графа. Вершины и дуги типизированы и могут характеризоваться определенными признаками, свойствами и величинами различной природы.

На сегодняшний день сетевое моделирование сложных систем различной природы является одним из наиболее эффективных, наглядных и универсальных способов изучения их структурно-функциональной организации. Этот подход успешно применяется во многих областях, таких как проектирование программного обеспечения, схемотехника, планирование и оптимизация бизнес-процессов и т.д.

Сетевое моделирование наиболее эффективно при условии использования средств компьютерной поддержки. В первую очередь это средства визуального моделирования, позволяющие наглядно представлять для пользователя, дополнять и модифицировать модель, а также различные средства анализа и симуляции. Разработка такого программного обеспечения стала возможна только в последние 10-15 лет вследствие развития информационных технологий и соответствующих аппаратных средств. Это в свою очередь явилось стимулом для развития самого подхода и его практических приложений.

Одна из областей применения сетевых моделей – это исследование биологических систем, по своей природе являющихся сложными. Сетевая модель биологической системы позволяет изучать ее в комплексе, учитывая непрямые и неочевидные влияния одних элементов на другие, выявлять значимые функциональные элементы и т.д. Например, применительно к молекулярно-генетическим системам, таким способом может исследоваться функционирование организмов на уровне молекулярных и суб-молекулярных элементов (генов, белков) в различных условиях *in silico* и без постановки дорогостоящих экспериментов *in vitro*, планироваться такие эксперименты, изучаться воздействие лекарственных препаратов, проектироваться трансгенные организмы с целевыми свойствами.

Наиболее трудоемкой стадией сетевого анализа биологических систем является реконструкция сетевой модели. Типичные размеры таких моделей для молекулярно-генетических

систем – от сотен до десятков тысяч элементов. Трудоемкость обусловлена необходимостью извлечения информации из большого количества публикаций и баз данных, причем в большинстве случаев это ручная работа. Данные разнородны, не полностью структурированы, представлены в различных форматах, имеют различную семантику. Часто в данных встречаются ошибки, пробелы противоречия между различными источниками. Вследствие чего их сбор и обработка требует значительных трудозатрат.

Таким образом, актуальна задача автоматизации процесса реконструкции сетевых моделей с помощью средств компьютерной поддержки. Такая автоматизация избавит исследователя от части рутинной работы и повысит эффективность исследований, в частности, при решении вышеперечисленных задач.

Цель работы

Целью работы является разработка методов и средств компьютерной поддержки, направленных на автоматизацию и повышение эффективности реконструкции сетевых моделей сложных биологических (молекулярно-генетических, экологических) систем.

Молекулярно-генетические системы рассматриваются, как основной объект исследования, на котором отрабатываются все принципы и методы и реконструкции. Далее методика обобщается и переносится на экологические системы.

Для достижения поставленной цели в работе решаются следующие задачи:

- анализ проблематики предметной области и потребностей пользователей с целью формулировки требований к программному обеспечению, предназначенному для поддержки реконструкции сетевых моделей;
- разработка методов универсального компьютерного представления и реконструкции сетевых моделей сложных биологических систем, обеспечивающих эффективный ввод, верификацию и хранение такого рода данных, а также настраиваемых на проблематику решаемых прикладных задач;
- разработка методов централизованного структурированного поиска данных об элементах молекулярно-генетических систем во внешних базах данных и других структурированных источниках, основанных на их семантической интеграции;
- разработка архитектуры программного обеспечения, основанного на созданных методах универсального

представления биологических данных, их семантической интеграции, а также реконструкции сетевых моделей;

- реализация программного обеспечения, позволяющего осуществлять структурированный централизованный поиск данных по расширяемому набору внешних источников информации с целью повышения эффективности процесса сбора первичных данных для реконструкции сетевых моделей;
- реализация универсального программного обеспечения для визуальной реконструкции сетевых моделей биологических систем, настраиваемого на проблематику различных задач, связанных с моделированием молекулярно-генетических и экологических систем;
- настройка и адаптация разработанных программных средств для решения следующих прикладных задач:
 - интеграция ряда баз данных по элементам молекулярно-генетических систем и их сетевым моделям;
 - реконструкция сетевых моделей молекулярно-генетических систем;
 - реконструкция сетевых моделей экосистем на примере иксодовых клещей.

Методы исследования

Для разработки методов представления и реконструкции сетевых моделей, а также проектирования программных средств применяется объектно-ориентированная методология, включая методы объектно-ориентированного анализа предметной области и архитектуры, проектирования и программирования. Для оптимизации производительности операций поиска и интеграции данных применяется реляционная алгебра, методы проектирования баз данных, в частности объектно-реляционного отображения. Также используются методы семантической интеграции структурированных данных (data warehousing) и методы представления онтологий.

Научная новизна

В работе получены следующие новые научные результаты.

1. Разработана и применена методика анализа требований к ПО для задач с нечетко определенной терминологической базой. В предложенной методике в первую очередь рассматриваются требования к представлению данных, включая требования к метамодели, отражающей семантику поддерживаемых онтологий. Далее, функциональные требования формулируются в терминах

метамодели. Стандартная методика, напротив, рассматривает в первую очередь функциональные требования, которые выражаются в терминах рассматриваемой предметной области.

2. Разработаны и реализованы в виде программного приложения, методы, обеспечивающие визуальную реконструкцию сетевых моделей биологических систем на основе задаваемой пользователем онтологии. Обеспечивается адаптивность под различные виды биологических систем (молекулярно-генетические и экологические) и строгая типизация вершин. Строгая типизация необходима для дальнейшего применения методов компьютерного анализа к реконструированной сетевой модели, в том числе проблемно-ориентированных, требующих биологической интерпретации входных данных. Наиболее распространенные аналоги (Cytoscape, CellDesigner, GeneNet) либо работают с фиксированной моделью предметной области, либо не обеспечивают необходимого уровня типизации данных.

3. Разработаны методы и основанные на них программные средства, обеспечивающие семантическую интеграцию слабоструктурированных биологических данных. В отличие от существующих аналогов (K2/Klesli, Biomediator) используется предварительная интеграция и индексация данных на основе онтологии, что обеспечивает:

- высокую производительность при выполнении запросов, не зависящую от качества канала связи с внешними источниками;
- идентификацию синонимичных объектов из различных источников, противоречий и пробелов в информации, а также возможность контроля релевантности источников.

4. Разработан язык запросов, обеспечивающих поиск данных, представленных в рамках разработанной метамодели. Принципиальным отличием от SQL-подобных языков является поддержка критериев запросов относительно многозначных (неатомарных) свойств.

Практическая значимость

Разработанное программное обеспечение внедрено:

- Для реконструкции сетевых моделей молекулярно-генетических систем. Используется в качестве клиентской части ранее эксплуатируемой системы.
- Для реконструкции сетевых моделей экосистем в ИЦиГ и Новосибирском государственном университете (НГУ). Применяется для решения задач, связанных с построением универсальной базовой онтологии экосистем, а также задач моделирования ряда экосистем.

- Для поиска данных по элементам молекулярно-генетических систем во внешних базах данных в ИЦиГ. Обеспечивается интеграция 7-ми баз.

Апробация работы

Основные защищаемые положения были доложены на международных конференциях Bioinformatics of Genome Regulation and Structure BGRS'06 и BGRS'08.

Автором опубликовано 11 печатных работ по теме диссертации.

Структура и объем работы

Диссертационная работа состоит из введения, четырех глав, заключения и списка литературы. Объем диссертации – 148 стр. текста. Список литературы содержит 77 наименований. Работа включает 28 рисунков.

СОДЕРЖАНИЕ РАБОТЫ

В первой главе описывается проблематика реконструкции сетевых моделей биологических систем, приводится обзор существующих программных решений, а также рассматривается ряд методов и технологий, необходимых для дальнейшего изложения.

В начале главы приводится общий обзор применения сетевых моделей в различных областях, а также выделяются общие характеристики этой методики, необходимые для применения в данной работе. В разделах «Понятие и проблематика генных (экологических) сетей» приводится характеристика существующих понятий о молекулярно-генетических и экологических системах, описывается их специфика с точки зрения сетевого моделирования, а также в общем виде формулируются требования к программным средствам реконструкции сетевых моделей типов живых систем.

В разделе «Обзор программных средств реконструкции сетевых моделей» приводится сравнительный анализ и классификация существующих программных средств. В первую очередь рассматриваются средства для исследования живых систем, но также приводятся несколько аналогов из других предметных областей. Отдельно и подробно рассматривается система GeneNet, являющаяся функциональным прототипом пользовательского интерфейса в данной работе. По результатам

анализа программных средств и предметной области формулируются следующие основные требования и ограничения для системы реконструкции:

- визуальное редактирование сетевых моделей;
- настраиваемость системы пользователем на предметную область включая типизацию вершин и связей сетевой модели, а также графическую нотацию;
- поддержка механизмов декомпозиции для работы со сложными сетевыми моделями;
- централизованное хранение данных и многопользовательский доступ к ним.

В заключении формулируется задача интеграции биологических данных, как одна из задач, связанных с реконструкцией сетевых моделей живых систем. Система интеграции данных должны обеспечивать следующую функциональность:

- семантическая интеграция данных из различных источников на основе заданной модели предметной области;
- контроль целостности, качества и происхождения данных;
- поддержка структурированных запросов по интегрируемым данным.

Также приводится описание и сравнительный анализ существующих методик и программных средств для интеграции биологических данных, в том числе семантической интеграции. Одной из методик является *data federation*, она основана на динамической трансляции запросов и делегировании их всем поддерживаемым базам. Альтернативой является методика *data warehousing*, обеспечивающая предварительную интеграцию данных и построение запросов к уже интегрированным данным. Другие методики не обеспечивают или существенно ограничивают анализ семантики данных, а частности не позволяют выполнять структурированный поиск.

Во второй главе проводится анализ требований к программному обеспечению. Приводится развернутая формулировка требований, включая основные варианты использования. Требования основаны на проведенном автором анализе проблематики предметной области, существующих методов реконструкции сетевых моделей и связанных с ними недостатков.

Терминологическая база для описания биологических систем, особенно экосистем, четко не определена. Это делает

неприемлемым применение классических подходов к проектированию ПО, где терминологическая база предметной области используется в явном виде для описания требований и архитектуры. Решением, предложенным и реализованным в данной работе, является замена естественной модели предметной области на модель с более высоким уровнем абстракции (метамодель), которая обладает следующими характеристиками: (1) является абстракцией, достаточной для решения задач реконструкции сетевых моделей и их анализа; (2) обеспечивает представление модели произвольной предметной области в рамках поставленных задач (в форме онтологии).

В работе предлагается и используется специализированная методика анализа требований, так как традиционные методики, делающие акцент на функциональных требованиях, в данной работе неприменимы. Суть примененной методики состоит в рассмотрении в первую очередь требований к представлению данных и семантике метамодели. Далее функциональные требования формулируются в терминах метамодели, а не модели предметной области, как в традиционных методиках.

Применение метамодели с одной стороны усложняет программные средства, однако в данном случае это необходимо для решения задач интеграции данных, а также реконструкции моделей экосистем. Действительно, методы семантической интеграции данных подразумевают наличие онтологического описания интегрируемых данных, причем это описание может расширяться при включении в систему новых внешних информационных источников (баз данных). При реконструкции моделей экологических систем (в отличие от молекулярно-генетических) на текущий момент невозможно создать достаточно универсальную модель предметной области, т.е. онтологию. В результате для каждой экосистемы такая онтология должна строиться пользователем, не имеющим возможности изменения программного кода.

В главе сформулированы и обоснованы требования к метамодели, операциям с ней, обеспечению целостности данных, представленных в рамках метамодели. Метамодель определяет семантику поддерживаемых онтологий и сетевых моделей. Основные требования следующие:

- онтология представляется в форме набора типов вершин и их свойств;
- связи в сетевой модели не типизируются, их семантика определяется вершинами;
- поддержка многозначной типизации ссылочных свойств;

- поддержка многозначных свойств, в том числе ссылочных (для реализации синонимии и отношений с мощностью «много-к-многим»);
- контроль целостности данных посредством проверки правил и ограничений, определяемых онтологией;
- поддержка транзакционности для серий операций, локально нарушающих целостность данных;
- расширяемость относительно новых типов свойств и ограничений.

Требования к метамодели являются общими для задач реконструкции сетевых моделей и интеграции данных и учитывают описанную выше проблематику этих задач.

Далее приводятся требования, специфические для отдельных задач, поставленных в работе. Требования к интеграции данных основаны на проведенном анализе существующих баз, качества данных, типичных сценариев их использования, поддерживаемой функциональности (в первую очередь семантики запросов). Приведем основные требования:

- набор интегрируемых источников должен быть расширяемым без изменения программного кода системы;
- запросы должны выполняться на основе заранее интегрированных данных, представленных в рамках метамодели и заданной онтологии;
- запросы должны быть структурированы, т.е. критерии запросов должны быть выражены в терминах типов и свойств (включая каскадное использование ссылочных свойств), задаваемых онтологией;
- при интеграции должны идентифицироваться эквивалентные объекты, полученные из разных источников, и объединяться в автоматическом режиме либо с участием пользователя (должны поддерживаться оба режима), результаты такого объединения должны учитываться при выполнении запросов;
- должна сохраняться информация о происхождении всех данных в системе, вплоть до отдельных значений свойств (необходимо для контроля достоверности данных).

Далее приводятся формулировки требований, связанных с визуальной реконструкцией сетевых моделей. Требования основаны на анализе опыта использования редактора GeneNet и других альтернативных программных средств. Кроме того, учитывается опыт, накопленный в других областях (проектирование ПО, анализ бизнес-процессов и т.д.) и

требования, сформулированные относительно метамоделей. Основные требования следующие:

- возможность задания онтологии и графической нотации пользователем (посредством специализированного редактора);
- поддержка графической нотации, параметризуемой относительно выбираемых пользователем свойств, заданных в онтологии;
- поддержка повторного использования объектов в различных моделях, при этом объект может иметь как общие для всех моделей свойства (надсистемные), так и специфичные для конкретной модели (системные);
- визуальное редактирование сетевой модели и отдельных ее элементов с учетом заданной онтологии, включая поддержку управляемой пользователем транзакционности при создании и редактировании объекта;
- поддержка структурной декомпозиции сетевых моделей на основе компартов.

В заключении приводится ряд требований, связанных с расширяемостью, адаптивностью и контролем доступа к данным.

Третья глава посвящена описанию архитектуры и реализации системы.

В начале главы проанализированы различные способы реализации требований, а также обоснован выбор технологической базы. Основные критерии выбора – это производительность и трудоемкость реализации. Для проведения такого анализа были разработаны ряд прототипов программных средств. На основе проведенного анализа и прототипирования были уточнены требования к семантике метамоделей, контролю целостности данных, а также сделан вывод о том, критическим элементом с точки зрения производительности является реализация персистентности и поиска данных. Показано, что единственным приемлемым решением является статическое отображение метамоделей на схему реляционной СУБД и использование оптимизированных с учетом функциональности и архитектуры системы механизмов объектно-реляционного отображения.

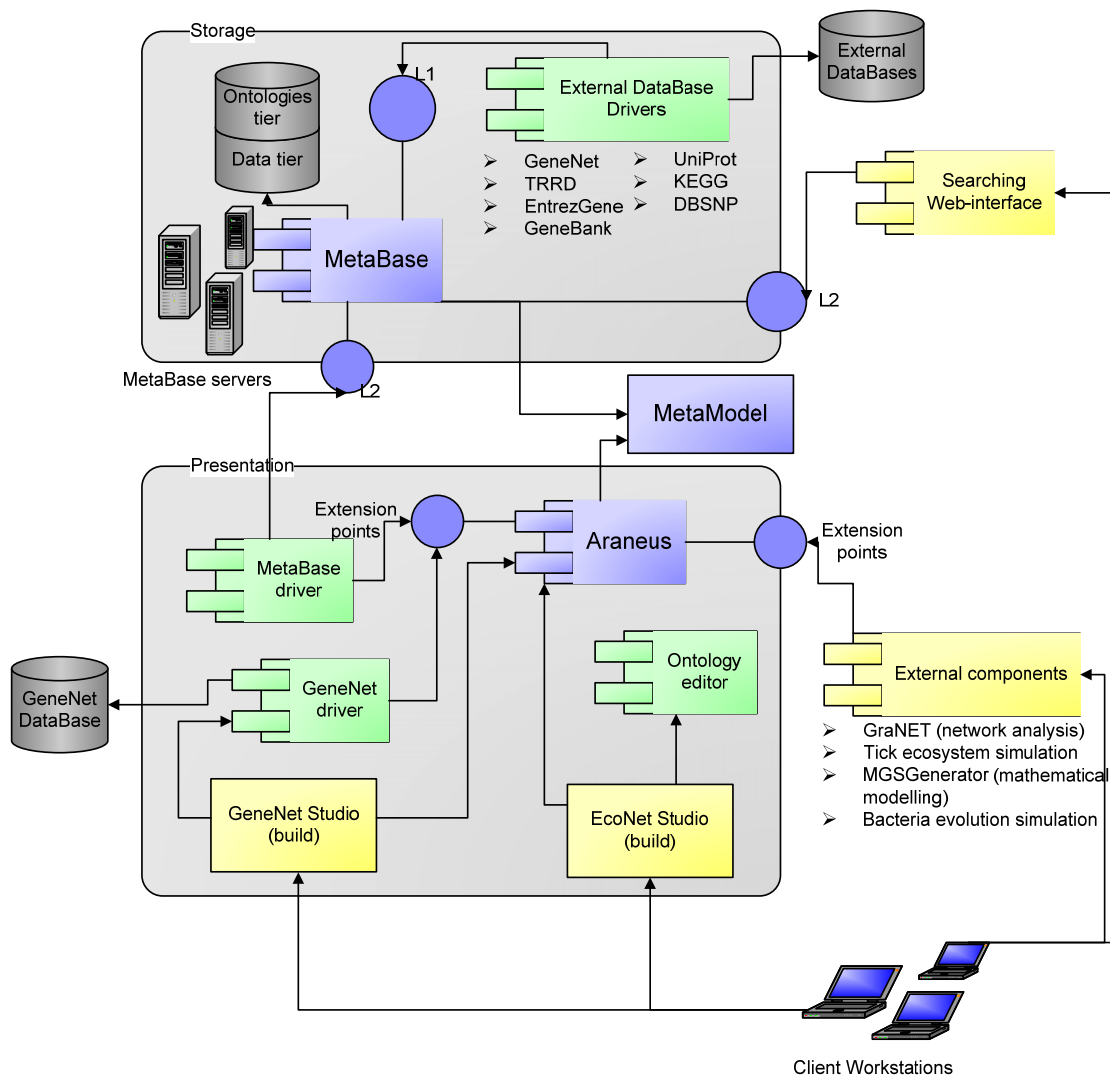


Рис 1. Общая архитектура системы

Далее описывается общая архитектура программных средств (рис. 1). Основными компонентами ее являются (1): метамодель (MetaModel), (2) сервер для хранения, поиска и интеграции данных MetaBase и (3) подсистема реконструкции сетевых моделей Araneus.

Метамодель является наиболее важным архитектурным элементом в разработанных методах и программных средствах. В работе описываются ее основные элементы, приводится сравнение с существующими аналогами (в первую очередь с метамоделью OWL), а также показывается, что она удовлетворяет всем предъявляемым к ней требованиям. Рассмотрим более подробно структуру базовой метамодели (рис. 2)

Для описания используется нотация UML, используются отношения генерализации (наследования), а также ассоциации и агрегации с указанием мощности отношений. Серым выделены ключевые классы. Класс Ontology (онтология) в данной

метамоделю является организационной сущностью, характеризуется названием и набором типов, в нее входящих. Классы *Type*, *Object*, *Property*, *Value* представляют классическую объектную модель (*Type* соответствует понятию класса). По умолчанию все свойства (если не наложены дополнительные ограничения) могут иметь множественное значение (что, в частности, является существенным отличием по отношению к реляционной модели данных). Свойства типизированы, что отражено деревьями наследования *Property* и *Value*. Ссылочные свойства представлены подклассом *ObjectProperty*, характеризуются набором типов, на экземпляры которых они могут ссылаться (в классической объектной модели такой тип может быть только один, если не считать его подклассы).

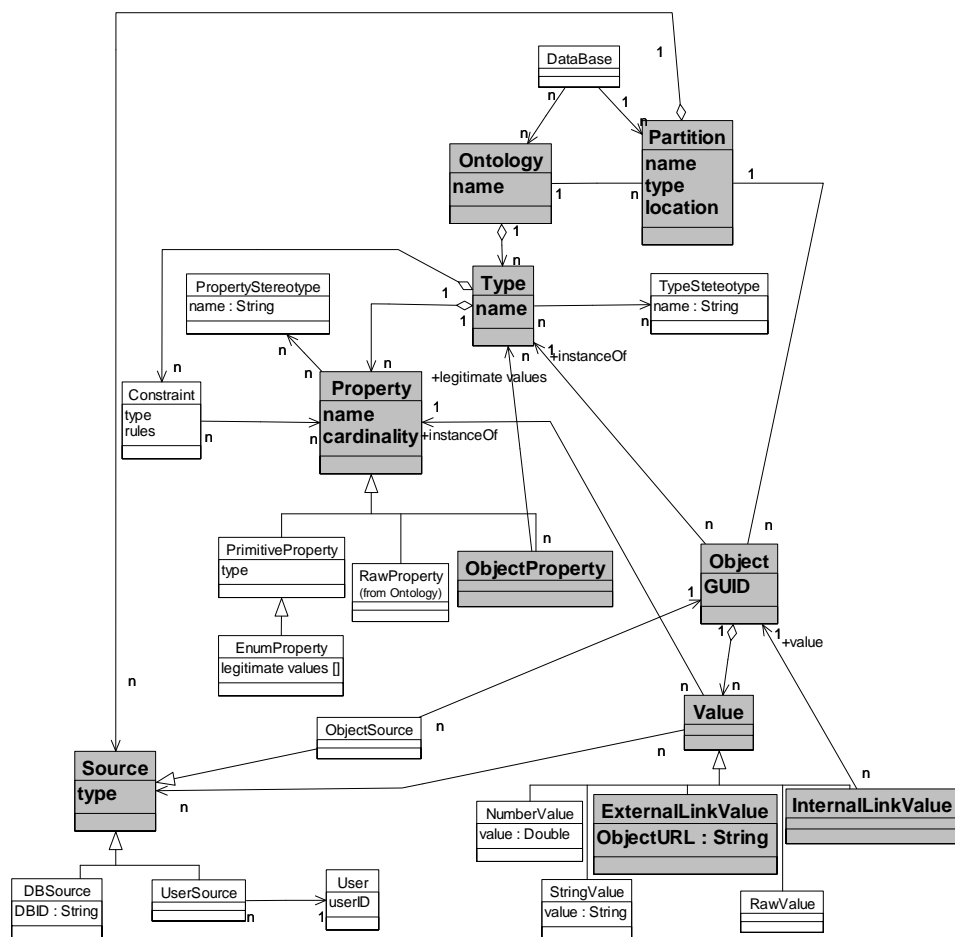


Рис. 2. Базовая метамодель

Раздел (*Partition*) играет роль организационной сущности для данных. Раздел может в себе содержать несколько онтологий и наоборот, одной онтологии могут соответствовать несколько разделов. Эта сущность может быть использована для создания

распределенных баз и реализации некоторых других возможностей. Значения ссылочных свойств подразделяются на два типа, аналогичные жестким и символическим ссылкам в файловых системах. Внутренние (аналог жестких) ссылки не могут пересекать границы раздела, при этом могут быть использованы в критериях запросов. Внешние ссылки могут ссылаться на объекты в других разделах, однако предназначены исключительно для навигации и целостность их не гарантируется системой.

Для обеспечения контроля происхождения и достоверности данных предназначен класс *Source* (источник). Каждое значение свойства должно быть подтверждено хотя бы одним источником. Подклассы *Source* соответствуют различным способам подтверждения информации. Это может быть литературный источник (ссылка на публикацию), web-ссылка, ссылка на внешнюю базу данных и т.д. Источники используются в первую очередь при интеграции внешних баз данных.

Кроме того, для представления сетевых моделей метамодель расширяется такими классам, как «схема» (сеть), «слой». Эти понятия необходимы в первую очередь для декомпозиции сложных моделей. Поддерживается структурная декомпозиция на основе компарментов (реализующих отношение вложенности) и функциональная на основе связанных слоев. Показано, что дополнительные элементы метамодели могут быть сведены к ранее определенным в рамках базовой метамодели.

Компонента *MetaBase* реализует требования, связанные с интеграцией и поиском данных. В тексте описан алгоритм интеграции данных, обеспечивающий выявление эквивалентных объектов в различных базах, а также пробелов и противоречий в данных. Описан способ интеграции новых баз, при котором от пользователя требуется реализация драйвера, осуществляющего преобразование форматов данных и, при необходимости, расширение онтологии. Также описывается реализация требований, связанных с обеспечением целостности данных и контролем доступа к ним.

В связи с тем, что структура метамодели не соответствует реляционной, потребовалось создание нового языка запросов. Ключевыми особенностями языка являются:

- поддержка кванторов существования и общности в критериях запросов;
- поддержка предикатов над множествами: отношения эквивалентности множеств, вложенности, пресечения;

- поддержка запросов относительно иерархических множеств, образуемых многозначными объектными ссылками.

В главе приведено описание его семантики и реализованного варианта синтаксиса, близкого к SQL. Также приведены примеры запросов для поиска молекулярно-генетических данных.

В разделе «Реализация» приводится описание реализации интерфейса компоненты *MetaBase*, технических решений, использованных при реализации, а также описание ряда периферийных компонент, взаимодействующих с ней: драйверов баз данных и пользовательского поискового интерфейса.

Компонента *Araneus* представляет собой библиотеку, которая обеспечивает всю необходимую функциональность для создания сетевых редакторов. Компонента спроектирована и реализована как набор подключаемых модулей, реализующих метамодель, а также различные функции визуализации и визуального моделирования. Это позволяет строить из них различные приложения, которые при необходимости могут использовать и внешние специализированные для решения определенных задач модули. В тексте приведена архитектура этой компоненты, которая обеспечивает основные функции для визуального редактирования сетевых моделей, а также три основных дистрибутива, собранных на основе этой компоненты: редакторы генных сетей, экосетей и онтологий. Описаны основные модули *Araneus*, их взаимосвязь и назначение, а также способы подключения дополнительных компонент.

В четвертой главе описано практическое применение разработанных программных средств.

Задача интеграции баз данных в области молекулярно-генетических систем решена для таких баз, как EntrezGene, UniProt, KEGG, DBSNP, части GeneBank, связанной с представлением информации об РНК, а также TRRD и базы системы GeneNet, разработанными в ИЦиГ СО РАН. Решение заключается в (1) разработке драйверов соответствующих баз, (2) разработке единой онтологии, (3) интеграции части данных из всех перечисленных баз (по отдельным наиболее актуальным с точки зрения исследований, проводимых в ИЦиГ видам организмов). Результатом является приложение с web-интерфейсом, обеспечивающее поиск по интегрированным данным из этих баз.

Для реконструкции сетевых моделей генных сетей было создано приложение *GeneNet Studio* (рис. 3) на основе компоненты

Araneus. В данной главе кратко описан ряд дополнительных реализованных компонент в составе данного приложения, предназначенных для анализа сетей, импорта и экспорта данных.

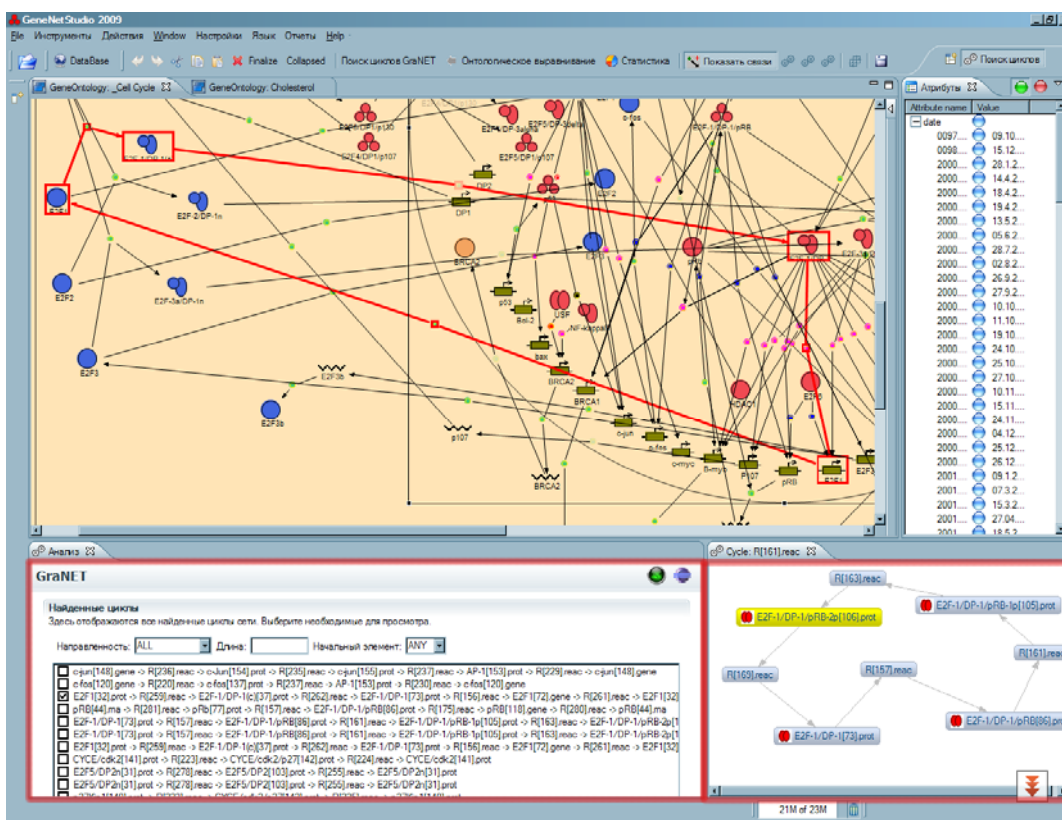


Рис. 3. GeneNet Studio – приложение для реконструкции и анализа сетевых моделей молекулярно-генетических систем

Также на основе *Araneus* был создан пользовательский интерфейс для среды моделирования эволюции популяций одноклеточных организмов. Программная среда также включает реализацию алгоритма моделирования на основе методики «Эволюционный конструктор». В тексте приведено краткое описание задачи, а также описаны особенности реализации данного приложения, в частности дополнительные модули, не входящие в состав *Araneus*.

Для приложения *EcoNet Studio* для визуальной реконструкции сетевых моделей экосистем проиллюстрировано практическое использование при решении задачи моделирования экосистемы иксодовых клещей (переносчиков клещевого энцефалита и боррелиоза). Приведено краткое описание предметной области и постановка задачи. Представлена реконструированная сетевая модель жизненного цикла клеща. Автором также была построена имитационная (дискретно-

событийная) модель поведения клеща. Предложен способ автоматической генерации имитационных моделей экосистем на основе сетевых моделей. В тексте описаны принципы построения онтологии и сетевой модели для решения такой задачи, а также предложено универсальное представление генерируемой имитационной модели, не привязанное к конкретной предметной области.

ЗАЩИЩАЕМЫЕ ПОЛОЖЕНИЯ

На защиту выносятся следующие положения:

1. Предложена и применена специализированная методика анализа требований к программному обеспечению для задач с нечетко определенной терминологической базой.
2. Впервые предложены методы универсального компьютерного представления сетевых моделей различных видов сложных биологических систем, основанные на метамодели и онтологическом описании. Методы включают в себя метамодель, правила представления онтологий и сетевых моделей, правила декомпозиции для оперирования сложными сетевыми моделями.
3. Впервые предложены методы семантической компьютерной интеграции слабоструктурированных биологических данных, обеспечивающие верификацию данных и структурированный поиск на основе онтологий.
4. На основе предложенных методов спроектировано и реализовано программное обеспечение для визуальной реконструкции сетевых моделей. В отличие от известных аналогов оно обеспечивает пользовательскую настройку на решаемую задачу путем задания онтологии и графической нотации.
5. На основе предложенных методов спроектировано и реализовано программное обеспечение для семантической интеграции биологических баз данных. Построена общая онтология и проведена интеграция для следующих баз: GenBank, KEGG, dbSNP, EntrezGene, UniProt, TRRD и GeneNet.
6. Предложен и реализован язык запросов, обеспечивающий структурированный поиск по интегрированным данным с учетом заданной онтологии. В отличие от SQL-подобных языков, предложенный язык оперирует критериями поиска, задаваемыми относительно многозначных свойств.
7. Разработанные методы и программное обеспечение применены для решения следующих биологических задач:

- реконструкция сетевых моделей сложных молекулярно-генетических систем;
- сетевое и имитационное моделирование экосистемы иксодовых клещей (переносчиков клещевого энцефалита).

ВКЛАД АВТОРА

Все основные результаты, включая постановки задач, методы, алгоритмы и основополагающие архитектурные решения получены автором, либо под его научным руководством и при непосредственном участии.

Автором лично проведен анализ проблематики предметной области, анализ требований, разработаны методы представления и семантической интеграции биологических данных, включая метамодель и семантику онтологий, а также математическую модель языка запросов. Также лично автором разработаны описанные в тексте диссертации прототипы программных средств.

В части связанной с практическим применением результатов работы (глава 4) лично автором была построена онтология для интеграции биологических баз данных, онтология и сетевая модель жизненного цикла клеща, разработана методика генерации имитационных моделей экосистем на основе их сетевого представления.

Проектирование и реализация компоненты Araneus и основанных на ней программных средств выполнены совместно с Тимоновым В.С.. Проектирование и реализация компоненты MetaBase, а также разработка языка запросов выполнены совместно с Лабужским В.В. и Соколовым С.А..

Пользовательский интерфейс для построения запросов разработан Морозовым А.В.. Компоненты анализа сетевых моделей, а также пользовательский интерфейс для задачи моделирования эволюции популяций одноклеточных организмов разработаны Тимоновым В.А. Методика «Эволюционный конструктор» и ее программная реализация являются разработками Лашина С.А.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Д.С. Мигинский, В.В. Лабужский, М.М. Лаврентьев-мл., А.В. Морозов, С.А. Соколов Технология семантической интеграции баз данных в системной биологии // Вычислительные технологии. 2008. Т. 13, N 6, С. 103-119
2. Miginsky D.S., Suslov V.V., Timonov V.S., Rasskazov D.A., Sournina N.Yu., Podkolodny N.L. Approaches to the Computer

- Reconstruction of the Biological Networks // *Intelligent Data Analysis*, Vol. 12, No. 5 (2008) P. 463-479.
3. E. A. Ananko, N. L. Podkolodny, I. L. Stepanenko, O. A. Podkolodnaya, D. A. Rasskazov, D. S. Miginsky, V. A. Likhoshvai, A. V. Ratushny, N. N. Podkolodnaya, and N. A. Kolchanov GeneNet in 2005 *Nucl. Acids Res.* 2005 33: D425-D427
 4. Колчанов Н.А., Мигинский Д.С., Подколотный Н.Л., Рассказов Д.А., Суслов В.В., Тимонов В.С., Сергеев М.Г. Сетевое описание биосистем: проблемы и подходы // *Вычислительные технологии.* 2007. Т. 12. Специальный выпуск 2: Информационные технологии для эколого-биологических исследований. Междисциплинарный интеграционный проект СО РАН. С. 107-122
 5. Сергеев М.Г., Суслов В.В., Мигинский Д.С., Юрлова Н.И., Колчанов Н.А. Опыт создания моделей на основе онтологии экосистем. В кн. *Биоразнообразие и динамика экосистем: информационные технологии и моделирование* (ред. В.К. Шумный, Ю.И. Шокин, Н.А.Колчанов, А.М. Федотов), Новосибирск, 2006, стр.95-118.
 6. Колчанов Н.А., Подколотная О.А., Ананько Е.А., Игнатьева Е.В., Степаненко И.Л., Хлебодарова Т.М., Меркулова Т.И., Меркулов В.М., Мищенко Е.Л., Ибрагимова С.М., Смирнова О.Г., Подколотный Н.Л., Ромашенко А.Г., Ощепков Д.Ю., Мигинский Д.С. Регуляторные последовательности ДНК: описание в базах данных. В кн. *Системная компьютерная биология* (ред. Н.А. Колчанов, С.С. Гончаров, В.А. Лихошвай, В.А. Иванисенко) Новосибирск, 2008, стр. 15-38.
 7. Khlebodarova, O. Podkolodnaya, D. Oshchepkov, D. Miginsky, E. Ananko, E. Ignatieva *Artsite Database: Comparison of in Vitro Selected and Natural Binding Sites of Eukaryotic Transcription Factors.* In: *Bioinformatics of Genome Regulation and Structure II.* (Eds. N.Kolchanov and R. Hofestaedt) Springer Science+Business Media, Inc. 2006, pp. 55-65.
 8. Miginsky D.S., Suslov V.V., Rasskazov D.A., Podkolodny N.L., Kolchanov N.A *Architecture of Software Toolkit for Storing and Operating with Biosystems Models // Proceedings of the Fifth International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2006), v. 3 pp.292-295.*
 9. Miginsky D.S., Sokolov S.A., Labuzhsky V.V., Nikitin A.G., Tarancev I.G. *Object-Oriented Approach to Bioinformatics Software Resources Integration // Proceedings Of The Fifth*

- International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2006), v. 3 pp.288-291.
10. Podkolodny N.L., Podkolodnaya N.N., Miginsky D.S., Poplavsky A.S., Likhoshvai V.A., Compani B., Mjolsness E. An Integration of the Descriptions of Gene Networks and Their Models Presented in Sigmoid (Cellerator) and Genenet // Proceedings Of The Fifth International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2006), v. 3 pp.86-90.
 11. Timonov V.S., Miginsky D.S. "PROMETHEUS" Toolkit for Agile Development of Biological Data Storing and Access Software // Proceedings of the Sixth International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2008), p. 237.