

На правах рукописи



Тимонов Владимир Сергеевич

**Программные технологии визуальной
реконструкции и анализа сетевых моделей
генетических, экологических и социальных
систем**

05.13.11 – Математическое и программное обеспечение вычислительных
машин, комплексов и компьютерных сетей

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата технических наук

Красноярск – 2011

Работа выполнена в *отделе системной биологии Учреждения Российской академии наук Института Цитологии и Генетики СО РАН*
(г. Новосибирск)

Научный руководитель: *академик РАН, д.б.н., профессор*
Колчанов Николай Александрович

Официальные оппоненты: *д.т.н., профессор*
Доррер Георгий Алексеевич

к.т.н., доцент
Швец Сергей Викторович

Ведущая организация: ***Учреждение Российской академии наук Институт вычислительной математики и математической геофизики СО РАН***
(г. Новосибирск)

Защита состоится «4» марта 2011 г. в 16-00 часов на заседании диссертационного совета ДМ 212.099.05 при Сибирском федеральном университете по адресу: 660074, г. Красноярск, ул. Киренского, 26, ауд. УЛК-115.

С диссертацией можно ознакомиться в библиотеке Сибирского федерального университета.

Автореферат разослан «_» февраля 2011 г.

Ученый секретарь
диссертационного совета



Непомнящий О.В.

Общая характеристика работы

Актуальность работы. Сетевое моделирование процессов различной природы является одним из наиболее эффективных и универсальных способов описания их структурно-функциональной организации (СФО). Визуальная сетевая реконструкция является разработкой компьютерной модели в виде ориентированного графа, семантически отражающего структурные и функциональные характеристики (СФХ) описываемого процесса или явления. Информативным и удобным для исследователей являются возможности визуального построения сетевых моделей. Они дают наглядное представление описываемого процесса или явления с возможностями модификации и дальнейшего анализа сети. Такие процессы могут включать в себя богатую терминологическую и семантическую составляющие, а также от сотен до десятков тысяч элементов, что также следует учитывать при дальнейшем анализе сети. Сетевое моделирование нашло широкое применение в таких областях, как проектирование программного обеспечения, схемотехника, планирование и оптимизация бизнес-процессов и др. Активные исследования, проводимые с применением современных информационных технологий, средств вычислительной техники и методов прикладной математики в области молекулярно-биологических, биомедицинских, экологических и социальных исследований способствовали развитию информационной биологии. В данной работе сетевое моделирование рассматривается по отношению к трем направлениям информационной биологии: (1) молекулярно-генетические сети; (2) экологические сети и (3) социальные сети. Выбор направлений связан с потенциальными возможностями получения полезных прикладных и фундаментальных результатов от исследований в каждой из областей в организации, где проводилась работа. Для реконструкции различных сетевых моделей в мире разработан ряд программных средств (ПС), каждое из которых обладает своими преимуществами и недостатками (Pajek, Cytoscape, Cell-Designer, Vizster и др.). Их отличительной особенностью является ориентированность на конкретную прикладную область с заранее определенной терминологией и семантикой либо полным их отсутствием, что накладывает свои ограничения на проводимые исследования. Этот критерий наиболее важен в областях, где предметная область слабо определена либо не определена вовсе, а формируется и уточняется в процессе проведения исследования. При стандартном подходе к разработке программного обеспечения (ПО), где проводится полное исследование предметной области, проектирование архитектуры и ее программная реализация практически с нуля, страдает эффективность исследований, в виду отсутствия своевременной инструментальной поддержки. При этом IT-специалисты вынуждены тесно взаимодействовать с исследователями, занимаясь анализом их требований и формализацией исследуемых

систем, необходимых для программной реализации, на что тратится время, которое можно бы было эффективно использовать для наукоемких задач. С учетом проведения широкомасштабных исследований в области сетевого моделирования в организации, где проводилась работа, возникла необходимость в подходе к быстрой разработке таких программных сред моделирования. К настоящему моменту отсутствовали эффективные программные технологии, предназначенные для быстрой разработки ПС визуальной реконструкции и анализа сетевых моделей, с возможностью гибкой настройки на конкретную предметную область с помощью определения терминологии и семантики непосредственно исследователями в рамках поставленных задач.

Актуальность работы заключается в потребности создания математического и программного обеспечения, повышающего эффективность разработки предметно-ориентируемых ПС, предназначенных для реконструкции и анализа сетевых моделей.

Целью диссертационной работы является разработка подхода, методов и алгоритмов для повышения эффективности процесса создания предметно-ориентируемых средств визуального сетевого моделирования с функциями анализа данных.

Для достижения поставленной цели были поставлены следующие **задачи**:

1. Обзор направлений сетевого моделирования и средств инструментальной поддержки исследований в его рамках, а также выявление проблем разработки таких инструментов.
2. Разработка подхода и алгоритмов, описывающих процесс унификации создания предметно-ориентируемого средства визуального сетевого моделирования.
3. Создание методов и алгоритмов для универсального компьютерного описания, визуального представления и анализа сетевых моделей.
4. Проектирование и реализация средств инструментальной поддержки для разработки сред визуального сетевого моделирования.
5. Выявление состава и структуры исследуемых систем, описание их предметных областей для проведения исследований в области информационной биологии.
6. Применение подхода и средств инструментальной поддержки для разработки программного обеспечения, предназначенного для реконструкции молекулярно-генетических, экологических и социальных систем.

Объектом исследования в работе являются методы и средства визуального сетевого моделирования.

Предметом исследования в работе являются подходы к унификации разработки предметно-ориентированных сред визуальной сетевой реконструкции и анализа с гибкой настройкой на предметную область.

Методы исследования. Для достижения цели и решения поставленных задач в работе использовались методы теории графов, теории множеств, статистики, сетевого моделирования, теории языков программирования, онтологического описания, объектно-ориентированного анализа и проектирования, а также дизайна визуальных интерфейсов.

Научная новизна.

1. Предложены алгоритмы и программные инструменты разработки проблемно-ориентируемых ПС в области сетевого моделирования, основанные на использовании готового каркаса приложения и позволяющие осуществлять гибкую настройку на предметную область без участия IT-специалистов, тем самым значительно снижая сроки разработки ПО.
2. Предложен научно-обоснованный метод каркасной сборки программ и программных систем, обеспеченный разработанной базовой инструментальной поддержкой, позволяющий реализовать разработку программных сред визуальной реконструкции и анализа сетевых моделей.
3. Предложен и реализован декларативный язык для описания предметных областей, основанный на гибкой параметризуемой графической нотации, позволяющий осуществлять построение семантических сетей прикладных систем.
4. Разработаны и реализованы методы и алгоритмы, направленные на обработку структурно-функциональной организации сетевых моделей, позволяющие повысить эффективность анализа данных и прогнозирования поведения исследуемых систем.

Практическая ценность. Предложенный подход и средства инструментальной поддержки можно применять для сокращения сроков разработки предметно-ориентированных сред в различных областях, где можно применить метод сетевого моделирования. Созданные алгоритмы и программные средства официально внедрены и используются:

- в научно-исследовательском процессе Института Цитологии и Генетики СО РАН (ИЦиГ СО РАН) для разработки предметно-ориентированных сред визуального сетевого моделирования в области информационной биологии («Araneus»);
- в научно-исследовательском и образовательном процессах ИЦиГ СО РАН в области моделирования молекулярно-генетических и экологических систем («GeneNetStudio», «EcoNetStudio»);
- в научно-исследовательском и образовательном процессах Новосибирского государственного университета (НГУ), в рамках применения для решения задач, связанных с моделированием экологических процессов («EcoNetStudio»);
- в научно-исследовательском процессе ООО «Генные сети» (ГеС) с целью проведения комплексных работ в области идентификации конфликт-

ных ситуаций в организациях (конфликтологии) («SocioTomStudio»).

Положения, выносимые на защиту:

1. Использование предложенных алгоритмов и программных инструментов позволяет значительно снизить сроки разработки сред сетевого моделирования.
2. Применение метода каркасной сборки сокращает сроки разработки и обеспечивает своевременную инструментальную поддержку при создании сред сетевого моделирования.
3. Использование языка описания предметной области, основанного на гибкой параметризуемой графической нотации, позволяет строить семантические сети прикладных систем.
4. Разработанные методы и алгоритмы для обработки структурно-функциональной организации сетевых моделей прикладных систем повышают эффективность их анализа.

Апробация работы. Материалы работы были представлены на отчетных сессиях ИЦиГ СО РАН с 2007 по 2009 год. Основные результаты работы докладывались и обсуждались: на конференции «Современная психотерапия и психологическое консультирование в контексте культуры» (2008, Новосибирск); на VI и VII международных конференциях по биоинформатике регуляции и структуры генома (2008, 2010, Новосибирск); на VII Всероссийской научно-практической конференции «Молодежь и современные информационные технологии» (2009, Томск); на VI съезде Вавиловского общества генетиков и селекционеров (2009, Москва); на XX юбилейной выставке SoftTool (2009, Москва); на II международной конференции «Молекулярная филогенетика» (2010, Москва); на XIV международной конференции «Evolutionary Biology Meeting» (2010, Марсель, Франция); на конференции «Фундаментальные и прикладные аспекты современной биологии» (2010, Томск); на XIII Российской конференции с международным участием "Распределенные информационно-вычислительные ресурсы" (2010, Новосибирск).

Публикации. По теме диссертации опубликовано 16 работ, в их числе 3 работы [2, 3, 4] опубликованы в изданиях по списку ВАК РФ. Получены авторские свидетельства об официальной регистрации программ для ЭВМ [13–16].

Личный вклад автора. Постановка задач, основополагающие методы и решения, а также основные методические и инженерные результаты получены непосредственно автором либо с его участием и под его руководством.

В части практического применения разработок на модельных объектах автором лично построены описания предметных областей сетевых моделей молекулярно-генетических, экологических и социальных систем. Автором реконструированы сетевые модели жизненного цикла трематоды и СФО одного из содово-соленых озер Новосибирской области, а также проведена филоге-

нетическая декомпозиция генной сети клеточного цикла млекопитающих.

Объем и структура диссертации. Работа состоит из списка используемых обозначений, введения, пяти глав с выводами по каждой главе, заключения, списка литературы из 136 наименований и 11 приложений. Основная часть работы изложена на 122 машинописных страницах и содержит 36 рисунков.

Содержание работы

Во введении обоснована актуальность диссертационной работы, сформулирована цель и аргументирована научная новизна исследований, показана практическая значимость полученных результатов, представлены выносимые на защиту научные положения.

В первой главе проведен обзор существующих методов и программных средств в области сетевого моделирования. Рассмотрены прикладные направления работы.

В исследовании рассмотрено понятие семантической сети, что позволило расширить и раскрыть понятие сетевой модели. *Семантическая сеть* является информационной моделью предметной области, имеющей вид ориентированного графа, вершины которого соответствуют объектам предметной области, а дуги задают отношения между ними. Семантическую сеть можно представить в виде:

$$N = \langle E, O_1, \dots, O_n \rangle, \quad (1)$$

где

E - множество информационных единиц (элементы);

$O_{1,\dots,n}$ - множество типов связей между элементами (отношения).

Элементами семантической сети могут быть понятия, события, свойства либо отдельные процессы. Семантика определяется применительно к области исследований. На эффективность процесса сетевого моделирования может влиять отсутствие: (1) гибкой формализации описываемой предметной области; (2) методов получения первичных данных; (3) удобных средств инструментальной поддержки; (4) четкой постановки задачи реконструкции; (5) правил и ограничений при построении сетевой модели. Для решения прикладных задач в разных областях сетевого моделирования используются разные программные пакеты (Pajek, Cytoscape, LabVIEW, Astah и прочие). Разработка подобного рода ПС трудоемка и занимает от нескольких месяцев до нескольких лет. Это негативно влияет на своевременное проведение исследований.

Для описания и классификации знаний также используют онтологии. *Онтология* – это надежный семантический базис в определении содержания,

основанный на общей логической теории и состоящий из словаря и набора утверждений, реализованного на специализированном языке. В работе считается, что онтология – есть описание предметной области исследуемой системы, с учетом поставленных задач реконструкции и анализа. Для решения задачи описания онтологий используют специализированные средства (Protege, OntoEdit). Основная функция любого редактора онтологий состоит в поддержке процесса формализации знаний и представлении онтологии как точного и полного описания. Под *формализмом* в работе понимается теоретический базис, лежащий в основе способа представления онтологических знаний. Примерами формализмов могут служить логика предикатов, концептуальные графы и т. п. Формализм, используемый редактором, влияет на внутренние структуры данных и определяет язык описания онтологии. Такой язык задает вид хранения и способ передачи знаний. Такими языками являются: RDF, OWL и др. Сложность языка описания онтологии накладывает свои ограничения и вносит специфику в описание предметной области. При проведении работы, автор рассматривает онтологию как: (1) онтология, ориентируемая на задачу, используемая конкретной прикладной программой и содержащая термины, которые используются при разработке и (2) предметная онтология, описывающая реальные предметы, участвующие в какой-либо деятельности.

С учетом прикладного аспекта диссертационной работы проведен анализ общих задач моделирования на различных иерархических уровнях организации живого мира. Рассмотрены способы визуального представления и методы описания сетевых моделей в нескольких направлениях сетевого моделирования информационной биологии.

Основным модельным объектом в работе являются сетевые модели молекулярно-генетических систем – генные сети. *Генная сеть* - это группа координировано работающих и взаимодействующих между собой генов, контролирующих формирование фенотипических признаков организмов на основе закодированной информации. Методы реконструкции генных сетей (ГС) дают исследователю важную информацию об их структурных элементах и процессах. В динамике функционирования генных сетей реализуются генетические программы организмов.

Вторым модельным объектом являются экологические сети (экосети), описывающие экологические системы. *Экосистема* является биологической системой, состоящей из сообщества живых организмов, среды их обитания, а также системы связей, осуществляющей обмен веществом и энергией между ними. Основными задачами, решаемыми с применением аппарата экосетей является анализ функционирования экосистемы в различных режимах, а также исследование устойчивого функционирования самой экосистемы.

Третьим модельным объектом являются социальные сети организаций.

Социальная сеть - множество социальных объектов и определенное на нем множество отношений между ними. Исследования проводятся с компьютерными аналогами групп живых людей, представляющими «искусственное общество». При построении таких сетей используют теоретические представления социологов о структуре и организации человеческого общества. Отклонения в развитии такой структуры позволяют своевременно выявлять причины ее поведения.

Проведен сравнительный анализ модельных объектов для более точного формирования требований к проектируемым средствам инструментальной поддержки. Оформлено описание основных этапов сетевого моделирования:

1. Постановка цели и задач исследования.
2. Обзор источников информации.
3. Формализация предметной области.
4. Реконструкция сетевой модели.
5. Анализ сетевой модели.
6. Применение полученной информации.

Этапы 1 и 6 не требуют каких-либо специализированных программных средств и напрямую зависят от компетенции исследователей. На этапе 2 возможно использование специализированных баз данных, рассмотренных в тексте работы. Этапы 3,4,5 требуют своевременной инструментальной поддержки, являясь наиболее трудоемкими.

Рассмотрена система GeneNet (Колчанов Н.А., Подколотный Н.Л., Ананько Е.А., Мигинский Д.С., Рассказов Д.А.), послужившая прикладным прототипом при формировании базовых концепций предлагаемого подхода. Она создана для описания СФО генных и метаболических сетей прокариот и эукариот. Наполнение базы данных и описание сетей ведется на основе аннотации экспериментальных данных из научных публикаций.

Приведен обзор специализированных сред сетевого моделирования, обеспечивающих реконструкцию представленных модельных объектов. Наиболее известными среди сред моделирования генных и метаболических сетей являются: Cytoscape, VirtualCell, E-Cell, CellDesigner, Sigmoid и Pathway Studio. Для моделирования экологических систем обычно применяют пакет Pajek. При моделировании социальных систем организаций применяют Vizster, Enronic и InFlow. Рассмотрение ПС проведено в рамках описанного процесса сетевого моделирования.

Проведен обзор средств инструментальной поддержки для быстрой разработки предметно-ориентированных сетевых редакторов. Близкими к цели работы явились Domain Specific Modeling (DSM) системы, разработанные в качестве платформ для предметно-ориентированного моделирования. В состав типового DSM-пакета входят: (1) среда, реализующая общие сервисные функции; (2) графический редактор, предоставляющий рабочую область для

рисования диаграмм; (3) репозиторий, являющийся единым хранилищем информации о моделях; (4) средства проверки корректности моделей; (5) прочие компоненты, предназначенные для смежной деятельности. Выделено два типа DSM пакетов:

1. Полноценные среды разработки DSM-пакетов (Microsoft DSL Tools, Eclipse GMF), ориентированные на конкретный язык или платформу разработки (.NET, JVM и прочие.).
2. Конфигурируемые пакеты (Microsoft Visio, MetaEdit+), не показывающие своей прямой зависимости от технологической платформы.

В результате анализа проблем, с которыми сталкиваются разработчики предметно-ориентированных сред визуального сетевого моделирования, основанного на стандарте жизненного цикла ПО ISO/IEC 12207:2008 «Information Technology — Software Life Cycle Processes», идентифицированы основные задачи, требующие методического решения.

Результаты первой главы опубликованы в работах [2, 3, 4].

Во второй главе предложен подход, упрощающий работу IT-специалистов и ускоряющий процесс разработки в области создания сред предметно-ориентированного визуального сетевого моделирования.

Большинство сред визуального сетевого моделирования выполняют общие функции на стадии реконструкции (и анализа) сетевых моделей. Соответственно они используют похожие принципы работы пользовательских подсистем, но различаются в части описания модели данных. Возникла идея разработки общего программного каркаса с готовыми к использованию компонентами, на основе которого в короткие сроки могло бы создаваться средство визуальной реконструкции и анализа сетевых моделей. В отличие от существующих аналогов (GMF, MetaEdit+), каркас должен включать в себя не только средства визуального построения сетевых моделей, но и различные средства анализа с целью изучения их СФО и прогнозирования поведения изучаемых систем. В каркас должны быть включены средства взаимодействия с популярными программными пакетами моделирования (Pajek, Cytoscape). Необходимо предусмотреть механизмы гибкого расширения функциональных возможностей каркаса и сред, созданных на его основе. Различные компоненты должны быть созданы как законченные модули с готовой к использованию функциональностью и связаны с подсистемой хранения данных. Преимуществом такого решения является экономия времени при создании и отладке средства сетевого моделирования. На рисунке 1 приведена концептуальная схема организации предметно-ориентированного средства сетевого моделирования на основе предлагаемого подхода. Ядром приложения является готовый к использованию инструментальный каркас, который дополняется предметно-ориентированными модулями. При создании среды моделирования необходимо учитывать разную терминологическую и семантическую базы, которые

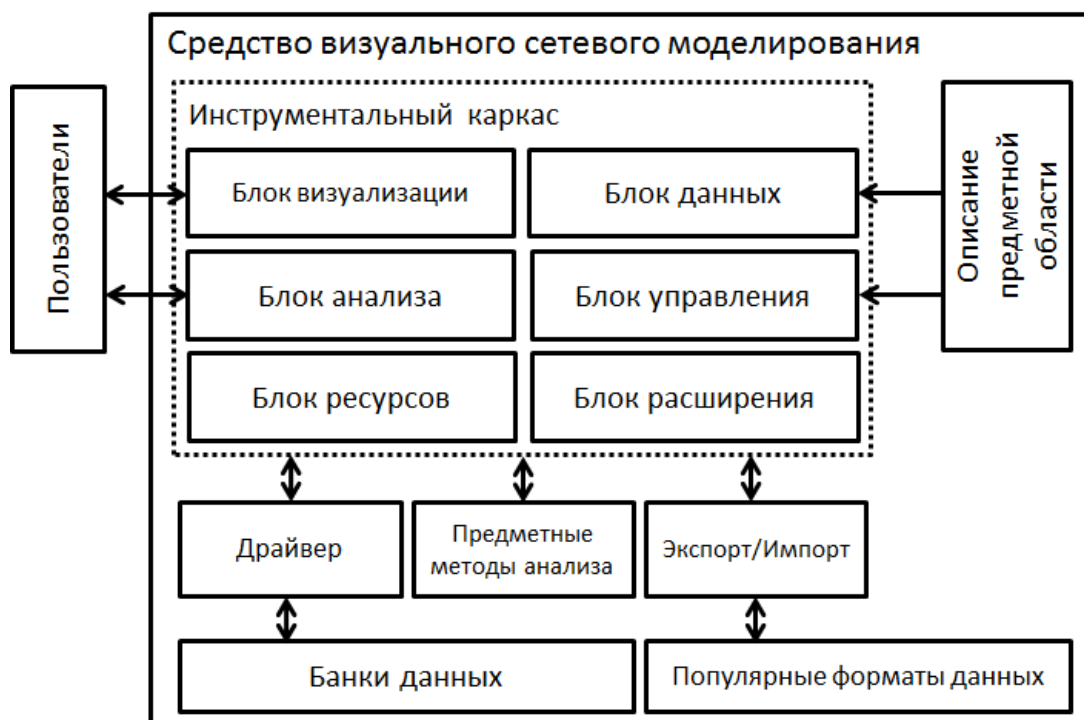


Рис. 1. Схема организации среды визуального сетевого моделирования

в процессе исследований могут изменяться и уточняться. В работу введено понятие мета-модели данных, которая является одним из ключевых элементов подхода. Мета-модель является абстрактным определением языка и процессов, формирующих модель данных конкретного средства визуального моделирования. Базовая модель данных (конкретного сетевого редактора) не закладывается жестко в структуру программы, а определяется внешним описанием. Такое описание названо “описанием предметной области” и должно создаваться либо изменяться непосредственно исследователями без участия IT-специалистов в удобной программной среде.

Внешнее описание предметной области строится на основе онтологического описания. Приведено расширение описания и свой формат, ориентированный на гибкую графическую нотацию при разработке каждого концепта, используемого далее для реконструкции сетевых моделей.

Приведена схема инструментального каркаса готового к работе средства визуального сетевого моделирования. На рисунке 2 описаны ключевые подсистемы каркаса, их назначение и особенности.

На схеме представлены основные подсистемы (блоки): (1) блок управления с базовыми программными интерфейсами, сервисами доступа к различным подсистемам каркаса, а также ключевую подсистему расширения функциональных возможностей; (2) блок визуализации, включающий базовые компоненты работы с сетевыми моделями и вспомогательные визуальные компоненты пользователя; (3) блок данных, включающий в себя модель

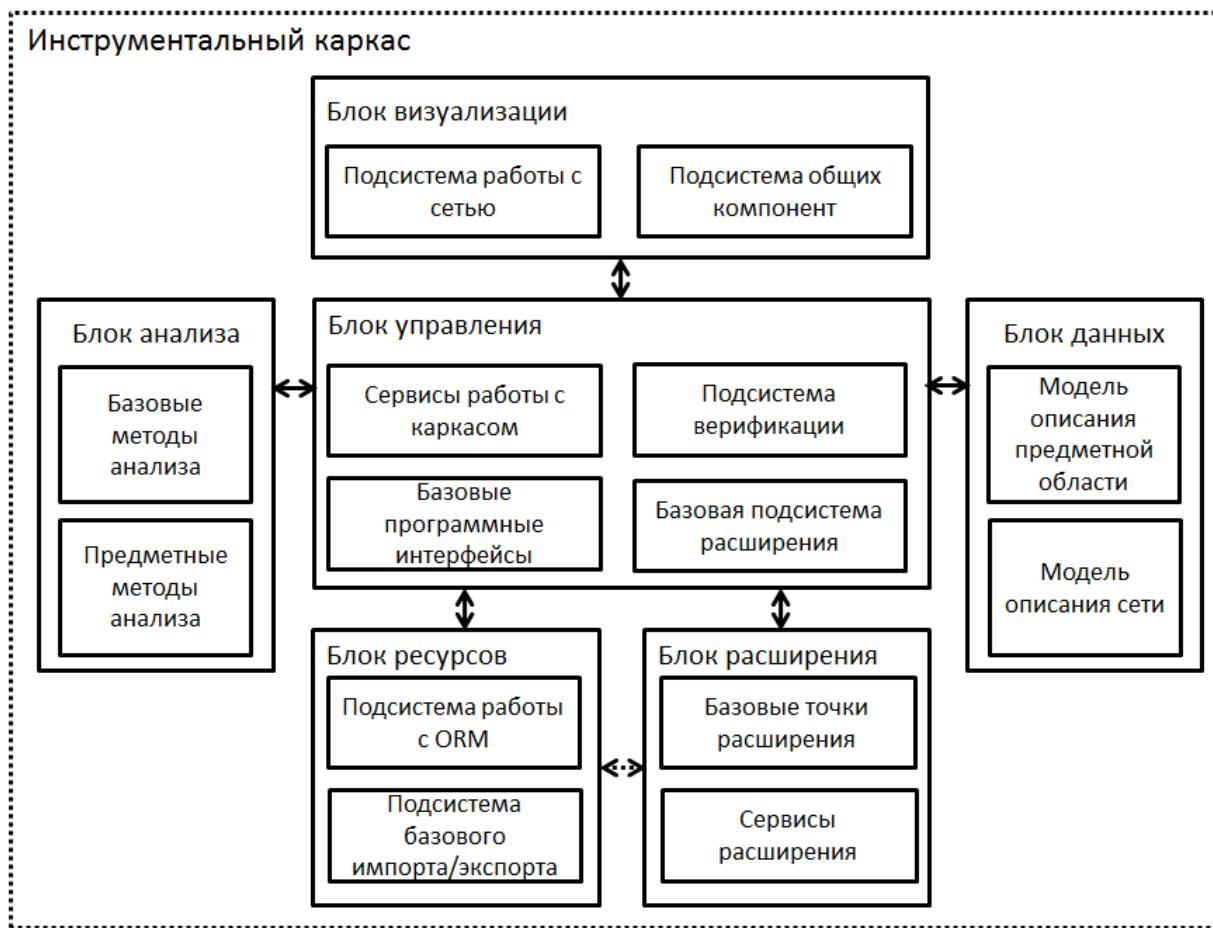


Рис. 2. Предложенный инструментальный каркас.

описания предметной области и модель хранения сетевой информации; (4) блок анализа, содержащий универсальные и проблемно-ориентированные методы анализа сетевых моделей; (5) блок ресурсов, включающий подсистему экспорта и импорта, а также взаимодействие с ORM системами и (6) блок расширения, включающий внешние сервисы расширения функциональных возможностей каркаса. При необходимости возможна разработка внешних модулей для хранения данных (СУБД, структурированные файлы). Описан ряд вспомогательных подсистем.

В терминах аналитической модели UML описана мета-модель данных, используемая для представления и хранения информации об описании предметной области и сетевых моделях. Диаграмма модели данных представлена на рисунке 3 и состоит из нескольких важных блоков: (1) блок описания предметной области предназначен для настройки средства визуальной сетевой реконструкции на определенную предметную область (данные подгружаются из внешнего описания); (2) блок хранения данных о сетевой модели предназначен для описания и хранения информации о сетевой модели конкретной системы, построенной в рамках определенной предметной области; (3) блок визуального представления и манипуляций непосредственно взаимодействует

с исследователем и предназначен для визуального представления сетевых моделей пользователю, согласно заранее заданной в блоке описания предметной области и графической нотации.

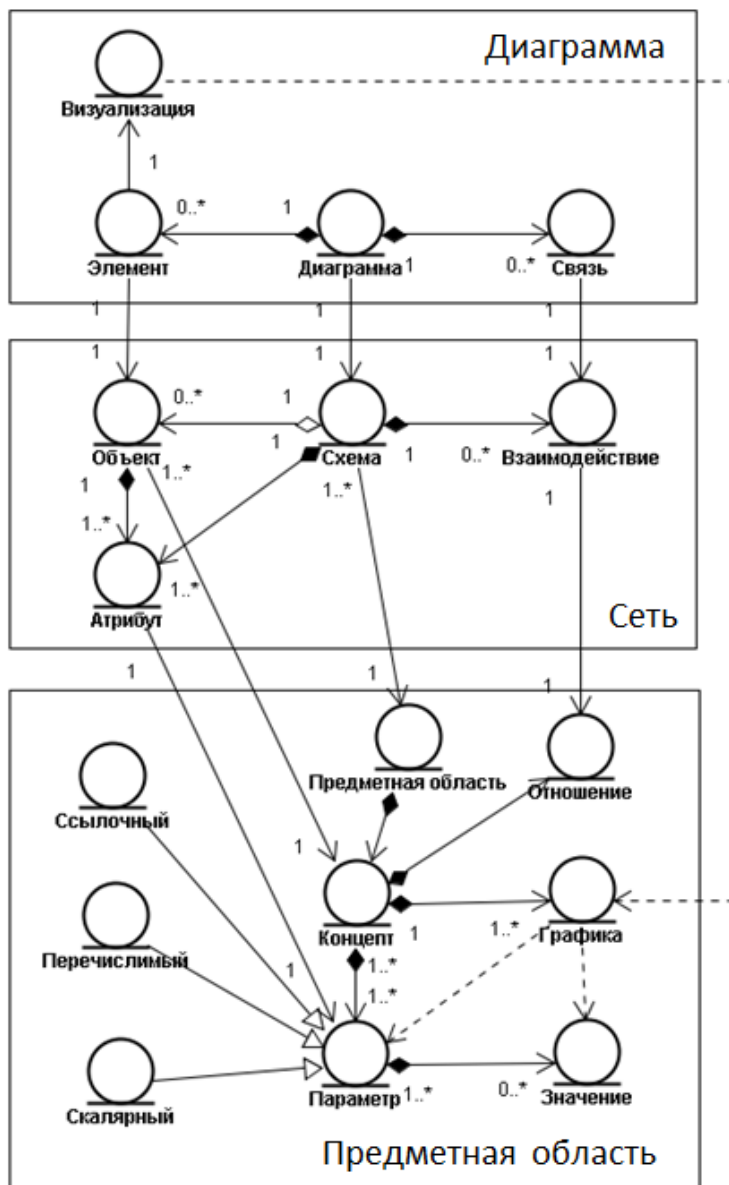


Рис. 3. Аналитическая диаграмма мета-модели данных.

В работе представлен процесс разработки предметно-ориентированного сетевого редактора на основе представленного подхода и программного каркаса. Описаны основные изменения на ключевых этапах разработки ПС и эффективность от внедрения предлагаемого подхода:

1. Анализ требований включает в себя постановку цели, задач и обзор предметной области, для которой разрабатывается ПС сетевого моделирования. После этого проводится описание предметной области в специализированном средстве.

2. Разработка архитектуры упрощается с учётом предоставления готового к использованию архитектурного решения (каркаса).
3. Проектирование практически можно исключить из деятельности IT-специалиста, что значительно может сократить время разработки.
4. Этап кодирования упрощается и заключается в генерации среды сетевого моделирования на основе предлагаемого каркаса и описания предметной области с подключением готовых к использованию компонентов пользователя. Возможна разработка и подключение предметно-ориентированных методов анализа.

Представлен ряд методов и алгоритмов, предназначенный для универсальной работы с различными сетевыми моделями. Одним из направлений является приведение сетевых моделей в двудольное представление. Только в таком виде динамика различных систем может быть проанализирована стандартными способами типа сетей Петри и т.п. Предложены два алгоритма, позволяющие преобразовать различные сетевые модели в двудольные орграфы. Первый является автоматическим и разбит на несколько шагов: (1) создание списка взаимодействия всех объектов множества 'А' друг с другом посредством ряда объектов множества 'В' (объект-объектное взаимодействие); (2) удаление всех объектов, входящих во множество 'В'; (3) в соответствии с ранее созданным списком для каждого объекта диаграммы, к которому вела дуга создать объект из множества 'В'; (4) в соответствии со списком происходит связка объектов из множества 'А' с помощью созданного на предыдущем этапе объекта; (5) поиск всех направленных циклов нечетной длины. Второй алгоритм является автоматизированным и подробно описан в диссертационной работе. После операций сетевая модель будет удовлетворять условию двудольности, если не будет обнаружено ни одного направленного цикла нечетной длины с ненарушенным принципом чередования объектов. В некоторых ситуациях при преобразовании необходимо использовать строго автоматизированный алгоритм с возможностью принятия исследователем ключевых решений.

Другим направлением является преобразование сетевых моделей в рамках смежных предметных областей, но с разным описанием предметной области. На основе опыта практического применения различных ПС сетевого моделирования (AND, Cytoscape, GeneNet) стало ясно, что тем или иным исследователям удобно работать в различных программных пакетах. За время развития сетевого моделирования в представленных областях различными группами исследователей по всему миру реконструировано большое количество сетевых моделей. Для экономии времени исследователей и IT-специалистов необходим механизм преобразования сетевых моделей между различными ПС. Следует учитывать описание предметной области таких пакетов и их прикладную направленность. Сформулированы ключевые положения,

необходимые для преобразования сетевых моделей: (1) разработка описания предметной области в рамках предлагаемого подхода и (2) разработка правил преобразования сетевых моделей (шаблона) описанных в декларативном виде. Предложен соответствующий формат, включающий в себя несколько блоков: (1) блок задания начальной и конечной онтологий; (2) блок описания правил преобразования типов объектов исходной сетевой модели в объекты конечной сетевой модели; (3) блок преобразования атрибутов. На вход подсистемы преобразования подаются данные об исходной сетевой модели. На основе множества правил преобразования, записанных с помощью определенного формата, происходит сопоставление онтологий двух сетевых моделей. Определяются соответствия между элементами и атрибутами онтологий, на основании которых осуществляется преобразование одной сетевой модели в другую. Пользователю предстает сетевая модель в рамках итоговой терминологической и семантической базы.

Заключительным направлением является поиск регуляторных контуров. Задача популярна и можно найти ряд специализированных ПС. Основываясь на одном из требований об удобной интеграции существующих разработок в разрабатываемый каркас, принято решение использовать разработанный в ИЦиГ СО РАН пакет GraNET (Гунбин К.В.). В рамках предложенного подхода было необходимо разработать модуль с графическим пользовательским интерфейсом, адаптируемый под различные прикладные области.

Результаты второй главы опубликованы в работах [2, 3, 4, 10 – 12].

В третьей главе представлена инженерная реализация предложенного подхода и описание основных функциональных возможностей разработок. Для программной реализации выбрана платформа JAVA и технология Eclipse Rich Client Platform с сопутствующими программными расширениями.

Описана технология и библиотека «Araneus», созданная на основе предложенного подхода и алгоритмов, предназначенная для быстрой разработки предметно-ориентированных ПС визуальной сетевой реконструкции с функциями анализа созданных моделей. Разработчикам предложен ряд программных сервисов, позволяющий взаимодействовать с ключевыми подсистемами каркаса приложения.

Созданы базовые пользовательские компоненты, составляющие основной каркас средства сетевого моделирования, позволяющие проводить: (1) загрузку сетевых моделей; (2) оперирование элементами сетевой модели; (3) просмотр уменьшенного отображения сетевой модели; (4) иерархическое отображение элементов сетевой модели; (5) просмотр и редактирование значений атрибутов.

Спроектирован и разработан предметно-ориентируемый язык описания предметной области. Его основной особенностью, в отличие от существующих языков (RDF, OWL) и с учетом специфики сетевого моделирования, являет-

ся наличие гибкой параметризуемой графической нотации, применяемой для описания различных состояний элемента сетевой модели. Описан подход создания элементов описания предметной области, используемых в дальнейшем при реконструкции и анализе сетевых моделей.

Разработан специализированный программный пакет «Ontology Manager», предназначенный для определения терминологии и семантики предметной области непосредственно исследователями без участия IT-специалиста. На рисунке 4 представлен визуальный интерфейс редактора. «Ontology Manager»

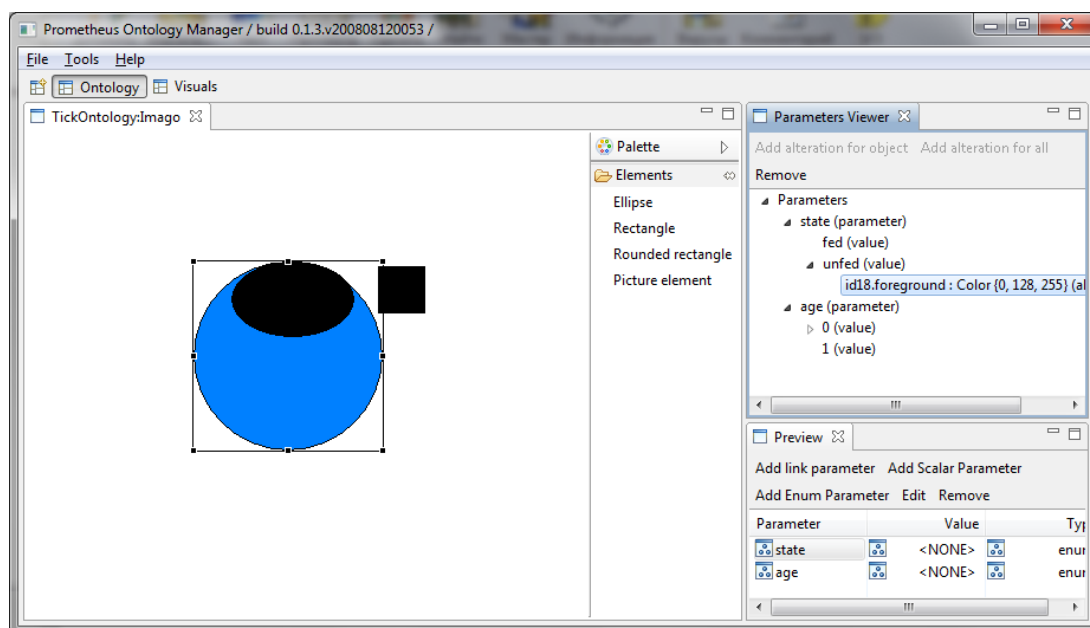


Рис. 4. Интерфейс программы «Ontology Manager».

оперирует созданным языком и обладает рядом ключевых возможностей: (1) описание графического отображения элемента будущей сети; (2) создание различных атрибутов элемента (скалярные, перечислимые, ссылочные); (3) задание отношений связывания между элементами; (4) наложение специальных ограничений; (5) параметризация изображения элемента с учетом его текущих значений атрибутов.

В работе описан подробный процесс создания программной среды визуального моделирования на основе «Araneus». При создании приложения применяется разработанный автором специализированный модуль для среды Eclipse, упрощающий разработку. Приведено описание процесса расширения функциональных возможностей средства визуального сетевого моделирования.

Разработана настраиваемая подсистема приведения сетевых моделей в двудольный вид. Создана подсистема «Regulatory Circuits», решающая задачи поиска регуляторных контуров в различных реконструированных сетевых моделях. Для решения задачи преобразования сетевых моделей с различным

описанием предметных областей разработана подсистема «NetConverter», а также расширен программный пакет «Ontology Manager» в рамках создания шаблонов преобразования. разработан ряд модулей для экспорта и импорта сетевых моделей других ПС (Pajek, Cytoscape, SBW) .

Разработанные средства инструментальной поддержки могут сократить время создания предметно-ориентированной среды сетевого моделирования с нескольких месяцев до нескольких дней.

Результаты третьей главы опубликованы в работах [3, 10 – 12].

В четвертой главе проведен анализ и описание прикладных систем, в рамках которых ведется апробация разработанного подхода и программных средств инструментальной поддержки. Для построения всех описаний предметных областей использован разработанный автором программный пакет «Ontology Manager». Для реализации программных сред планировалось использование разработанных средств инструментальной поддержки.

Для решения задач реконструкции и анализа генных сетей проведен анализ системы GeneNet и построено проблемно-ориентированное описание предметной области генных сетей. Описание включает в себя элементы, из которых в дальнейшем создаются сетевые модели: ген, белок, мРНК, реакция/регуляция и др. Элементы обладают: (1) набором уникальных графических отображений, (2) атрибутами и (3) параметризацией графического отображения в зависимости от значений ключевых атрибутов. Определены взаимодействия между элементами и верификационная политика.

Изучение СФО генных сетей является одним из приоритетных направлений информационной биологии. Одним из методов анализа является поиск регуляторных контуров. С помощью замкнутых регуляторных контуров возможно поддержание определенного состояния сети или ее переход в другой режим функционирования. Для решения задачи целесообразно применить подсистему «Regulatory Circuits», включенную в «Araneus» с ее настройкой на предметную область.

Эффективность анализа генных сетей значительно зависит от их размеров (достигают десятков тысяч элементов и взаимодействий). В работе описан метод филогенетической декомпозиции, позволяющий редуцировать сеть на основе эволюционных данных. Суть метода заключается в последовательном объединении пар белков генной сети, связанных общим происхождением. Редуцированная сеть может включать основные первоначальные регуляторные контуры, при меньшем размере сети, которые в процессе эволюции претерпели изменения.

При работе с реконструкцией экологических систем исследователи часто сталкиваются с отсутствием универсального компьютерного описания предметной области. Без четко формализованной базы изучение экосистем затруднено либо является достаточно трудоемким. Для решения этой проблемы про-

веден анализ концептуальных онтологий описания макро- и микродинамики экологических систем, разработанных в ИЦиГ СО РАН (Суслов В.В., Сергеев М.Г.). На основе анализа построено компьютерное описание предметной области и разработана уникальная графическая нотация и набор рангов (атрибутов). Анализ сетевых моделей экологических систем преимущественно основан на адаптируемой подсистеме «Regulatory Circuits».

Для решения задач описания и анализа социальных систем коллективов организаций разработана компьютерная методика «СоциоТом». Она предназначена для инструментальной поддержки в рамках конфликтологии. За основу методики взят метод психотерапии «Дианализ» (Завьялов В.Ю.). В работе описывается процесс проведения исследования организации, разбитый на несколько стадий. Для получения первичных данных предложены концепции сбора информации на основе семантического дифференциала Осгуда и проксемики. Разработаны методы анализа сетевых моделей: (1) гистограммы отношения испытуемого к объектам реального мира; (2) индивидуальный кластерный анализ результатов тестирования; (3) многомерное распределение сотрудников коллектива на основе ключевых объектов кластерного анализа.

Результаты четвертой главы опубликованы в работах [1, 3 – 9].

В пятой главе описаны программные пакеты, разработанные на основе предложенного подхода и библиотеки «Araeus», а также созданных описаний предметных областей.

Программный пакет «GeneNetStudio» создан для реконструкции и анализа генных сетей. Его визуальный интерфейс представлен на рисунке 5.

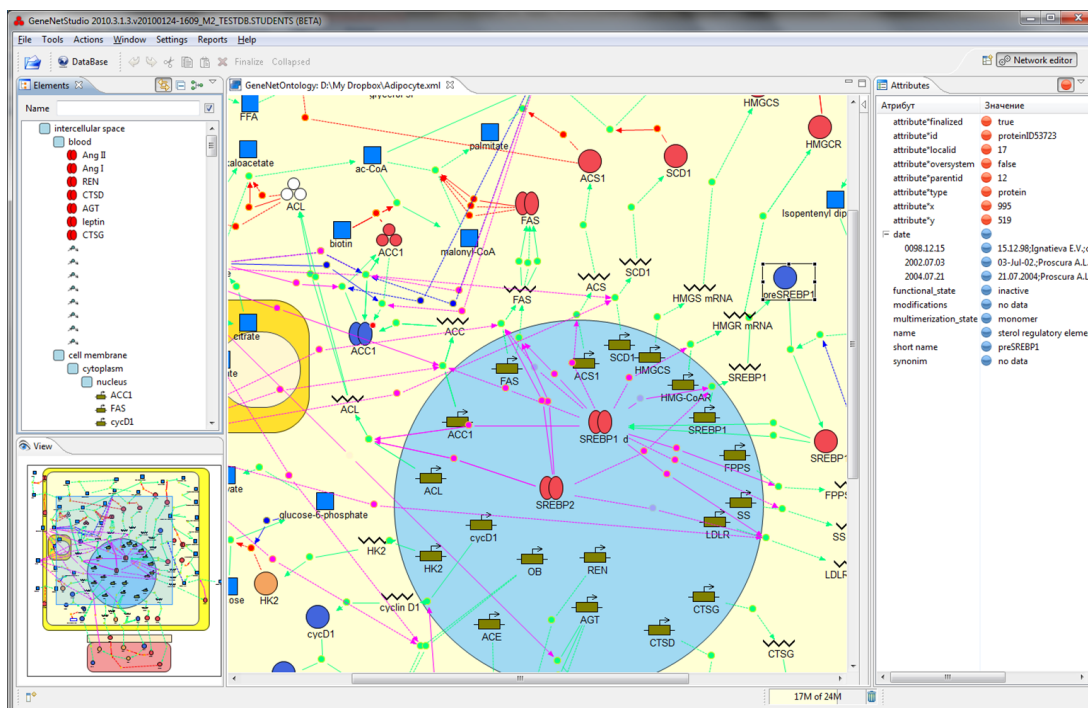


Рис. 5. Визуальный интерфейс «GeneNetStudio».

В программный пакет включены и адаптированы подсистемы: (1) поиска регуляторных контуров; (2) филогенетической декомпозиции; (3) преобразования ассоциативных семантических сетей молекулярно-генетических взаимодействий из системы AND (Деменков П.С., Иванисенко В.А.) в генные сети; (4) приведения сетевых моделей в двудольное представление. Проведена успешная апробация «GeneNetStudio» на ряде прикладных задач, в их числе реконструкция генной сети гомеостаза холестерина и изучение эволюции генной сети клеточного цикла млекопитающих.

Программный пакет «EcoNetStudio» предназначен для реконструкции и анализа сетевых моделей экологических систем. Визуальный интерфейс пакета представлен на рисунке 6. Проведена успешная апробация на задачах

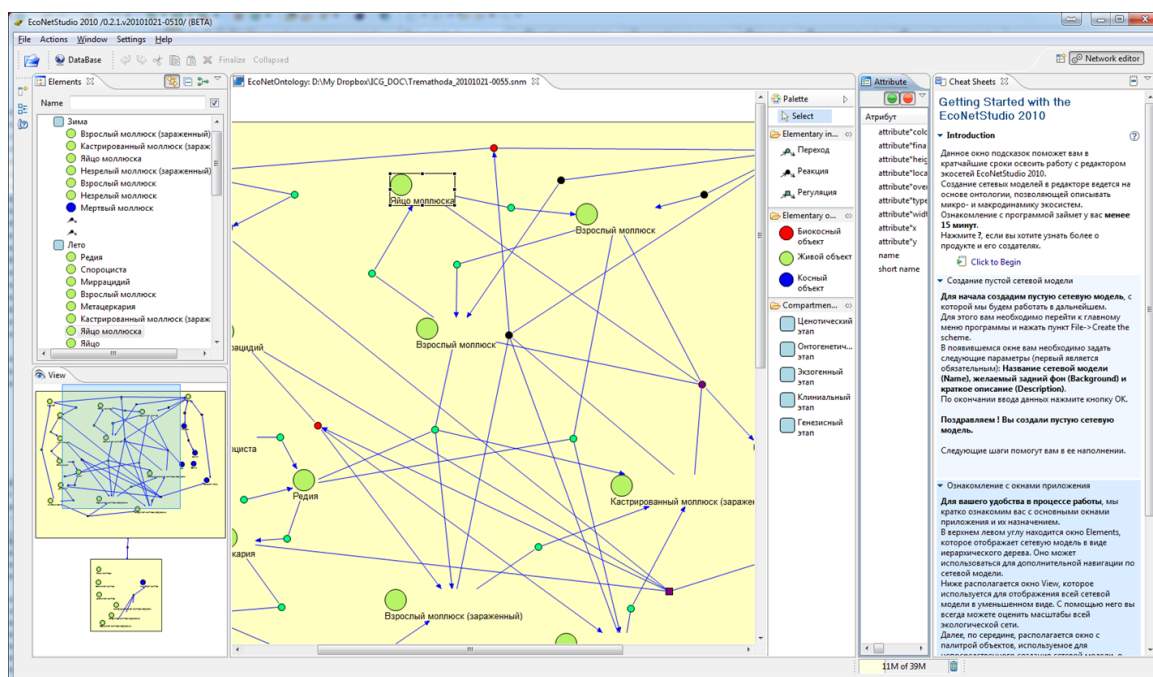


Рис. 6. Визуальный интерфейс «EcoNetStudio».

реконструкции сетевой модели жизненного цикла трематоды и описании сетевой модели одного из содово-соленых озер Новосибирской области.

Программный пакет «EVAConstructor» разработан в качестве прототипа для методики «Эволюционный конструктор» (Лашин С.А., Матушкин Ю.Г.), созданной в ИЦиГ СО РАН. Он позволяет строить сетевые модели жизненных циклов сообществ бактерий с определением изначальных численностей популяций, величины притока и оттока определенных субстратов и пр.

Создан ряд ПС, составляющий компьютерную методику «СоциоТом» и предназначенный для исследований в области конфликтологии. Разработан ключевой пакет методики - «SocioTomStudio». Он предназначен для описания сетевой модели структуры организации, проведения анализа тестов сотрудников организации, а также идентификации конфликтных ситуаций.

Разработаны подсистемы, включенные в «SocioTomStudio»: «P-Builder» и «P-Analyser», предназначенные для создания тестов и анализа результатов тестирования. Для индивидуального тестирования создан программный пакет «P-Test». Программный пакет «SocioTomStudio» с подсистемами успешно апробирован при анализе коллективов 7 организаций в РФ и за ее рубежами. В диссертационной работе частично приведено описание одного из исследований.

Результаты пятой главы опубликованы в работах [1, 3–9, 12].

В заключении изложены **выводы**, отражающие результаты диссертационного исследования:

1. Анализ направления сетевого моделирования и средств инструментальной поддержки привел к идентификации и актуализации проблем, требующих методического решения.
2. Предложен подход, методы и алгоритмы для быстрой разработки предметно-ориентированных сетевых редакторов, на основе готового к использованию инструментального каркаса и гибкого описания предметной области без участия IT-специалистов.
3. Созданы методы и алгоритмы для универсального компьютерного описания, визуального представления и анализа сетевых моделей на основе параметризуемой графической нотации.
4. Спроектированы и реализованы средства инструментальной поддержки для сокращения сроков разработки предметно-ориентированных сред визуальной реконструкции и анализа сетевых моделей.
5. Анализ состава и структуры прикладных исследуемых систем привел к созданию описаний предметных областей и проблемно-ориентированных методов обработки данных.
6. Подход и средства инструментальной поддержки успешно применены для разработки проблемно-ориентированных сред визуального моделирования и апробированы в области изучения молекулярно-генетических, экологических и социальных систем.

Предлагаемый подход имеет естественные ограничения, связанные со сложностью анализируемых систем, а также с объемом, неточностью и возможной противоречивостью исходных данных.

В приложениях приведены материалы, относящиеся к применению разработанных средств инструментальной поддержки, прикладные результаты работы созданных сред визуального сетевого моделирования, а также акты о внедрении разработок в научно-исследовательские и учебные процессы.

Список работ автора

1. **Тимонов В.С.**, Турнаев И.И., Генаев М.А., Гунбин К.В. Компьютерная система филогенетического анализа генных сетей: изучение эволюции генной сети клеточного цикла животных // Труды ТГУ. 2010.— Т. 275. С. 412–414.
2. **Timonov V.S.**, Miginsky D.S. Prometheus: Technology for Rapid Development of Biological Databases // Intelligent Data Analysis.— 2010.— Vol. 14, no. 2.— Pp. 159–171.
3. 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.— 2008.— Vol. 12, no. 5.— Pp. 463–479.
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. 173–180.
5. **Тимонов В.С.**, Турнаев И.И., Генаев М.А., Гунбин К.В. Компьютерная система филогенетического анализа генных сетей: изучение эволюции генной сети клеточного цикла животных // Тезисы всероссийской молодёжной научной конференции «Фундаментальные и прикладные аспекты современной биологии». Томск, 6–9 октября 2010 г.— 2010.— С. 145.
6. **Timonov V.S.**, Gunbin K.V., Turnaev I.I., Genaev M.A., Kolchanov N.A. Method and Software Tools for Gene Networks Evolution Research // Proceedings of the 14th Evolutionary Biology Meeting at Marseilles. Marseille, 21-24 September.— 2010.— P. 74.
7. **Timonov V.S.**, Gunbin K.V., Turnaev I.I., Genaev M.A., Miginsky D.S. Regulatory Circuits and Phylogenetic Decomposition in Gene Networks Evolution Research // Proceedings of the 7th International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2010). Novosibirsk, 20-27 June.— 2010.— P. 286.
8. Suslov V.V., Gunbin K.V., **Timonov V.S.** Phylogenetic decomposition of the gene networks regulating the level of thyroid hormones // Proceedings of the 7th International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2010). Novosibirsk, 20-27 June.— 2010.— P. 282.
9. **Timonov V.S.**, Gunbin K.V., Turnaev I.I., Miginsky D.S. Tool and methods for analysis of the gene networks evolution // Proceedings of the 2nd Moscow International Conference «Molecular Phylogenetics MolPhy-2». Moscow,

- 18–21 May.— 2010.— P. 74.
10. **Тимонов В.С.**, Мигинский Д.С., Сергеев М.Г. ARANEUS: Технология визуальной реконструкции и анализа сетевых моделей биологических систем // Тезисы VI съезда вавиловского общества генетиков и селекционеров. Москва, 21-28 июня 2009 г. Т. 2. С. 34.
 11. **Timonov V.S.**, Miginsky D.S. «PROMETHEUS» Toolkit for Agile Development of Biological Data Storing and Access Software // Proceedings of the 6th International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2008). Novosibirsk, June 22-28.—2008. P. 237.
 12. **Тимонов В.С.**, Мигинский Д.С., Колчанов Н.А. Технология реконструкции и анализа сетевых моделей экологических систем // Молодежь и современные информационные технологии. Сборник трудов VII Всероссийской научно практической конференции студентов, аспирантов и молодых ученых. Томск, 25-27 февраля 2009 г. Т. 1. Томск: Изд-во СПб Графикс. С. 335–336.
 13. **Тимонов В.С.** Библиотека для разработки сетевых редакторов (ARANEUS) / Мигинский Д.С., Тимонов В.С., Сурнина Н.Ю. // Свидетельство об официальной регистрации программы для ЭВМ №2009611761 / Федеральная служба по интеллектуальной собственности, патентам и товарным знакам. – 03.04.2009.
 14. **Тимонов В.С.** Библиотека быстрой разработки баз данных (ПРОМЕТЕЙ) / Тимонов В.С., Мигинский Д.С., Сурнина Н.Ю. // Свидетельство об официальной регистрации программы для ЭВМ 2008615514 / Федеральная служба по интеллектуальной собственности, патентам и товарным знакам. – 27.11.2008.
 15. **Тимонов В.С.** Программный пакет анализа портрета личности P-analyser / Мигинский Д.С., Завьялов В.Ю., Ефимов В.М., Тимонов В.С., Сурнина Н.Ю. // Свидетельство об официальной регистрации программы для ЭВМ 2008611605 / Федеральная служба по интеллектуальной собственности, патентам и товарным знакам. – 27.03.2008.
 16. **Тимонов В.С.** Программный пакет для описания личности P-test / Мигинский Д.С., Завьялов В.Ю., Тимонов В.С., Сурнина Н.Ю. // Свидетельство об официальной регистрации программы для ЭВМ 2008611336 / Федеральная служба по интеллектуальной собственности, патентам и товарным знакам. – 17.03.2008.

Тимонов Владимир Сергеевич

**Программные технологии визуальной реконструкции и анализа
сетевых моделей генетических, экологических и социальных
систем**

Автореферат диссертации на соискание
ученой степени кандидата технических наук