

# Информатика и её применения

Том 10 Выпуск 1 Год 2016

## СОДЕРЖАНИЕ

Проблемы доступа к данным в исследованиях с интенсивным использованием данных в России <i>Л. А. Калиниченко, А. А. Вольнова, Е. П. Гордов, Н. Н. Киселева, Д. А. Ковалева, О. Ю. Малков, И. Г. Окладников, Н. Л. Подколотный, А. С. Позаненко, Н. В. Пономарева, С. А. Ступников, А. З. Фазлиев</i>	2
Анализ системного риска совместного кредитования над неоднородными коллекциями данных <i>С. А. Ступников, Д. О. Брюхов, Н. А. Скворцов</i>	23
Ортогональные субоптимальные фильтры для нелинейных стохастических систем на многообразиях <i>И. Н. Сеницын</i>	34
Аналитическое моделирование распределений в нелинейных стохастических системах на многообразиях методом эллипсоидальной аппроксимации <i>И. Н. Сеницын, В. И. Сеницын</i>	45
Применение метапрограммирования для повышения технологичности больших автоматизированных систем <i>С. П. Ковалёв</i>	56
Байесовские модели массового обслуживания и надежности: априорные распределения с компактным носителем <i>А. А. Кудрявцев</i>	67
Концепция онлайн-комплекса для стохастического моделирования реальных процессов <i>А. К. Горшенин</i>	72
Разработка алгоритма численного решения задачи оптимального управления инвестициями в закрытой динамической модели трехсекторной экономики <i>П. В. Шнурков, В. В. Засыпко, В. В. Белоусов, А. К. Горшенин</i>	82
Мелкозернистые гибридные интеллектуальные системы. Часть 2: Двухнаправленная гибридизация <i>И. А. Кириков, А. В. Колесников, С. В. Листопад, С. Б. Румовская</i>	96
Представление кроссязыковых знаний о коннекторах в надкорпусных базах данных <i>И. М. Зацман, О. Ю. Инькова, М. Г. Кружков, Н. А. Попкова</i>	106
Применение контекстно-свободных грамматик для извлечения онтологии из текстов коротких описаний статей биологической тематики <i>Д. А. Алексеевский</i>	119
Complexity and its information content <i>N. Callaos and R. Seyful-Mulyukov</i>	129
Об авторах	140
Правила подготовки рукописей	142
Requirements for manuscripts	145

## ПРОБЛЕМЫ ДОСТУПА К ДАННЫМ В ИССЛЕДОВАНИЯХ С ИНТЕНСИВНЫМ ИСПОЛЬЗОВАНИЕМ ДАННЫХ В РОССИИ\*

Л. А. Калиниченко<sup>1</sup>, А. А. Вольнова<sup>2</sup>, Е. П. Гордов<sup>3</sup>, Н. Н. Киселева<sup>4</sup>, Д. А. Ковалева<sup>5</sup>,  
О. Ю. Малков<sup>6</sup>, И. Г. Окладников<sup>7</sup>, Н. Л. Подколотный<sup>8</sup>, А. С. Позаненко<sup>9</sup>,  
Н. В. Пономарева<sup>10</sup>, С. А. Ступников<sup>11</sup>, А. З. Фазлиев<sup>12</sup>

**Аннотация:** Целью данного обзора является анализ глобальных тенденций создания массивных коллекций данных в мире и обеспечения возможности совместного использования таких коллекций при решении задач исследования и принятия решений в различных областях с интенсивным использованием данных (ОИИД) в России. Конкретный набор ОИИД, отобранный для обзора, включает астрономию, материаловедение, науки о Земле, геномику и протеомику, нейронауку. По каждой из рассмотренных ОИИД представлены крупные стратегические инициативы США и ЕС, примеры крупных коллекций данных в мире до 2025 г., известные проекты информационных и телекоммуникационных инфраструктур и центров данных. Включенный в обзор набор массивных коллекций данных, планируемых к получению в мире, предлагается использовать в качестве ориентира при планировании и развитии исследовательских инфраструктур для накопления и анализа данных, совместимых с зарубежными открытыми инфраструктурами в науке. В частности, рассматриваемые в обзоре коллекции данных, цели их создания и научные исследования, планируемые к осуществлению с их помощью, позволяют перейти к постановке и решению задач создания компонентов перспективных информационных и телекоммуникационных инфраструктур, таких как, например, средства концептуализации ОИИД, необходимые метамодели, средства обеспечения возможности повторного использования коллекций данных, воспроизводимости программ и потоков работ и др.

**Ключевые слова:** 4-я парадигма; области с интенсивным использованием данных; исследовательские инфраструктуры; коллекции данных; большие данные

**DOI:** 10.14357/19922264160101

### 1 Введение

Исследования и принятие решений в различных областях деятельности людей реализуются на основе анализа данных, накопленных в соответ-

ствующих областях, объем и разнообразие которых в наши дни растут экспоненциально.

В соответствии с 4-й парадигмой научных исследований [1], проведение исследований, движи-

\* Подготовка настоящего обзора была частично поддержана различными грантами, полученными группами из вовлеченных в эту работу исследовательских организаций: для ИПИ ФИЦ ИУ РАН грантами 13-07-00579, 14-07-00548 и 16-07-01028; для ИМКЭС СО РАН грантами РФФИ 13-05-12034 и 14-05-00502; для ИМЕТ РАН грантами РФФИ 14-07-00819 и 15-07-00980; для ИНАСАН РАН грантом РФФИ 15-02-04053 и грантом Президиума РАН по программе П-41; для ИЦИГ СО РАН грантом РФФИ 14-24-00123; для ИКИ РАН грантом РФФИ 15-02-10203-К; для ИЦН грантами РФФИ 15-04-08744 и 15-04-05066; для ИОА СО РАН грантом РФФИ 13-07-00411.

<sup>1</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, leonidandk@gmail.com

<sup>2</sup> Институт космических исследований Российской академии наук, alinusss@gmail.com

<sup>3</sup> Международный исследовательский центр климатологических исследований Института мониторинга климатических и экологических систем Сибирского отделения Российской академии наук, gordov@scert.ru

<sup>4</sup> Институт металлургии и материаловедения им. А. А. Байкова Российской академии наук, kis@imet.ac.ru

<sup>5</sup> Институт астрономии Российской академии наук, dana@inasan.ru

<sup>6</sup> Институт астрономии Российской академии наук, malkov@inasan.ru

<sup>7</sup> Международный исследовательский центр климатологических исследований Института мониторинга климатических и экологических систем Сибирского отделения Российской академии наук, igor.okladnikov@gmail.com

<sup>8</sup> Центр коллективного пользования «Биоинформатика» Федерального исследовательского центра Института цитологии и генетики Сибирского отделения Российской академии наук, pnl@bionet.nsc.ru

<sup>9</sup> Институт космических исследований Российской академии наук, arozanen@iki.rssi.ru

<sup>10</sup> Научный центр неврологии, ropomare@yandex.ru

<sup>11</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, sstupnikov@ipiran.ru

<sup>12</sup> Центр интегрированных информационных систем Института оптики атмосферы Сибирского отделения Российской академии наук, faz@iao.ru

мых данными, становится неотъемлемой частью различных областей науки, экономики, бизнеса (далее — областей с интенсивным использованием данных — ОИИД, или *data intensive domains* — DID). Без обеспечения все новыми данными, являющимися результатом наблюдений, измерений в природе и обществе, развитие исследований в различных ОИИД становится немыслимым.

По мере развития ОИИД извлечение данных из природы становится все более сложным и изолированным из-за необходимости проникновения во все более масштабные микро-, мезо- и макроявления. Организуются глобальные проекты и миссии (в том числе космические) по сбору и накоплению данных при помощи специализированных новейших высокотехнологичных инструментов, размещаемых не только на Земле, но и в космосе. Получение данных становится все более сложным, дорогостоящим делом, требующим развития специальных технологий и серьезных капиталовложений. В результате удается получать сырые данные, подлежащие дальнейшей обработке и анализу. Сам процесс сбора данных при изучении определенного вида явлений в конкретной ОИИД может занимать многие годы.

Наряду со сбором данных беспрецедентно быстро развиваются методы и средства накопления, обработки, анализа и управления накапливаемыми данными в разнообразных ОИИД, происходит быстрое расширение спектра задач, требующих решения на основе полученных данных, накопление опыта решения подобных задач и обеспечение возможности его междисциплинарного использования.

Главный побудительный мотив настоящей инициативной работы<sup>1</sup> заключается в необходимости положить начало систематическому анализу развития массивных коллекций данных в различных ОИИД в мире, создания и развития инфраструктур для накопления и использования больших коллекций данных, систематизации опыта решения задач в ОИИД и пр. Некоторыми прагматическими целями такого анализа являются выявление технических, правовых и финансовых проблем на пути обеспечения доступа ученых России в различных ОИИД к уже накопленным и ожидаемым коллекциям данных в мире<sup>2</sup>, определение потребности создания специальных инфраструктур технических и программных средств в России для поддержки та-

кой возможности, а также способности России заметным образом участвовать во вкладе в мировую сокровищницу данных, в создание соответствующих инфраструктур, методов и средств решения задач анализа данных.

Предварительный анализ показывает, что западный мир весьма озабочен проблемами, порожденными все возрастающим «наводнением» ОИИД большими данными, проблемами их анализа (включая анализ публикаций в виде текстов на естественном языке как части когнитивного процесса), накопления и совместного (в том числе междисциплинарного) использования данных и опыта решения задач, планированием специальных инфраструктур, позволяющих справиться с подобным наводнением по мере ввода в действие новых инструментов получения данных. Для этого организуются крупные совместные проекты, рабочие группы, обсуждаются возможные решения, планируются новые инфраструктуры и уже тестируются их фрагменты, ориентированные на получение конкретных коллекций данных после 2020 г., создаются методы и средства поддержки таких инфраструктур, отрабатываются характерные примеры будущих задач, проводятся конференции, специализированные симпозиумы рабочих групп и пр.

Вместе с тем в ряде ОИИД в России обстановка такова, что если своевременно не позаботиться о том, чтобы иметь эффективный доступ к данным (наиболее важные из которых, увы, собираются и накапливаются за пределами России), то исследования по ряду направлений во многих ОИИД можно просто прекратить.

Все это служит мотивацией для проведения анализа глобальных тенденций создания массивных коллекций данных в мире и обеспечения возможности совместного использования таких коллекций при решении задач исследования и принятия решений в различных ОИИД в России. Анализ проведен на выборке коллекций, отражающей современные тенденции научных исследований.

Отбор глобальных коллекций данных и примеров их использования в данной работе ограничен временными рамками (учитываются крупные проекты накопления и использования данных в различных ОИИД, выполняемые до 2025 г.), а также конкретным набором предварительно определен-

<sup>1</sup> Данная статья является расширенным русскоязычным вариантом работы *Kalinichenko L., Fazliev A., Gordov E., Kiselyova N., Kovaleva D., Malkov O., Okladnikov I., Podkolodny N., Ponomareva N., Pozanenko A., Stupnikov S., Volnova A.* New data access challenges for data intensive research in Russia // 17th Conference (International) on Data Analytics and Management in Data Intensive Domains Proceedings, 2015. P. 215–237.

<sup>2</sup> Следует заметить, что на этом пути из-за высокой технологической сложности и стоимости средств извлечения данных во многих ОИИД, высокой стоимости самого процесса получения конкретных коллекций данных в течение ближайших 10 лет ни о каком «импортозамещении» не может быть и речи.

ных ОИИД, включающим астрономию, материаловедение, науки о Земле, геномику и протеомику, нейронауку. Этот список дополнен информатикой, в силу того что все инфраструктурные проекты в значительной мере посвящены проблемам информационно-коммуникационных технологий (ИКТ).

По каждой из рассмотренных ОИИД авторы стремились представить в нижеследующих разделах работы следующую информацию:

- крупные стратегические инициативы США и ЕС по направлению;
- примеры крупных коллекций данных в мире до 2025 г. по направлению;
- известные проекты ИКТ-инфраструктур и центров данных;
- сравнимые проекты в России, при наличии таковых.

## 2 Астрономические данные

### 2.1 Большой обзорный телескоп

Обзорный широкоугольный (поле зрения — около 10 кв. град.) 8-метровый телескоп (Large Synoptic Survey Telescope, LSST) строится в Чили на высоте 2700 м и начнет функционировать в начале 2020-х гг. Он предназначен для регистрации объектов всей доступной полусферы неба и прежде всего для обнаружения темной материи и темной энергии, поиска околоземных астероидов, изучения природы транзиентных явлений и картографирования Млечного Пути.

Поток данных, ожидаемый от проекта LSST начиная с 2020 г., будет поступать с беспрецедентной скоростью. Телескоп будет собирать данные о более чем 40 млрд объектов, а также проводить исследования переменных источников и транзиентных событий. Предполагается, что доступная для наблюдений полусфера будет полностью покрываться наблюдениями LSST в 6 фотометрических фильтрах (ugrizu) не реже, чем раз в неделю.

Объем наблюдений за ночь достигнет 15 ТБ (терабайт), что за 10 лет приведет к суммарному объему около 60 ПБ (петабайт) сжатых сырых данных, 15 ПБ баз данных, 0,5 ЭБ (эксабайт) в коллекции изображений. Эти данные соответствуют каталогу, содержащему 20 млрд галактик и 17 млрд звезд, 7 трлн детектируемых источников и около 30 трлн измерений.

Две группы данных планируется сделать публичными: автоматическая система будет посылать оповещения о транзиентных событиях, а также публично доступными будут данные верхнего уровня (каталоги). Для институтов, сотрудничающих

в рамках миссии, гарантирован доступ к вычислительным ресурсам для эффективного поиска, запуска программ над базой данных в 15 ПБ и обработки изображений в базе данных в 100 ПБ [2]. Институты России не участвуют в этом проекте.

### 2.2 Массив квадратного километра

Массив квадратного километра (Square Kilometer Array, SKA) — наиболее амбициозный проект в радиоастрономии. Радиотелескоп, расположенный в Южной Африке, содержит тысячи отдельных антенн, занимающих площадь около квадратного километра. Он работает в широком диапазоне частот (от 50 МГц до 14 ГГц), имеет чувствительность, в пятьдесят раз превышающую возможности современных радиотелескопов, и способен производить обзор неба в 10 тыс. раз быстрее.

Количество собранной информации ставит сложную задачу хранения и потребует обработки данных в реальном времени. По оценкам, SKA может создавать эксасбайт сырых данных ежедневно, который после обработки в режиме реального времени можно будет сжимать до 10 ПБ [3]. Требования к мощности компьютеров для обработки данных превышают характеристики имеющихся самых быстрых компьютеров в 2015 г., а передача данных в Интернете требует нового вида высокоскоростных сетей. Ученые России не участвуют в этом проекте.

### 2.3 Космическая обсерватория Гайя

Гайя (Gaia) является космической обсерваторией Европейского космического агентства (European Space Agency, ESA), созданной для астрометрии и выведенной на орбиту в 2013 г. Цель — создание трехмерного каталога 1 млрд астрономических объектов, главным образом звезд, позволяющего понять образование и эволюцию нашей Галактики. Этот каталог станет основой для нового взгляда на Галактику и ключом для решения фундаментальных астрономических проблем. Дополнительно ожидается обнаружение от тысяч до десятков тысяч планет, подобных Юпитеру, за пределами Солнечной системы (экзопланет), около полумиллиона квазаров и десятков тысяч астероидов и комет в Солнечной системе. За пять лет миссии общий объем данных составит 20 ТБ. Окончательная версия каталога будет доступна в 2020 г. Доступ к полученным данным ограничен. Российские ученые участвуют в поддержке проекта наземными наблюдениями (*Gaia Follow-Up Network*).

## 2.4 Детекторы гравитационных волн (gravitational wave astronomy)

Проекты LIGO (Laser Interferometer Gravitational-Wave Observatory) или Advanced LIGO и Virgo ориентированы на экспериментальное подтверждение поступления гравитационных волн от их наиболее мощных источников — взрывов коллапсирующих сверхновых и слияния нейтронных звезд в тесных двойных системах. Исходные данные проекта LIGO составляют 1 ПБ, и к ним предоставлен публичный доступ. В рамках проектов в реальном времени работает система оповещения об обнаружении и возможных областях локализации транзитных источников гравитационных волн. Российские ученые участвуют в коллаборации LIGO. Система оповещения для поддержки проекта наблюдениями областей локализации в оптическом диапазоне доступна всем ученым после подписания соглашения с коллаборацией LIGO.

## 2.5 Публичные коллекции данных

*Sloan Digital Sky Survey (SDSS)* — один из основных продолжающихся проектов астрономических наблюдений, более 15 лет его поддержки посвящены созданию карты Вселенной. Каждую ночь широкоугольный телескоп производит более 200 ГБ многоцветных фотометрических обзоров и спектроскопических данных (для сравнения, LSST за ночь будет производить 15 ТБ). В настоящее время фотометрическими наблюдениями в пяти фильтрах покрыто около 37% всего неба. На их основе создаются каталоги, включающие звезды, звездные скопления, галактики, экзопланеты и др.

*NASA/IPAC Extragalactic Database (NED)* — это база данных внегалактических объектов, обеспечивающая систематический анализ интегрируемой информации из сотен обзоров неба и десятков тысяч публикаций. Диапазон наблюдаемых спектров — от гамма- до радиочастотного излучения. По мере опубликования наблюдения кросс-идентифицируются с предшествующими данными и интегрируются в базу данных для упрощения запросов и извлечения требуемых данных. Приблизительный объем базы данных NED около 20 ТБ.

*Mikulski Archive for Space Telescope (MAST)*. Основой MAST является архив научных данных, полученных от чрезвычайно успешно до сих пор функционирующего космического телескопа им. Э. Хаббла. В него включены также данные таких космических проектов, как Kepler, IUE (International Ultraviolet Explorer), GALEX (Galaxy Evolution

Explorer) и др. Объем данных составляет чуть более 100 ТБ, и они публично доступны [4].

*The ESO (European South Observatory) Science Archive* представляет коллекцию данных Европейской южной обсерватории. Месячный поток данных составляет 7 ~ 8 ТБ, а полный объем превышает 100 ТБ за последние несколько лет. Данные после научной обработки в большей части становятся доступными через год. Публичная часть архива доступна для зарегистрированных пользователей из международного сообщества. России пока не удалось стать членом ESO.

## 2.6 Примеры астрономических миссий и коллекций данных в России

В России наиболее близким аналогом архива ESO является архив общих наблюдений Специальной астрофизической обсерватории РАН, содержащий в 2010 г. данные объемом 250 ГБ [5].

Данные международной космической обсерватории INTEGRAL составляют несколько десятков терабайт и являются публично доступными по прошествии одного года, в течение которого исключительные права на данные принадлежат заявителям ежегодных открытых программ.

В России запланировано несколько космических проектов: наряду с уже работающим с 2011 г. орбитальным радиотелескопом «Радиоастрон» [6] это Спектр-Рентген-Гамма (<http://hea.iki.rssi.ru/ru/index.php?page=srg>), WSO-UV [7], а также «Миллиметр» [8].

## 3 Данные в исследованиях мозга

Нейронаука — это совокупность анатомии, физиологии, генетики, биохимии, патологии нервной системы, психологии. Она является передним краем изучения мозга и мышления. Изучение мозга важно для понимания того, как мы воспринимаем и взаимодействуем с внешним миром.

Количество данных, генерируемых в типовой лаборатории, проводящей исследования в нейронауке, растет с поразительной быстротой. Интеграция полученных данных в единую картину является сложной задачей. Для ее решения необходима нейроинформатика, возникающая при сотрудничестве исследователей в нейронауке с информатиками, для того чтобы как новые, так и ранее известные данные стали доступнее сообществу исследователей для ускорения нашего понимания работы мозга [9].

### 3.1 Исследование мозга в рамках стратегической инициативы развития инновационных нейротехнологий (BRAIN)

Инициатива Белого дома BRAIN, объявленная в апреле 2013 г., — это десятилетняя программа, нацеленная на создание динамического понимания функций мозга и демонстрацию того, как отдельные клетки и сложные нейросети взаимодействуют в здоровом или больном организме.

Главные цели анализа сетей взаимодействующих нейронов:

- идентификация и описание компонентов нейронов, определяющих их (клеток) синаптические связи друг с другом, на основе изучения динамики активности во время функционирования нейросетей в живом организме;
- понимание алгоритмов управления обработкой информации внутри нейросетей и между взаимодействующими нейросетями в мозге в целом.

Ожидается, что в результате данного исследования появится концептуальная база понимания биологической основы ментальных процессов вследствие развития новых теоретических инструментов, а также инструментов обработки данных. Теоретические и статистические исследования, а также моделирование способствуют пониманию комплексных, нелинейных функций мозга.

Программе BRAIN необходима инфраструктура для обобщения и обмена релевантными наборами данных, а также методами анализа данных. Значительным препятствием, которое затрудняет понимание работы мозга, является раздробленность исследований мозга и получаемых в результате этих исследований данных. Основной целью является согласование международных усилий по интеграции этих данных в единую картину мозга как отдельной многоуровневой системы. Объем данных о мозге на клеточном уровне имеет порядок эксабайтов. Планируется построить комплексную систему исследовательских платформ, основанных на ИКТ, которая позволила бы нейробиологам, медикам-исследователям и разработчикам новых технологий ускорить темпы их исследований.

### 3.2 Проект Европейского Союза по исследованию человеческого мозга

Human Brain Project (НВР) — это главный десятилетний проект Европейского Союза с бюджетом в 1 млрд Евро, нацеленный на ускорение процесса понимания работы человеческого мозга. Данный проект включает исследования по диагностике

и определению расстройств мозга, а также по разработке новых технологий, основанных на принципах работы мозга [10].

Human Brain Project состоит из 13 подпроектов, охватывающих стратегические данные нейробиологии, когнитивную архитектуру, теорию, этику, менеджмент, а также развитие новых платформ, основанных на информатике.

Основной целью НВР является создание реалистичной симуляции человеческого мозга. Для этого потребуются молекулярная и клеточная информация, позволяющая моделировать и понять биологические процессы в норме и патологии. Это позволит использовать данную информацию для разработки и применения новых типов компьютеров и робототехники, т. е. для применения полученных результатов для разработки новых технологий (создания нейроморфных устройств).

Планируется построить управляемые данными модели, которые отображают то, что удалось узнать о мозге экспериментальным путем, его глубинную механику, а также познать основные принципы, на которых основан мыслительный процесс. Модели мозга будут создаваться при помощи правил обучения, максимально приближенных к реальным закономерностям, которые использует мозг. Ожидается, что подобные модели будут обучаться с помощью тех же механизмов, которые используются человеческим мозгом, и что они будут проявлять подобное интеллектуальное поведение.

Проект НВР развивает 6 новых платформ, основанных на информатике:

- (1) нейроинформатика (поисковые атласы и анализ данных мозга);
- (2) симуляция мозга (построение и симуляция многоуровневых моделей нервных сетей и церебральных функций);
- (3) медицинская информатика (анализ клинических данных для лучшего понимания болезней мозга);
- (4) нейроморфные вычисления (применение функций, подобных функциям мозга, в аппаратном обеспечении);
- (5) нейроробототехника (тестирование моделей мозга и их симуляция в виртуальной среде);
- (6) высокопроизводительные вычисления (обеспечивающие необходимую вычислительную способность).

### 3.3 Проект коннектома человека

Структурные (анатомические) связи мозга (его коннектом) могут быть отображены на нескольких

уровнях: макро- (в сантиметровом и миллиметровом масштабе), мезо- (в миллиметровом и микронном масштабе) и микромасштабе (в микронном и нанометровом разрешении). Текущие разработки по человеческому коннектому (отображению всех нейронных связей в нервной системе) проводятся только в макромасштабе [11]. Данные по Human Connectome Project (HCP) (объемом в десятки терабайт) уже доступны для анализа.

### 3.4 Нейробиологические базы данных

*Атласы мозга (Аллена)* — это проекты по совмещению геномики и нейроанатомии при помощи создания карт экспрессии генов для мозга мыши и человека [12]. Данные этих проектов будут способствовать развитию различных областей нейронаук, они помогут выяснить роль определенного гена в том или ином заболевании мозга. Разные типы клеток центральной нервной системы возникают в связи с изменением экспрессии генов. Карта экспрессии генов в мозге позволяет исследовать отношения между формой и функцией. Атлас мозга дает исследователю вид областей с отличиями экспрессии генов в мозге, которые позволяют исследовать пути формирования нейронных связей. Изучение этих путей, также и с помощью методов нейровизуализации, позволит установить отношения между экспрессией генов, типами клеток и функцией различных путей мозга в организации поведения и фенотипами. Атлас позволит показать, какие гены и области мозга связаны с неврологическими и психическими расстройствами.

### 3.5 Данные в нейронауке в России

В ряде российских исследовательских центров (Научный центр неврологии, Институт высшей нервной деятельности и нейрофизиологии РАН, Институт мозга человека им. Н. П. Бехтеревой и др.) накоплены большие коллекции данных по анатомии, гистологии, генетике и биохимии нервной системы, компьютерной томографии, структурной и функциональной магнитно-резонансной томографии мозга, электроэнцефалографии и вызванным потенциалам при нормальном развитии и старении, а также при неврологических и психических заболеваниях. В настоящее время эти данные доступны в центрах, где коллекции были получены, и в сотрудничающих с ними организациях. Разработка открытых коллекций данных будет способствовать повышению эффективности исследований в области нейронаук.

## 4 Данные в геномике и протеомике

Для современной молекулярной генетики характерно появление качественно новых возможностей, связанных с использованием в исследованиях высокопроизводительных экспериментальных технологий, которые привели к беспрецедентному объему накопленных данных и знаний [13]. Эти данные используются для сравнительного анализа геномов, поиска генетических вариаций и биомаркеров, которые применяются в биотехнологии, сельском хозяйстве, фармакологии, клинических исследованиях, персонализированной медицине и т. д.

Прогнозные оценки указывают на то, что общий объем геномных данных по всем проектам ежегодно будет увеличиваться в 3 раза и достигнет к 2018 г. объема 3300 ПБ.

В настоящее время существует около 7400 высокопроизводительных геномных секвенаторов, которые работают в 1027 центрах по всему миру. В России находится только 14 геномных секвенаторов в 6 научных центрах. Поэтому большая часть геномных данных генерируется за рубежом: в США, Европе, Китае, Южной Корее и др.

Мультиомодальность, многоуровневость и широкомасштабность биологических систем порождают огромный объем неоднородных и распределенных данных, для которых характерна изменчивость и несогласованность, необходимость контроля точности этих данных.

Как и в других предметных областях, отсутствие технологий поиска, распределения, хранения, поддержки целостности, передачи, интеграции и визуализации больших данных существенно затрудняют анализ и систематизацию больших данных [14–16].

### 4.1 Коллекции геномных данных

Целью проекта «1000 геномов» является создание наиболее детализированного каталога генетического разнообразия человеческого генома, основанного на результатах секвенирования геномов более чем 2600 человек из 26 популяций по всему миру.

Проект «1001 геном» ориентирован на поиск генетических вариаций в геномах различных штаммов растения *Arabidopsis thaliana*, которое используется как модель для детального изучения молекулярно-генетических механизмов у растений. Эта информация открывает новые возможности в генетике, определяя аллели, ответственные за фенотипическое разнообразие целого генома одного вида как на разных уровнях, включая биохимический,

метаболический, физиологический, морфологический, так и на уровне целого растения. Результаты исследования проекта «1001 геном» важны для развития таких наук, как селекция растений, биотехнология и медицина.

Проект «Геном 10К» содержит коллекцию более чем 16 000 последовательностей геномов позвоночных, включая ныне живущих и недавно вымерших млекопитающих, птиц, рептилий, амфибий, рыб и многих других видов, находящихся под угрозой исчезновения или вымирающих [17].

Целью проекта «Человеческий микробиом» является описание метагенома микробных сообществ, найденных во многих частях человеческого тела, а также поиск соотношений между изменениями в микробиоме и здоровьем человека.

Проект «Атлас генома рака» содержит исследование геномов пациентов, страдающих от более 33 видов рака. В настоящий момент накоплена информация о более чем 7000 вариантах рака [18]. Эта информация важна для поиска генетических маркеров рака и использования их для диагностики.

## 4.2 Атлас протеомы человека

Интерактивный атлас протеом человека, созданный в Стокгольме, в Royal Institute of Technology, ориентирован на фундаментальные исследования в области биологии человека и применение в трансляционной медицине. В настоящее время атлас содержит 13 млн аннотированных изображений человеческих тканей.

В рамках этого проекта могут изучаться различного типа протеомы человека, например протеома домашнего хозяйства, включающая белки, экспрессирующиеся во всех типах тканей, тканеспецифические протеомы, включающие белки, которые показывают повышенную экспрессию только в одной или нескольких типах тканей, или протеомы, связанные с определенными функциями, такими как лекарственные протеомы, включающие все белки-мишени лекарств, раковая протеома, включающая белки, участвующие в патогенезе рака, а также секретома — все белки, которые секретируются, и т. д.

## 4.3 ELIXIR — Европейская медико-биологическая инфраструктура биологической информации

ELIXIR — это проект Европейской молекулярной биологической обсерватории (European Molecular Biology Laboratory, EMBL), реализуется как

панъевропейская исследовательская инфраструктура. Целью ELIXIR является предоставление средств, необходимых для всех исследователей в области медицины и биологии, начиная с полевых биологов и заканчивая химико-информатиками, позволяя им получить полную информацию из быстрорастущего хранилища информации о живых системах. Эти данные являются основой, на которой базируется наше понимание жизни.

Задачей ELIXIR является управление сбором, контролем качества и архивированием больших объемов биологических данных, полученных вследствие биологических и медицинских экспериментов. Некоторые из этих наборов данных ранее были слишком специализированными и доступными лишь для ученых той страны, в которой они были получены.

## 4.4 Интеграция BILS-ProteomeXchange на основе ресурсов EUDAT

Этот пилотный проект нацелен на интеграцию хранилищ сырых данных масс-спектропии, данных протеомики, собираемых в BILS (Швеция) и ProteomeXchange (через базу данных PRIDE (Proteomics Identifications), EMBL-EBI, U.K.), используя Европейскую инфраструктуру EUDAT. Проект служит примером объединения национальных хранилищ данных и международных репозиторий посредством ELIXIR.

## 4.5 Проект BD2K

Проект *От больших данных к знаниям* (Big Data to Knowledge, BD2K) позволяет использовать биомедицинские большие данные для укрепления человеческого здоровья посредством создания, индексирования и распространения методов, инструментов и обучающих материалов. Проект BD2K (начатый в 2012 г.) имеет четыре главные цели, которые в совокупности расширят использование биомедицинских больших данных:

- упростить широкое использование биомедицинских цифровых ресурсов, сделав их более доступными, распространенными и цитируемыми;
- проводить исследования и разрабатывать методы, программное обеспечение и инструменты, необходимые для анализа биомедицинских больших данных;
- усилить обучение развитию и использованию методов и инструментов, необходимых для науки биомедицинских больших данных;

– поддержать экосистему данных, ускоряющую открытия.

В проекте участвуют 185 институтов, 11 BD2K центров мастерства (centers of excellence).

## 5 Данные в материаловедении

Современные материалы во многом определяют развитие человеческой цивилизации. Они широко применяются в промышленности, включая те отрасли, которые непосредственно связаны с национальной безопасностью, разработкой источников чистой энергии и обеспечением высокого уровня жизни людей. Особенностью данных в неорганической химии и материаловедении является то, что они представляют собой результат обработки и систематизации больших (сотни петабайт) объемов исходных экспериментальных данных. В связи с этим создание инфраструктуры для хранения и поиска данных — одна из важнейших проблем разработки информационных систем для материаловедения.

### 5.1 Инициатива генома материалов

Согласно Инициативе генома материалов (Materials Genome Initiative, MGI), объявленной Белым домом в 2011 г., ускоренное создание новых материалов, обладающих заданными свойствами, критично для достижения высокого уровня конкурентоспособности промышленности США [19]. Цель MGI — обеспечение разработки и внедрения новых материалов за счет координации исследований и предоставления доступа к расчетным моделям и инструментарию для оценки свойств и поведения материалов, а также использования прорывных методов моделирования и анализа данных. Главной целью MGI является создание механизмов, способствующих обмену данными и знаниями о материалах не только между исследователями, но и между академической наукой и промышленностью.

Инициатива генома материалов будет способствовать поддержке лидирующей роли США во многих секторах современного материаловедения и промышленности: от энергетики до электроники, от обороны до здравоохранения, а также поддержке недавних прорывов в теории, моделировании свойств материалов и data mining для существенного прогресса в материаловедении, что приведет к снижению затрат на разработку, исследование и получение новых материалов. Основой MGI является *Инфраструктура инноваций в материаловедении* (Materials Innovation Infrastructure),

которая обеспечит интеграцию методов и средств современного моделирования, включающего данные, а также экспериментальный и теоретический инструментарий.

### 5.2 Средства организации данных о материалах

В июне 2014 г. консорциум Национальных сервисов данных (National Data Service, NDS) объявил о первом показательном проекте разработки средств для организации данных, выбрав для этого область материаловедения (Materials Data Facility, MDF) [20]. Этот проект является реакцией на инициативу Белого дома MGI по ускорению разработки современных материалов. MDF обеспечит материаловедов масштабируемым репозиторием для хранения экспериментальных и расчетных данных, в том числе и до их публикации, снабженных ссылками на соответствующие библиографические источники. MDF станет рычагом для создания национальной инфраструктуры коллективного использования информации, включая разработанные в мире базы данных по свойствам материалов и информационные системы для расчета и моделирования, а также будет способствовать организации обмена данными о материалах, в том числе и еще не опубликованными. Доступность данных и средств расчета обеспечивается современной информационной и телекоммуникационной инфраструктурой, которая позволяет предоставить данные исследователям материалов для многоцелевого использования, дополнительного анализа и проверки.

### 5.3 Программа VAMAS

Versailles Project on Advanced Materials and Standards (VAMAS) [21] — это программа международного сотрудничества, призванная продвигать исследования и разработки, которые обеспечивают подготовку новых стандартов для современных материалов. Предполагается, что эта программа приведет к согласованию стандартов по всему миру. Предварительные исследования при разработке стандартов особенно необходимы в случае современных материалов, поскольку традиционные тесты не всегда подходят для них. VAMAS создан для преодоления барьеров в обмене новыми технологиями, необходимыми для исследований на базе международных стандартов.

### 5.4 Коллекции данных в материаловедении

Коллекция данных Национального института стандартов и технологии (National Institute of Stan-

dards and Technology, NIST) США содержит информацию о широком наборе веществ и материалов: неорганических и органических веществах, включая пластмассы, углеродные нанотрубки, высокопрочные сплавы, искусственные кости и т. д., для которых в институте развиваются стендовые испытания и определяются эталонные тесты.

Коллекция данных Национального института материаловедения (Япония) содержит информацию о веществах и материалах разной природы: неорганических веществах, композитах, промышленных сплавах и т. д.

Немецкая сеть научно-технической информации STN (Scientific and Technical Network) предоставляет доступ к опубликованным экспериментальным данным о структуре и свойствах материалов, патентам и иной информации.

Коллекция данных Springer Materials (Германия) обеспечивает доступ к данным о 3000 физических и химических свойств более 250 000 материалов и веществ.

## 5.5 Проекты информационных систем в области материаловедения в России

Развитие информационных систем по материаловедению в России является инициативой разработчиков. Наиболее известные информационные системы разработаны в Объединенном институте высоких температур РАН [22] и Институте металлургии и материаловедения РАН [23]. Базы данных в этих системах объединены с подсистемами расчета термодинамических свойств веществ [22] и системами data mining, позволяющими конструировать еще не полученные неорганические соединения [22].

## 6 Коллекции данных в науках о Земле

Объектами исследования наук о Земле являются планета Земля и ее атмосфера. Комплексные исследования процессов, происходящих в литосфере, атмосфере, гидросфере, биосфере и криосфере, направлены на понимание функционирования Земли как системы. Особенностью наук о Земле является сложная иерархия предметных областей, включающих в себя как фундаментальные, так и прикладные науки. Эта иерархия накладывает жесткие ограничения как на данные отдельных предметных областей наук о Земле, так и на структуры интегрированных данных наук о Земле, предназначенных для использования в таких приложениях, как метеорология, климатология, океанология, экология.

Основные массивы данных в науках о Земле получаются в результате локальных и дистанционных наблюдений, а также численного моделирования изучаемых процессов. При этом объемы соответствующих архивов, например для данных дистанционного зондирования, достигают десятков петабайт, а данных климатических вычислительных экспериментов — единиц петабайт (CMIP5 — Coupled Model Intercomparison Project Phase 5, ERA CLIM). В области климатологии основные усилия направлены на выяснение причин и последствий происходящих сейчас и возможных в будущем глобальных климатических изменений. Прикладной целью этих исследований является создание «службы климата» как аналога службы погоды. Инструментами получения данных здесь являются сети метеостанций, сети плавающих в океанах буев, сети наземных измерительных комплексов, осуществляющие наблюдения за локальными климатическими и экологическими характеристиками, системы спутников, осуществляющих наблюдения за атмосферой и поверхностью Земли, и климатические модели.

Основным источником больших массивов данных являются спутники. Петабайтные коллекции данных формируются, поддерживаются и обслуживаются в США профильными национальными ведомствами (NOAA — National Oceanic and Atmospheric Administration, NASA — National Aeronautics and Space Administration, DoE — Department of Energy), а в Европе — либо наднациональными тематическими структурами (ECMWF — European Center for Medium range Weather Forecasting, ESA), либо консорциумами ведущих по теме университетов и исследовательских центров. Эти структуры активно участвуют в реализации указанных программ.

### 6.1 Примеры крупных проектов получения и накопления данных в науках о Земле

В области наук о Земле, точнее об окружающей среде, наибольший прогресс в деле обеспечения всего комплекса работ, связанных с большими объемами данных, достигнут в области дистанционного зондирования Земли. Примеры соответствующих программ рассматриваются далее.

В ЕС наиболее амбициозную программу *Copernicus* возглавляют Европейская комиссия и ESA. Европейское космическое агентство координирует доставку данных с 30 спутников, а комиссия отвечает за проект, устанавливает требования и управляет сервисами.

Европейское космическое агентство создает семейство спутников (Sentinels) для оперативных нужд этой программы. Спутники будут проводить уникальный набор наблюдений, таких как всепогодные круглосуточные радарные изображения, получать оптические изображения высокого разрешения для наземных сервисов, данные для сервисов, относящихся к океану и приземному слою, данные по мониторингу состава атмосферы с геостационарных и полярных орбит, а также данные с радарного высотомера для измерения высоты морской поверхности в океанографии.

Программа *Copernicus* [24] обеспечит сервисы для предсказания качества воздуха, предупреждения наводнений, раннего обнаружения засух и опустынивания, оценки качества морской воды и анализа урожая зерновых, мониторинга лесов, контроля изменений землепользования, предсказания катастрофических погодных условий, фиксации разливов нефти, слежения за отклонением кораблей от курса и т. д.

Программа *Copernicus* (получая 5 млрд Евро за период 2014–2020 гг.) предусматривает доставку высококачественных данных (до 8 ТБ в день) в рамках политики, основанной на полном и открытом доступе к данным. Предоставляя данные высокого разрешения о приземном слое, океане и атмосфере, *Copernicus* получит возможность управлять развитием исследований и сотрудничества в новых приложениях наук о Земле.

Развиваемая в США *Earth Observing System* (EOS) — это координированный набор спутников для долговременных глобальных наблюдений приземного слоя Земли, биосферы, земной поверхности, атмосферы и океанов, позволяющих улучшить понимание Земли как сложной интегрированной системы. Информационная инфраструктура EOS содержит 12 национальных центров в США, которые хранят и обеспечивают непрерывный доступ к широкому разнообразию геофизической информации о Земле и космосе: полярных и приземных процессах; верхней атмосфере, глобальной биосфере, атмосферной динамике и геофизике; физической океанографии, радиационному бюджету, тропосферной химии, облакам и аэрозолям; глобальном распределении снега и льда; криосфере; биохимической динамике; воздействию человека на окружающую среду; гидрологическом цикле; климате и погоде; геофизике земной тверди, геологии и геофизике морей, солнечно-земной физике, палеоклиматологии; спутниковом дистанционном зондировании. Данные и средства работы с ними объединены в *Earth Observing System Data and Information System* (EOSDIS) [25].

В настоящее время этот опыт активно используется в мире для создания национальных сегментов такой глобальной информационной системы и для глобальной инфраструктуры международного проекта по наблюдению Земли из космоса *GEOSS* (*Global Earth Observation System of Systems*) — международного проекта, рассчитанного на несколько десятилетий. Объем финансирования — десятки миллиардов долларов.

Проект *Data Observation Network for Earth* (DataONE, <https://www.dataone.org>) является основой для создания науки об окружающей среде в форме распределенной базы и устойчивой киберинфраструктуры для открытого, постоянного, устойчивого и безопасного доступа к качественным описаниям и легкодоступным данным наблюдений о Земле. Проект не предназначен для хранения данных. Он является основой для соединения многочисленных репозиториях в федеральных сетях для поиска, извлечения и обеспечения репликаций на репозиториях данных внутри сетей.

В проекте будет создано легкое и просто устанавливаемое программное обеспечение и развитая совместимость программного обеспечения, уже развернутого в репозиториях по всему миру. Новыми чертами проекта будут:

- семантический поиск результатов измерений;
- отслеживание всех этапов жизненного цикла данных;
- сервисы обработки данных, дающие исследователям возможность простыми способами обращаться к большим данным.

Проект *Satellite Observations for Climate Modeling* (SOCM) посвящен интеграции спутниковой информации и моделированию процессов. Новое поколение инфраструктуры будет поддерживать сравнение спутниковых наблюдений с климатическими моделями. Публикация данных дистанционного зондирования вместе с результатами моделирования климата будет способствовать их сравнению и пониманию. Кроме того, лица, принимающие ключевые решения о будущем климата, состоянии регионального уровня туризма, водных ресурсах и управлении питанием: штаты, федеральное правительство и иностранные структуры, — будут использовать эту более полную информацию.

Следующим шагом является преобразование климатической аналитики в сервисы [26]. Например, сервис CAaaS (*Continuous Analytics as a Service*) сочетает вычисления высокой производительности и аналитику данных с масштабируемым управлением данными, виртуализацией облачных вычислений, представлением адаптивной аналитики

и API (Application Programming Interface), связанными с предметными областями для улучшения доступа к большим коллекциям климатических данных.

В рамках международного сотрудничества *Earth System Grid Federation* (ESGF) созданы порталы, интегрирующие коллекции научных данных, распределенные по всему миру. В рамках этого сотрудничества развивается виртуальная среда *Earth System Grid* (ESG) для содействия анализу глобальных климатических изменений и обеспечивается доступ к предсказанным климатическим данным. В частности, исторические климатические данные и результаты моделирования по климатическим сценариям, выполненные при подготовке недавнего доклада Межправительственной группы экспертов по изменению климата, распространялись через ESGF. В настоящее время около 2 ПБ данных архивированы в узлах ESGF, распределенных по всему свету.

В Европе скоординированный подход к созданию глобальной инфраструктуры данных был разработан в ходе выполнения проекта 7-й Рамочной программы ЕС «Global Research Data Infrastructures: The Big Data Challenges». Реализующие эту программу проекты в Европе можно разделить на три группы: проекты наднациональных структур (ESA, ECMWF), внутригосударственные проекты развития ведомств, о которых имеется очень скудная информация, и межгосударственные научные проекты в рамках программ ЕС, информация о которых доступна в Сети (<http://cordis.europa.eu>). В частности, список инфраструктурных проектов 7-й Рамочной программы включает более чем 350 проектов. Не менее десятой части этих проектов связано с науками о Земле.

Следует добавить, что основой многих прикладных направлений наук о Земле являются результаты фундаментальных наук. В частности, существенную роль играют количественные данные, полученные в таких фундаментальных науках, как спектроскопия, химия атмосферы и др. Учитывая, что число молекул, рассматриваемых при решении задач, например, прогноза качества воздуха в регионе, достигает почти тысячи, а с учетом их изотопов — более двух тысяч, объем спектральных данных и затраты на анализ их качества, с учетом постоянного потока данных в новых спектральных интервалах, делают такие задачи чрезвычайно трудозатратными. Одним из выполненных в Европе проектов, относящихся к фундаментальным наукам, стал проект VAMDC — Virtual Atomic and Molecular Data Center [27, 28]. Этот проект ориентирован на исследовательские группы и институты, играющие центральную роль в производстве атомных и моле-

кулярных данных, которые критичны для использования в широкой области применений.

## 6.2 Сравнимые проекты в России

В области создания информационных ресурсов для наук о Земле инфраструктурных научных проектов, сравнимых по масштабу с названными в подразд. 6.1, в России не было.

Крупным ведомственным проектом является ЕСИМО (Единая государственная система информации об обстановке в Мировом океане).

Более мелкие проекты связаны с пространственными данными субъектов РФ. Проекты в области наук о Земле, финансируемые РФФИ и РНФ, не являются частью каких-либо долгосрочных государственных программ.

## 7 Инфраструктуры данных и проекты для доступа к данным и анализа перспективных источников информации

### 7.1 Проекты исследовательских инфраструктур в Европейском Союзе

*Исследовательские инфраструктуры*, создаваемые в ЕС, представляют собой средства, ресурсы или сервисы уникальной природы, которые были идентифицированы в различных областях сообществами исследователей Европы для поддержки соответствующей деятельности на высоком уровне. Подобное определение *исследовательской инфраструктуры*, включая ассоциированные с ней людские ресурсы, охватывает крупное оборудование или наборы инструментов вместе с содержащими знания ресурсами, такими как коллекции данных, архивы или банки данных.

Европейский стратегический форум исследовательских инфраструктур (European Strategy Forum on Research Infrastructures, ESFRI) является стратегическим механизмом, образованным в 2002 г. странами — членами ЕС и Еврокомиссией, чтобы способствовать научной интеграции Европы и усилению ее международного влияния. Члены ESFRI назначаются министрами науки стран — членов или ассоциированных членов ЕС, а также включают представителей Еврокомиссии. Они работают совместно для определения объединенного видения и общей стратегии, включающих в качестве инструментов планирования и реализации новых

панъевропейских исследовательских инфраструктур регулярно обновляемые дорожные карты, отчеты и критерии. Подобный стратегический подход нацелен на обеспечение Европы наиболее современными исследовательскими инфраструктурами, отвечающими нуждам быстро развивающихся областей науки, продвижение основанных на знаниях технологий и расширение их применений.

Ряд примеров исследовательских инфраструктур, деятельность которых приводит к образованию новых коллекций данных и знаний, к их совместному использованию, рассматривается ниже.

ЦЕРН (<http://home.web.cern.ch>) — наибольшая в мире лаборатория ядерной физики частиц; именно ЦЕРН стал родоначальником идеи исследовательской инфраструктуры.

GEANT (<http://www.geant.net/pages/home.aspx>) — проект высокоскоростной сети, является примером инфраструктуры, способствующей совместному использованию данных и знаний учеными.

ЕММА — Европейский архив мышиных мутантов (European Mouse Mutant Archive, <http://www.emmanet.org>) — типичный пример распределенной инфраструктуры с узлами в шести странах, представленной для пользователей в виде единственного центра.

SIOS (Svalbard Integrated Arctic Earth Observation System, <http://www.sios-svalbard.org/servlet/Satellite?c=Page&pagename=sios/Hovedsidemal&cid=1234130481072>) — интегрированная система наблюдений Арктики на Шпицбергене, предназначена для изучения геофизических, химических и биологических процессов, охватывая всю арктическую систему, начиная от верхних уровней атмосферы до процессов в морских глубинах и земной коре.

BBMRI-LPC (Biobanking and Biomolecular Resources Research Infrastructure — Large Prospective Cohorts, <http://www.bbmri-lpc.org/about>) — исследовательская инфраструктура для получения биобанков биомолекулярных ресурсов — одна из крупнейших сетей поддержки биобанков в Европе; целью проекта является изучение подобных коллекций и связи накопленных данных со здоровьем людей.

EMbaRC (European Consortium of Microbial Resource Centres, <http://www.embarc.eu>) — Европейский консорциум центров микробиомных ресурсов, служит для координации обеспечения информационными микробиомными ресурсами исследователей в Европе и в мире.

SYNTHESYS (Synthesis of Systematic Resources, <http://www.synthesys.info>) — проект создания интегрированной европейской инфраструктуры для поддержки коллекций естественной истории.

## 7.2 Панъевропейская инфраструктура данных EUDAT

Европейская комиссия поддерживает развитие панъевропейской междисциплинарной инфраструктуры данных в рамках программы Horizon 2020, следуя нескольким ведущим принципам.

*Федерализация.* Предполагается, что основные действия над данными реализуются в федерациях данных. Они являются сетями репозиторий и центров данных, которые предоставляют структуры для обработки данных и действуют на основе соглашений о легальных или этических правилах, интерфейсах и спецификациях протоколов, а также стека общих сервисов манипулирования данными. Такие центры могут являться членами многих федераций. Координированный подход предполагает, что каждый центр создает описание своих возможностей, а каждая федерация может использовать одни и те же описания для извлечения необходимой информации. Такой подход способствует открытому представлению исследовательских данных и помогает изменять существующую культуру исследований для поддержки совместного использования данных.

*Открытое совместное использование данных.* Поскольку научные дисциплины интернациональны по своей природе, то критичным является следование международным подходам к снижению барьеров при обмене данными или при их повторном использовании. На этом пути основными препятствиями являются неоднородность данных и языков запросов, способность к пониманию и обнаружению данных, перемещение данных сквозь семантические границы между многозначными контекстами, а также проблемы рассогласования данных (относительно качества, неполноты, абстракции данных).

*Европейская инфраструктура данных EUDAT* является начальным шагом в этих направлениях. EUDAT (<http://www.eudat.eu>) объединяет 25 европейских партнеров, включающих центры данных, провайдеры технологий, сообщества исследователей и фондовые агентства из 15 стран. EUDAT предлагает общие сервисы данных в рамках географически распределенной сети, связывающей центры данных и специализированные репозитории, а также решения для поиска, совместного использования, хранения, репликации, стабильности первичных и вторичных данных исследований и выполнения их анализа. Такая сеть образует Совместную инфраструктуру данных (Collaborative Data Infrastructure), обозначаемую далее СИД, которая развивается как сервис-ориентированная, междисциплинарная и устойчивая инфраструктура.

*EUDAT2020* — новый трехлетний большой проект развития СИД, начатый в 2015 г., целями которого являются: поддержка политики Европейской комиссии открытого доступа к данным исследований, достижение интероперабельности существующих в Европе инфраструктур научных исследований (ИНИ) для доступа ученых к сетевым, вычислительным ресурсам и ресурсам данных в различных ИНИ, включая гриды и облачные инфраструктуры. Так, например, будут достигнуты возможности подключения данных в СИД к высокопроизводительным ресурсам, организуемым в рамках PRACE (Partnership for Advanced Computing in Europe), для их анализа или в качестве входных данных моделей и репликации полученных результатов в систему хранения EUDAT; подключения данных в СИД к гридам и облачным ресурсам, поддерживаемым EGI (European Grid Infrastructure); а также федерализации данных при их подключении к ряду европейских инициатив (таких как Nebula, GEANT, TERENA, OpenAIR и др.).

При организации EUDAT2020 достигнута договоренность о партнерстве с NDS по образованию совместных пилотных проектов (междисциплинарных и межконтинентальных). В СИД будет поддерживаться функция долгосрочного архивирования данных, репликации, каталогизации, цитируемости данных наряду с обеспечением обнаружения, доступа, повторного использования коллекций и отдельных объектов данных. Функции анализа данных будут поддерживаться ресурсами EGI и PRACE, а также средствами, образуемыми на основе виртуализации вычислительного оборудования центров данных и кластерных платформ.

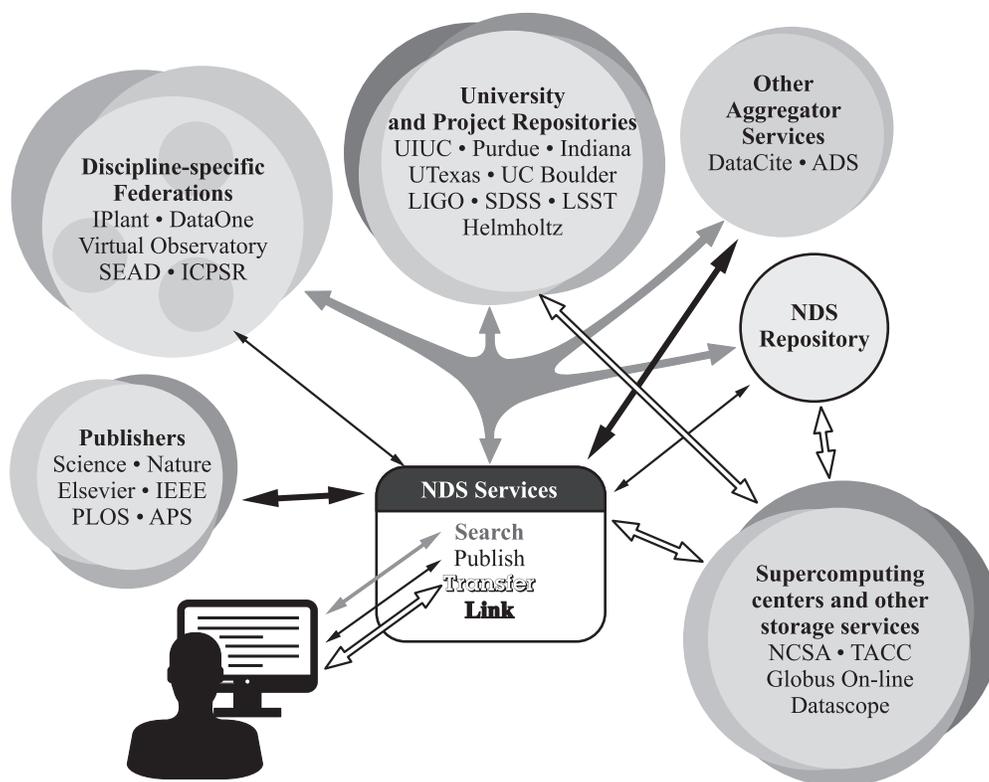
Специальная программа в рамках EUDAT2020 ориентирована на создание средств оценки качества данных и сертификации репозиторий данных в СИД. EUDAT2020 развивает мультидисциплинарный подход, охватывая сообщества исследователей в гуманитарных областях и в социальных сетях (CLARIN — Common Language Resources and Technology Infrastructure, DARIAH, CESSDA), в науках о Земле и атмосфере (EPOS — European Plate Observing System, ICOS, EMSO, VERCE, IAGOS, DRIMM), науке о климате (ENES — European Network for Earth System), биоразнообразии (LifeWatch, LTER, iMarine), науке о жизни (VPH, ELIXIR, BMMRI, ECRIN, INCF, DiXA) и физике (EISCAT, EURO-VO, ISIS, WLCG, PaNdata). Значительное внимание в проекте будет уделено динамическим данным и научным потокам работ, созданию сервисов управления динамическими данными, оставаясь в рамках СИД. Эти исследования будут опираться на сценарии динамического использования данных из ENES и EPOS и обобщения их для ана-

лиза будущих динамических данных при решении реальных научных задач. Одним из планируемых результатов будет создание модели и языка представления жизненного цикла данных, сервисных инфраструктур и происхождения данных. Одновременно будут происходить исследования инфраструктурных операций более эффективных, надежных, устойчивых и близких к потребностям научных сообществ. Примерами планируемых задач являются следующие: оценка объектно-ориентированной среды хранения для машин баз данных центров данных при создании масштабируемой и интероперабельной СИД на основе облачных решений (одной из целей этого анализа является определение возможности реализации B2SHARE без использования iRODS); расширение возможностей уровня долговременного хранения путем применения распределенной графовой базы данных для поддержки отношений между объектами данных вместо собственной базы метаданных, используемой в настоящее время в B2SHARE-сервисе (по замыслу это должно сблизить подходы в СИД с применениями семантического веба, поддержкой происхождения данных и семантического аннотирования).

### 7.3 Инфраструктура проекта «Национальные сервисы данных»

США и ряд международных научных сообществ нуждаются в унификации структур и сервисов для хранения, совместного использования, публикации, размещения и верификации данных. Нужны стандартные средства доступа к данным, программному обеспечению, метаданным, инструментам и иным компонентам, характерным для многих дисциплин. Отсутствие таких стандартных средств создает трудности при проведении исследований и репродуцировании опубликованных научных результатов. США планируют открытую инфраструктуру для поддержки интегрированного набора сервисов национального масштаба для эффективного, удобного и безопасного хранения, совместного использования, публикации, обнаружения, верификации и атрибуции данных на уровне индивидуальных, групповых и кооперативных потребностей. Именно такую инфраструктуру и сервисы формирует проект NDS [29] (см. рисунок).

Международные партнеры, в особенности Research Data Alliance (Альянс исследовательских данных) — RDA, будут способствовать NDS в обеспечении прозрачного, глобального доступа к данным.



Среда NDS

## 7.4 Альянс исследовательских данных

Альянс исследовательских данных был образован для поддержки совместного использования данных сквозь барьеры в 2013 г. Ядро организаторов включало Европейскую комиссию, National Science Foundation, NIST, Министерство инноваций Австралии. В настоящее время число членов альянса превышает 2600 из 90 стран. В рамках альянса образовано большое число рабочих групп и групп по интересам. Дважды в год организуются пленарные совещания в различных местах мира. Например, в марте 2015 г. на совещании в Сан-Диего рассматривались крупномасштабные инфраструктурные проекты организации и анализа данных (включая EUDAT, DataOne, CLARIN, Supercomputing and Big Data, ELIXIR, NDS и др.). Пока еще RDA находится в состоянии обсуждения и уточнения целей альянса.

## 7.5 Проекты обеспечения доступа к ожидаемым данным (на примере астрономии)

Разнообразные проекты (миссии) в мире в различных предметных областях, рассматриваемые

в настоящем обзоре, недавно начали получать данные или планируют начать получать их до либо после 2020 г. В разных странах исследователи в соответствующих областях X-информатики уже начали или подготавливают исследования инфраструктур, поддерживающих доступ к данным, их анализ и управление данными в подобных проектах (миссиях). В настоящем обзоре астрономия выбрана для того, чтобы показать примеры подобных исследований, относящихся к проекту LSST, миссии Gaia, а также к проекту ASTERICS,

### 7.5.1 Подготовка к доступу к данным в проекте LSST

В марте 2015 г. заключено партнерское соглашение между Institut National de Physique Nucléaire et de Physique des Particules (IN2P3), Корпорацией LSST, проектом LSST, а также NCSA (Национальным центром суперкомпьютерных приложений Иллинойского университета) о вкладе в доступ и обработку версий данных LSST во время формирования телескопом обзоров неба [30]. По этому соглашению IN2P3 должен реализовать операции обработки данных LSST посредством коммуникаций, средств обработки данных, а также персонала для образования годичных версий обзора в узле об-

работки, спутниковом по отношению к Архивному центру NCSA LSST. Цель проекта CNRS/LSST заключается в том, чтобы предоставить получаемые LSST данные ученым и более широкой аудитории в мире в виде двух видов данных: (а) извещения о транзиентах, посылаемые в течение 60 с после завершения формирования изображения; (б) годовые релизы данных, которые будут содержать наиболее полно обработанные данные обзора. Каталог годового релиза будет состоять из более 100 таблиц, самыми важными из которых будут каталог объектов, суммирующий для каждого физического источника всю информацию, накопленную за время действия проекта, а также исходный каталог, обеспечивающий доступ к данным каждого конкретного наблюдения одного объекта за одну экспозицию. Данные будут представлены в виде, при котором алгоритмы их анализа смогут сосредоточиться на извлечении знаний из каталогов, накопленных в базе данных без необходимости доступа к первоначальным пикселям. Согласно проекту, предполагается применить массивно-параллельную реляционную технологию баз данных (основанную на принципах архитектуры shared nothing), которая при текущем уровне развития оценивается как более эффективная, чем Map-Reduce. Предварительные измерения производительности и масштабируемости были проведены в проекте LSST на различных кластерах: от 20 узловых 100-терабайтных кластеров до 300 узловых 30-терабайтных кластеров с таблицами, содержащими порядка 50 млрд строк.

### 7.5.2 Подготовка к доступу к данным миссии Gaia

Космический аппарат Gaia в среднем за день передает 40 ГБ данных. К концу миссии ожидается накопление данных, превышающих 1 ПБ. Для доступа к данным созданный для обработки и анализа данных Европейский научный консорциум (DPAC) образовал шесть центров обработки данных (DPC), разбросанных по Европе: Мадрид (DPCE), Тулуза (DPCC — Data Processing and Coordinating Center), Кембридж (DPCI), Турин (DPCT), Барселона (DPCB) и Женева (DPCG) [31].

Различаются два вида DPC: (а) основанные на «инфраструктурном пакете», соединенном с централизованной файловой системой (DPCE, DPCT, DPCB, DPCG); (б) использующие Hadoop (DPCC и DPCI). DPCC ориентирован на обработку спектров, в конце миссии планируется установка в нем 6000 ядер в кластере с 2-гигабитной сетью для связи между узлами и 10-гигабитной сетью между стойками (racks).

### 7.5.3 Подготовка к доступу к разнообразным данным в комплексном проекте исследовательской инфраструктуры ASTERICS

В 2015 г. стартовал координируемый ESFRI и финансируемый Horizon 2020 комплексный проект ASTERICS (<http://www.asterics2020.eu>). Это первый проект, в котором совместно рассматриваются проблемы астрономии, астрофизики и физики космических частиц. В проекте будут использованы различные координируемые ESFRI инфраструктуры (включая SKA, массив телескопов Черенкова CTA (<https://portal.cta-observatory.org/Pages/Home.aspx>), глубоководный нейтринный телескоп KM3NeT (<http://www.km3net.org>), гигантский, создаваемый ESO телескоп E-ELT (<http://www.eso.org/public/unitedkingdom/teles-instr/e-elt>) и другие проекты). Основные цели ASTERICS — поддержка и ускорение реализации телескопов, находящихся в ведении ESFRI, и обеспечение их интероперабельности в рамках интегрированного, многочастотного и многоцелевого посредника. Основные ожидаемые результаты четырехлетнего проекта включают: создание технологий для обеспечения надежного и гибкого манипулирования гигантскими потоками данных, генерируемыми названными выше инфраструктурами, охватываемыми ASTERICS, адаптацию и оптимизацию систем управления огромными базами данных для нужд создаваемой инфраструктуры, адаптацию средств виртуальной обсерватории IVOA для использования в результирующей инфраструктуре. Кроме того должны быть проведены исследования возможности анализа данных в создаваемой инфраструктуре, применяя средства статистического анализа и data mining над коллекциями данных петабайтного масштаба.

## 8 Заключение

Практически во всех ОИИД данные становятся стратегическим ресурсом, затрагивающим все сферы деятельности людей и определяющим конкурентоспособность, уровень развития науки, промышленности, здравоохранения, обороноспособности страны. Анализ состояния пяти представительных областей науки в обзоре показал следующее.

Новизна ситуации заключается в том, что повсеместно в мире развивается процесс образования петабайтных коллекций данных как результат применения новых высокотехнологичных наземных или космических миссий и инструментов в крупных программах (инициативах) исследований, посвященных изучению разнообразных явлений окру-

жающей среды в различных ОИИД. В некоторых областях массивные коллекции данных образуются как результат интеграции большого числа относительно небольших баз данных, создаваемых в различных исследовательских лабораториях мира. В США получение петабайтных коллекций данных часто является одним из естественных результатов стратегических инициатив, объявляемых на уровне Президента США, вовлекающих большое число государственных ведомств и ведущих исследовательских центров в выполнение соответствующих проектов. В Европейском Союзе подобные программы являются межгосударственными.

В России практически нет межведомственных крупных исследовательских программ, которые требовали бы создания новейших инструментов изучения природных явлений, а также крупных международных информационных инфраструктур для накопления и анализа данных (например, со странами БРИКС и ШОС). Большая часть исследовательских проектов организуется инициативно в рамках межличностных, академических и университетских связей. Потребности в научных данных в стране не формируются системно государственными органами науки, ими не регулируются процессы дублирования действий разных ведомств, научных институтов и университетов в области накопления, стандартизации и контроля качества данных.

В результате вклад России в мировые коллекции данных незначительный; более того, трудно прогнозировать изменение ситуации в ближайшие 10 лет ввиду неразвитости соответствующих технологий в стране и отсутствия возможности образования адекватных программ, требующих значительных ассигнований. Таким образом, одной из важнейших проблем сохранения уровня научных исследований в России является обеспечение возможности эффективного доступа исследовательских организаций России к данным, накапливаемым в мире. Доступ к центрам данных, размещенным на территории иностранных государств, требует решения ряда серьезных технических проблем, а также преодоления политических и финансовых ограничений (часто требующих заключения международных соглашений). Эффективный доступ означает возможность проведения анализа данных с темпом их предоставления для ученых в мире. При этом следует понимать, что совершенно недостаточно создания методов решения типовых классических проблем — статистических, машинного обучения, data mining и пр. Опыт показывает, что в конкретных ОИИД каждая конкретная задача анализа данных, особенно больших, требует проведения исследований и экспериментов для

создания специального подхода к решению задачи, по возможности опираясь на типовые методы.

Анализ показал, что, в отличие от России, за рубежом идет активная подготовка к использованию новых источников данных (примеры подготовки даны в разд. 7), включая обсуждение и планирование проектов новых информационных инфраструктур (таких как, например, ASTERICS, NDS, EUDAT2020, RDA, DataONE, MDF, ELIXIR и др.), создание и отработка элементов таких инфраструктур (например, для анализа данных, которые начнут поступать в ближайшее время (миссия Gaia), или не ранее чем через пять лет (телескоп LSST), или по завершении проекта (ASTERICS)). Для этого в каждом крупном проекте образуются большие международные междисциплинарные сообщества специалистов, рабочие группы для спецификации новых функций, которые должны поддерживаться новыми инфраструктурами.

Приведенные в обзоре примеры коллекций данных, создаваемых в мире, инфраструктур формирования новых коллекций данных в процессе исследований предполагается использовать в качестве ориентира при планировании создания и развитии исследовательских инфраструктур для накопления и анализа данных в России, совместимых с зарубежными открытыми инфраструктурами данных в науке. В частности, рассматриваемые в обзоре коллекции данных, цели их создания и научные исследования, планируемые к осуществлению с их помощью, позволяют планировать создание в России компонентов перспективных ИКТ-инфраструктур, таких как, например, средства концептуализации ОИИД, необходимые метамоделли, средства обеспечения возможности повторного использования коллекций данных, воспроизводимости программ и потоков работ и др.

Для достижения эффективного доступа исследовательских организаций России к данным, накапливаемым в мире, с целью их использования в исследовательских проектах России представляется целесообразной организация целевой междисциплинарной программы для реализации пилотного проекта распределенной инфраструктуры для накопления и анализа данных, совместимой с зарубежными открытыми инфраструктурами в науке. Предполагается, что программа должна включать решение следующих основных задач:

- анализ и выбор вариантов инфраструктур и платформ для поддержки решения задач анализа больших данных в различных ОИИД, а также для обеспечения доступа исследователей к разнообразным видам данных в мире и совместного междисциплинарного их исполь-

- зования (наряду с техническими проблемами, в том числе коммуникационными, предполагается решение на международном уровне правовых и финансовых проблем, вызываемых установленными ограничениями доступа к конкретным коллекциям данных);
- организация рабочих групп и формирование сообществ в различных областях с интенсивным использованием данных, принятие мер для установления контактов с международными сообществами аналогичного назначения;
  - создание высокопроизводительного междисциплинарного центра интенсивного использования данных (МЦИИД) для исследователей и практиков из разнообразных ОИИД, накопление междисциплинарного опыта создания подходов к решению конкретных задач анализа данных в конкретных ОИИД, реализация проектов с интенсивным использованием данных в МЦИИД, выработка предложений по тиражированию МЦИИД в стране, их интероперабельности и размещению в составе распределенной междисциплинарной инфраструктуры совместного использования данных.

## Литература

1. The fourth paradigm: Data-intensive scientific discovery / Eds. T. Hey, S. Tansley, K. Tolle. — Redmond, WA, USA: Microsoft Research, 2009. 284 p. <http://go.gl/edvr6W>.
2. *Juric M., Tyson T.* LSST data management: Entering the era of petascale optical astronomy // *High. Astron.*, 2015. Vol. 16. P. 675.
3. *Taylor A. R.* Data intensive radio astronomy en route to the SKA: The rise of big radio data // *High. Astron.*, 2015. Vol. 16. P. 677.
4. *Fleming S. W., Abney F., Donaldson T., et al.* Beyond the prime directive: The MAST discovery portal and high level science products // *American Astronomical Society Meeting (AAS 225)*, 2015. #336.59.
5. *Zhelenkova O., Vitkovsky V., Plyaskina T.* Electronic archive of observational data of astrophysical observatory // *Russ. J. Digital Libraries*, 2010. Vol. 13. Iss. 4. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2010/part4/ZVP>.
6. *Kardashev N. S., Khartov V. V., Abramov V. V., et al.* “RadioAstron” — a telescope with a size of 300 000 km: Main parameters and first observational results // *Astron. Rep.*, 2013. Vol. 57. Iss. 3. P. 153–194.
7. *Shustov B. M., Gomez de Castro A. I., Sachkov M., et al.* WSO-UV progress and expectations // *Astrophys. Space Sci.*, 2014. Vol. 354. Iss. 1. P. 155–161.
8. *Кардашёв Н. С., Новиков И. Д., Лукаш В. Н. и др.* Обзор научных задач для обсерватории Миллиметрон // *УФН*, 2014. Т. 184. № 12. С. 1319–1352.
9. Why neuroinformatics? International Neuroinformatics Coordinating Facility. <http://www.incf.org/about/why-neuroinformatics>.
10. Human Brain Project. <https://www.humanbrainproject.eu>.
11. Human Connectome Project. WU-Minn HCP 500 Subjects Data Release: Reference manual. 2014. 166 p. [http://www.humanconnectome.org/documentation/S500/HCP\\_S500\\_Release\\_Reference\\_Manual.pdf](http://www.humanconnectome.org/documentation/S500/HCP_S500_Release_Reference_Manual.pdf).
12. *Hawrylycz M. J., Lein E. S., Guillozet-Bongaarts A. L., et al.* An anatomically comprehensive atlas of the adult human brain transcriptome // *Nature*, 2012. Vol. 489. P. 391–399.
13. *Gomez-Cabrero D., Abugessaisa I., Maier D., Teschendorff A., Merckenschlager M., Gisel A., Ballestar E., Bongcam-Rudloff E., Conesa A., Tegner J.* Data integration in the era of omics: Current and future challenges // *BMC Syst. Biol.*, 2014. Vol. 8. Suppl. 2. P. 11.
14. *Greene C. S., Tan J., Ung M., Moore J. H., Cheng C.* Big data bioinformatics // *J. Cell. Physiol.*, 2014. Vol. 229. Iss. 12. P. 1896–1900.
15. *Herland M., Khoshgoftaar T. M., Wald R.* A review of data mining using big data in health informatics // *J. Big Data*, 2014. Vol. 1. Iss. 2. 35 p.
16. *Kamesh D. B. K., Neelima V., Ramya Priya R.* A review of data mining using bigdata in health informatics // *Int. J. Sci. Res. Publ.*, 2015. Vol. 5. Iss. 3. 35 p.
17. Genome 10K community of scientists. Genome 10K: A proposal to obtain whole-genome sequence for 10 000 vertebrate species // *J. Heredity*, 2009. Vol. 100. Iss. 6. P. 659–674.
18. *Davis-Dusenbery B., Onder Z., Locke D., Kural D.* Petabyte-scale cancer genomics in the cloud // *TCGA Symposium Oral Presentations*, 2015. P. 34.
19. Materials Genome Initiative for Global Competitiveness. 2011. [http://www.whitehouse.gov/sites/default/files/microsites/ostp/materials\\_genome\\_initiative-final.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/materials_genome_initiative-final.pdf).
20. The Materials Data Facility. <http://www.nationaldataservice.org/mdf>.
21. Versailles Project on Advanced Materials and Standards (VAMAS). <http://www.vamas.org>.
22. *Belov G. V., Iorish V. S., Yungman V. S.* IVTANTHERMO for Windows — database on thermodynamic properties and related software // *CALPHAD*, 1999. Vol. 23. Iss. 2. P. 173–180.
23. *Киселева Н. Н., Дударев В. А., Земсков В. С.* Компьютерные информационные ресурсы неорганической химии и материаловедения // *Усп. хим.*, 2010. Т. 79. Вып. 2. С. 162–188.
24. Copernicus. Observing the Earth. [http://www.esa.int/Our\\_Activities/Observing\\_the\\_Earth/Copernicus/Overview3](http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Overview3).
25. *Ramapriyan H. K., Behnke J., Sofinowski E., Lowe D., Esfandiari M. A.* Evolution of the Earth Observing System (EOS) data and Information System (EOSDIS) // *Standard-based data and information systems for Earth*

- observation / Eds. L. Di, H. K. Ramapriyan. — Lecture notes in geoinformation and cartography ser. — Berlin–Heidelberg: Springer, 2010. P. 63–92.
26. *Schnase J. L., Duffy D. Q., McInerney M. A., et al.* Climate analytic as a service // Conference on Big Data from Space (BiDS'14) Proceedings. — Luxembourg: Publications Office of the European Union, 2014. P. 90–93.
  27. *Dubernet M. L., Boudon V., Culhane J. L., et al.* Virtual atomic and molecular data centre // *J. Quant. Spectrosc. Ra. Transfer*, 2010. Vol. 111. Iss. 15. P. 2151–2159.
  28. *Rixon G., Dubernet M.-L., Piskunov N., et al.* VAMDC — the Virtual Atomic and Molecular Data Centre — a new way to disseminate atomic and molecular data — VAMDC Level 1 Release // *J. Phys. Conf. Ser.*, 2011. Vol. 1344. P. 107–115.
  29. National Data Service (NDS). <http://www.nationaldataservice.org>.
  30. *Gangler E.* Big data challenge posed by the Large Synoptic Survey Telescope // Conference on Big Data from Space (BiDS'14) Proceedings. — Luxembourg: Publications Office of the European Union, 2014. P. 194–197.
  31. *Frezouls B., Brunet P.-M.* Big data technology in the service of the Gaia data processing // Conference on Big Data from Space (BiDS'14) Proceedings. — Luxembourg: Publications Office of the European Union, 2014. P. 198–201.

*Поступила в редакцию 02.12.15*

---



---

## DATA ACCESS CHALLENGES FOR DATA INTENSIVE RESEARCH IN RUSSIA

L. A. Kalinichenko<sup>1,2</sup>, A. A. Volnova<sup>3</sup>, E. P. Gordov<sup>4</sup>, N. N. Kiselyova<sup>5</sup>, D. A. Kovaleva<sup>6</sup>,  
O. Yu. Malkov<sup>6</sup>, I. G. Okladnikov<sup>4</sup>, N. L. Podkolodnyy<sup>7</sup>, A. S. Pozanenko<sup>3</sup>, N. V. Ponomareva<sup>8</sup>,  
S. A. Stupnikov<sup>1</sup>, and A. Z. Fazliev<sup>9</sup>

<sup>1</sup>Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

<sup>2</sup>Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation

<sup>3</sup>Space Research Institute of the Russian Academy of Sciences, 84/32 Profsoyuznaya Str., Moscow 117997, Russian Federation

<sup>4</sup>Siberian Center for Environmental Research and Training, Institute of Monitoring of Climatic and Ecological Systems of the Siberian Branch of the Russian Academy of Sciences, 10/3 Akademicheskii Av., Tomsk 634055, Russian Federation

<sup>5</sup>A. A. Baikov Institute of Metallurgy and Materials Science of the Russian Academy of Sciences, 49 Leninsky Av., GSP-1, Moscow 119991, Russian Federation

<sup>6</sup>Institute of Astronomy of the Russian Academy of Sciences, 48 Pyatnitskaya Str., Moscow 119017, Russian Federation

<sup>7</sup>Center for Bioinformatics, Federal Research Center Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, 10 Acad. Lavrentyeva Av., Novosibirsk 630090, Russian Federation

<sup>8</sup>Research Center of Neurology, 80 Volokolamskoe Shosse, Moscow 125367, Russian Federation

<sup>9</sup>Integrated Information Systems Center, Institute of Atmospheric Optics of the Siberian Branch of the Russian Academy of Sciences, 1 Acad. Zuev Sq., Tomsk 634055, Russian Federation

**Abstract:** The goal of this survey is to analyze the global trends of development of massive data collections and related infrastructures in the world aimed at the evaluation of the opportunities for the shared usage of such collections during research, decision making, and problem solving in various data intensive domains (DIDs) in Russia. The representative set of DIDs selected for the survey includes astronomy, genomics and proteomics, neuroscience (human brain investigation), materials science, and Earth sciences. For each of such DIDs, the strategic initiatives (or large projects) in the USA and Europe aimed at creation of big data collections and the

respective infrastructures planned up to 2025 are briefly overviewed. The information technology projects aimed at the development of the infrastructures supporting access to and analysis of such data collections are also briefly overviewed. The set of large data collections included into the survey and expected to be created soon is planned to be used as a reference point for the design and development of the research infrastructures for data management and analysis making them compatible with the foreign open research infrastructures. In particular, the data collections considered in the survey, the goals of their creation and the researches planned to be accomplished based on them make it possible to proceed to the design and implementation of the advanced components of the research infrastructures, such as, for example, conceptualization facilities of the application domains to be investigated in data intensive research, respective metamodels, components intended for data reuse and reproducing of programs and workflows, etc.

**Keywords:** fourth paradigm; data intensive domains; research infrastructures; data collections; big data

**DOI:** 10.14357/19922264160101

## Acknowledgments

This survey was partially supported by different grants for groups from participating research institutes: for IPI FRC CSC RAS by RFBR grants 13-07-00579, 14-07-00548, and 16-07-01028; for IOA SB RAS by RFBR grant 13-07-00411; for IMCES SB RAS by RFBR grants 13-05-12034 and 14-05-00502; for IMET RAS by RFBR grants 14-07-00819 and 15-07-00980; for INASAN RAS by RFBR grant 15-02-04053 and by the Presidium of RAS Program P-41; for ICG SB RAS by RSF grant 14-24-00123; for RCN by RFBR grants 15-04-08744 and 15-04-05066; and for SRI (IKI) RAS by RFBR grant 15-02-10203-K.

## References

- Hey, T., S. Tansley, and K. Tolle, eds. 2009. The fourth paradigm: Data-intensive scientific discovery. Redmond, WA: Microsoft Research. 284 p. Available at: <http://goo.gl/edvr6W> (accessed February 1, 2016).
- Juric, M., and T. Tyson. 2015. LSST data management: Entering the era of petascale optical astronomy. *High. Astron.* 16:675.
- Taylor, A. R. 2015. Data intensive radio astronomy en route to the SKA: The rise of big radio data. *High. Astron.* 16:677.
- Fleming, S. W., F. Abney, T. Donaldson, et al. 2015. Beyond the Prime Directive: The MAST discovery portal and high level science products. *American Astronomical Society (AAS) Meeting #225.* #336.59.
- Zhelenkova, O., V. Vitkovsky, and T. Plyaskina. 2010. Electronic archive of observational data of astrophysical observatory. *Russ. J. Digital Libraries* 13(4). Available at: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2010/part4/ZVP> (accessed February 1, 2016).
- Kardashev, N. S., V. V. Khartov, V. V. Abramov, et al. "RadioAstron" — a telescope with a size of 300 000 km: Main parameters and first observational results. *Astron. Rep.* 57(3):153–194.
- Shustov, B. M., A. I. Gomez de Castro, M. Sachkov, et al. 2014. WSO-UV progress and expectations. *Astrophys. Space Sci.* 354(1):155–161.
- Kardashev, N. S., I. D. Novikov, V. N. Lukash, et al. 2014. Review of scientific topics for the Millimetron space observatory. *Physics-Uspekhi* 57(12):1199–1228.
- Why neuroinformatics? International Neuroinformatics Coordinating Facility. Available at: <http://www.incf.org/about/why-neuroinformatics> (accessed February 1, 2016).
- Human Brain Project. Available at: <https://www.humanbrainproject.eu> (accessed February 1, 2016).
- Human Connectome Project. WU-Minn HCP 500 Subjects Data Release: Reference manual. Available at: <http://goo.gl/FsfmUb> (accessed February 1, 2016).
- Hawrylycz, M. J., E. S. Lein, A. L. Guillozet-Bongaarts, et al. 2012. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489:391–399.
- Gomez-Cabrero, D., I. Abugessaisa, D. Maier, A. Teschendorff, M. Merckenschlager, A. Gisel, E. Ballestar, E. Bongcam-Rudloff, A. Conesa, and J. Tegner. 2014. Data integration in the era of omics: Current and future challenges. *BMC Syst. Biol.* 8(2):11.
- Greene, C. S., J. Tan, M. Ung, J. H. Moore, and C. Cheng. 2014. Big data bioinformatics. *J. Cell. Physiol.* 229(12):1896–1900.
- Herland, M., T. M. Khoshgoftaar, and R. Wald. 2014. A review of data mining using big data in health informatics. *J. Big Data* 1(2). 35 p.
- Kamesh, D. B. K., V. Neelima, and R. R. Priya. 2015. A review of data mining using bigdata in health informatics. *Int. J. Sci. Res. Publ.* 5(3).
- Genome 10K community of scientists. 2009. Genome 10K: A proposal to obtain whole-genome sequence for 10 000 vertebrate species. *J. Heredity* 100(6):659–674.
- Davis-Dusenbery, B., Z. Onder, D. Locke, and D. Kural. 2015. Petabyte-scale cancer genomics in the cloud. *TCGA Symposium Oral Presentations.* 34.
- Materials Genome Initiative for Global Competitiveness. Available at: [http://www.whitehouse.gov/sites/default/files/microsites/ostp/materials\\_genome\\_initiative-final.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/materials_genome_initiative-final.pdf) (accessed February 1, 2016).

20. The Materials Data Facility. Available at: <http://www.nationaldataservice.org/mdf/> (accessed February 1, 2016).
21. Versailles Project on Advanced Materials and Standards (VAMAS). Available at: <http://www.vamas.org/> (accessed February 1, 2016).
22. Belov, G. V., V. S. Iorish, and V. S. Yungman. 1999. IVTANTHERMO for Windows — database on thermodynamic properties and related software. *CALPHAD* 23(2):173–180.
23. Kiselyova, N. N., V. A. Dudarev, and V. S. Zemskov. 2010. Computer information resources in inorganic chemistry and materials science. *Russ. Chem. Rev.* 79(2):145–166.
24. Copernicus. Observing the Earth. Available at: [http://www.esa.int/Our\\_Activities/Observing\\_the\\_Earth/Copernicus/Overview3](http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Overview3) (accessed February 1, 2016).
25. Ramapriyan, H. K., J. Behnke, E. Sofinowski, D. Lowe, and M. A. Esfandiari. 2010. Evolution of the Earth Observing System (EOS) data and Information System (EOSDIS). *Standard-based data and Information systems for Earth observation*. Eds. L. Di and H. K. Ramapriyan. Lecture notes in geoinformation and cartography ser. Berlin–Heidelberg: Springer. 63–92.
26. Schnase, J. L., D. Q. Duffy, M. A. McInerney, *et al.* 2014. Climate analytic as a service. *Conference on Big Data from Space BiDS'14 Proceedings*. Luxembourg: Publications Office of the European Union. 90–93.
27. Dubernet, M. L., V. Boudon, J. L. Culhane, *et al.* 2010. Virtual atomic and molecular data centre. *J. Quant. Spectrosc. Ra. Transfer* 111(15):2151–2159.
28. Rixon, G., M.-L. Dubernet, N. Piskunov, *et al.* 2011. VAMDC — the Virtual Atomic and Molecular Data Centre — a new way to disseminate atomic and molecular data — VAMDC Level 1 Release. *J. Phys. Conf. Ser.* 1344:107–115.
29. National Data Service (NDS). Available at: <http://www.nationaldataservice.org/> (accessed February 1, 2016).
30. Gangler, E. 2014. Big data challenge posed by the Large Synoptic Survey Telescope. Big data technology in the service of the Gaia data processing. *Conference on Big Data from Space BiDS'14 Proceedings*. Luxembourg: Publications Office of the European Union. 194–197.
31. Frezouls, B., and P.-M. Brunet. 2014. Big data technology in the service of the Gaia data processing. *Conference on Big Data from Space BiDS'14 Proceedings*. Luxembourg: Publications Office of the European Union. 198–201.

*Received December 2, 2015*

## Contributors

**Kalinichenko Leonid A.** (b. 1937) — Doctor of Science in physics and mathematics, professor; Head of Laboratory, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; professor, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; [leonidandk@gmail.com](mailto:leonidandk@gmail.com)

**Volnova Alina A.** (b. 1986) — scientist, Space Research Institute of the Russian Academy of Sciences, 84/32 Profsoyuznaya Str, Moscow 117997, Russian Federation; [alinuss@gmail.com](mailto:alinuss@gmail.com)

**Gordov Evgeny P.** (b. 1946) — Doctor of Science in physics and mathematics, Head of Siberian Center for Environmental Research and Training, Institute of Monitoring of Climatic and Ecological Systems of the Siberian Branch of the Russian Academy of Sciences, 10/3 Akademicheskii Av., Tomsk 634055, Russian Federation; [gordov@scert.ru](mailto:gordov@scert.ru)

**Kiselyova Nadezhda N.** (b. 1949) — Doctor of Science in chemistry, Head of Laboratory, A. A. Baikov Institute of Metallurgy and Materials Science of the Russian Academy of Sciences, 49 Leninsky Av., GSP-1, Moscow 119991, Russian Federation; [kis@imet.ac.ru](mailto:kis@imet.ac.ru)

**Kovaleva Dana A.** (b. 1973) — Candidate of Science (PhD) in physics and mathematics, scientist, Institute of Astronomy of the Russian Academy of Sciences, 48 Pyatnitskaya Str., Moscow 119017, Russian Federation; [dana@inasan.ru](mailto:dana@inasan.ru)

**Malkov Oleg Yu.** (b. 1961) — Doctor of Science in physics and mathematics, Head of Department, Institute of Astronomy of the Russian Academy of Sciences, 48 Pyatnitskaya Str., Moscow 119017, Russian Federation; [malkov@inasan.ru](mailto:malkov@inasan.ru)

**Okladnikov Igor G.** (b. 1978) — Candidate of Science (PhD) in technology, senior scientist, Siberian Center for Environmental Research and Training, Institute of Monitoring of Climatic and Ecological Systems of the Siberian Branch of the Russian Academy of Sciences, 10/3 Akademicheskii Av., Tomsk 634055, Russian Federation; [oig@scert.ru](mailto:oig@scert.ru)

**Podkolodnyy Nikolay L.** (b. 1952) — Head of Center for Bioinformatics, Federal Research Research Center Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, 10 Acad. Lavrentyeva Av., Novosibirsk 630090, Russian Federation; pnl@bionet.nsc.ru

**Pozanenko Alexey S.** (b. 1962) — Candidate of Science (PhD) in physics and mathematics, Head of Laboratory, Space Research Institute of the Russian Academy of Sciences, 84/32 Profsoyuznaya Str, Moscow, Russian Federation; apozenen@iki.rssi.ru

**Ponomareva Natalya V.** (b. 1956) — Doctor of Science in medicine, Head of Group and leading scientist, Research Center of Neurology, 80 Volokolamskoe Shosse, Moscow 125367, Russian Federation; ponomare@yandex.ru

**Stupnikov Sergey A.** — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sstupnikov@ipi.ac.ru

**Fazliev Alexander Z.** (b. 1953) — Candidate of Science (PhD) in physics and mathematics, Head of Integrated Information Systems Center, Institute of Atmospheric Optics of the Siberian Branch of the Russian Academy of Sciences, 1 Acad. Zuev Sq., Tomsk 634055, Russian Federation; faz@iao.ru

# АНАЛИЗ СИСТЕМНОГО РИСКА СОВМЕСТНОГО КРЕДИТОВАНИЯ НАД НЕОДНОРОДНЫМИ КОЛЛЕКЦИЯМИ ДАННЫХ\*

С. А. Ступников<sup>1</sup>, Д. О. Брюхов<sup>2</sup>, Н. А. Скворцов<sup>3</sup>

**Аннотация:** Рассматривается подход к решению задачи анализа системного риска совместного кредитования в области финансового макро моделирования — одной из областей с интенсивным использованием данных — над неоднородными коллекциями данных в виртуально-материализованной среде интеграции. Виртуальная интеграция в среде осуществляется с использованием технологии предметных посредников. Материализованная интеграция реализуется с использованием свободно распространяемой платформы распределенного хранения и обработки данных Hadoop, а также системы Hive, предназначенной для организации реляционных хранилищ данных над Hadoop.

**Ключевые слова:** системный риск совместного кредитования; решение задач; интеграция данных; неоднородные коллекции данных

**DOI:** 10.14357/19922264160102

## 1 Введение

Рост объема и разнообразия данных в науке и бизнесе в последние годы ведет к необходимости следования *Четвертой научной парадигме*, подчеркивающей роль данных в исследованиях с интенсивным использованием данных. Эта роль заключается в том, что новые знания, как и принятие решений, являются результатом анализа данных [1]. В различных областях, называемых *областями с интенсивным использованием данных*, происходит накопление массивных коллекций разнородных данных, представленных в различных моделях данных.

Спектр используемых моделей данных необычайно широк: он включает традиционные реляционные модели, объектные модели, основанные на многомерных массивах модели, графовые модели, модели ключ–значение, документные модели, семантические модели (RDF — Research Description Framework, OWL — Web Ontology Language) и др. Такое разнообразие моделей данных, увеличивающееся со временем, приводит к необходимости создания подходов к интеграции моделей и коллекций данных, представленных в этих моделях, разработки подходов к решению задач над неоднородными коллекциями.

В работе [2] была предложена архитектура комбинированной виртуально-материализованной среды интеграции неоднородных коллекций структурированных, слабо структурированных и не-

структурированных данных (рис. 1). Среда поддерживает как виртуальную, так и материализованную интеграцию коллекций данных, представленную как в традиционных (реляционных), так и в нетрадиционных моделях данных.

Виртуальная интеграция в среде осуществляется с использованием технологии предметных посредников [3]. Посредники образуют промежуточный слой между пользователем (приложением) и неоднородными информационными ресурсами; данные из ресурсов в посреднике не материализуются.

При материализованной интеграции предполагается создание хранилища данных. В хранилище загружаются подлежащие интеграции коллекции данных, при этом данные преобразуются из схемы коллекции в общую схему хранилища. Материализованная интеграция реализуется с использованием свободно распространяемой платформы распределенного хранения и обработки данных Hadoop [4], а также системы Hive [5], предназначенной для организации реляционных хранилищ данных над Hadoop.

В настоящей статье рассматривается подход к решению задач над неоднородными коллекциями данных в виртуально-материализованной среде интеграции. Подход иллюстрируется на примере задачи из области финансового макро моделирования — одной из областей с интенсивным использованием данных.

\* Работа выполнена при поддержке РФФИ (проекты 13-07-00579, 14-07-00548 и 16-07-01028).

<sup>1</sup>Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, sstupnikov@ipi.ru

<sup>2</sup>Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, brd@ipi.ac.ru

<sup>3</sup>Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, nskv@ipi.ac.ru

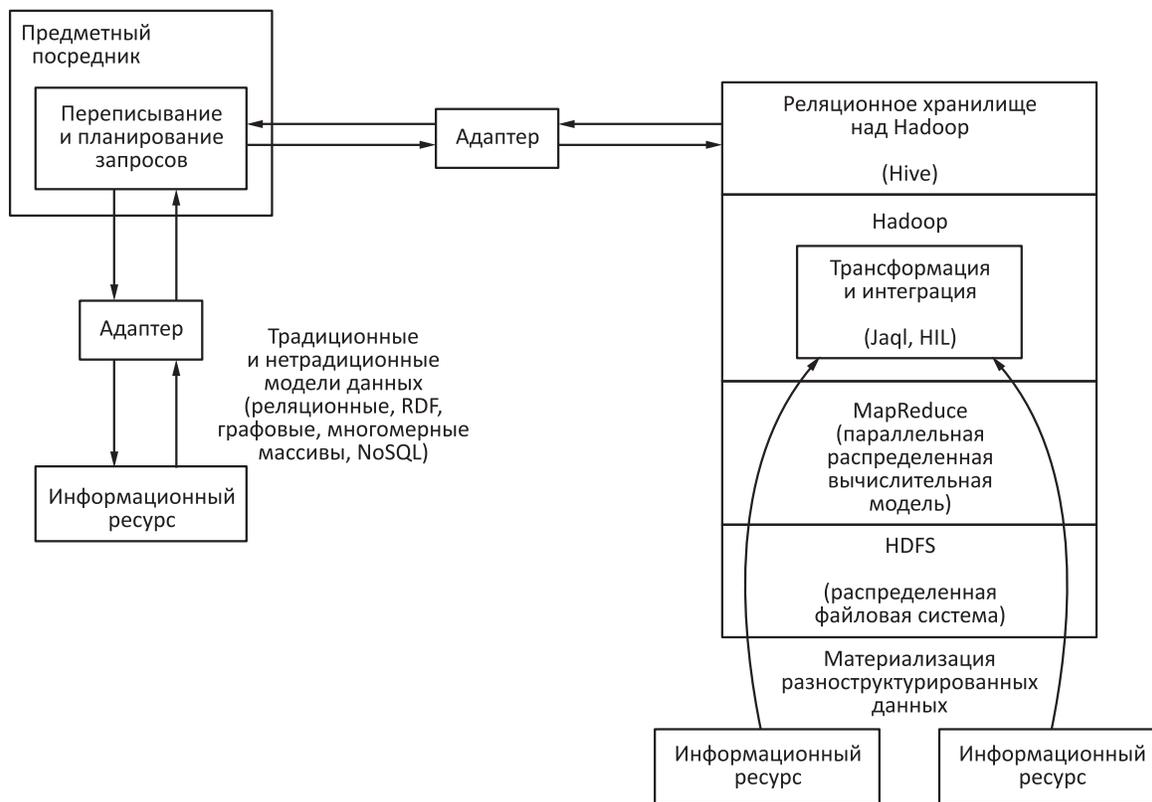


Рис. 1 Архитектура среды виртуально-материализованной интеграции

Статья структурирована следующим образом. В разд. 2 рассмотрена постановка задачи анализа системного риска совместного кредитования. В разд. 3 рассмотрены основные решения задачи в среде интеграции: определение концептуальной схемы предметной области задачи и декларативной спецификации задачи (подразд. 3.1 и 3.2), выбор релевантных предметной области ресурсов и способов их интеграции (подразд. 3.3), определение схем ресурсов, подлежащих виртуальной интеграции, и взглядов, связывающих их со схемой посредника (подразд. 3.5), определение схемы хранилища для материализации коллекций данных и взглядов, связывающих ее со схемой посредника (подразд. 3.6), а также архитектура среды решения задачи (подразд. 3.7).

## 2 Интенсивное использование данных в задаче анализа системного риска совместного кредитования

В данной статье в качестве примера рассматривается конкретная задача анализа системного риска

совместного кредитования. Она состоит в выявлении ведущих игроков на кредитном рынке, банкротство которых может вызвать системный финансовый кризис, оказать влияние на финансовое положение множества других игроков [6, 7].

Исходными данными для задачи являются документы (записи) о совместных (синдицированных) кредитах. Записи содержат даты подписания и погашения кредита, объем кредита, состав участников синдиката, предоставляющего кредит, и другую информацию. На основании исходных данных формируется граф совместного кредитования. Вершинами графа являются организации. Две вершины соединяются ненаправленным ребром, если соответствующие организации предоставляют некоторый совместный кредит (возможно, вместе с другими организациями). Так, если кредит предоставлен совместно пятью банками, ребрами соединяются всевозможные пары, образуемые этими пятью банками. Ребру приписывается вес в зависимости от того, сколько совместных кредитов предоставлено парой организаций. Банкротство некоторого банка, очевидно, окажет прямое влияние на банки, связанные с ним ребрами в графе совместного кредитования, а также окажет опосредованное влияние на банки, связанные с ним в графе путями [6].

Определение того, банкротство каких банков имеет наибольший вклад в риск системного финансового кризиса, базируется на вычислении центральности вершин в графе совместного кредитования.

Существуют различные варианты центральности вершин в графе. *Степенная центральность* вершины вычисляется как сумма весов ребер, связанных с вершиной. *Центральность по посредничеству (betweenness centrality)* [8] вычисляется как мера отношения числа кратчайших путей, проходящих через вершину, к общему числу кратчайших путей в графе. *Центральность по собственному значению (eigenvalue centrality)* [9] вычисляется как мера связи вершины с другими вершинами с высокой центральностью. Вершины с наивысшими показателями центральности с высокой вероятностью являются *критическими хабами* для сети совместного кредитования.

После того как в графе совместного кредитования определены вершины с наибольшей центральностью, представитель финансового регулятора может быть заинтересован в дополнительной информации о соответствующих компаниях. К такой информации относятся связи с другими компаниями (владение, дочерняя компания), ключевые лица компании (директора), агрегированные финансовые данные и т. д.

### 3 Решение задач в виртуально-материализованной среде интеграции

Решение задачи (класса задач) в виртуально-материализованной среде интеграции включает следующие этапы:

- определение *концептуальной схемы* предметной области задачи; такая схема становится спецификацией предметного посредника для виртуальной интеграции релевантных ресурсов;
- *описание задачи* в виде декларативной программы над концептуальной схемой;
- определение *ресурсов, релевантных предметной области* (содержащих данные, необходимые для решения задачи), определение способа их интеграции (виртуальная или материализованная);
- создание *отображений моделей данных информационных ресурсов*, подлежащих виртуальной интеграции, в каноническую информационную модель предметных посредников;

- создание *адаптеров* информационных ресурсов, подлежащих виртуальной интеграции;
- определение *схем информационных ресурсов*, подлежащих виртуальной интеграции, и *отображение* этих схем в каноническую модель (создание *локальных схем ресурсов*);
- определение *взглядов* (представлений), связывающих элементы схемы посредника и схем ресурсов для обеспечения возможности переписывания запросов при виртуальной интеграции ресурсов в посреднике [3];
- определение *схемы хранилища* для материализации коллекций данных и ее отображение в каноническую модель;
- определение *взглядов* (представлений), связывающих элементы схемы посредника и схемы хранилища для обеспечения возможности переписывания запросов при виртуальной интеграции хранилища в посреднике;
- создание *преобразований коллекций данных*, подлежащих материализованной интеграции, в реляционную модель хранилища;
- создание приложения, связывающего исполнительную среду предметных посредников [3], адаптеры информационных ресурсов и хранилище материализованных данных.

Все эти этапы более подробно будут рассмотрены ниже на примере задачи анализа системного риска совместного кредитования.

#### 3.1 Концептуальная схема предметной области задачи

Определение концептуальной схемы производится с использованием языка СИНТЕЗ [10], используемого в качестве канонической информационной модели предметных посредников.

Концептуальная схема предметной области задачи системного риска совместного кредитования представляется в виде модуля (модуль является основной единицей спецификации канонической модели) ColendingSystemicRisk:

```
{ ColendingSystemicRisk; in: module;
class_specification: ...
function: ...
}
```

Модуль содержит секцию классов (моделирующих множества объектов предметной области) и секцию функций. Секция классов включает, в частности, классы `companies`, `persons`, `colendings`:

```

{ companies; in: class;
  instance_section: {
    id: string;
    names: {set; type_of_element: string;};
    ownedBy: {set; type_of_element: Company;};
    ownerOf: {set; type_of_element: Company;};
    keyPersons: {set; type_of_element: Person;};
  }},
{ persons; in: class;
  instance_section: {
    id: string;
    names: {set; type_of_element: string;};
    keyPersonOf: {set; type_of_element:
      Company;};
  }},
{ colendings; in: class;
  instance_section: {
    id: string;
    colender1: string;
    colender2: string;
    numberOfColendings: integer;
  }};

```

Класс `companies` отвечает компаниям, представляющим кредиты (например, банкам). В секции описания экземпляров класса (`instance_section`) определены атрибуты `id` (уникальный идентификатор компании), `names` (различные варианты названия компании), `ownedBy` (множество компаний, владеющих долей данной компании), `ownerOf` (множество компаний, долями которых владеет данная компания), `keyPersons` (множество ключевых лиц компании — директоров, управляющих).

Класс `persons` отвечает лицам, принимающим участие в управлении компаниями. Для экземпляров класса определены атрибуты `id` (уникальный идентификатор персоны), `names` (различные варианты имени персоны), `keyPersonOf` (множество компаний, в которых персона занимает управляющую должность).

Класс `colendings` отвечает отношению совместного кредитования между компаниями. Для экземпляров класса определены атрибуты `id` (уникальный идентификатор отношения кредитования), `colender1` и `colender2` (идентификаторы пары компаний, участвующих в выдаче совместных кредитов), `numberOfColendings` (количество совместно выданных кредитов). Таким образом, совокупность экземпляров класса `colendings` задает граф совместного кредитования, вершинами которого являются компании. Атрибут `numberOfColendings` задает вес ребра в графе.

Секция функций включает, в частности, функцию `isValidColending`:

```

{ isValidColending; in: function;
  params: {+rel/integer, +clnd1/integer,
    +clnd2/integer, -returns/boolean};

```

```

  predicative: {
    ex c/colendings.inst, cmp1/companies.inst,
    cmp2/companies.inst(
      is_in(c, colending) & is_in(cmp1, companies)
      & is_in(cmp2, companies) &
      c.id = rel & cmp1.id = clnd1 &
      cmp2.id = clnd2 &
      (clnd1 = rel.colender1 &
      clnd2 = rel.colender2 -> returns = true) &
      (clnd1 <> rel.colender1 | clnd2 <>
      rel.colender2 -> returns = false) )
    };};

```

Функция принимает на вход идентификатор экземпляра класса `colendings` (`rel`), а также два идентификатора экземпляров класса `companies` (`clnd1` и `clnd2`). Предикативная спецификация функции задается формулой типизированной логики предикатов первого порядка [10]. Функция возвращает значение `true` в том случае, если компании с идентификаторами `clnd1` и `clnd2` являются сокредиторами в отношении совместного кредитования с идентификатором `rel`.

### 3.2 Описание задачи в виде декларативной программы

Первая часть задачи заключается в вычислении центральности компаний в графе совместного кредитования. Например, для вершинной центральности такое вычисление может быть описано в виде одного Дatalog-подобного правила канонической модели [10]:

```

degreeCentrality(x/[companyId,
  sumColendings]) :-
  companies(comp/[companyId: id]) &
  companies(neighbour/[neighbourId:id]) &
  colendings(clnd/[colendId:id,
    numberOfColendings]) &
  isValidColending(colendId, companyId,
    neighbourId) &
  group_by(comp) &
  sumColendings = sum(colend.numberOfColendings).

```

Предикат `degreeCentrality` в голове правила содержит атрибуты `companyId` и `sumColendings`, т. е. для каждой компании (вершины) возвращается суммарное число кредитов, выданных ей совместно с другими компаниями (сумма весов ребер). Переменная `comp` объявляется пробегаящей по экземплярам класса `companies` с использованием одноименного предиката-класса. Атрибут `id` переименовывается в `companyId` (конструкция `companyId: id`) для предотвращения смешивания с одноименными атрибутами других переменных. Аналогично объявляется переменная `neighbour`. Переменная `clnd` пробегает по экземплярам класса `colendings`. Предикат-функция `isValidColending`

устанавливает связь между значениями `comp`, `neighbour` и `colend` через идентификаторы: компания `comp1` должна быть связана ребром `colend` с компанией `comp2`. Операция `group_by` производит группировку по имени компании, а функция `sum` суммирует веса ребер в соответствующей группе.

Вторая часть задачи заключается в извлечении дополнительной информации о компаниях с наибольшей центральностью. Пусть, например, наивысшей центральностью обладает компания `ING Group`. Запрос, выдающий идентификаторы всех компаний (`owned`), в которых у `ING` есть доля, представляется следующим декларативным правилом:

```
ownedByING([owned]) :-
  companies(x/[names]) & companies(y/[owned: id])
  & is_in('ING', x.names) & is_in(y, x.ownerOf).
```

Здесь встроенный предикат `is_in` означает принадлежность элемента ('`ING`') множеству (`x.names`).

Запрос, выдающий всех лиц (`pers`), принимающих участие в управлении компанией `ING` одновременно с управлением некоторой другой компанией (`cmpn`), представляется следующим декларативным правилом:

```
overlappedPersonsOfING([pers, cmpn]) :-
  companies(x) & companies(y/[cmpn: iri]) &
  persons(z/[pers: iri]) &
  is_in('ING', x.names) &
  is_in(z, x.keyPersons) &
  is_in(z, y.keyPersons).
```

### 3.3 Релевантные ресурсы и способы их интеграции

Для решения задачи могут быть использованы, например, следующие ресурсы:

- база данных на основе графовой системы управления базами данных (СУБД) `Neo4j` [11], содержащая информацию о совместном кредитовании компаний, представленную в виде графа;
- триплетная (RDF [12]) база данных `DBpedia`, содержащая структурированную информацию,

извлеченную из Википедии (в частности, информацию о компаниях и персонах).

`Neo4j` представляет собой популярную графовую СУБД, поддерживающую декларативный язык запросов `Cypher` [11]. Естественным представляется использование такой СУБД для решения задач на графах и ее виртуальная интеграция в посреднике. База данных формируется на основе информации о совместных кредитах, публикуемой на сайтах финансовых новостных агентств. Пример того, как может выглядеть информация о кредите, приведен в табл. 1.

В этом случае в графе совместного кредитования появятся вершины, соответствующие компаниям `ABN Amro`, `Citigroup`, `ING`, и ребра, попарно соединяющие эти вершины.

Дополнительную информацию о компаниях и персонах можно извлечь из `DBpedia`. Эта информация представлена в структурированном виде, удобном для материализации в `Nadoop` и преобразования к реляционному виду.

Доступ к `DBpedia` осуществляется посредством запросов на языке `SPARQL` через точку доступа <http://dbpedia.org/sparql>. Например, уникальные идентификаторы (URI — Uniform Resource Identifier) кредитных организаций могут быть извлечены при помощи следующего запроса:

```
SELECT DISTINCT ?bank
WHERE {
  {?bank rdf:type dbo:Bank} UNION
  {?bank dbp:industry dbr:Bank} }
```

Уникальные идентификаторы (URI) персон могут быть извлечены при помощи следующего запроса:

```
SELECT DISTINCT ?person
WHERE {
  {?person rdf:type dbo:Person} UNION
  {?person rdf:type foaf:Person} }
```

Далее, конкретный RDF-документ, описывающий банк или персону, можно извлечь из `DBpedia` в необходимом формате с использованием найденного URI. Так, для компании `BNP Paribas` с идентификатором [http://dbpedia.org/resource/BNP\\_Paribas](http://dbpedia.org/resource/BNP_Paribas), RDF-документ с описанием в формате JSON [13] доступен по ссылке <http://dbpedia.org/>

Таблица 1 Пример данных о совместном кредите

Кредит	ВТБ, 2.2005
Страна	Россия
Объем	450 000 000 USD
Ставка по кредиту	LIBOR + 120.BP
Дата подписания	Февраль 2005
Период	36 месяцев
Дата погашения	Февраль 2008
Уполномоченные ведущие организаторы (MLAs)	ABN Amro, Citigroup и ING

data/BNP\_Paribas.json (приведена лишь часть документа):

```
{ "http://dbpedia.org/resource/ING_Group" : {
  "http://dbpedia.org/ontology/industry" : [
    {"type": "uri", "value": "http://dbpedia.org/resource/Financial_services"},
    {"type": "uri", "value": "http://dbpedia.org/resource/Bank" } ],
  "http://dbpedia.org/property/assets" :
  [ { "type" : "literal",
    "value" : "1.169E12" , "datatype" :
    "http://dbpedia.org/datatype/euro" } ] ,
  "http://dbpedia.org/property/keyPeople" :
  [ { "type" : "literal",
    "value": "Ralph Hamers Patrick Flynn Jeroen van der Veer", "lang": "en" } ],
  "http://dbpedia.org/resource/Bank_Mendes_Gans" : {
    "http://dbpedia.org/ontology/parentCompany" :
    [ { "type" : "uri", "value" :
      "http://dbpedia.org/resource/ING_Group" } ] },
  "http://dbpedia.org/resource/Voya_Financial" : {
    "http://dbpedia.org/ontology/parentCompany" :
    [ { "type" : "uri", "value" : "http://dbpedia.org/resource/ING_Group" } ] }
}
```

В документе содержится, в частности, информация о ключевых лицах компании (`keyPeople`), объеме имущества (`assets`), дочерних компаниях (связь, обратная `parentCompany`) и т. д.

Таким образом, все необходимые дополнительные данные о компаниях извлекаются из DBpedia и загружаются в Nadoor в виде RDF-документов.

### 3.4 Отображение моделей данных информационных ресурсов в каноническую модель и создание адаптеров

Необходимым предусловием виртуальной интеграции модельно однородного класса информационных ресурсов (представленных с использованием одной модели данных) в предметных посредниках является *унификация модели данных ресурсов* — ее отображение в каноническую информационную модель (служащую общим языком в среде разнообразных моделей ресурсов), сохраняющее информацию и семантику операций языка манипулирования данными, а также разработка адаптера для сопряжения ресурсов данного класса со средой исполнения предметных посредников.

Основные принципы отображения модели данных атрибутированных графов и вопросы доказательств сохранения информации и семантики

операций при этом отображении рассмотрены в работе [14]. Общие принципы построения адаптеров изложены в работе [15].

Для реализации конкретного адаптера СУБД Neo4j необходима разработка трансформации запросов (программ) канонической модели в запросы на языке Cypher, основывающейся на упомянутом отображении моделей. Вопросы построения такой трансформации будут вынесены в отдельную статью.

### 3.5 Схема ресурса, подлежащего виртуальной интеграции, и взгляды, связывающие ее со схемой посредника

Схемой ресурса, подлежащего виртуальной интеграции — СУБД Neo4j, содержащей информацию о совместном кредитовании компаний, можно считать шаблоны операций языка Cypher, порождающие экземпляры базы данных. Например, операция вида

```
merge (c:Organization{id: URI});
```

создает в базе данных вершину с меткой `Organization` и атрибутом `id`.

Следующая операция:

```
match (c1: Organization{id: URI1}),
      (c2: Organization{id: URI2})
merge (n1)-[e:colends]-(n2)
on create set e.numberofColendings = 1
on match set e.numberofColendings =
  e.numberofColendings + 1;
```

создает в базе данных ребро с меткой `colends` между двумя вершинами, помеченными как `Organization`, и устанавливает значение атрибута `numberofColendings`, равным 1 (либо увеличивает значение на 1, если ребро уже существует в базе).

Из этих операций можно заключить, что в базе данных имеются вершины типа `Organization` с атрибутом `id` и их соединяют ребра типа `colends` с атрибутом `numberofColendings`.

Согласно принципам отображения модели данных атрибутированных графов в каноническую модель [14] схема базы данных, порождаемой приведенными выше операциями, представляется в канонической модели следующим образом (приведено упрощенное подмножество соответствующей спецификации):

```
{ PropertyGraph; in: module;
  class_specification:
  { vertices; in: class;
    instance_section: {
      id: string;
    }
  }
},
```

```
{ edges; in: class;
  instance_section: {
    id: string;
    startVertex: string;
    endVertex: string;
  }};
function:
isValidEdge: { in: predicate;
  params: {+edg/string, +stVtx/string,
    +endVtx/string, returns/boolean};
};
}
{ Colending; in: module; import: PropertyGraph;
{ Organization; in: class; superclass: vertices;
  instance_section: {
    id: string;
  }};
};
{ colends; in: class; superclass: edges;
  instance_section: {
    numberOfColendings: integer;
  }};
}
```

Здесь модуль PropertyGraph, содержащий классы vertices (вершины) и edges (ребра), задает граф общего вида [14], а модуль Colending, содержащий классы Organization и colends, задает атрибуты вершин и ребер конкретных типов. Вышеприведенная схема, представленная в канонической модели, называется *локальной схемой* ресурса (базы данных о совместном кредитовании компаний).

Для обеспечения возможности переписывания запросов при виртуальной интеграции ресурсов в посреднике необходимо определение взглядов, связывающих элементы схемы посредника и схем ресурсов [16, 17]. Взгляды представляются Даталог-подобными декларативными правилами канонической модели. Для сопоставления вышеприведенных фрагментов схем достаточно трех следующих взглядов вида LAV (Local As View), определяющих, как элемент (класс или функция) локальной схемы выражается через элементы схемы посредника:

```
Colending.organization(x/[id]):-
  companies(x/[id]).

Colending.colends(x/[id, colender1,
  colender2, numberOfColendings]):-
colendings(x/[id, colender1: startVertex,
  colender2: endVertex,
  numberOfColendings]).

Colending.isValidEdge(colend, comp1, comp2):-
isValidColending(colend, comp1, comp2).
```

Первый взгляд задает выражение класса organization через класс companies, второй — класса colends через класс colendings, третий — функции isValidEdge через функцию isValidColending.

Применение алгоритма переписывания запросов [16] с использованием вышеприведенных взглядов к правилу вычисления вершинной центральности (см. разд. 3.2) позволяет получить запрос в терминах локальной схемы:

```
degreeCentrality(x/[companyId,
sumColendings]) :-
organization(comp/[companyId: id]) &
organization(neighbour/[neighbourId: id]) &
colends(cInd/[colendId: id,
  numberOfColendings]) &
isValidEdge(colendId, companyId, neighbourId) &
group_by(comp) &
sumColendings = sum(colend.numberOfColendings).
```

Согласно принципам отображения языка правил канонической модели в язык Cypher [14] адаптер Neo4j должен преобразовывать такое правило в следующий запрос на языке Cypher:

```
match(comp: Organization)-[cInd:
  colends]-(neighbour: Organization)
return comp.id as companyId,
  sum(cInd.numberOfColendings) as sumColendings
```

### 3.6 Схема хранилища для материализации коллекций данных и взгляды, связывающие ее со схемой посредника

Фрагмент реляционной схемы для представления информации об именах банков (отношение banks) и персон (отношение persons), отношении владения между компаниями (bankOwners), ключевых лицах в компаниях (отношение keyPersons) выглядит следующим образом:

```
banks(iri STRING, name STRING)
persons(iri STRING, name STRING)
bankOwners(owner STRING, owned STRING)
bankKeyPersons(person_iri STRING,
  bank_iri STRING)
```

Локальная схема в канонической модели, соответствующая данной реляционной схеме, выглядит следующим образом:

```
{ BanksPersons; in: module;
class_specification:
{ banks; in: class;
  instance_section:{
    iri: string;
    name: string;
  }};
{ persons; in: class;
  instance_section:{
    iri: string;
    name: string;
  }};
}
```

```

{ bankOwners; in: class;
  instance_section:{
    owner: string;
    owned: string;
  }},
{ keyPersons; in: class;
  instance_section:{
    person_iri: string;
    bank_iri: string;
  }},
}

```

Схема представляется модулем `BanksPersons`, каждому отношению реляционной схемы соответствует одноименный класс модуля, каждому атрибуту отношения — одноименный атрибут типа экземпляров соответствующего класса.

Вопросы построения преобразования RDF-коллекций компаний и персон в реляционное представление при помощи языка Jaql [18] в `Nadoop`, необходимого для загрузки данных в реляционное хранилище `Hive`, будут рассмотрены в отдельной статье.

После того как данные загружены в `Hive` (осуществлена материализованная интеграция), необходимо осуществить виртуальную интеграцию хранилища в посреднике в соответствии с подходом, изложенным в [2]. Для этого, как и в случае с интеграцией СУБД `Neo4j`, требуется установить соответствие между элементами схемы хранилища и схемы посредника при помощи взглядов. LAV-взгляды для приведенных фрагментов схем выглядят следующим образом:

```
BanksPersons.banks(b/[iri, name]) :-
companies(c/[iri, names]) & is_in(name, names).
```

```
BanksPersons.people(p/[iri, name]) :-
persons(p/[iri, names]) & is_in(name, names).
```

```
BanksPersons.bankOwners(bo/[owner, owned]) :-
companies(x/[owner: iri]) &
companies(y/[owned: iri]) &
is_in(y, x.ownerOf).
```

```
BanksPersons.bankKeyPersons(kp/[person_iri,
bank_iri]) :-
persons(p/[person_iri: iri]) &
companies(b/[bank_iri: iri]) &
is_in(p, b.keyPersons).
```

Здесь встроенный предикат `is_in(x, y)` обозначает принадлежность элемента `x` множеству `y`. Первый взгляд выражает класс `banks` схемы хранилища через класс `companies` схемы посредника, второй — класс `people` через класс `persons`, третий — класс `bankOwners` через класс `companies`, четвертый — класс `bankKeyPersons` через классы `persons` и `companies`.

Применение алгоритма переписывания запросов [16] с использованием вышеприведенных взглядов к правилу вычисления лиц с конфликтом интересов в компании `ING` (см. разд. 3.2) позволяет получить запрос в терминах локальной схемы:

```
ownedByING([owned]) :-
banks(x/[iri, name]) & name = 'ING' &
bankOwners(y/[iri, owned]).
```

Адаптер `Hive` должен преобразовывать такое правило в следующий запрос на языке `HiveQL`:

```
SELECT owned
FROM banks x JOIN bankOwners y ON x.iri = y.iri
WHERE x.name like "ING";
```

### 3.7 Архитектура среды решения задачи анализа системного риска совместного кредитования

Архитектура среды решения задачи представлена на рис. 2.

Среда включает СУБД `Neo4j`; RDF-базу данных `DBpedia`; хранилище на основе `Hive` над `Nadoop`; а также приложение, связывающее среду исполнения предметных посредников и адаптеры ресурсов.

Для решения задачи анализа системного риска совместного кредитования в СУБД `Neo4j` были загружены данные о 1500 совместных кредитах, выданных компаниям России и стран ближнего зарубежья. Данные были получены с веб-страниц сайта одного из финансовых новостных агентств. Полученный граф совместного кредитования включает около 400 вершин (компаний) и 4500 ребер.

Также из базы данных `DBpedia` были извлечены RDF-документы о финансовых организациях и персонах, принимающих участие в управлении такими организациями. RDF-документы были преобразованы к реляционному виду и помещены в хранилище `Hive`.

Пример данных, полученных в результате решения задачи в комбинированной среде интеграции, приведен в табл. 2. Были обнаружены четыре компании, нормализованная вершинная центральность которых превышает 0,5 (кандидаты в критические хабы). На основании данных, извлеченных из `DBpedia`, были найдены компании, владельцами или совладельцами которых являются критические хабы (приведена лишь часть данных).

## 4 Заключение

В статье рассмотрен подход к решению задачи анализа системного риска совместного кредитования в области финансового макромоделирования над неоднородными коллекциями данных

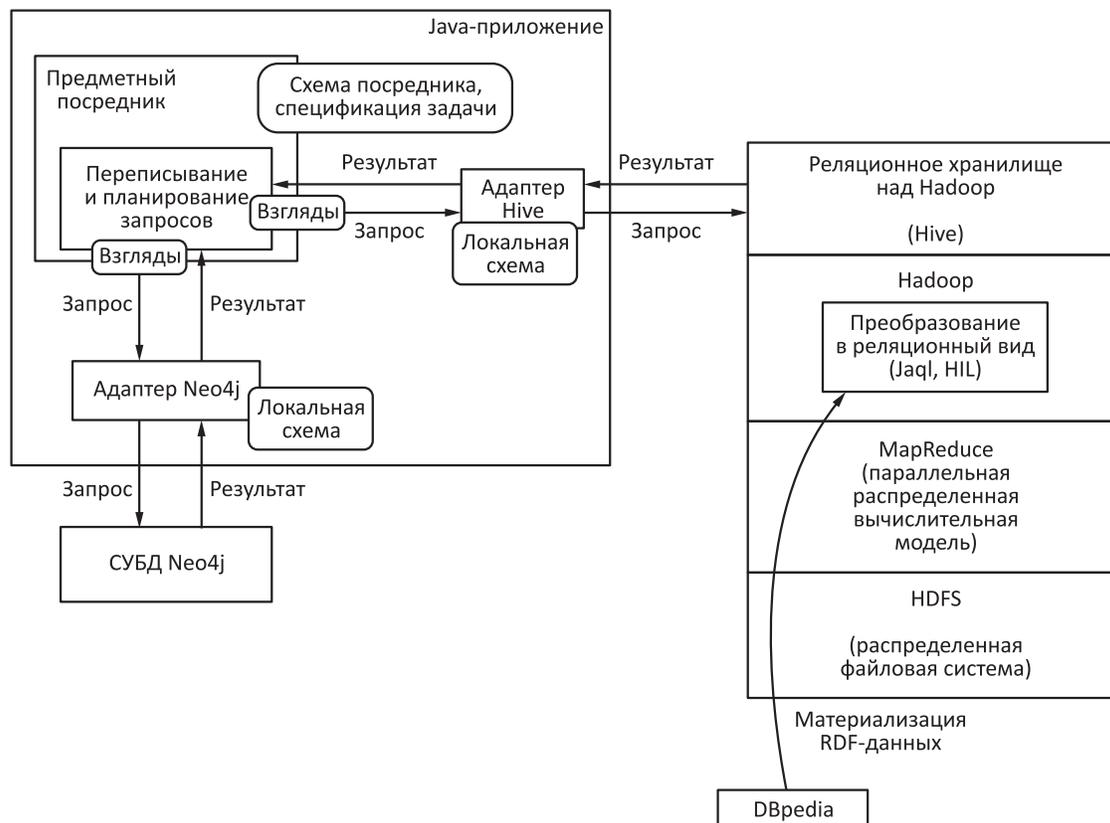


Рис. 2 Архитектура среды решения задачи анализа системного риска совместного кредитования

Таблица 2 Пример результата решения анализа задачи риска совместного кредитования

Название	Нормализованная вершинная центральность	Владение другими компаниями
ING Group	1	Bank Mendes Gans Voya Financial
UniCredit	0,608	Bank Austria Pioneer Investments
HSBC Bank	0,581	HSBC Bank Canada HSBC Bank Australia
BNP Paribas	0,504	SBI Life Insurance Company Bank Insinger de Beaufort

в виртуально-материализованной среде интеграции. Основные этапы решения иллюстрируются на примере задачи анализа системного риска совместного кредитования.

В статье остались нераскрытыми следующие важные вопросы: создание трансформации графовой модели данных Neo4j в каноническую информационную модель и построение адаптера Neo4j, а также создание преобразования коллекции RDF-документов из DBpedia в реляционные данные, пригодные для загрузки в хранилище Hive над

Hadoop. Эти вопросы станут предметом отдельной статьи.

## Литература

1. Hey T., Tansley S., Tolle K. The fourth paradigm — data intensive scientific discovery. 2009. <http://goo.gl/edvr6W>.
2. Ступников С.А., Вовченко А.Е. Комбинированная виртуально-материализованная среда интеграции больших неоднородных коллекций данных // Электронные библиотеки: перспективные методы и тех-

- нологии, электронные коллекции (RCDL 2014): Тр. 16-й Всеросс. науч. конф. — Дубна: ОИЯИ, 2014. С. 339–348.
3. Брюхов Д. О., Вовченко А. Е., Захаров В. Н., Желенкова О. П., Калинин Л. А., Мартынов Д. О., Скворцов Н. А., Ступников С. А. Архитектура промежуточного слоя предметных посредников для решения задач над множеством интегрируемых неоднородных распределенных информационных ресурсов в гибридной грид-инфраструктуре виртуальных обсерваторий // Информатика и её применения, 2008. Т. 2. Вып. 1. С. 2–34.
  4. Apache Hadoop project. 2014. <http://hadoop.apache.org>.
  5. Capriolo E., Wampler D., Rutherglen J. Programming Hive: Data warehouse and query language for Hadoop. — O'Reilly Media, 2012. 350 p.
  6. Burdick D., Hernández M. A., Ho H., Koutrika G., Krishnamurthy R., Popa L., Stanoi I., Vaithyanathan S., Das S. R. Extracting, linking and integrating data from public sources: A financial case study // IEEE Data Eng. Bull., 2011. Vol. 34. No. 3. P. 60–67.
  7. Burdick D., Evfimievski A., Krishnamurthy R., Lewis N., Popa L., Rickards S., Williams P. Financial analytics from public data // SIGMOD/PODS'2014: Workshop on Data Science for Macro-Modeling Conference (International) Proceedings. — New York, NY, USA: ACM, 2014. P. 1–6.
  8. Freeman L. C. A set of measures of centrality based on betweenness // Sociometry, 1977. Vol. 40. No. 1. P. 35–41.
  9. Bonacich P. Power and centrality: A family of measures // Am. J. Sociol., 1987. Vol. 92. No. 5. P. 1170–1182.
  10. Kalinichenko L. A., Stupnikov S. A., Martynov D. O. SYNTHESIS: A language for canonical information modeling and mediator definition for problem solving in heterogeneous information resource environments. — Moscow: IPI RAN, 2007. 171 p.
  11. The Neo4j manual. 2014. <http://goo.gl/cHiOGF>.
  12. RDF 1.1 concepts and abstract syntax / Eds. R. Cyganiak, D. Wood, M. Lanthaler. — W3C Recommendation, February 25, 2014. <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225>.
  13. Introducing JSON. 2014. <http://www.json.org>.
  14. Ступников С. А. Отображение графовых моделей данных в каноническую модель в системах с интенсивным использованием данных // Системы высокой доступности, 2014. Вып. 2. С. 13–31.
  15. Вовченко А. Е. Рассредоточенная реализация приложений в среде предметных посредников: Дис. . . . канд. техн. наук. — М.: ИПИ РАН, 2012. 216 с.
  16. Kalinichenko L. A., Martynov D. O., Stupnikov S. A. Query rewriting using views in a typed mediator environment // Advances in databases and information systems / Eds. G. Gottlob, A. Benczur, J. Demetrovics. — Lecture notes in computer science ser. — Berlin–Heidelberg: Springer-Verlag, 2004. Vol. 3255. P. 37–53.
  17. Briukhov D. O., Kalinichenko L. A., Martynov D. O. Source registration and query rewriting applying LAV/GLAV techniques in a typed subject mediator // RCDL'2007: 9th Russian Conference on Digital Libraries Proceedings. — Pereslavl'-Zalesskiy: Pereslavl' University, 2007. P. 253–262.
  18. Beyer K. S., Ercegovic V., Gemulla R., Balmin A., Eltabakh M., Kanne C.-Ch., Ozcan F., Shekita E. J. Jaql: A scripting language for large scale semistructured data analysis // VLDB Endowment Proceedings, 2011. Vol. 4. No. 12. P. 1272–1283.

Поступила в редакцию 17.11.15

## CO-LENDING SYSTEMIC RISK ANALYSIS OVER HETEROGENEOUS DATA COLLECTIONS

S. A. Stupnikov, D. O. Briukhov, and N. A. Skvortsov

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** The paper considers an approach to solving the problem of co-lending systemic risk analysis over heterogeneous data collections in a combined virtual and materialized integration environment. The problem belongs to the data-intensive domain of financial macromodeling. Virtual integration is implemented using the subject mediation technology. Materialized integration is implemented using the Hadoop open-source software framework for distributed storage and processing of large datasets accompanied by the Hive system intended for relational warehousing over Hadoop.

**Keywords:** co-lending systemic risk; problem solving; data integration; heterogeneous data collections

**DOI:** 10.14357/19922264160102

### Acknowledgments

The work was supported by the Russian Foundation for Basic Research (projects 13-07-00579, 14-07-00548, and 16-07-01028).

## References

- Hey, T., S. Tansley, and K. Tolle. 2009. The fourth paradigm — data intensive scientific discovery. Available at: <http://goo.gl/edvr6W> (accessed November 17, 2015).
- Stupnikov, S.A., and A.E. Vovchenko. 2014. Kombinirovannaya virtual'no-materializovannaya sreda integratsii bol'shikh neodnorodnykh kollektiy dannykh [Combined virtual and materialized environment for integration of large heterogeneous data collections]. *Tr. 16-y Vseross. Konferentsii "Elektronnye Biblioteki: Perspektivnye Metody i Tekhnologii, Elektronnye Kollektii"* [16th Russian Conference on Digital Libraries Proceedings, CEUR Workshop Proceedings]. 1297:339–348.
- Briukhov, D. O., A. E. Vovchenko, V. N. Zakharov, O. P. Zhelenkova, L. A. Kalinichenko, D. O. Martynov, N. A. Skvortsov, and S. A. Stupnikov. 2008. Arhitektura promezhutochnogo sloya predmetnykh posrednikov dlya resheniya zadach nad mnozhestvom integriruemyykh neodnorodnykh raspredelennykh informatsionnykh resursov v gibridnoy grid-infrastrukture virtual'nykh observatoriy [The middleware architecture of the subject mediators for problem solving over a set of integrated heterogeneous distributed information resources in the hybrid grid-infrastructure of virtual observatories]. *Informatika i ee Primeneniya — Inform. Appl.* 2(1):2–34.
- Apache Hadoop project. 2015. Available at: <http://hadoop.apache.org> (accessed November 17, 2015).
- Capriolo, E., D. Wampler, and J. Rutherglen. 2012. *Programming Hive: Data warehouse and query language for Hadoop*. O'Reilly Media. 350 p.
- Burdick, D., M. A. Hernández, H. Ho, G. Koutrika, R. Krishnamurthy, L. Popa, I. Stanoi, S. Vaithyanathan, and S. R. Das. 2011. Extracting, linking and integrating data from public sources: A financial case study. *IEEE Data Eng. Bull.* 34(3):60–67.
- Burdick, D., A. Evfimievski, R. Krishnamurthy, N. Lewis, L. Popa, S. Rickards, and P. Williams. 2014. Financial analytics from public data. *Workshop on Data Science for Macro-Modeling, SIGMOD/PODS'2014 Conference (International) Proceedings*. 1–6.
- Freeman, L. C. 1977. A set of measures of centrality based on betweenness. *Sociometry* 40(1):35–41.
- Bonacich, P. 1987. Power and centrality: A family of measures. *Am. J. Sociol.* 92(5):1170–1182.
- Kalinichenko, L. A., S. A. Stupnikov, and D. O. Martynov. 2007. *SYNTHESIS: A language for canonical information modeling and mediator definition for problem solving in heterogeneous information resource environments*. Moscow: IPI RAN. 171 p.
- The Neo4j manual. 2015. Available at: <http://goo.gl/cHiOGF> (accessed November 17, 2015).
- Cygniak, R., D. Wood, and M. Lanthaler, eds. February 25, 2014. RDF 1.1 concepts and abstract syntax. W3C Recommendation. Available at: <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225> (accessed November 17, 2015).
- Introducing JSON. 2015. Available at: <http://www.json.org/> (accessed November 17, 2015).
- Stupnikov, S. A. 2014. Otobrazhenie grafovyykh modeley dannykh v kanonicheskuyu model' v sistemakh s intensivnym ispol'zovaniem dannykh [Mapping of graph data models into a canonical model for the development of data intensive systems]. *Sistemy Vysokoy Dostupnosti* [Systems of High Availability] 2:13–31.
- Vovchenko, A. E. 2012. Rassredotochennaya realizatsiya prilozheniy v srede predmetnykh posrednikov [Distributed implementation of the applications in the subject mediation environment]. Moscow: IPI RAN. PhD Thesis. 216 p.
- Kalinichenko, L. A., D. O. Martynov, and S. A. Stupnikov. 2004. Query rewriting using views in a typed mediator environment. *Advances in databases and information systems*. Eds. G. Gottlob, A. Benzur, and J. Demetrovics. Lecture notes in computer science ser. — Berlin—Heidelberg: Springer-Verlag. 3255:37–53.
- Briukhov, D. O., L. A. Kalinichenko, and D. O. Martynov. 2007. Source registration and query rewriting applying LAV/GLAV techniques in a typed subject mediator. *9th Russian Conference on Digital Libraries RCDL'2007 Proceedings*. Pereslavl'-Zalesskiy: Pereslavl University. 253–262.
- Beyer, K. S., V. Ercegovic, R. Gemulla, A. Balmin, M. Eltabakh, C.-Ch. Kanne, F. Ozcan, and E. J. Shekita. 2011. Jaql: A scripting language for large scale semistructured data analysis. *VLDB Endowment Proceedings* 4(12):1272–1283.

Received November 17, 2015

## Contributors

**Stupnikov Sergey A.** (b. 1978) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; [sstupnikov@ipi.ac.ru](mailto:sstupnikov@ipi.ac.ru)

**Briukhov Dmitry O.** (b. 1971) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; [brd@ipi.ac.ru](mailto:brd@ipi.ac.ru)

**Skvortsov Nikolay A.** (b. 1973) — scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; [nskv@ipi.ac.ru](mailto:nskv@ipi.ac.ru)

# ОРТОГОНАЛЬНЫЕ СУБОПТИМАЛЬНЫЕ ФИЛЬТРЫ ДЛЯ НЕЛИНЕЙНЫХ СТОХАСТИЧЕСКИХ СИСТЕМ НА МНОГООБРАЗИЯХ\*

И. Н. Сеницын<sup>1</sup>

**Аннотация:** Для нелинейных дифференциальных стохастических систем на гладких многообразиях с винеровскими и пуассоновскими шумами в уравнениях состояния и винеровскими шумами в наблюдениях разработана теория синтеза ортогональных субоптимальных фильтров (ОСОФ) по среднеквадратическому критерию. Получены точные фильтрационные уравнения для стохастических систем на многообразиях (МСтС). Обсуждаются вопросы упрощения точных фильтрационных уравнений. Приводятся уравнения субоптимальных фильтров (СОФ) на основе методов нормальной аппроксимации (МНА) и статистической линеаризации (МСЛ). Для решения задач в реальном времени использование нормальных СОФ (НСОФ) не обеспечивает необходимой точности, поэтому в основу синтеза положены методы ортогональных разложений (МОР) и квазимоментов (МКМ) для апостериорной одномерной плотности. Получены уравнения точности и чувствительности алгоритмов. В качестве тестового примера рассмотрена одномерная нелинейная стохастическая система с аддитивным и мультипликативным белым шумом. Рассмотрены некоторые обобщения разработанных алгоритмов.

**Ключевые слова:** апостериорное одномерное распределение; винеровский шум; квазимомент (КМ); коэффициент ортогонального разложения (КОР); метод квазимоментов (МКМ); метод ортогональных разложений (МОР); нормальный фильтр; ортогональный СОФ (ОСОФ); первая функция чувствительности; пуассоновский шум; стохастическая система на многообразиях (МСтС); субоптимальный фильтр (СОФ)

DOI: 10.14357/19922264160103

## 1 Введение

В [1, 2] МОР и МКМ развиты для аналитического моделирования одно- и многомерных распределений в стохастических системах на гладких многообразиях и дано их применение для задач надежности и безопасности технических систем.

Рассмотрим развитие теории СОФ на базе МОР и МКМ для МСтС.

Раздел 2 посвящен точным фильтрационным уравнениям для МСтС, а также СОФ на основе МНА и МСЛ.

В разд. 3 рассматриваются ОСОФ на основе МОР и МКМ аппроксимаций нормированной апостериорной одномерной плотности.

В разд. 4 обсуждаются вопросы точности и чувствительности ОСОФ.

В разд. 5 приводится тестовый пример.

Заключение содержит выводы и некоторые обобщения.

## 2 Точные фильтрационные уравнения. Нормальный субоптимальный фильтр

### 2.1 Уравнения процессов. Вспомогательные формулы

Пусть векторный стохастический процесс (СтП)  $[X_t^T Y_t^T]^T$  определяется системой векторных стохастических дифференциальных уравнений Ито:

$$dX_t = \varphi(X_t, Y_t, \Theta, t) dt + \psi'(X_t, Y_t, \Theta, t) dW_0 + \int_{R_0^q} \psi''(X_t, Y_t, \Theta, t, v) P^0(dt, dv), \quad X(t_0) = X_0; \quad (1)$$

$$dY_t = \varphi_1(X_t, Y_t, \Theta, t) dt + \psi'_1(X_t, Y_t, \Theta, t) dW_0 + \int_{R_0^q} \psi''_1(X_t, Y_t, \Theta, t, v) P^0(dt, dv), \quad Y(t_0) = Y_0. \quad (2)$$

\* Работа выполнена при поддержке РФФИ (проект 15-07-02244).

<sup>1</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, sinitsin@dol.ru

Здесь  $Y_t = Y(t)$  —  $n_y$ -мерный наблюдаемый СтП,  $Y_t \in \Delta^y$  ( $\Delta^y$  — гладкое многообразие наблюдений);  $X_t = X(t)$  —  $n_x$ -мерный ненаблюдаемый СтП (вектор состояния),  $X_t \in \Delta^x$  ( $\Delta^x$  — гладкое многообразие состояний);  $W_0 = W_0(t)$  —  $n_w$ -мерный винеровский СтП ( $n_w \geq n_y$ );  $P^0(\Delta, A) = P(\Delta, A) - \mu_P(\Delta, A)$ ,  $P(\Delta, A)$  — представляет собой для любого множества  $A$  простой пуассоновский СтП, а  $\mu_P(\Delta, A)$  — его математическое ожидание, причем

$$\mu_P(\Delta, A) = MP(\Delta, A) = \int_{\Delta} \nu_P(\tau, A) d\tau;$$

$\nu_P(\Delta, A)$  — интенсивность соответствующего пуассоновского потока событий,  $\Delta = (t_1, t_2]$ ; интегрирование по  $\nu$  распространяется на все пространство  $R^q$  с выколотым началом координат;  $\Theta$  — вектор случайных параметров размерности  $n_\Theta$ ;  $\varphi = \varphi(X_t, Y_t, \Theta, t)$ ,  $\varphi_1 = \varphi_1(X_t, Y_t, \Theta, t)$ ,  $\psi' = \psi'(X_t, Y_t, \Theta, t)$  и  $\psi'_1 = \psi'_1(X_t, Y_t, \Theta, t)$  — известные функции, отображающие  $R^{n_x} \times R^{n_y} \times R$  соответственно в  $R^{n_x}$ ,  $R^{n_y}$ ,  $R^{n_x n_w}$  и  $R^{n_y n_w}$ ;  $\psi'' = \psi''(X_t, Y_t, \Theta, t, v)$  и  $\psi''_1 = \psi''_1(X_t, Y_t, \Theta, t, v)$  — известные функции, отображающие  $R^{n_x} \times R^{n_y} \times R^q$  в  $R^{n_x}$  и  $R^{n_y}$ . Требуется найти оценку  $\hat{X}_t$  СтП  $X_t$  в каждый момент времени  $t$  по результатам наблюдения СтП  $Y(\tau)$  до момента  $t$ ,  $Y_{t_0}^t = \{Y(\tau) : t_0 \leq \tau < t\}$ .

Предположим, что

- уравнение состояния имеет вид (1);
- уравнение наблюдения (2), во-первых, не содержит пуассоновского шума ( $\psi'_1 \equiv 0$ ), во-вторых, коэффициент при винеровском шуме  $\psi'_1$  в уравнениях наблюдения не зависит от состояния ( $\psi'_1(X_t, Y_t, \Theta, t) = \psi'_1(Y_t, \Theta, t)$ ).

В этом случае уравнения задачи нелинейной фильтрации имеют следующий вид:

$$dX_t = \varphi(X_t, Y_t, \Theta, t) dt + \psi'(X_t, Y_t, \Theta, t) dW_0 + \int_{R_0^q} \psi''(X_t, Y_t, \Theta, t, v) P^0(dt, dv); \quad (3)$$

$$dY_t = \varphi_1(X_t, Y_t, \Theta, t) dt + \psi_1(Y_t, \Theta, t) dW_0. \quad (4)$$

Будем считать, что выполнены условия существования и единственности СтП  $[X_t^T Y_t^T]^T$ , определяемого (3) и (4) при соответствующих начальных условиях [3–5].

В дальнейшем потребуется для стохастического уравнения

$$dZ = a dt + b dW_0 + \int_{R_0^q} c P^0(dt, dv) \quad (5)$$

следующая обобщенная формула Ито [5, 6] для дифференциала  $U = U(Z, t)$ :

$$dU = \left\{ U_t + U_z^T a + \frac{1}{2} \text{tr} [U_{zz} b v b^T] \right\} dt + \int_{R_0^q} [U(Z + c, t)^T - U(Z, t)^T - U_z^T c] \mu_P(dt, dv) + U_z^T b dW_0 + \int_{R_0^q} [U(Z + c, t) - U(Z, t)] P^0(dt, dv). \quad (6)$$

Здесь  $a$ ,  $b$  и  $c$  — известные функции  $Z$  и  $t$ .

## 2.2 Фильтрационные уравнения

Как известно [4, 6, 7], для любых СтП  $X_t$  и  $Y_t$  оптимальная оценка  $\hat{X}_t$ , минимизирующая средний квадрат ошибки в каждый момент времени  $t$  представляет собой апостериорное математическое ожидание СтП  $X_t$ :  $\hat{X}_t = M[X_t | Y_{t_0}^t]$ . Чтобы найти это условное математическое ожидание, необходимо знать  $p_t = p_t(x)$  — апостериорное одномерное распределение СтП  $X_t$ .

В основе уравнений оптимальной (в смысле минимума средней квадратической ошибки) фильтрации для уравнений (3) и (4) в силу (6) лежит следующая формула для стохастического дифференциала апостериорного математического ожидания скалярной функции  $f = f(X, t)$  вектора состояния:

$$d\hat{f} = M_{\Delta^x}^{p_t} [f_t(X, t) + f_x(X, t)^T \varphi(X, Y, t) + \frac{1}{2} \text{tr} \{ f_{xx}(X, t) (\psi' \nu \psi'^T) (X, Y, t) \} + \int_{R_0^q} \{ f(X + \psi'', t) - f(X, t) - f_x(X, t)^T \psi''(X, Y, t) \} \nu_P(t, dv) | Y_{t_0}^t] dt + M_{\Delta^x}^{p_t} \{ f(X, t) [\varphi_1(X, Y, t)^T - \hat{\varphi}_1^T] + f_x(X, t)^T (\psi \nu \psi_1^T) (X, Y, t) | Y_{t_0}^t \} \times (\psi_1 \nu \psi_1^T)^{-1} (Y, t) (dY - \hat{\varphi}_1 dt). \quad (7)$$

Здесь для краткости аргумент  $\Theta$  опущен;  $X = X_t$ ,  $Y = Y_t$ ;  $\nu = \nu_0$  и  $\nu_P$  — интенсивности  $W_0$  и  $P^0$ ;  $\hat{\varphi}_1$  — апостериорное математическое ожидание  $\varphi_1$  при заданной условной плотности  $p_t = p_t(x, \Theta)$ :

$$\hat{\varphi}_1 = M_{\Delta^x}^{p_t} [\varphi_1(X, Y, t)].$$

Полагая в (5)

$$f(X, t) \equiv g_t(\lambda, \Theta) = M_{\Delta^x}^{p_t} \exp(i\lambda^T X),$$

получим точное нелинейное фильтрационное уравнение для характеристической функции  $g_t(\lambda, \Theta)$ :

$$\begin{aligned}
dg_t(\lambda, \Theta) = & M_{\Delta^x}^{p_t} \left[ \left\{ i\lambda^T \varphi(X_t, Y_t, \Theta, t) - \right. \right. \\
& \left. \left. - \frac{1}{2} \lambda^T (\psi \nu \psi^T)(X_t, Y_t, \Theta, t) \lambda + \gamma(\lambda, X_t, Y_t, \Theta, t) \right\} \times \right. \\
& \left. \times e^{i\lambda^T X_t} \mid Y_{t_0}^t \right] dt + M_{\Delta^x}^{p_t} \left\{ \varphi_1(X_t, Y_t, \Theta, t)^T - \hat{\varphi}_1^T + \right. \\
& \left. + i\lambda^T (\psi \nu \psi^T)(X_t, Y_t, \Theta, t) \times \right. \\
& \left. \times e^{i\lambda^T X_t} \mid Y_{t_0}^t \right\} (\psi_1 \nu \psi_1^T)^{-1}(Y_t, \Theta, t) (dY_t - \hat{\varphi}_1 dt) \quad (8)
\end{aligned}$$

где

$$\begin{aligned}
\gamma = & \gamma(\lambda, X_t, Y_t, \Theta, t) = \int_{R_0^q} [e^{i\lambda^T \psi''(X_t, Y_t, \Theta, t, v)} - \\
& - 1 - i\lambda^T \psi''(X_t, Y_t, \Theta, t, v)] \nu_P(\Theta, t, dv). \quad (9)
\end{aligned}$$

Функции  $g_t(\lambda, \Theta)$  и  $p_t(x, \Theta)$  связаны между собой преобразованием Фурье.

Отсюда для гауссовской МСтС (3) и (4) ( $\psi'' \equiv 0$ ) уравнение (7) при  $\gamma = 0$  упрощается и приобретает вид:

$$\begin{aligned}
dg_t(\lambda, \Theta) = & M_{\Delta^x}^{p_t} \left[ \left\{ i\lambda^T \varphi(X_t, Y_t, \Theta, t) - \right. \right. \\
& \left. \left. - \frac{1}{2} \lambda^T (\psi \nu \psi^T)(X_t, Y_t, \Theta, t) \lambda \right\} e^{i\lambda^T X_t} \mid Y_{t_0}^t \right] dt + \\
& + M_{\Delta^x}^{p_t} \left\{ \varphi_1(X_t, Y_t, \Theta, t)^T - \hat{\varphi}_1^T + \right. \\
& \left. + i\lambda^T (\psi \nu \psi^T)(X_t, Y_t, \Theta, t) e^{i\lambda^T X_t} \mid Y_{t_0}^t \right\} \times \\
& \times (\psi_1 \nu \psi_1^T)^{-1}(Y_t, \Theta, t) (dY_t - \hat{\varphi}_1 dt), \quad (10)
\end{aligned}$$

Таким образом, в основу синтеза СОФ для МСтС (3) и (4) могут быть положены следующие утверждения.

**Теорема 2.1.** Пусть для МСтС (3) и (4) выполнены условия существования и единственности решения, а матрица  $\sigma_1 = \psi_1 \nu \psi_1^T$  не вырождена. Тогда при условии ограниченности соответствующих условных математических ожиданий в (9) точное фильтрационное уравнение для условной одномерной характеристической функции имеет вид (8).

**Теорема 2.2.** Пусть для гауссовской МСтС (3) и (4) ( $\psi'' \equiv 0$ ) выполнены условия существования и единственности решения, а матрица  $\sigma_1 = \psi_1 \nu \psi_1^T$  не вырождена. Тогда точное фильтрационное уравнение для условной одномерной характеристической функции имеет вид (10).

**Замечание 2.1.** Точное решение фильтрационных уравнений теорем 2.1 и 2.2 возможно только в случаях, когда уравнения гауссовской дифференциальной МСтС линейны или линейны лишь относительно вектора состояния  $X_t$  при независимой

от состояния функции  $\psi$ . Эти уравнения дают точное решение задачи оптимальной нелинейной фильтрации. Однако это решение не может быть реализовано практически. Для нахождения оптимальной оценки вектора состояния необходимо решить фильтрационное уравнение для апостериорной характеристической функции (или фильтрационное уравнение для апостериорной плотности вектора состояния  $X_t$ ) после получения результатов наблюдений, затем вычислить оптимальную оценку вектора  $X_t$ . Но методов точного решения этих уравнений в общем случае пока еще не существует.

**Замечание 2.2.** Численное решение фильтрационных уравнений в задачах реального времени (или онлайн-оценивания) тоже невозможно, так как для этого требуется много времени, а решать их необходимо каждый раз после получения результатов наблюдений. Кроме того, практическое применение точной теории оптимальной нелинейной фильтрации имеет смысл только в тех случаях, когда оценки можно вычислять в реальном масштабе времени по мере получения результатов наблюдений. Точная теория дает оптимальные оценки в каждый момент  $t$  по результатам наблюдений, полученным к этому моменту, без использования последующих результатов наблюдений. Если эти оценки не могут быть вычислены в тот же момент  $t$  или хотя бы с фиксированным приемлемым запаздыванием и их вычисление приходится откладывать на будущее, то нет никакого смысла отказываться от использования наблюдений, получаемых после момента  $t$ , для оценивания состояния системы в момент  $t$ . Поэтому для статистической обработки результатов после окончания наблюдений, т.е. для офлайн-оценивания, целесообразно применять известные из математической статистики методы постобработки информации [4].

## 2.3 О приближенных методах нелинейной фильтрации

Необходимость обработки результатов наблюдений в реальном масштабе времени непосредственно в процессе эксперимента привела к появлению ряда приближенных методов оптимальной нелинейной фильтрации, называемых обычно методами субоптимальной фильтрации [6]. Одни приближенные методы основаны на приближенном решении фильтрационных уравнений, а другие — на превращении формул для стохастических дифференциалов оптимальной оценки  $\hat{X}_t$  и апостериорной ковариационной матрицы ошибки  $R_t$  в стохастические дифференциальные уравнения для  $\hat{X}_t$  и  $R_t$  путем разложения функций  $\varphi$ ,  $\varphi_1$  и  $\psi_1$  или  $\varphi$ ,

$\varphi_1, \psi', \psi''$  и  $\psi, \psi_1$  в степенные ряды и отбрасывания остаточных членов.

Для приближенного решения уравнения для апостериорной одномерной характеристической функции  $g_1(\lambda, \Theta)$  вектора  $X_t$  можно использовать методы, основанные на параметризации одномерных распределений СтП, определяемого стохастическим дифференциальным уравнением [6, 7]. Эти методы позволяют изучить стохастические дифференциальные уравнения для параметров апостериорного распределения. Простейшим таким методом является МНА апостериорного распределения. Исключительно важное практическое значение имеют квазилинейные фильтры, получаемые с помощью методов эквивалентной линеаризации [5–7].

## 2.4 Субоптимальный фильтр на основе метода нормальной аппроксимации

Так как нормальное (гауссовское) распределение, аппроксимирующее апостериорное одномерное распределение  $X_t$ , полностью определяется математическим ожиданием  $\hat{X}_t$  и ковариационной матрицей  $R_t$  вектора  $X_t$ , то при аппроксимации апостериорного одномерного распределения вектора  $X_t$  нормальным распределением все математические ожидания в правых частях формул для  $d\hat{X}_t$  и  $dR_t$  будут определенными функциями  $\hat{X}_t, R_t$  и  $t$ . Для гауссовских МСтС ( $\psi'' = 0$  и  $\psi_1'' = 0$ ), пользуясь формулой (7), можно показать, что фильтрационные уравнения будут представлять собой стохастические дифференциальные уравнения, определяющие  $\hat{X}_t$  и  $R_t$ :

$$\left. \begin{aligned} d\hat{X}_t &= f(\hat{X}_t, Y_t, R_t, \Theta, t) dt + \\ &+ h(\hat{X}_t, Y_t, R_t, \Theta, t) \times \\ &\times [dY_t - f^{(1)}(\hat{X}_t, Y_t, R_t, \Theta, t) dt]; \\ dR_t &= \left\{ f^{(2)}(\hat{X}_t, Y_t, R_t, \Theta, t) - \right. \\ &- h(\hat{X}_t, Y_t, R_t, \Theta, t) (\psi_1 \nu \psi_1^T)(Y_t, \Theta, t) \times \\ &\times h(\hat{X}_t, Y_t, R_t, \Theta, t)^T \left. \right\} dt + \\ &+ \sum_{r=1}^{n_y} \rho_r(\hat{X}_t, Y_t, R_t, \Theta, t) \times \\ &\times [dY_r - f_r^{(1)}(\hat{X}_t, Y_t, R_t, \Theta, t) dt]. \end{aligned} \right\} \quad (11)$$

Здесь введены следующие обозначения:

$$\begin{aligned} f &= f(\hat{X}_t, Y_t, R_t, t) = \\ &= M_{\Delta^x}^N [\varphi(Y_t, X_t, \Theta, t)] = \hat{\varphi}; \quad (12) \end{aligned}$$

$$\begin{aligned} f^{(1)} &= f^{(1)}(\hat{X}_t, Y_t, R_t, \Theta, t) = \\ &= \left\{ f_r^{(1)}(\hat{X}_t, Y_t, R_t, \Theta, t) \right\} = \\ &= M_{\Delta^x}^N [\varphi_1(Y_t, X, \Theta, t)] = \hat{\varphi}_1^T; \quad (13) \end{aligned}$$

$$\begin{aligned} h &= h(\hat{X}_t, Y_t, R_t, t) = \\ &= \left\{ M_{\Delta^x}^N \left[ \hat{X}_t \varphi_1(Y_t, X_t, \Theta, t)^T + \right. \right. \\ &+ \left. \left. \psi \nu \psi_1^T(Y_t, X_t, \Theta, t) \right] - \right. \\ &- \left. \hat{X}_t f^{(1)}(\hat{X}_t, Y_t, R_t, \Theta, t)^T \right\} \times \\ &\times (\psi_1 \nu \psi_1^T)^{-1}(Y_t, \Theta, t); \quad (14) \end{aligned}$$

$$\begin{aligned} f^{(2)} &= f^{(2)}(\hat{X}_t, Y_t, R_t, \Theta, t) = \\ &= M_{\Delta^x}^N \left\{ (X_t - \hat{X}_t) \varphi(Y_t, X_t, \Theta, t)^T + \right. \\ &+ \varphi(Y_t, X_t, \Theta, t) (X_t^T - \hat{X}_t^T) + \\ &+ \left. \psi \nu \psi^T(Y_t, X_t, \Theta, t) \right\}; \quad (15) \end{aligned}$$

$$\begin{aligned} \rho_r &= \rho_r(\hat{X}_t, Y_t, R_t, \Theta, t) = \\ &= M_{\Delta^x}^N \left\{ (X_t - \hat{X}_t) (X_t^T - \hat{X}_t^T) \times \right. \\ &\times a_r(Y_t, X_t, \Theta, t) + (X_t - \hat{X}_t) \times \\ &\times b_r(Y_t, X_t, \Theta, t)^T (X_t^T - \hat{X}_t^T) + \\ &+ \left. b_r(Y_t, X_t, \Theta, t) (X_t^T - \hat{X}_t^T) \right\} \\ &(r = 1, \dots, n_y), \quad (16) \end{aligned}$$

где  $a_r$  —  $r$ -й элемент матрицы-строки  $(\varphi_1^T - \hat{\varphi}_1^T)(\psi_1 \nu \psi_1^T)$ ;  $b_{kr}$  — элемент  $k$ -й строки и  $r$ -го столбца  $(\psi \nu \psi_1^T)(\psi_1 \nu \psi_1^T)^{-1}$ ;  $b_r$  —  $r$ -й столбец матрицы  $(\psi \nu \psi_1^T)(\psi_1 \nu \psi_1^T)$ ,  $b_r = [b_{1r} \dots b_{n_x r}]^T$ .

Число уравнений для апостериорного одномерного распределения определяется по формуле:

$$Q_{\text{МНА}} = n_x + \frac{n_x(n_x + 1)}{2} = \frac{n_x(n_x + 3)}{2}.$$

За начальные значения  $\hat{X}_t$  и  $R_t$  при интегрировании уравнений (11), естественно, следует принять условные математическое ожидание и ковариационную матрицу величины  $X_0$  относительно  $Y_0$ :

$$\left. \begin{aligned} \hat{X}_0 &= M_{\Delta^x}^N [X_0 | Y_0]; \\ R_0 &= M_{\Delta^x}^N \left[ (X_0 - \hat{X}_0) (X_0^T - \hat{X}_0^T) | Y_0 \right]. \end{aligned} \right\} \quad (17)$$

Если нет информации об условном распределении  $X_0$  относительно  $Y_0$ , то начальные условия можно взять в виде:  $\hat{X}_0 = M X_0$  и  $R_0 = M(X_0 - M X_0)(X_0^T - M X_0^T)$ . Если же и об этих величинах

нах нет никакой информации, то начальные значения  $\hat{X}_t$  и  $R_t$  приходится задавать произвольно.

Таким образом, имеем утверждение.

**Теорема 2.3.** Пусть МСтС (3) и (4) — гауссовская ( $\psi'' = 0$ ), выполнены условия существования и единственности решения, а матрица  $\sigma_1 = \psi_1 \nu \psi_1^T$  не вырождена. Тогда алгоритм СОФ на основе МНА определяется уравнениями (11) и (17) при условиях ограниченности функций (12)–(16).

В основе соответствующей теоремы для МСтС (3) и (4) с пуассоновскими шумами в (3) и невырожденной матрицей  $\sigma = \psi_1 \nu \psi_1^T$  лежат уравнения теоремы 2.3. При этом, если учесть формулу (7), потребуется ограниченность функций  $f$ ,  $f^{(1)}$ ,  $h$  и  $\rho_r$ , определяемых (12)–(16), и функции

$$\bar{f}^{(2)} = f^{(2)} + M_{\Delta x}^N \left[ \int_{R_0^g} \psi'' \psi''^T \nu_P(\Theta, t, dv) \right]. \quad (18)$$

Таким образом, приходим к утверждению.

**Теорема 2.4.** Пусть МСтС (3), (4) удовлетворяет условиям существования и единственности решения, а матрица  $\sigma = \psi_1 \nu \psi_1^T$  не вырождена. Тогда СОФ на основе МНА определяется уравнениями (11) и (17) при условиях ограниченности функций  $f$ ,  $f^{(1)}$ ,  $\bar{f}^{(2)}$ ,  $h$  и  $\rho_r$ .

**Замечание 2.3.** Для гладких функций  $\varphi$ ,  $\varphi_1$ ,  $\psi'$  и  $\psi'_1$  и гауссовских МСтС (3) и (4) СОФ на основе МНА называется гауссовским фильтром [5–7].

## 2.5 Квазилинейный субоптимальный фильтр на основе метода статистической линеаризации

Для МСтС (1), (2) при  $\psi' = \psi'(\Theta, t)$ ,  $\psi'' = \psi''(\Theta, t, v)$ ,  $\psi'_1 = \psi'_1(\Theta, t)$  и  $\psi''_1 = \psi''_1(\Theta, t, v)$  (т. е. с аддитивными винеровскими и пуассоновскими шумами) уравнения НСОФ получаются проще, если нелинейные функции  $\varphi$  и  $\varphi_1$  на основе гауссовского (нормального) распределения заменить на статистически линеаризованные [4, 5]:

$$\left. \begin{aligned} \varphi &= \varphi(X_t, Y_t, \Theta, t) \approx \\ &\approx \varphi_0 + k_x^\varphi (X_t - m_t^x) + k_y^\varphi (Y_t - m_t^y); \\ \varphi_1 &= \varphi_1(X_t, Y_t, \Theta, t) \approx \\ &\approx \varphi_{10} + k_x^{\varphi_1} (X_t - m_t^x) + k_y^{\varphi_1} (Y_t - m_t^y), \end{aligned} \right\} \quad (19)$$

а затем использовать уравнения линейной фильтрации [4]. Входящие в (19) коэффициенты статистической линеаризации зависят от математических ожиданий, дисперсий и ковариаций:

$$Z_t = \begin{bmatrix} X_t \\ Y_t \end{bmatrix}; \quad m_t^z = \begin{bmatrix} m_t^x \\ m_t^y \end{bmatrix}; \quad K_t^z = \begin{bmatrix} K_t^{xx} & K_t^{xy} \\ K_t^{xy} & K_t^{yy} \end{bmatrix}.$$

Они определяются из уравнений

$$\left. \begin{aligned} \dot{Z}_t &= A^z Z_t + A_0^z + B_0^z V, \quad V = \dot{W}; \\ \dot{m}_t^z &= A^z m_t^z + A_0^z, \quad m_{t_0}^z = m_0^z; \\ \dot{K}_t^z &= B^z K_t^z + K_t^z (B^z)^T + B_0^z \nu^m (B_0^z)^T; \\ &K_{t_0}^z = K_0^z. \end{aligned} \right\} \quad (20)$$

Здесь введены следующие обозначения:

$$A_0^z = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix}; \quad A^z = \begin{bmatrix} a_1 & a \\ b_1 & b \end{bmatrix}; \quad B_0^z = \begin{bmatrix} \bar{\psi} \\ \bar{\psi}_1 \end{bmatrix},$$

где

$$\left. \begin{aligned} a_0 &= \varphi_0 - k_x^\varphi m_t^x - k_y^\varphi m_t^y; \quad a_1 = k_x^\varphi; \quad a = k_y^\varphi; \\ b_0 &= \varphi_0 - k_x^{\varphi_1} m_t^x - k_y^{\varphi_1} m_t^y; \quad b_1 = k_x^{\varphi_1}; \quad b = k_y^{\varphi_1}; \\ \psi dW_0 + \int_{R_0^g} \psi'' P^0(dt, dv) &= \bar{\psi} dW_0; \\ \psi'_1 dW_0 + \int_{R_0^g} \psi''_1 P^0(dt, dv) &= \bar{\psi}_1 dW. \end{aligned} \right\} \quad (21)$$

Тогда уравнения квазилинейного НСОФ будут иметь вид:

$$\dot{\hat{X}}_t = a Y_t + a_1 \hat{X}_t + a_0 + \beta_t \left[ Z_t - (b Y_t + b_1 \hat{X}_t + b_0) \right]; \quad (22)$$

$$\beta_t = (R_t b_1^T + \bar{\psi} \nu^W \bar{\psi}_1^T) (\bar{\psi}_1 \nu^W \bar{\psi}_1^T)^{-1}; \quad (23)$$

$$\dot{R}_t = a_1 R_t + R_t a_1^T + \bar{\psi} \nu^W \bar{\psi}^T - (R_t b_1^T + \bar{\psi} \nu^W \bar{\psi}_1^T) \times \\ \times (\bar{\psi}_1 \nu^W \bar{\psi}_1^T)^{-1} (b_1 R_t + \bar{\psi}_1 \nu^W \bar{\psi}^T), \quad (24)$$

где  $\nu^W$  — интенсивность СтП с независимыми приращениями, состоящего из винеровской и пуассоновской частей (21).

**Теорема 2.5.** Пусть МСтС (1), (2) содержит только аддитивные винеровские и пуассоновские шумы и допускает замену статистически линеаризованной системой, а матрица  $\sigma_1 = \bar{\psi}_1 \nu^W \bar{\psi}_1^T$  не вырождена. Тогда в основе алгоритма квазилинейного НСОФ лежат уравнения (22)–(24) при начальных условиях (17).

**Замечание 2.4.** Уравнения теоремы 2.3 сохраняют свой вид, если коэффициенты статистической линеаризации в (19) вычислять для известного эквивалентного (негауссовского) распределения. При этом уравнения (20), (22)–(24), как известно [6], имеют место для любого негауссовского СтП.

**Замечание 2.5.** Из теоремы 2.3 немедленно следуют уравнения НСОФ для фильтрации стационарных процессов в установившемся режиме для стационарных МСтС, если приравнять нулю правые части уравнений (20), (22) и (24).

### 3 Ортогональные субоптимальные фильтры

#### 3.1 Гауссовские шумы

При аппроксимации апостериорной одномерной плотности отрезком ее ортогонального разложения [1, 2]:

$$p_t(x, \Theta) = p^*(x; \Theta, \vartheta) = w(x; \Theta) \left[ 1 + \sum_{k=3}^N \sum_{|\nu|=k} c_\nu p_\nu(x) \right] \quad (25)$$

естественно принять за параметры, образующие вектор  $\vartheta$ , апостериорные математическое ожидание  $\hat{X}_t$ , ковариационную матрицу  $R_t$  вектора  $X_t$ , а также коэффициенты ортогонального разложения (КОР)  $c_\nu$  ( $|\nu| = 3, \dots, N$ ). Здесь КОР определяется формулой:

$$c_\kappa = \left[ q_\kappa \left( \frac{\partial}{i\partial\lambda} \right) g_t(\lambda, \Theta) \right]_{\lambda=0}. \quad (26)$$

Заметим, что полином  $q_\kappa$  зависит от  $\hat{X}_t$  и  $R_t$ .

На основе (7) и (8) для гауссовской МСтС (3) и (4) при  $\psi'' = 0$  получим, что ОСОФ определяется следующими уравнениями:

$$\left. \begin{aligned} d\hat{X}_t &= f dt + h (dY_t - f^{(1)} dt); \\ dR_t &= (f^{(2)} - h\psi_1\nu\psi_1^T h^T) dt + \\ &+ \sum_{r=1}^{n_y} \rho_r (dY_r - f_r^{(1)} dt). \end{aligned} \right\} \quad (27)$$

Здесь введены обозначения:

$$\left. \begin{aligned} f &= f(Y_t, \vartheta, \Theta, t) = M_{\Delta x}^* [\varphi(Y_t, X, \Theta, t)]; \\ f^{(1)} &= \{f_r^{(1)}\} = f^{(1)}(Y_t, \vartheta, \Theta, t) = \\ &= M_{\Delta x}^* [\varphi_1(Y_t, X, \Theta, t)]; \\ f^{(2)} &= f^{(2)}(Y_t, \vartheta, \Theta, t) = \\ &= M_{\Delta x}^* \left[ (X - \hat{X}_t) \varphi(Y_t, X, \Theta, t)^T + \right. \\ &+ \varphi(Y_t, X, \Theta, t) (X^T - \hat{X}_t^T) + \\ &+ (\psi\nu\psi^T)(Y_t, X, \Theta, t) \left. \right]; \end{aligned} \right\} \quad (28)$$

$$\begin{aligned} h &= h(Y_t, \vartheta, \Theta, t) = \left\{ M_{\Delta x}^* \left[ X\varphi_1(Y_t, X, \Theta, t)^T + \right. \right. \\ &+ (\psi\nu\psi_1^T)(Y_t, X, \Theta, t) \left. \right] - \\ &\left. - \hat{X}_t f^{(1)T} \right\} (\psi_1\nu\psi_1^T)^{-1}(Y_t, \Theta, t); \end{aligned} \quad (30)$$

$$\begin{aligned} \rho_r &= \rho_r(Y_t, \vartheta, \Theta, t) = \\ &= M_{\Delta x}^* \left[ (X - \hat{X}_t) (X^T - \hat{X}_t^T) a_r(Y_t, X, \Theta, t) + \right. \\ &+ (X - \hat{X}_t) b_r(Y_t, X, \Theta, t)^T + \\ &+ b_r(Y_t, X, \Theta, t) (X^T - \hat{X}_t^T) \left. \right] \\ &(r = 1, \dots, n_y). \end{aligned} \quad (31)$$

Далее перепишем (27) покоординатно:

$$\left. \begin{aligned} d\hat{X}_s &= f_s dt + h_s (dY_t - f^{(1)} dt) = \\ &= A^{\hat{X}_s} dt + B^{\hat{X}_s} dY_t \quad (s = 1, \dots, n_x); \\ dR_{sq} &= (f_{sq}^{(2)} - h_s \psi_1 \nu \psi_1^T h_q^T) dt + \\ &+ \eta_{sq} (dY_t - f^{(1)} dt) = A^{R_{sq}} dt + B^{R_{sq}} dY_t, \end{aligned} \right\} \quad (32)$$

где  $\hat{X}_s(t_0) = X_{s0}$ ;  $R_{sq}(t_0) = R_{sq0}$ ;  $s, q = 1, \dots, n_x$ ;  $\eta_{sq}$  — матрица-строка, элементами которой служат соответствующие элементы матрицы  $\rho_1, \dots, \rho_{n_1}$ :

$$\eta_{sq} = \eta_{e_s + e_q} = [\rho_{1sq} \dots \rho_{msq}] \quad (s, q = 1, \dots, n_x).$$

Здесь и далее для краткости индекс  $t$  сохраним только у  $Y_t$ . По формуле дифференцирования Ито для винеровского СтП [3, 5], учитывая (32), находим в силу (26) стохастический дифференциал:

$$\begin{aligned} dc_\kappa &= \left[ d \left\{ q_\kappa \left( \frac{\partial}{i\partial\lambda} \right) g_t(\lambda, \Theta) \right\} \right]_{\lambda=0} = \\ &= \sum_{s=1}^{n_x} \left[ \partial q_\kappa \left( \frac{\partial}{i\partial\lambda} \right) \partial \hat{X}_s \cdot g_t(\lambda, \Theta) \right]_{\lambda=0} d\hat{X}_s + \\ &+ \sum_{s,u=1}^{n_x} \left[ \partial q_\kappa \left( \frac{\partial}{i\partial\lambda} \right) \partial R_{su} \cdot g_t(\lambda, \Theta) \right]_{\lambda=0} dR_{su} + \\ &+ \left[ q_\kappa \left( \frac{\partial}{i\partial\lambda} \right) dg_t(\lambda, \Theta) \right]_{\lambda=0} + \\ &+ \left\{ \frac{1}{2} \sum_{s,u=1}^{n_x} \left[ \frac{\partial^2 q_\kappa (\partial/(i\partial\lambda)) \cdot g_t(\lambda, \Theta)}{\partial \hat{X}_s \partial \hat{X}_u} \right]_{\lambda=0} h_s \psi_1 \nu \psi_1^T h_u^T + \right. \\ &+ \frac{1}{2} \sum_{s,u,k,l=1}^{n_x} \left[ \frac{\partial^2 q_\kappa (\partial/(i\partial\lambda)) \cdot g_t(\lambda, \Theta)}{\partial R_{su} \partial R_{kl}} \right]_{\lambda=0} \eta_{su} \psi_1 \nu \psi_1^T \eta_{kl}^T + \\ &\left. + \sum_{s,k,l=1}^{n_x} \left[ \frac{\partial^2 q_\kappa (\partial/(i\partial\lambda)) \cdot g_t(\lambda, \Theta)}{\partial \hat{X}_s \partial R_{kl}} \right]_{\lambda=0} h_s \psi_1 \nu \psi_1^T \eta_{kl}^T \right\} dt. \end{aligned}$$

Подставив сюда выражения (32) и (7) дифференциалов  $d\hat{X}_s$ ,  $dR_{sq}$  и  $dg_t(\lambda, \Theta)$  и вспомнив, что для любого полинома  $P(x)$   $P[(\partial/(i\partial\lambda))g_t(\lambda)]_{\lambda=0} = P(\alpha)$ , получаем стохастические дифференциальные уравнения:

$$\begin{aligned}
dc_\kappa = & \left\{ F_\kappa + \sum_{s=1}^{n_x} \frac{\partial q_\kappa(\alpha)}{\partial \hat{X}_s} f_s + \right. \\
& + \sum_{s,u=1}^{n_x} \frac{\partial q_\kappa(\alpha)}{\partial R_{su}} \left( f_{su}^{(2)} - h_s \psi_1 \nu \psi_1^T h_u^T \right) + \\
& + \frac{1}{2} \sum_{s,u=1}^{n_x} \frac{\partial^2 q_\kappa(\alpha)}{\partial \hat{X}_s \partial \hat{X}_u} h_s \psi_1 \nu \psi_1^T h_u^T + \\
& + \frac{1}{2} \sum_{s,u,k,l=1}^{n_x} \frac{\partial^2 q_\kappa(\alpha)}{\partial R_{su} \partial R_{kl}} \eta_{su} \psi_1 \nu \psi_1^T \eta_{kl}^T + \\
& \left. + \sum_{s,k,l=1}^{n_x} \frac{\partial^2 q_\kappa(\alpha)}{\partial \hat{X}_s \partial R_{kl}} h_s \psi_1 \nu \psi_1^T \eta_{kl}^T \right\} dt + \\
& + \left\{ H_\kappa + \sum_{s=1}^{n_x} \frac{\partial q_\kappa(\alpha)}{\partial \hat{X}_s} h_s + \sum_{s,u=1}^{n_x} \frac{\partial q_\kappa(\alpha)}{\partial R_{su}} \eta_{su} \right\} \times \\
& \times \left( dY_t - f^{(1)} dt \right) = A^{c_\kappa} dt + B^{c_\kappa} dY_t, \quad c_\kappa(t_0) = c_{\kappa 0} \\
& \quad (|\kappa| = 3, \dots, N). \quad (33)
\end{aligned}$$

Здесь в дополнение к прежним обозначениям принято:

$$\left. \begin{aligned}
F_\kappa &= F_\kappa(Y_t, \Theta, \vartheta, t) = \\
&= \sum_{s=1}^{n_x} M_{\Delta^x}^{p*} \left[ \varphi_s(Y_t, X, \Theta, t) \frac{\partial q_\kappa(X)}{\partial X_s} \right] + \\
&+ \frac{1}{2} \sum_{s,u=1}^{n_x} M_{\Delta^x}^{p*} \left[ \sigma_{su}(Y_t, X, \Theta, t) \frac{\partial^2 q_\kappa(X)}{\partial X_s \partial X_u} \right]; \\
H_\kappa &= H_\kappa(Y_t, \vartheta, \Theta, t) = \\
&= \left\{ M_{\Delta^x}^{p*} \left[ \varphi_1(Y_t, X, \Theta, t)^T q_\kappa(X) \right] + \right. \\
&+ \sum_{s=1}^{n_x} M_{\Delta^x}^{p*} \left[ (\psi \nu \psi_1^T)_s(Y_t, X, \Theta, t) \frac{\partial q_\kappa(X)}{\partial X_s} \right] - \\
&\left. - c_\kappa f^{(1)T} \right\} (\psi_1 \nu \psi_1^T)^{-1}(Y_t, \Theta, t),
\end{aligned} \right\} (34)$$

где через  $(\psi \nu \psi_1^T)_s$  обозначена  $s$ -я строка матрицы  $\psi \nu \psi_1^T$ ;  $\sigma = \psi \nu \psi_1^T = \{\sigma_{su}\}$ .

Функции  $f_s$ ,  $f^{(1)}$ ,  $f_{su}^{(2)}$ ,  $h_s$ ,  $\eta_{su}$ ,  $F_\kappa$  и  $H_\kappa$  в уравнениях (32) и (33) представляют собой линейные комбинации величин  $c_\nu$  ( $|\nu| = 3, \dots, N$ ) с коэффициентами, зависящими от  $\hat{X}_t$  и  $R_t$ . Величины  $\partial q_\kappa(\alpha)/\partial \hat{X}_s$ ,  $\partial q_\kappa(\alpha)/\partial R_{su}$ ,  $\partial^2 q_\kappa(\alpha)/(\partial \hat{X}_s \partial \hat{X}_u)$ ,  $\partial^2 q_\kappa(\alpha)/(\partial R_{su} \partial R_{kl})$  и  $\partial^2 q_\kappa(\alpha)/(\partial \hat{X}_s \partial R_{kl})$  после замены моментов их выражениями через  $c_\nu$  тоже будут линейными комбинациями величин  $c_\nu$  с коэффициентами, зависящими от  $\hat{X}_t$  и  $R_t$ .

В частном случае разложений (25) по полиномам Эрмита КОР  $c_\nu$  представляют собой квазимоменты (КМ). В этом случае, как показано в [6, 7], для производных полиномов Эрмита  $G_\nu$ , формулы (34) приводятся к виду:

$$\left. \begin{aligned}
F_\kappa &= \\
&= \sum_{s=1}^{n_x} \kappa_s M_{\Delta^x}^{p*} [\varphi_s(Y_t, X, \Theta, t) G_{\kappa-e_s}(X-m)] + \\
&+ \frac{1}{2} \sum_{s=1}^{n_x} \kappa_s (\kappa_s - 1) M_{\Delta^x}^{p*} [\sigma_{ss}(Y_t, X, \Theta, t) \times \\
&\quad \times G_{\kappa-2e_s}(X-m)] + \sum_{u=2}^{n_x} \sum_{s=1}^{u-1} \kappa_s \kappa_u \times \\
&\quad \times M_{\Delta^x}^{p*} [\sigma_{su}(Y_t, X, \Theta, t) G_{\kappa-e_s-e_u}(X-m)]; \\
H_\kappa &= \left\{ M_{\Delta^x}^{p*} [\varphi_1(Y_t, X, \Theta, t)^T G_\kappa(X-m)] + \right. \\
&+ \sum_{s=1}^{n_x} \kappa_s M_{\Delta^x}^{p*} [(\psi \nu \psi_1^T)(Y_t, X, t) \times \\
&\quad \times G_{\kappa-e_s}(X-m)] - \\
&\left. - f^{(1)T} c_\kappa \right\} (\psi_1 \nu \psi_1^T)^{-1}(Y_t, t),
\end{aligned} \right\} (35)$$

где

$$\begin{aligned}
\frac{\partial q_\kappa(\alpha)}{\partial \hat{X}_s} &= -\kappa_s c_{\kappa-e_s}; \\
\frac{\partial q_\kappa(\alpha)}{\partial R_{ss}} &= -\frac{1}{2} \frac{\partial q_\kappa(\alpha)}{\partial \hat{X}_s^2} = -\frac{1}{2} \kappa_s (\kappa_s - 1) c_{\kappa-2e_s}; \\
\frac{\partial q_\kappa(\alpha)}{\partial R_{su}} &= -\frac{\partial^2 q_\kappa(\alpha)}{\partial \hat{X}_s \partial \hat{X}_u} = -\kappa_s \kappa_u c_{\kappa-e_s-e_u}; \\
\frac{\partial^2 q_\kappa(\alpha)}{\partial R_{ss}^2} &= \frac{1}{4} \kappa_s (\kappa_s - 1) (\kappa_s - 2) (\kappa_s - 3) c_{\kappa-4e_s}; \\
\frac{\partial^2 q_\kappa(\alpha)}{\partial R_{ss} \partial R_{kk}} &= \frac{1}{4} \kappa_s (\kappa_s - 1) \kappa_k (\kappa_s - 1) c_{\kappa-2e_s-2e_k}; \\
\frac{\partial^2 q_\kappa(\alpha)}{\partial R_{ss} \partial R_{sl}} &= \frac{1}{2} \kappa_s (\kappa_s - 1) (\kappa_s - 2) \kappa_l c_{\kappa-3e_s-e_l}; \\
\frac{\partial^2 q_\kappa(\alpha)}{\partial R_{ss} \partial R_{kl}} &= \frac{1}{2} \kappa_s (\kappa_s - 1) \kappa_k \kappa_l c_{\kappa-2e_s-e_k-e_l}; \\
\frac{\partial^2 q_\kappa(\alpha)}{\partial R_{su} \partial R_{sl}} &= \kappa_s (\kappa_s - 1) \kappa_u \kappa_l c_{\kappa-2e_s-e_u-e_l}; \\
\frac{\partial^2 q_\kappa(\alpha)}{\partial R_{su} \partial R_{kl}} &= \kappa_s \kappa_u \kappa_k \kappa_l c_{\kappa-e_s-e_u-e_k-e_l}; \\
\frac{\partial^2 q_\kappa(\alpha)}{\partial \hat{X}_s \partial R_{ss}} &= \frac{1}{2} \kappa_s (\kappa_s - 1) (\kappa_s - 2) c_{\kappa-3e_s}; \\
\frac{\partial^2 q_\kappa(\alpha)}{\partial \hat{X}_s \partial R_{sl}} &= \kappa_s (\kappa_s - 1) \kappa_l c_{\kappa-2e_s-e_l}; \\
\frac{\partial^2 q_\kappa(\alpha)}{\partial \hat{X}_s \partial R_{kk}} &= \frac{1}{2} \kappa_s \kappa_k (\kappa_k - 1) c_{\kappa-e_s-2e_k}; \\
\frac{\partial^2 q_\kappa(\alpha)}{\partial \hat{X}_s \partial R_{kl}} &= \kappa_s \kappa_k \kappa_l c_{\kappa-e_s-e_k-e_l}.
\end{aligned}$$

Таким образом, имеем следующие утверждения.

**Теорема 3.1.** Пусть МСтС (3), (4) — гауссовская ( $\psi'' = 0$ ), выполнены условия существования и единственности решения, а матрица  $\sigma_1 = \psi_1 \nu \psi_1^T$  не вырождена. Тогда в основе алгоритма ОСОФ по МОР лежат уравнения (25), (32) и (33) при условии ограниченности функций (34).

**Теорема 3.2.** В условиях теоремы 3.1 алгоритм ОСОФ по МКМ определяется уравнениями (25), (32) и (33) при условии ограниченности функций (35).

### 3.2 Негауссовские шумы

Пользуясь формулами (7) и (8), устанавливаем, что наличие пуассоновского шума влияет только на функцию  $f^{(2)}$ . В результате заменим  $f^{(2)}$  в (29) на  $\tilde{f}^{(2)}$  согласно (18) и придем к следующим утверждениям.

**Теорема 3.3.** Пусть для МСтС (3), (4) выполнены условия существования и единственности решения, а матрица  $\sigma_1 = \psi_1 \nu \psi_1^T$  не вырождена. Тогда алгоритм ОСОФ, согласно МОР, задается уравнениями (25), (32) и (33) при условии ограниченности функций  $f$ ,  $f^{(1)}$ ,  $\tilde{f}^{(2)}$ ,  $h$ ,  $\rho_r$ ,  $F_\kappa$  и  $H_\kappa$ , определяемых (28)–(31), (18) и (34).

**Теорема 3.4.** Пусть для МСтС (3), (4) выполнены условия существования и единственности решения, а матрица  $\sigma_1 = \psi_1 \nu \psi_1^T$  не вырождена. Тогда алгоритм ОСОФ, согласно МКМ, задается уравнениями (25), (32) и (33) при условии ограниченности функций  $f$ ,  $f^{(1)}$ ,  $\tilde{f}^{(2)}$ ,  $h$ ,  $\rho_r$ ,  $F_\kappa$  и  $H_\kappa$ , определяемых (28)–(31), (18) и (35).

## 4 Точность и чувствительность ортогонального субоптимального фильтра

Точность СОФ на базе МНА (МСЛ) оценивается на основе ОСОФ по МОР или МКМ путем удержания конечного числа членов в разложении (25).

Применяя методы теории чувствительности [8, 9] для приближенного анализа фильтрационных уравнений в теоремах 3.1–3.4 и учитывая случайность параметров  $\Theta$ , придем к следующим уравнениям для функций чувствительности первого порядка:

$$\begin{aligned} d\nabla^\Theta \hat{X}_s &= \nabla^\Theta A^{\hat{X}_s} dt + \nabla^\Theta B^{\hat{X}_s} dY_t, \\ \nabla^\Theta B^{\hat{X}_s}(t_0) &= 0; \\ d\nabla^\Theta R_{sq} &= \nabla^\Theta A^{R_{sq}} dt + \nabla^\Theta B^{R_{sq}} dY_t, \\ \nabla^\Theta R_{sq}(t_0) &= 0; \\ d\nabla^\Theta c_\kappa &= \nabla^\Theta A^{c_\kappa} dt + \nabla^\Theta B^{c_\kappa} dY_t, \quad \nabla^\Theta c_\kappa(t_0) = 0. \end{aligned}$$

Здесь вычисление взятия производных ведется по всем входящим переменным, а коэффициенты чувствительности вычисляются при  $\Theta = m^\Theta$ . При этом предполагается малость дисперсий по сравнению с их математическими ожиданиями. Очевидно, что при дифференцировании по  $\Theta$  ( $\nabla^\Theta = \partial/\partial\Theta$ ) порядок уравнений возрастает пропорционально числу производных. Аналогично составляются уравнения для элементов матриц вторых функций чувствительности.

Следуя [1, 2], для оценки качества ОСОФ, определяемых теоремами 3.1–3.4 при гауссовских  $\Theta$  с математическим ожиданием  $m^\Theta$  и ковариационной матрицей  $K^\Theta$ , введем условную функцию потерь, допускающую квадратичную аппроксимацию:

$$\begin{aligned} \rho^{\hat{X}_s} &= \rho^{\hat{X}_s}(\Theta) = \rho(m^\Theta) + \sum_{ii=1}^{n^\Theta} \rho'_i(m^\Theta) \Theta_s^0 + \\ &+ \sum_{i,j=1}^{n^\Theta} \rho''_{ij}(m^\Theta) \Theta_i^0 \Theta_j^0, \end{aligned}$$

а также показатель  $\varepsilon$ , определяемый формулой:

$$\varepsilon = \varepsilon_2^{1/4}.$$

Здесь

$$\varepsilon_2 = M^N [\rho(\Theta)^2] - \rho(m^\Theta)^2,$$

где

$$\begin{aligned} M^N [\rho(\Theta)^2] &= \rho(m^\Theta)^2 + \rho'(m^\Theta)^T K^\Theta \rho'(m^\Theta) + \\ &+ 2\rho(m^\Theta) \text{tr} [\rho''(m^\Theta) K^\Theta] + \\ &+ \{\text{tr} [\rho''(m^\Theta) K^\Theta]\}^2 + 2\text{tr} [\rho''(m^\Theta) K^\Theta]^2, \end{aligned}$$

а функции  $\rho'$  и  $\rho''$  по известным формулам [8, 9] определяются на основе первых и вторых функций чувствительности.

Изложенные выше методы синтеза ОСОФ дают принципиальную возможность получить фильтр, близкий к оптимальному, по оценке с любой степенью точности. Чем выше максимальный порядок  $N$  учитываемых моментов, КОР и КМ, тем выше будет точность приближения к оптимальной оценке. Однако число уравнений, определяющих параметры апостериорного одномерного распределения, быстро растет с увеличением числа учитываемых параметров. Соответствующие оценки можно найти в [5–7].

## 5 Тестовый пример

Рассмотрим нелинейную с мультипликативным шумом гауссовскую стохастическую систему [7]:

$$\begin{aligned}\dot{X}_t &= -X_t^3 + X_t V_1(\Theta); \quad Z_t(\Theta) = \dot{Y}_t = X_t + V_2(\Theta); \\ \hat{X}_{t_0} &= MX(t_0); \quad R_{t_0} = DX(t_0).\end{aligned}$$

Здесь предполагается, что интенсивности гауссовских белых шумов  $\nu_1$  и  $\nu_2$  зависят от одного скалярного параметра  $\Theta$ , т. е.  $\nu_{1,2} = \nu_{1,2}(\Theta)$ .

Положим  $\nabla = \nabla^\Theta = (\partial/\partial\Theta)$ . Тогда уравнения точности и чувствительности СОФ на основе МНА имеют вид:

$$\left. \begin{aligned}\dot{\hat{X}}_t &= -\hat{X}_t \left( \hat{X}_t^2 + 3R_t \right) + \nu_2^{-1} R_t \left( Z_t - \hat{X}_t \right); \\ \dot{R}_t &= (\nu_1 - 6R_t) \left( \hat{X}_t^2 + R_t \right) - \nu_2^{-1} R_t^2;\end{aligned} \right\} (36)$$

$$\left. \begin{aligned}\nabla \hat{X}_t &= - \left[ 3\hat{X}_t + (3 + \nu_2^{-1}) \right] \nabla \hat{X}_t + \\ &\quad + \nu_2^{-1} R_t \nabla Z_t - \nu_2^{-2} R_t \left( Z_t - \hat{X}_t \right) \nabla \nu_2, \\ &\quad \nabla \hat{X}_{t_0} = 0; \\ \nabla \dot{R}_t &= 2(\nu_1 - 6R_t) \hat{X}_t \nabla \hat{X}_t + \left[ (\nu_1 - 6R_t) - \right. \\ &\quad \left. - 6 \left( \hat{X}_t^2 + R_t \right) - 2\nu_2^{-1} R_t \right] \nabla R_t + \\ &\quad \left. + \left( \hat{X}_t^2 + R_t \right) \nabla \nu_1 + \nu_2^{-2} R_t \nabla \nu_2, \quad \nabla R_{t_0} = 0.\end{aligned} \right\} (37)$$

Уравнения (36) нелинейны относительно  $\hat{X}_t$  и  $R_t$ , причем фильтр существует только при наличии шума в наблюдениях  $\nu_2 \neq 0$  и произвольном шуме интенсивности  $\nu_1$  в уравнении состояния. Уравнения (37) для первых функций чувствительности являются линейными неоднородными уравнениями вследствие зависимости  $\nu_{1,2} = \nu_{1,2}(\Theta)$  и  $Z_t = Z_t(\Theta)$ .

В соответствии с теоремой 3.1 (с точностью до вероятностных моментов третьего порядка) имеем следующие уравнения ОСОФ:

$$\left. \begin{aligned}\dot{\hat{X}}_t &= -\hat{X}_t \left( \hat{X}_t^2 + 3R_t \right) + \\ &\quad + \nu_2^{-1} R_t \left( Z_t - \hat{X}_t \right) - c_3; \\ \dot{R}_t &= \left[ \nu_1 - 6 \left( \hat{X}_t^2 + R_t \right) \right] R_t - \nu_2^{-1} R_t^2 - \\ &\quad - 6\hat{X}_t c_3 + \nu_2^{-1} c_3 \left( Z_t - \hat{X}_t \right); \\ \dot{c}_3 &= -18\hat{X}_t R_t^2 - 9 \left( \hat{X}_t^2 + 3R_t \right) c_3 + \\ &\quad + 3\nu_1 \left( 2\hat{X}_t R_t + c_3 \right) - \frac{3}{2} \nu_2^{-1} c_3.\end{aligned} \right\} (38)$$

Уравнениям (38) отвечают следующие уравнения для первых функций чувствительности:

$$\begin{aligned}\nabla \hat{X}_t &= - \left( 3\hat{X}_t^2 + 3R_t + \nu_2^{-1} R_t \right) \nabla \hat{X}_t + \\ &\quad + \left[ \nu_2^{-1} \left( Z_t - \hat{X}_t \right) - 3\hat{X}_t \right] \nabla R_t + \nu_2^{-1} R_t \nabla Z_t - \\ &\quad - \nu_2^{-2} R_t \left( Z_t - \hat{X}_t \right) \nabla \nu_2 - \nabla c_3, \quad \nabla \hat{X}(t_0) = 0;\end{aligned}$$

$$\begin{aligned}\nabla \dot{R}_t &= - \left( 12\hat{X}_t R_t + 6c_3 + \nu_2^{-1} c_3 \right) \nabla \hat{X}_t - \\ &\quad - \left( \nu_1 + \nu_2^{-1} R_t + 6\hat{X}_t^2 + 12R_t \right) \nabla R_t + \nu_2^{-1} c_3 \nabla Z_t + \\ &\quad + \left[ \nu_2^{-1} \left( Z_t - \hat{X}_t \right) - 6\hat{X}_t \right] \nabla c_3 + R_t \nabla \nu_1 + \\ &\quad + \nu_2^{-2} \left[ R_t^2 - c_3 \left( Z_t - X_t \right) \right] \nabla \nu_2, \quad \nabla R(t_0) = 0;\end{aligned}$$

$$\begin{aligned}\nabla \dot{c}_3 &= -9 \left( c_3 + 2\hat{X}_t c_3 + 2R_t^2 \right) \nabla \hat{X}_t + \\ &\quad + 3 \left( 2\nu_1 \hat{X}_t - 9c_3 - 2\hat{X}_t R_t \right) \nabla R_t + \\ &\quad + 3 \left( \nu_1 + \frac{9}{2} \nu_2^{-1} - 3\hat{X}_t - 9R_t - 3\hat{X}_t^2 \right) \nabla c_3 + \\ &\quad + 3 \left( 2\hat{X}_t R_t + c_3 \right) \nabla \nu_1 + \frac{3}{2} \nu_2^{-2} \nabla \nu_2, \quad \nabla c_3(t_0) = 0.\end{aligned}$$

Уравнения точности этого ОСОФ нелинейны относительно  $\hat{X}_t$ ,  $R_t$  и  $c_3$  и справедливы только при  $\nu_2 \neq 0$  и произвольном  $\nu_1$ . Уравнения для первых функций чувствительности являются линейными неоднородными уравнениями вследствие зависимости  $Z_t = Z_t(\Theta)$ ,  $\nu_{1,2} = \nu_{1,2}(\Theta)$ .

## 6 Заключение

Для нелинейных дифференциальных систем с винеровскими и пуассоновскими шумами в уравнениях состояния и винеровскими шумами в уравнениях наблюдения, понимаемых в смысле Ито (в том числе и на многообразиях), разработаны методы синтеза ОСОФ на основе аппроксимации апостериорного одномерного распределения МНА и МСЛ, а также МОР и МКМ. Получены уравнения точности и чувствительности СОФ. Приведен тестовый пример для одномерной нелинейной системы с параметрическим шумом. Алгоритмическое обеспечение положено в основу инструментального программного обеспечения в библиотеке «StS-Filter».

В качестве дальнейших обобщений можно рассмотреть дискретные и непрерывно-дискретные МСтС, а также модифицированные СОФ на основе ненормированной апостериорной плотности. Развития требуют также методы экстраполяции и интерполяции в таких системах.

## Литература

1. Сеницын И. Н. Аналитическое моделирование распределений на основе ортогональных разложений в нелинейных стохастических системах на многообразиях // Информатика и её применения, 2015. Т. 9. Вып. 3. С. 17–24.

2. Синицын И. Н. Применение ортогональных разложений для аналитического моделирования многомерных распределений в нелинейных стохастических системах на многообразиях // Системы и средства информатики, 2015. Т. 25. № 3. С. 3–22.
3. Ватанабэ С., Икэда Н. Стохастические дифференциальные уравнения и диффузионные процессы / Пер. с англ. — М.: Наука, 1986. 448 с. (*Watanabe S., Ikeda N. Stochastic differential equations and diffusion processes.* — Amsterdam — Oxford — New York: North-Holland Publishing Co.; Tokyo: Kodansha Ltd., 1981. 476 p.)
4. Королюк В. С., Портенко Н. И., Скороход А. В., Турбин А. Ф. Справочник по теории вероятностей и математической статистике. — М.: Наука, 1985. 640 с.
5. Пугачёв В. С., Синицын И. Н. Теория стохастических систем. — М.: Логос, 2000; 2004. 1000 с.
6. Синицын И. Н. Фильтры Калмана и Пугачёва. — 2-е изд. — М.: Логос, 2007. 776 с.
7. Пугачёв В. С., Синицын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — М.: Наука, 1990. 632 с. (*Pugachev V. S., Sinitsyn I. N. Stochastic differential systems. Analysis and filtering.* — Chichester — New York, NY, USA: John Wiley, 1987. 549 p.)
8. Евланов А. Г., Константинов В. М. Системы со случайными параметрами. — М.: Наука, 1987. 568 с.
9. Справочник по теории автоматического управления / Под ред. А. А. Красовского. — М.: Наука, 1987. 712 с.

*Поступила в редакцию 29.10.15*

---



---

## ORTHOGONAL SUPTIMAL FILTERS FOR NONLINEAR STOCHASTIC SYSTEMS ON MANIFOLDS

I. N. Sinitsyn

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** The authors developed the synthesis theory of suboptimal filters (SOF) based on normal approximation method (NAM), statistical linearization method (SLM), orthogonal expansions method (OEM), and quasi-moment method (QMM) for nonlinear differential stochastic systems on manifolds (MStS) with Wiener and Poisson noises. Exact optimal (for mean square error criteria) equations for MStS with Gaussian noises in observation equations for the one-dimensional *a posteriori* characteristic function are derived. Problems of approximate solving of exact equations are discussed. Accuracy and sensitivity equations are presented. A test example for the nonlinear scalar differential equation with additive and multiplicative noises is given. Some generalizations are mentioned.

**Keywords:** *a posteriori* one-dimensional distribution; coefficient of orthogonal expansion; first sensitivity function; normal approximation method; normal suboptimal filter; orthogonal expansion method; orthogonal suboptimal filter; quasi-moment method; quasi-moment; statistical linearization method; stochastic system on manifolds; suboptimal filter; Wiener white noise

**DOI:** 10.14357/19922264160103

### Acknowledgments

The research was supported by the Russian Foundation for Basic Research (project 15-07-002244).

### References

1. Sinitsyn, I. N. 2015. Analiticheskoe modelirovanie raspredeleniy na osnove ortogonal'nykh razlozheniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraziyakh [Analytical modeling in stochastic systems on manifolds based on orthogonal expansions]. *Informatika i ee Primeneniya — Inform. Appl.* 9(2):17–24.
2. Sinitsyn, I. N. 2015. Primenenie ortogonal'nykh razlozheniy dlya analiticheskogo modelirovaniya mnogomernykh raspredeleniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraziyakh [Applications of orthogonal expansions for analytical modeling of multidimensional distributions in stochastic systems on manifolds]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(3):3–22.
3. Watanabe, S., and N. Ikeda. 1981. *Stochastic differential equations and diffusion processes.* Amsterdam — Oxford — New York: North-Holland Publishing Co.; Tokyo: Kodansha Ltd. 476 p.
4. Korolyuk, V. S., N. I. Portenko, A. V. Skorokhod, and A. F. Turbin, eds. 1985. *Spravochnik po teorii veroyatnostey i matematicheskoy statistike* [Probability theory and

- mathematical statistics: Handbook]. Moscow: Nauka. 640 p.
5. Pugachev, V. S., and I. N. Sinitsyn. 2001. *Stochastic systems. Theory and applications*. Singapore: World Scientific. 908 p.
  6. Sinitsyn, I. N. 2007. *Fil'try Kalmana i Pugacheva* [Kalman and Pugachev filters]. 2nd ed. Moscow: Logos. 776 p.
  7. Pugachev, V. S., and I. N. Sinitsyn. 1987. *Stochastic differential systems. Analysis and filtering*. Chichester – New York, NY: John Wiley. 549 p.
  8. Evlanov, A. G., and V. M. Konstantinov. 1976. *Sistemy so slozhnymi parametrami* [Systems with random parameters]. Moscow: Nauka. 568 p.
  9. Krasovskii, A. A., ed. 1987. *Spravochnik po teorii avtomaticheskogo upravleniya* [Handbook for automatic control]. Moscow: Nauka. 712 p.

Received October 29, 2015

## Contributor

**Sinitsyn Igor N.** (b. 1940) — Doctor of Science in technology, professor, Honored scientist of RF, Head of Department, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sinitsin@dol.ru

# АНАЛИТИЧЕСКОЕ МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЙ В НЕЛИНЕЙНЫХ СТОХАСТИЧЕСКИХ СИСТЕМАХ НА МНОГООБРАЗИЯХ МЕТОДОМ ЭЛЛИпсоИДАЛЬНОЙ АППРОКСИМАЦИИ\*

И. Н. Сеницын<sup>1</sup>, В. И. Сеницын<sup>2</sup>

**Аннотация:** Рассматриваются вопросы оценки точности и чувствительности алгоритмов структурного аналитического моделирования одно- и многомерных распределений методами эллипсоидальной аппроксимации (ЭА) и эллипсоидальной линеаризации (ЭЛ). Разработаны общие алгоритмы эллипсоидального анализа распределений в нелинейных стохастических системах (СтС) на многообразиях (МСтС) с винеровскими и пуассоновскими шумами. Особое внимание уделено алгоритмам для МСтС с аддитивными негауссовскими шумами. Получены уравнения точности и чувствительности к параметрам. В качестве иллюстрации рассмотрена нелинейная двумерная угловая система с аддитивным угловым белым шумом. Составлены уравнения точности и чувствительности, позволяющие оценивать моменты до четвертого порядка включительно. Приведены результаты численных расчетов, показывающие эффективность метода ЭА (ЭЛ) по сравнению с общим методом ортогональных разложений одномерной плотности. Сформулированы некоторые возможные обобщения.

**Ключевые слова:** метод аналитического моделирования (МAM); метод эллипсоидальной аппроксимации (МЭА); метод эллипсоидальной линеаризации (МЭЛ); плотность одно- и многомерного распределения; стохастическая система на многообразиях (МСтС); уравнения точности МЭА и МЭЛ; уравнения чувствительности МЭА и МЭЛ

**DOI:** 10.14357/19922264160104

## 1 Введение

В [1, 2] подробно описаны известные методы аналитического моделирования (МAM) распределений процессов в СтС, описываемых дифференциальными стохастическими уравнениями Ито с винеровскими и пуассоновскими шумами, основанные на параметризации их распределений. Обобщение результатов [1, 2] на случай многоканальных круговых и сферических СтС выполнено в [3–7]. Статья [8] посвящена развитию дискретных методов параметрического статистического и аналитического моделирования в МСтС. В ней рассмотрены уравнения МСтС, приближенные методы статистического моделирования (МСМ) различной точности и методы аналитического моделирования (МAM). Подробно развита нелинейная корреляционная теория МСМ и МAM. В [9, 10] представлены МAM одно- и многомерных распределений в МСтС, основанные на методе ортогональных разложений (МОР) плотностей. Рассмотрим развитие [9, 10] на случай, когда плотности

допускают ЭА [1, 2, 11]. В [9–12] развиты методы и алгоритмы аналитического моделирования гауссовских (нормальных) процессов в МСтС.

Статья включает в себя: введение, 5 разделов, заключение и список литературы. В разд. 2 представлены уравнения МСтС. Общие уравнения метода ЭА (МЭА) даны в разд. 3, а для МСтС с аддитивными шумами разработан метод ЭЛ (МЭЛ) в разд. 4. В разд. 5 получены уравнения точности и чувствительности МЭА и МЭЛ. Раздел 6 содержит иллюстративный пример. В заключении кратко сформулированы основные результаты и указаны некоторые возможные обобщения.

## 2 Уравнения нелинейных стохастических систем на многообразиях

Рассмотрим сначала дифференциальную СтС в конечномерном пространстве следующего вида:

\* Работа выполнена при финансовой поддержке РФФИ (проект 15-07-02244).

<sup>1</sup>Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, sinitsin@dol.ru

<sup>2</sup>Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, vsinitsin@ipiran.ru

$$dY_t = a(Y_t, \Theta, t) dt + b(Y_t, \Theta, t) dW_0 + \int_{R_0^q} c(Y_t, \Theta, t, v) dP^0(\Theta, t, dv), \quad Y(t_0) = Y_0. \quad (1)$$

Здесь  $Y_t$  —  $p$ -мерный вектор состояния,  $Y_t \in \Delta^y$  ( $\Delta^y$  — многообразие состояний);  $\Theta$  — вектор случайных параметров размерности  $p^\Theta$ ;  $a = a(y_t, \Theta, t)$  и  $b = b(y_t, \Theta, t)$  — известные  $(p \times 1)$ -мерная и  $(p \times r)$ -мерная функции вектора  $Y_t$  и времени  $t$ ;  $W_0 = W_0(\Theta, t)$  —  $r$ -мерный винеровский стохастический процесс (СтП) интенсивности  $\nu_0 = \nu_0(\Theta, t)$ ;  $c(y_t, \Theta, t, v)$  —  $(p \times 1)$ -мерная функция  $y_t$ ,  $t$  и вспомогательного  $(q \times 1)$ -мерного параметра  $v$ ;  $\int dP^0(\Theta, t, A)$  — центрированная пуассоновская мера:

$$\int_{\Delta} dP^0(\Theta, t, A) = \int_{\Delta} dP(\Theta, t, A) - \int_{\Delta} \nu_P(\Theta, t, A) dt,$$

где  $\int_{\Delta} dP(\Theta, t, A)$  — число скачков пуассоновского СтП в интервале времени  $\Delta$ ;  $\nu_P(\Theta, t, A)$  — интенсивность пуассоновского СтП  $P(t, A)$ ;  $A$  — некоторое борелевское множество пространства  $R_0^q$  с выколотым началом координат. Интеграл (1) в общем случае распространяется на все пространство  $R_0^q$  с выколотым началом координат. Начальное значение  $Y_0$  вектора  $Y_t$  представляет собой случайную величину (с.в.), не зависящую от приращений винеровского СтП  $W_0(\Theta, t)$  и пуассоновского СтП  $P(\Theta, t, A)$  на интервалах времени  $\Delta = (t_1, t_2]$ , следующих за  $t_0$ ,  $t_0 \leq t_1 \leq t_2$ , для любого множества  $A$ .

Теперь выделим важный частный случай (1):

$$\dot{Y}_t = a(Y_t, \Theta, t) + b_0(\Theta, t)V(\Theta) \quad (2)$$

при следующих допущениях

$$b(Y_t, \Theta, t) = b_0(\Theta, t); \quad c(Y_t, \Theta, t, v) = c_0(v, \Theta);$$

$$V = \dot{W}; \quad W = W_0 + W_P;$$

$$W_P = \int_{R_0^q} c_0(v, \Theta) P^0(\Theta, t, dv);$$

$$\chi^{W_0} = \chi^{W_0}(\mu; \Theta, t) = -\frac{1}{2} \mu^T \nu_0(\Theta, t) \mu;$$

$$\chi^{W_P} = \chi^{W_P}(\mu; \Theta, t) = \int_{R_0^q} \left[ e^{i\lambda^T c_0(v, \Theta)} - 1 - i\mu^T c_0(v, \Theta) \right] \nu_P(\Theta, t, v) dv.$$

Здесь  $\nu_0 = \nu_0(\Theta, t)$  и  $\nu_P = \nu_P(\Theta, t, v)$  — интенсивности СтП и потока скачков, равных  $c_0(v, \Theta)$ ;  $\chi^{W_0}$  и  $\chi^{W_P}$  — логарифмические производные одномерных характеристических функций СтП  $W_0$  и  $W_P$ .

Для вычисления вероятностей событий, связанных с СтП, в прикладных задачах достаточно знания многомерных распределений. Поэтому центральной задачей теории МСтС является задача вероятностного анализа одномерных распределений СтП, удовлетворяющих дифференциальному стохастическому уравнению Ито (1) при соответствующих начальных условиях. В теории МСтС различают два принципиально разных подхода к вычислению распределений. Первый общий подход основан на статистическом моделировании, т. е. на прямом численном решении (1) с последующей статистической обработкой результатов. Второй общий подход основан на теории непрерывных марковских СтП и предполагает аналитическое моделирование, т. е. решение детерминированных уравнений в функциональных пространствах (уравнений Фоккера–Планка–Колмогорова, Феллера–Колмогорова, Пугачёва и др.) для одномерных распределений. В практических задачах часто используют и комбинированные методы. При этом будем предполагать, что существуют одно- и многомерные плотности СтП в МСтС (1) и (2). Достаточные условия их существования можно найти, например, в [1, 2, 12].

### 3 Эллипсоидальная аппроксимация одно- и многомерных распределений в стохастической системе на многообразиях (1)

Как известно [11], для конечномерных МСтС часто оказывается полезной структурная аппроксимация распределений посредством эллипсоидальных распределений. Следуя [11], для структурной аппроксимации плотностей вероятности случайных векторов будем использовать плотности, имеющие эллипсоидальную структуру, т. е. плотности, у которых поверхностями уровней равной вероятности являются подобные концентрические эллипсоиды (эллипсы для двумерных векторов, эллипсоиды для трехмерных векторов, гиперэллипсоиды для векторов размерности больше трех). В частности, эллипсоидальную структуру имеет нормальное распределение в любом конечномерном пространстве. Характерная особенность таких распределений состоит в том, что их плотности вероятности являются функциями положительно определенной квадратичной формы  $u = u(y) = (y^T - m^T)C(y - m)$ , где  $m$  — математическое ожидание случайного вектора  $Y$ ;  $C$  — некоторая положительно определенная матрица.

Для нахождения ЭА плотности вероятности  $r$ -мерного случайного вектора будем пользоваться конечным отрезком разложения по биортонормальной системе полиномов  $\{p_{r,\nu}(u(y)), q_{r,\nu}(u(y))\}$ , которые зависят только от квадратичной формы  $u = u(y)$  и функцией веса для которых служит некоторая плотность вероятности эллипсоидальной структуры  $w(u(y))$ :

$$M_{\Delta y}^w [p_{r,\nu}(u(Y)) q_{r,\mu}(u(Y))] = \delta_{\nu\mu}. \quad (3)$$

Индексы  $\nu$  и  $\mu$  у полиномов означают их степени относительно переменной  $u$ . Конкретный вид и свойства полиномов определены ниже. Однако без потери общности можно принять, что  $q_{r,0}(u) = p_{r,0}(u) = 1$ . Тогда плотность вероятности вектора  $Y$  может быть приближенно представлена в виде:

$$f(y) \approx f^*(u) = w(u) \sum_{\nu=0}^N c_{r,\nu} p_{r,\nu}(u), \quad (4)$$

где коэффициенты  $c_{r,\nu}$  определяются по формуле

$$c_{r,\nu} = M_{\Delta y}^{\text{ЭА}} [q_{r,\nu}(U)], \quad (\nu = 1, \dots, N). \quad (5)$$

Учитывая, что  $p_{r,0}(u)$  и  $q_{r,0}(u)$  — взаимно обратные постоянные (полиномы нулевой степени), то всегда  $c_{r,0} p_{r,0} = 1$ . Поэтому из формулы (4) следует, что

$$f(y) \approx f^*(u) = w(u) \left[ 1 + \sum_{\nu=2}^N c_{r,\nu} p_{r,\nu}(u) \right]. \quad (6)$$

Для приложений большое значение имеет случай, когда за распределение  $w(u)$  выбирается нормальное (гауссовское) распределение

$$w(u) = w(y^T C y) = \frac{1}{\sqrt{(2\pi)^r |K|}} \exp\left(-\frac{y^T K^{-1} y}{2}\right).$$

Учитывая, что  $C = K^{-1}$ , приведем условие биортонормальности (3) к виду:

$$\frac{1}{2^{r/2} \Gamma(r/2)} \int_0^\infty p_{r,\nu}(u) q_{r,\mu}(u) u^{r/2-1} e^{-u/2} du = \delta_{\nu\mu}.$$

Задача выбора системы полиномов  $\{p_{r,\nu}(u), q_{r,\mu}(u)\}$ , используемой при ЭА плотностей (4) и (5), сводится к нахождению биортонормальной системы полиномов, для которой весом служит  $\chi^2$ -распределение с  $r$  степенями свободы. При этом используются следующие формулы:

$$p_{r,\nu}(u) = S_{r,\nu}(u);$$

$$q_{r,\nu}(u) = \frac{(r-2)!!}{(r+2\nu-2)!!(2\nu)!!} S_{r,\nu}(u), \quad r \geq 2,$$

где

$$S_{r,\nu}(u) = S_{r/2-1}^{\nu}(u) = \sum_{\mu=0}^{\nu} (-1)^{\nu+\mu} C_{\nu}^{\mu} \frac{(r+2\nu-2)!!}{(r+2\mu-2)!!} u^{\mu}.$$

При разложении по полиномам  $S_{r,\nu}(u)$  плотности вероятности случайного вектора  $Y$  и всех его возможных проекций согласованы.

Пользуясь обобщенной формулой Ито [1, 2] для дифференциала нелинейной функции  $\varphi = \varphi(Y_t, \Theta, t)$

$$\begin{aligned} d\varphi(Y_t, \Theta, t) = & \left\{ \varphi_t(Y_t, \Theta, t) + \varphi_Y(Y_t, \Theta, t)^T a(Y_t, \Theta, t) + \right. \\ & \left. + \frac{1}{2} \text{tr} \left[ \varphi_{YY}(Y_t, \Theta, t) b(Y_t, \Theta, t) \nu_0(\Theta, t) \times \right. \right. \\ & \left. \left. \times b(Y_t, \Theta, t)^T \right] \right\} dt + \\ & + \int_{R_0^q} \left[ \varphi(Y_t + c(Y_t, \Theta, t, v)) - \varphi(Y_t, \Theta, t) - \right. \\ & \left. - \varphi_Y(Y_t, \Theta, t)^T c(Y_t, \Theta, t, v) \right] \nu_P(\Theta, dv) dt + \\ & + \varphi_Y(Y_t, \Theta, t)^T b(Y_t, \Theta, t) dW_0(\Theta, t) + \\ & + \int_{R_0^q} \left[ \varphi(Y + c(Y, \Theta, t, v), \Theta, t) - \right. \\ & \left. - \varphi(Y_t, \Theta, t) \right] P^0(\Theta, dt, dv), \end{aligned}$$

получаем, что уравнения для нахождения одномерной плотности вероятности  $f_1(y; t)$   $p$ -мерного случайного процесса  $Y_t$ , определяемого стохастическим дифференциальным уравнением Ито (1), с помощью МЭА имеют вид:

$$\begin{aligned} f_1(y; \Theta, t) & \cong f_1^{\text{ЭА}}(\Theta, u) = \\ & = w_1(\Theta, u) \left[ 1 + \sum_{\nu=2}^N c_{p,\nu} p_{p,\nu}(\Theta, u) \right], \quad (7) \end{aligned}$$

где

$$\begin{aligned} u & = (y - m)^T K^{-1} (y - m); \\ c_{p,\nu} & = M_{\Delta y}^{\text{ЭА}} [q_{p,\nu}(U_t)] \quad (\nu = 1, \dots, N). \end{aligned}$$

Здесь  $m = m_t$ ,  $K = K_t$  удовлетворяют уравнениям

$$\begin{aligned} \dot{m} & = A^m(m_t, K_t, \Theta, t) = \varphi_{10}(m_t, K_t, \Theta, t) + \\ & + \sum_{\nu=2}^N c_{p,\nu} \varphi_{1\nu}(m_t, K_t, \Theta, t); \quad (8) \end{aligned}$$

$$\dot{K}_t = A^K(m_t, K_t, \Theta, t) = \varphi_{20}(m_t, K_t, \Theta, t) + \sum_{\nu=2}^N c_{p,\nu} \varphi_{2\nu}(m_t, K_t, \Theta, t), \quad (9)$$

где введены следующие обозначения:

$$\left. \begin{aligned} \varphi_{10}(m_t, K_t, t) &= M_{\Delta y}^{w_1} [a(Y, \Theta, t)]; \\ \varphi_{1\nu}(m_t, K_t, t) &= M_{\Delta y}^{p_{p,\nu} w_1} [a(Y, \Theta, t)]; \\ \varphi_{20}(m_t, K_t, t) &= \\ &= M_{\Delta y}^{w_1} [a(Y, \Theta, t) (Y^T - m_t^T) + \\ &\quad + (Y - m_t) a(Y, \Theta, t)^T + \bar{\sigma}(Y, \Theta, t)]; \\ a_{2\nu}(m_t, K_t, \Theta, t) &= \\ &= M_{\Delta y}^{p_{p,\nu} w_1} [a(Y, \Theta, t) (y^T - m_t^T) + \\ &\quad + (Y - m_t) a(Y, \Theta, t)^T + \bar{\sigma}(Y, \Theta, t)]. \end{aligned} \right\} \quad (10)$$

Коэффициенты  $c_{p,\kappa} = c_{p,\kappa,t}$  определяются уравнением

$$\begin{aligned} \dot{c}_{p,\kappa} &= A^{c_{p,\kappa}}(m_t, K_t, \Theta, t) = \\ &= - \left( \frac{c_{p,\kappa}-1}{2p} - \frac{\kappa c_{p,\kappa}}{p} \right) \text{tr} \left\{ K_t^{-1} \varphi_{20}(m_t, K_t, \Theta, t) + \right. \\ &\quad \left. + K_t^{-1} \sum_{\nu=2}^N c_{p,\nu} \varphi_{2\nu}(m_t, K_t, \Theta, t) \right\} + \\ &\quad + \psi_{\kappa 0}(m_t, K_t, \Theta, t) + \sum_{\nu=2}^N c_{p,\nu} \psi_{\kappa \nu}(m_t, K_t, \Theta, t), \end{aligned} \quad (11)$$

$\kappa = 2, \dots, N,$

где введены следующие обозначения:

$$\begin{aligned} \psi_{\kappa 0}(m_t, K_t, \Theta, t) &= \\ &= M_{\Delta y}^{w_1} \left[ q'_{p,\kappa}(U) \left( 2(Y - m_t)^T K_t^{-1} a(Y, \Theta, t) + \right. \right. \\ &\quad \left. \left. + \text{tr} K_t^{-1} \sigma(Y, \Theta, t) \right) + \right. \\ &\quad \left. + 2q''_{\kappa}(U) (Y - m_t)^T K^{-1} \sigma(Y, \Theta, t) (Y - m_t) + \right. \\ &\quad \left. + \int_{R_0^g} \left\{ q_{p,\kappa} \left[ (Y^T + c^T - m_t^T) K_t^{-1} (Y + c - m_t) \right] - \right. \right. \\ &\quad \left. \left. - q_{p,\kappa}(U) - \right. \right. \\ &\quad \left. \left. - 2q'_{p,\kappa}(U) (Y - m_t)^T K_t^{-1} c \right\} \nu_P(\tau, dv) \right]; \quad (12) \\ \psi_{\kappa \nu}(m_t, K_t, \Theta, t) &= \\ &= M_{\Delta y}^{p_{p,\nu} w_1} \left[ q'_{p,\kappa}(U) \left( 2(Y - m_t)^T K_t^{-1} a(Y, \Theta, t) + \right. \right. \\ &\quad \left. \left. + \text{tr} K_t^{-1} \sigma(Y, \Theta, t) \right) + \right. \\ &\quad \left. + 2q''_{\kappa}(U) (Y - m_t)^T K_t^{-1} \sigma(Y, \Theta, t) K_t^{-1} (Y - m_t) + \right. \\ &\quad \left. + \int_{R_0^g} \left\{ q_{p,\kappa} \left( (Y^T + c^T - m_t^T) K_t^{-1} (Y + c - m_t) \right) - \right. \right. \end{aligned}$$

$$\left. - q_{p,\kappa}(U) - 2q'_{p,\kappa}(U) (Y - m_t)^T K_t^{-1} c \right\} \nu_P(t, dv) \Big]. \quad (13)$$

Таким образом, уравнения (8), (9) и (11) при начальных условиях

$$\left. \begin{aligned} m(\Theta, t_0) &= m_0; \quad K(\Theta, t_0) = K_0; \\ c_{p,\kappa}(\Theta, t_0) &= c_{p,\kappa}^0 \quad (\kappa = 2, \dots, N) \end{aligned} \right\} \quad (14)$$

определяют  $m_t, K_t, c_{p,2}, \dots, c_{p,N}$  как функции времени. Для нахождения величин  $c_{p,\kappa}^0$  следует аппроксимировать плотность начального значения  $Y_0$  вектора состояния системы формулой (7).

Количество уравнений МЭА лишь на  $N/2 - 1$  больше числа уравнений метода нормальной аппроксимации (МНА):  $Q^{\text{МЭА}} = Q^{\text{МНА}} + N/2 - 1$ .

Теперь представим  $n$ -мерную плотность вероятности процесса  $Y_t$  в форме

$$\begin{aligned} f_n(y_1, \dots, y_n; \Theta, t_1, \dots, t_n) &\approx f_n^{\text{ЭА}}(\Theta, u) = \\ &= w_n(\Theta, u) \left[ 1 + \sum_{\nu=2}^N c_{np,\nu} p_{np,\nu}(\Theta, u) \right], \end{aligned} \quad (15)$$

где

$$c_{np,\nu} = M_{\Delta y}^{f_n} [q_{np,\nu}(U_t)].$$

Здесь использованы следующие обозначения:

$$\begin{aligned} U_t &= (\tilde{Y}_n - \tilde{m}_n)^T C_n (\tilde{Y}_n - \tilde{m}_n); \quad C_n = K_n^{-1}; \\ \tilde{Y}_n &= [Y_1^T \ Y_2^T \ \dots \ Y_n^T]^T; \quad Y_k = Y(t_k); \\ \tilde{m}_n &= [m_1^T \ m_2^T \ \dots \ m_n^T]^T; \quad m_k = m(t_k); \\ K_n &= \begin{bmatrix} K(t_1, t_1) & K(t_1, t_2) & \dots & K(t_1, t_n) \\ K(t_2, t_1) & K(t_2, t_2) & \dots & K(t_2, t_n) \\ \vdots & \vdots & \ddots & \vdots \\ K(t_n, t_1) & K(t_n, t_2) & \dots & K(t_n, t_n) \end{bmatrix}, \end{aligned}$$

где  $K(t_i, t_j)$  — ковариационная функция процесса  $Y_t$ ;  $C_n$  — матрица, состоящая из блоков  $C_{kl}^{(n)}$  размерности  $p \times p$ :

$$C_n = \begin{bmatrix} C_{11}^{(n)} & C_{12}^{(n)} & \dots & C_{1n}^{(n)} \\ C_{21}^{(n)} & C_{22}^{(n)} & \dots & C_{2n}^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1}^{(n)} & C_{n2}^{(n)} & \dots & C_{nn}^{(n)} \end{bmatrix};$$

$\bar{C}_k = [C_{k1}^{(n)} \ \dots \ C_{kn}^{(n)}]$  — ее  $k$ -я блочная строка.

В качестве  $w_n(\Theta, u)$  возьмем гауссовскую плотность вероятности

$$w_n(\Theta, u) = [(2\pi)^n |K_n|]^{-1/2} e^{-u/2}.$$

Полиномы  $p_{np,\nu}(u)$  и  $q_{np,\nu}(u)$  определим формулами:

$$\begin{aligned} p_{np,\nu}(u) &= S_{np,\nu}(u) = \\ &= \sum_{\mu=0}^{\nu} (-1)^{\nu-\mu} C_{\nu}^{\mu} \frac{(np+2\nu-2)!!}{(np+2\mu-2)!!} u^{\mu}; \\ q_{np,\nu}(u) &= \frac{(np-2)!!}{(np+2\nu-2)!!(2\nu)!!} p_{np,\nu}(u). \end{aligned}$$

Символ  $n$  в индексе коэффициентов разложения и полиномов означает принадлежность к  $n$ -мерному распределению.

Уравнение для коэффициентов разложения  $c_{np,\kappa}$  примет вид:

$$\begin{aligned} \frac{\partial c_{np,\kappa}}{\partial t_n} &= A^{c_{np,\kappa}}(\tilde{m}_n, C_n, K_n, \Theta, t_n) = \\ &= - \left( \text{frc}_{np,\kappa-1} 2np + \frac{\kappa c_{np,\kappa}}{np} \right) \text{tr} \left[ \frac{\partial K_n}{\partial t_n} C_n \right] + \\ &+ M_{\Delta y}^{\text{ЭА}n} \left\{ 2q'_{np,\kappa}(U) \left( \tilde{Y}_n - \tilde{m}_n \right)^{\text{T}} \bar{C}_n^{\text{T}} a(Y_n, \Theta, t_n) + \right. \\ &\quad + 2q''_{np,\kappa}(U) \left( \tilde{Y}_n - \tilde{m}_n \right)^{\text{T}} \bar{C}_n^{\text{T}} \sigma(Y_n, \Theta, t_n) \times \\ &\quad \times \bar{C}_n \left( \tilde{Y}_n - \tilde{m}_n \right) + q'_{np,\kappa}(U) \text{tr} \left[ C_{nn}^{(n)} \sigma(Y_n, \Theta, t_n) \right] + \\ &\quad \left. + \int_{R_0^a} [q_{np,\kappa}(U) - 2q'_{np,\kappa}(U) \left( \tilde{Y}_n - \tilde{m}_n \right)^{\text{T}} \times \right. \\ &\quad \left. \times \bar{C}_n^{\text{T}} c(Y_n, \Theta, t_n, v) \nu_P(t, dv) \right] \left\} \right. \\ &\quad (\kappa = 2, 3, \dots, N). \quad (16) \end{aligned}$$

Будем требовать согласованности  $n$ - и  $(n-1)$ -мерного распределений при  $t_n = t_{n-1} + \tau_n$ , где  $\tau_n$  — малая величина, достаточная для обеспечения существования матрицы  $C_n$ . Тогда начальные условия для уравнений (16), определяющих коэффициенты разложения, можно записать в виде:

$$\begin{aligned} c_{np,\nu}(t_1, t_2, \dots, t_{n-1}, t_{n-1} + \tau_n) = \\ = c_{(n-1)p,\nu}(t_1, \dots, t_{n-1}). \quad (17) \end{aligned}$$

В случае двумерного распределения  $n = 2$  к уравнениям для коэффициентов разложения (16) следует добавить уравнение для ковариационной функции процесса  $Y(t)$ :

$$\frac{\partial K(t_1, t_2)}{\partial t_2} = M_{\Delta y}^{\text{ЭА}2} \left[ (Y_1 - m_1) a(Y_2, \Theta, t_2)^{\text{T}} \right], \quad (18)$$

где  $M_{\Delta y}^{\text{ЭА}2}$  вычисляются на основе

$$f_2^{\text{ЭА}}(u) = w_2(\Theta, u) \left[ 1 + \sum_{\nu=2}^N c_{2p,\nu}(\Theta, t_1, t_2) p_{2p,\nu}(u) \right];$$

$$\begin{aligned} w_2(\Theta, u) &= [(2\pi)^2 |K_n|]^{-1/2} \times \\ &\quad \times \exp \left\{ - (\tilde{y}_n - \tilde{m}_n)^{\text{T}} K_n^{-1} (\tilde{y}_n - \tilde{m}_n) \right\}; \\ p_{2p,\nu}(u) &= S_{2p,\nu}(u) = \\ &= \sum_{\mu=0}^{\nu} (-1)^{\nu-\mu} C_{\nu}^{\mu} \frac{(2p+2\nu-2)!!}{(2p+2\mu-2)!!} u^{\mu}. \end{aligned}$$

К этому уравнению следует добавить начальное условие  $K(t_1, t_1) = K(t_1)$ . Ввиду вырожденности двумерного распределения при  $t_2 = t_1$  уравнение для ковариационной функции до момента  $t_1 + \tau_2$  интегрируется при вырожденном распределении

$$\begin{aligned} f_2^{\text{ЭА}}(\Theta, u) &\approx \\ &\approx w_1(\Theta, u) \left[ 1 + \sum_{\nu=2}^N c_{1p,\nu}(\Theta, t_1) p_{1p,\nu}(u) \right] \delta(y_1 - y_2), \end{aligned}$$

а начиная с момента  $t_1 + \tau_2$  — совместно с уравнениями (18). Величина  $\tau_2$  определяется в процессе интегрирования, когда матрица  $K_2$  станет невырожденной.

Для приближенного определения МЭА одномерного распределения стационарного в узком смысле процесса в стационарной нелинейной негауссовской СтС (1) следует положить в уравнениях (8), (11) и (16)

$$\dot{m} = 0; \quad \dot{K} = 0; \quad \dot{c}_{p,\kappa} = 0, \quad \kappa = 1, \dots, N. \quad (19)$$

Если полученные таким путем уравнения имеют решение, которое может служить вектором параметров соответствующей ЭА одномерного распределения, то можно предположить, что стационарный в узком смысле процесс в системе существует. В данном случае для определения других многомерных распределений этого стационарного процесса следует заменить в уравнениях (16) производные по  $t_n$  производными по  $\tau_{n-1} = t_n - t_1$ , а начальные условия (17) принять в виде:

$$\begin{aligned} c_{np,\kappa}(\Theta, \tau_1, \dots, \tau_{n-2}, \tau_{n-2} + \Delta) = \\ = c_{(n-1)p,\kappa}(\Theta, \tau_1, \dots, \tau_{n-2}), \quad (20) \end{aligned}$$

где  $\Delta$  — малая величина, обеспечивающая невырожденность матрицы  $C$  в уравнениях для параметров распределения.

Таким образом, в основе эллипсоидальных МАМ одно- и  $n$ -мерных распределений лежат следующие утверждения.

**Теорема 3.1.** Пусть существует одномерное распределение СтП  $Y_t$  в МСтС (1). Тогда при фиксированном векторе параметров  $\Theta$  и полиномиальных  $\{p_{\nu}, q_{\nu}\}$  в основе алгоритма аналитического моделирования по

МЭА лежат уравнения (7)–(9) и (11) при условии конечности интегралов (10), (12) и (13) и начальных условиях (14). Количество уравнений МЭА на  $N/2 - 1$  больше числа уравнений МНА  $Q^{МЭА} = Q^{МНА} + N/2 - 1$ ,  $Q^{МНА} = p(p + 3)/2$ . В стационарном случае используются конечные уравнения (19).

**Теорема 3.2.** Пусть существует  $n$ -мерное распределение СтП  $Y_t$  в МСтС (1). Тогда при фиксированном векторе параметров  $\Theta$  и полиномиальных  $\{p_{np,\nu}, q_{np,\nu}\}$  в основе алгоритма аналитического моделирования по МЭА при  $n = 1$  лежат уравнения теоремы 3.1, а для  $n > 1$  — уравнения (15) и (16) при условиях (17). В стационарном случае используются конечные уравнения (20).

#### 4 Эллипсоидальная аппроксимация одно- и многомерных распределений в стохастической системе на многообразиях (2)

Пусть в условиях теорем 3.1 и 3.2 дополнительно выполнены следующие условия:

1. Нелинейная векторная функция  $a(Y_t, \Theta, t)$  допускает ЭЛ согласно

$$a(Y_t, \Theta, t) \approx a_0(m_t, K_t, c_1, \Theta, t) + a_1(m_t, K_t, c_1, \Theta, t)(Y_t - m_t),$$

где  $a_0 = a_0(m_t, K_t, c_1, \Theta, t)$  и  $a_1 = a_1(m_t, K_t, c_1, \Theta, t)$  называются соответственно вектором смещения нуля и матричным коэффициентом усиления ЭЛ, а  $c_1 = \{c_{p,\nu}\}$  — вектором структурных коэффициентов эллипсоидального распределения.

2. Одномерная плотность  $f_1(y; \Theta, t)$  существует и имеет эллипсоидальную структуру вида (6):

$$f_1^{\text{ЭЛ}}(y; \Theta, t) = w_1(\Theta, u) \left[ 1 + \sum_{\nu=2}^N c_{p,\nu} p_{p,\nu}(u) \right], \quad (21)$$

где

$$w_1(\Theta, u) = \frac{1}{\sqrt{(2\pi)^p |K_t|}} e^{-u/2};$$

$$c_{p,\nu} = M_{\Delta y}^{\text{ЭЛ}} [q_{p,\nu}(U_t)];$$

$$U_t = (Y_t - m_t)^T K_t^{-1} (Y_t - m_t),$$

причем первые и вторые моменты  $m_t$  и  $K_t$  неизвестны, а структурные коэффициенты одномерного эллипсоидального распределения  $c_1 = \{c_{p,\nu}\}$  известны или известны функции от  $m_t$  и  $K_t$ .

При этих условиях и соответствующих начальных условиях нелинейная дифференциальная МСтС (2) будет эквивалентна следующей системе уравнений для  $m_t$  и  $Y_t^0 = Y_t - m_t$ :

$$\begin{aligned} \dot{m}_t &= A^m(m_t, K_t, c_1, \Theta, t) = \\ &= a_0(m_t, K_t, c_1, \Theta, t), \quad m(t_0) = m_0; \quad (22) \\ \left. \begin{aligned} \dot{Y}_t^0 &= a_1(m_t, K_t, c_1, \Theta, t) Y_t^0 + V; \\ Y^0(t_0) &= Y_0^0. \end{aligned} \right\} \quad (23) \end{aligned}$$

Здесь  $V$  — составной белый шум, равный

$$V = b'_0(\Theta, t)V_0 + V_P.$$

Для вычисления параметров МЭЛ используются формулы:

$$a_0 = a_0(m_t, K_t, c_1, \Theta, t) = a_{01}(m_t, K_t, \Theta, t) + \sum_{\nu=2}^N c_{p,\nu} a_{1\nu}(m_t, K_t, \Theta, t),$$

где

$$a_{01}(m_t, K_t, \Theta, t) = M_{\Delta y}^{w_1} [a(Y, \Theta, t)];$$

$$a_1 = a_1(m_t, K_t, c_1, \Theta, t) = a_{10}(m_t, K_t, \Theta, t) + \sum_{\nu=2}^N c_{p,\nu} a_{1,p,\nu}(m_t, K_t, \Theta, t), \quad (24)$$

где

$$a_{10}(m_t, K_t, \Theta, t) = M_{\Delta y}^{w_1} [a(Y, \Theta, t)(Y - m_t)^T K_t^{-1}],$$

$$a_{1,p,\nu}(m_t, K_t, \Theta, t) = M_{\Delta y}^{p_{p,\nu} w_1} [a(Y, \Theta, t)(Y - m_t)^T K_t^{-1}];$$

$$\begin{aligned} \bar{\sigma}_{0t} &= \\ &= \sigma_0(\Theta, t) + \int_{R_0^q} c_0(\Theta, v) c_0(\Theta, v)^T \nu_P(\Theta, t) dv, \end{aligned}$$

где

$$\sigma_0(\Theta, t) = b_0(\Theta, t) \nu_0(\Theta, t) b_0(\Theta, t)^T. \quad (25)$$

В силу линейности (23) уравнение для ковариационной матрицы  $K_t$  при соответствующих начальных условиях имеет вид:

$$\begin{aligned} \dot{K}_t &= A^K(m_t, K_t, \Theta, t) = a_{1,t} K_t + K_t a_{1,t}^T + \bar{\sigma}_{0,t}, \\ K(t_0) &= K_0. \quad (26) \end{aligned}$$

Отсюда следует утверждение.

**Теорема 4.1.** В основе МЭЛ нахождения одномерной плотности (21) с известными структурными коэффициентами  $c_1 = \{c_{p,\nu}\}$  для МСтС (2) при услови-

ях 1–2, лежат уравнения (22) и (26) при условиях (24) и (25). Количество уравнений для параметров одномерного распределения  $Q_1^{МЭЛ} = Q_1^{МНА} = p(p+3)/2$ . В основе МЭЛ нахождения эллипсоидальной стационарной плотности вида (21) с известными структурными коэффициентами  $c_1 = \{c_{p,\nu}\}$  для стационарного (в узком смысле) процесса в стационарной МСтС (2) при условиях 1–2 лежат уравнения (22) и (26) при  $\dot{m}_t = 0$  и  $\dot{K}_t = 0$ :

$$\left. \begin{aligned} A^m(m, K, c_1, \Theta) &= a_0(m, K, c_1, \Theta) = 0; \\ A^K(m, K, c_1, \Theta) &= a_1(m, K, c_1, \Theta) K + \\ &+ K a_1(m, K, c_1, \Theta)^T + \bar{\sigma}_0 = 0. \end{aligned} \right\} (27)$$

Теперь получим основные уравнения МЭЛ для  $n$ -мерных распределений. Дополнительно к условиям 1–2 предположим:

3.  $n$ -мерная плотность  $f_n = f_n(y_1, \dots, y_n; \Theta, t_1, \dots, t_n)$  существует и имеет эллипсоидальную структуру вида

$$f_n^{\text{ЭЛ}}(y_1, \dots, y_n; \Theta, t_1, \dots, t_n) = f_n^{\text{ЭЛ}}(u, c_n) = w_n(\Theta, u) \left[ 1 + \sum_{\nu=2}^N c_{np,\nu} p_{np,\nu}(u) \right],$$

где

$$\begin{aligned} w_n(\Theta, u) &= [(2\pi)^n |K_n|]^{-1/2} e^{-u/2}; \\ c_{np,\nu} &= M_{\Delta y}^{\text{ЭЛ}}[q_{np,\nu}(U)]; \\ U &= (\tilde{Y}_n - \tilde{m}_n) K_n^{-1} (\tilde{Y}_n - \tilde{m}_n). \end{aligned}$$

Здесь

$$\begin{aligned} \tilde{Y}_n &= [Y_1^T Y_2^T \dots Y_n^T]^T, \quad Y_l = Y(t_l); \\ \tilde{m}_n &= [m_1^T m_2^T \dots m_n^T]^T, \quad m_l = m(t_l); \\ K_n &= \begin{bmatrix} K(t_1, t_1) & K(t_1, t_2) & \dots & K(t_1, t_n) \\ K(t_2, t_1) & K(t_2, t_2) & \dots & K(t_2, t_n) \\ \vdots & \vdots & \ddots & \vdots \\ K(t_n, t_1) & K(t_n, t_2) & \dots & K(t_n, t_n) \end{bmatrix}. \end{aligned}$$

Структурные коэффициенты  $c_n = \{c_{np,\nu}\}$  предполагаются, во-первых, известными, во-вторых, удовлетворяющими условиям согласованности.

В условиях 1–3 для получения уравнений МЭЛ применительно к  $f_n^{\text{ЭЛ}}$  достаточно воспользоваться известными уравнениями теории линейных СтС, основываясь на уравнениях (22) и (23).

**Теорема 4.2.** В основе МЭЛ  $n$ -мерной плотности при условиях 1–3 лежат уравнения квазилинейной теории для (22) и (23).

В случае двумерного эллипсоидального распределения ( $n = 2$ ), полагая

$$a(Y(t_2), \Theta, t_2) = a_{0,\Theta,t_2} + a_{1,\Theta,t_2} (Y_{t_2} - m_{t_2}),$$

где

$$\begin{aligned} a_{0,\Theta,t_2} &= a_0(m_{t_2}, K_{t_2}, c_{1,t_2}, t_2); \\ a_{1,\Theta,t_2} &= a_1(m_{t_2}, K_{t_2}, c_{1,t_2}, \Theta, t_2), \end{aligned}$$

для стационарного случая, когда

$$\begin{aligned} K(\Theta, t_1, t_2, a_{1,t_2}) &= k(\tau, \alpha_1), \\ \tau = t_1 - t_2, \quad a_1 &= a_1(m_t, K_t, c_1, \Theta), \end{aligned}$$

уравнения теоремы 4.2 приобретают вид:

$$\frac{dk(\Theta, \tau, a_1)}{d\tau} = a_1 k(\Theta, \tau, a_1),$$

$$k(0, a_1) = K_t, t = t_1 = t_2 \text{ при } \tau > 0,$$

$$k(\Theta, \tau, a_1) = k(\Theta, -\tau, \varphi_1)^T \text{ при } \tau < 0.$$

Таким образом, квазилинейная эллипсоидальная спектрально-корреляционная теория процессов в стационарной МСтС (2) основана на уравнениях (27) и следующих соотношениях для эквивалентной передаточной функции, спектральной плотности и ковариационной функции:

$$\begin{aligned} Y^0 &= \Phi(\Theta, s, a_1) V, \quad \Phi(\Theta, s, a_1) = -(a_1 - I_p s)^{-1}; \\ s_x(\Theta, \omega, a_1) &= \Phi(\Theta, i\omega, a_1) s_v \Phi(\Theta, i\omega, a_1)^*; \\ k_x(\Theta, \tau, a_1) &= \int_{-\infty}^{\infty} s_x(\Theta, \omega, a_1) e^{i\omega\tau} d\tau \end{aligned}$$

при условии асимптотической устойчивости функции  $\Phi(\Theta, s, \varphi_1)$ .

## 5 Точность и чувствительность алгоритмов, основанных на методах эллипсоидальной аппроксимации и эллипсоидальной линеаризации

В задачах надежности и безопасности технических систем [13–15] точность алгоритмов, основанных на теоремах 3.1, 3.2, 4.1 и 4.2, оценивается исходя из выбранного критерия качества. Широкое распространение получили метод сравнения вероятностей попадания на множества (в данном случае эллипсоида) и метод оценки вероятностных моментов четвертого порядка.

В инженерных приложениях, как правило, в основе оценок точности и чувствительности ал-

горитмов к параметрам  $\Theta$  лежат методы теории чувствительности [13, 14]. В условиях разд. 2–4 соответствующие уравнения для первых функций чувствительности получаются путем дифференцирования по  $\Theta$  правых и левых частей уравнений МЭА (МЭЛ).

В силу (8), (9) и (11) приходим к следующим уравнениям для функций чувствительности первого порядка  $\nabla m_t, \nabla K_t$  и  $\nabla c_{p,\kappa,t}$  ( $\nabla = \partial/\partial\Theta$ ):

$$\left. \begin{aligned} \nabla \dot{m}_t &= \nabla A^m, & \nabla m_{t_0} &= 0; \\ \nabla \dot{K}_t &= \nabla A^K, & \nabla K_{t_0} &= 0; \\ \nabla \dot{c}_{p,\kappa,t} &= \nabla A^{c_{p,\kappa}}, & \nabla c_{p,\kappa,t_0} &= 0. \end{aligned} \right\} \quad (28)$$

Заметим, что при дифференцировании по  $\Theta$  порядок уравнений возрастает пропорционально числу производных.

Аналогично [9, 10] выписываются уравнения для функций чувствительности второго порядка  $\nabla(\nabla)^T m_t, \nabla(\nabla)^T K_t$  и  $\nabla(\nabla)^T c_{p,\kappa,t}$  для одномерных распределений, а также  $n$ -мерных распределений.

Алгоритмы МАМ, основанные на МЭА (МЭЛ), содержат только на  $N/2 - 1$  уравнений больше, чем алгоритмы, основанные на МНА (МСЛ). В этом их существенное преимущество по сравнению с алгоритмами МОР и метода квази моментов [9, 10].

## 6 Пример

Следуя [1, 2, 11], рассмотрим двумерную угловую МСтС вида

$$\begin{aligned} \dot{Y}_1 &= -Y_1 Y_2; & \dot{Y}_2 &= -a Y_2 + h V(\Theta); \\ Y_1(t_0) &= Y_{10}; & Y_2(t_0) &= Y_{20}, \end{aligned}$$

где  $V = V(\Theta)$  — угловой гауссовский белый шум, зависящий от скалярного параметра  $\Theta$  с интенсивностью  $\nu = \nu(\Theta)$ . В силу теоремы 3.1, ограничиваясь моментами до четвертого порядка включительно, получим следующие уравнения для  $m_1, m_2, K_{11}, K_{12}, K_{22}$  и  $c_{2,2}$  параметров ЭА одномерной плотности (7):

$$\left. \begin{aligned} \dot{m}_1 &= -m_1 m_2 - K_{12}, & \dot{m}_2 &= -a m_2; \\ \dot{K}_{11} &= -2(m_2 K_{11} + m_1 K_{12}); \\ \dot{K}_{12} &= -(m_2 + a) K_{12} - m_1 K_{22}; \\ \dot{K}_{22} &= \nu h^2 - 2a K_{22}; \\ \dot{c}_{2,2} &= 4h^2 \nu C_{22} + \\ &+ c_{2,2} [(m_2 + a)(6 + 8C_{12} K_{12}) + \\ &+ 8m_1 C_{12} K_{12} - 3h^2 \nu C_{22}]. \end{aligned} \right\} \quad (29)$$

Здесь  $C = [C_{ij}]$ ,  $C = K^{-1}$ ,  $K = [K_{ij}]$  ( $i, j = 1, 2$ ). Уравнениям (29) в силу (28) отвечают следующие уравнения для функций чувствительности параметров ЭА:

$$\left. \begin{aligned} \nabla \dot{m}_1 &= -m_2 \nabla m_1 - m_1 \nabla m_2 - \nabla K_{12}; \\ \nabla m_{10} &= 0; & \nabla \dot{m}_2 &= -a \nabla m_2; & \nabla m_{20} &= 0; \\ \nabla \dot{K}_{11} &= -2K_{12} \nabla m_1 - 2K_{11} \nabla m_2 - \\ &- 2m_2 \nabla K_{11} - 2m_1 \nabla K_{12}; & \nabla K_{11,0} &= 0; \\ \nabla \dot{K}_{12} &= -K_{22} \nabla m_1 - K_{12} \nabla m_2 - \\ &-(m_2 + a) \nabla K_{12} - m_1 \nabla K_{22}; & \nabla K_{12,0} &= 0; \\ \nabla \dot{K}_{22} &= h^2 - 2a K_{22}; & \nabla K_{22,0} &= 0; \\ \nabla \dot{c}_{2,2} &= 8c_{2,2} C_{12} K_{12} \nabla m_1 + \\ &+ (6 + 8C_{12} K_{12}) c_{2,2} \nabla m_2 + \\ &+ 8C_{12} c_{2,2} (m_1 + m_2 + a) \nabla K_{12} + \\ &+ 8K_{12} c_{2,2} (m_1 + m_2 + a) \nabla C_{12} + \\ &+ h^2 (4 + 3\nu c_{2,2}) \nabla C_{22} + \\ &+ [(m_2 + a)(6 + 8C_{12} K_{12}) + \\ &+ 8C_{12} K_{12} - 3h^2 \nu C_{22}] \nabla c_{2,2}, \\ \nabla c_{2,2,0} &= 0. \end{aligned} \right\} \quad (30)$$

Для стационарного случая в уравнениях (29) и (30) надо приравнять нулю их правые части.

При вычислении вероятностных моментов второго порядка МЭА совпадает с МНА. Если, следуя [11], ограничиться вероятностными моментами до четвертого порядка включительно, положив  $t \in [0, 1]$ ,  $\alpha = 5$ ,  $h = 1$ ,  $m_1(0) = 0,5$ ,  $m_2(0) = 0,5$ ,  $K_{11}(0) = D_1(0) = 0,1$ ,  $K_{22}(0) = D_2(0) = 0,1$ ,  $K_{12}(0) = 0$ , то точность МЭА при вычислении начального момента  $\alpha_4$  составит около 2% по отношению к точному решению. Такую же чувствительность к параметру  $\nu = \Theta$  проявляют  $D_1, D_2, K_{12}$  и  $c_{2,2}$ .

В [11] показано, что точность вычисления начального момента  $\alpha_6$  составляет 8%, а начального момента  $\alpha_8$  — 20%. Этого обычно достаточно для приложений в задачах надежности и безопасности технических систем.

## 7 Заключение

Даны обобщения алгоритмов эллипсоидального анализа и моделирования одно- и многомерных распределений в нелинейных МСтС с винеровскими и пуассоновскими шумами. Особое внимание уделено алгоритмам для систем с аддитивными негауссовскими шумами.

Получены уравнения точности и чувствительности к параметрам. В качестве иллюстрации рассмотрена нелинейная двумерная угловая система с аддитивным угловым белым шумом. Составлены уравнения точности и чувствительности, позволяющие оценивать моменты до четвертого порядка включительно. Приведены результаты численных расчетов, показывающие эффективность МЭА (МЭЛ) по сравнению с общим МОР одномерной плотности. Сформулированы некоторые возможные обобщения.

В качестве обобщений можно рассмотреть задачи аналитического моделирования процессов в дискретных МСтС, а также задачи синтеза субоптимальных фильтров для обработки информации в таких стохастических нелинейных системах.

## Литература

1. Пугачёв В. С., Сеницын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — М.: Наука, 1990. 632 с. (*Pugachev V. S., Sinitsyn I. N. Stochastic differential systems. Analysis and filtering.* — Chichester — New York, NY, USA: John Wiley, 1987. 549 p.)
2. Пугачёв В. С., Сеницын И. Н. Теория стохастических систем. — М.: Логос, 2000; 2004. 1000 с.
3. Сеницын И. Н. Стохастические информационные технологии для исследования нелинейных круговых стохастических систем // Информатика и её применения, 2011. Т. 5. Вып. 4. С. 78–89.
4. Sinitsyn I. N., Belousov V. V., Konashenkova T. D. Software tools for circular stochastic systems analysis // 29th Seminar (International) on Stability Problems for Stochastic Models: Abstracts. — Svetlogorsk, Russia, 2011. P. 86–87.
5. Сеницын И. Н. Математическое обеспечение для анализа нелинейных многоканальных круговых стохастических систем, основанное на параметризации распределений // Информатика и её применения, 2012. Т. 6. Вып. 1. С. 12–18.
6. Сеницын И. Н., Корепанов Э. Р., Белоусов В. В., Конашенкова Т. Д. Развитие математического обеспечения для анализа нелинейных многоканальных круговых стохастических систем // Системы и средства информатики, 2012. Вып. 22. № 1. С. 29–40.
7. Sinitsyn I. N., Belousov V. V., Konashenkova T. D. Software tools for spherical stochastic systems analysis and filtering // Прикладные задачи теории вероятности и математической статистики, связанные с моделированием информационных систем. — М.: ИПИ РАН, 2012. С. 91–93.
8. Сеницын И. Н. Параметрическое статистическое и аналитическое моделирование распределений в нелинейных стохастических системах на многообразиях // Информатика и её применения, 2013. Т. 7. Вып. 2. С. 4–16.
9. Сеницын И. Н. Аналитическое моделирование распределений методом ортогональных разложений в нелинейных стохастических системах на многообразиях // Информатика и её применения, 2015. Т. 9. Вып. 3. С. 17–24.
10. Сеницын И. Н. Применение ортогональных разложений для аналитического моделирования многомерных распределений в нелинейных стохастических системах на многообразиях // Системы и средства информатики, 2015. Т. 25. № 3. С. 3–22.
11. Сеницын И. Н., Сеницын В. И. Лекции по теории нормальной и эллипсоидальной аппроксимации распределений в стохастических системах. — М.: ТОРУС ПРЕСС, 2013. 488 с.
12. Ватанабэ С., Икэда Н. Стохастические дифференциальные уравнения и диффузионные процессы / Пер. с англ. — М.: Наука, 1986. 448 с. (*Watanabe S, Ikeda N. Stochastic differential equations and diffusion processes.* — Amsterdam — Oxford — New York: North-Holland Publishing Co.; Tokyo: Kodansha Ltd., 1981. 476 p.)
13. Евланов А. Г., Константинов В. М. Системы со случайными параметрами. — М.: Наука, 1976. 568 с.
14. Справочник по теории автоматического управления / Под ред. А. А. Красовского. — М.: Наука, 1987. 712 с.
15. Александровская Л. Н., Аронов И. З., Круглов В. И. и др. Безопасность и надежность технических систем. — М.: Университетская книга, Логос, 2008. 376 с.

Поступила в редакцию 02.12.15

# ANALYTICAL MODELING OF DISTRIBUTIONS IN STOCHASTIC SYSTEMS ON MANIFOLDS BASED ON ELLIPSOIDAL APPROXIMATION

I. N. Sinitsyn and V. I. Sinitsyn

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** Accuracy and sensitivity problems for algorithms of structural analytical modeling of one- and multidimensional distributions based on the method of ellipsoidal approximation (MEA) and the method of ellipsoidal linearization (MEL) are considered. General algorithms of ellipsoidal stochastic analysis in nonlinear stochastic systems on manifolds (MStS) with Wiener and Poisson noises are developed. Special attention is paid to MStS with additive noises. Accuracy and sensitivity equations are presented. For two-dimensional nonlinear circular MStS, accuracy and sensitivity equations are derived. Equations make it possible to calculate probability moments up to the fourth order. Algorithms based on orthogonal expansion of one-dimensional density are compared with algorithms based on MEA (MEL). Some generalizations are given.

**Keywords:** accuracy equations; analytical modeling; method of ellipsoidal linearization (MEL); method of ellipsoidal approximation (MEA); sensitivity equations; stochastic system on manifold (MStS)

**DOI:** 10.14357/19922264160104

## Acknowledgments

The work was supported by the Russian Foundation for Basic Research (project 15-07-02244).

## References

1. Pugachev, V. S., and I. N. Sinitsyn. 1987. Stochastic differential systems. Analysis and filtering. — Chichester — New York, NY: John Wiley, 1987. 549 p.
2. Pugachev, V. S., and I. N. Sinitsyn. 2001. *Stochastic systems. Theory and applications*. Singapore: World Scientific. 908 p.
3. Sinitsyn, I. N. 2011. Stokhasticheskie informatsionnye tekhnologii dlya issledovaniya nelineynykh krugovykh stokhasticheskikh sistem [Stochastic informational technologies for circular stochastic systems investigation]. *Informatika i ee Primeneniya — Inform. Appl.* 5(4):78–89.
4. Sinitsyn, I. N., V. V. Belousov, and T. D. Konashenkova. Software tools for circular stochastic systems analysis. *29th Seminar (International) on Stability Problems for Stochastic Models: Abstracts*. Svetlogorsk, Russia. 86–87.
5. Sinitsyn, I. N. 2012. Matematicheskoe obespechenie dlya analiza nelineynykh mnogokanal'nykh krugovykh stokhasticheskikh sistem, osnovannoe na parametrizatsii raspredeleniy [Mathematical software for analysis of nonlinear multichannel circular stochastic systems based on distributions parametrization]. *Informatika i ee Primeneniya — Inform. Appl.* 6(1):12–18.
6. Sinitsyn, I. N., E. R. Korepanov, V. V. Belousov, and T. D. Konashenkova. 2012. Razvitie matematicheskogo obespecheniya dlya analiza nelineynykh mnogokanal'nykh krugovykh stokhasticheskikh sistem [Development of mathematical software for analysis of nonlinear multichannel circular stochastic systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 22(1):29–40.
7. Sinitsyn, I. N., V. V. Belousov, and T. D. Konashenkova. 2012. Software tools for spherical stochastic systems analysis and filtering. *Prikladnye zadachi teorii veroyatnosti i matematicheskoy statistiki, svyazannye s modelirovaniem informatsionnykh sistem* [Applied problems of probability theory and mathematical statistics related to modeling of information systems]. Moscow: IPI RAN. 91–93.
8. Sinitsyn, I. N. 2013. Parametricheskoe statisticheskoe i analiticheskoe modelirovanie raspredeleniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraznykh [Parametric statistical and analytical modeling of distributions in nonlinear stochastic systems on manifolds]. *Informatika i ee Primeneniya — Inform. Appl.* 7(2):4–16.
9. Sinitsyn, I. N. 2015. Analiticheskoe modelirovanie raspredeleniy metodom ortogonal'nykh razlozheniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraznykh [Analytical modeling in stochastic systems on manifolds based on orthogonal expansions]. *Informatika i ee Primeneniya — Inform. Appl.* 9(3):17–24.
10. Sinitsyn, I. N. 2015. Primenenie ortogonal'nykh razlozheniy dlya analiticheskogo modelirovaniya mnogomernykh raspredeleniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraznykh [Applications of orthogonal

- expansions for analytical modeling of multi-dimensional distributions in nonlinear stochastic systems on manifolds]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(3):3–22.
11. Sinitsyn, I. N., and V. I. Sinitsyn. 2013. *Lektsii po teorii normal'noy i ellipsoidal'noy approksimatsii raspredeleniy v stokhasticheskikh sistemakh* [Lectures on normal and ellipsoidal approximation of distributions in stochastic systems]. Moscow: TORUS PRESS. 488 p.
  12. Watanabe, S., and N. Ikeda. 1981. *Stochastic differential equations and diffusion processes*. Amsterdam – Oxford – New York: North-Holland Publishing Co.; Tokyo: Kodansha Ltd. 476 p.
  13. Evlanov, A. G., and V. M. Konstantinov. 1976. *Sistemy so slozhnymi parametrami* [Systems with random parameters]. Moscow: Nauka. 568 p.
  14. Krasovskii, A. A., ed. 1987. *Spravochnik po teorii avtomaticheskogo upravleniya* [Handbook for automatic control]. Moscow: Nauka. 712 p.
  15. Aleksandrovskaia, L. N., I. Z. Aronov, V. I. Kruglov, et al. 2008. *Bezopasnost' i nadezhnost' tekhnicheskikh sistem* [Security and reliability of technical systems]. Moscow: Universitetskaya Kniga, Logos. 376 p.

*Received December 2, 2015*

## Contributors

**Sinitsyn Igor N.** (b. 1940) — Doctor of Science in technology, professor, Honored scientist of RF, Head of Department, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sinitsin@dol.ru

**Sinitsyn Vladimir I.** (b. 1968) — Doctor of Science in physics and mathematics, associate professor, Head of Department, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vsinitsin@ipiran.ru

# ПРИМЕНЕНИЕ МЕТАПРОГРАММИРОВАНИЯ ДЛЯ ПОВЫШЕНИЯ ТЕХНОЛОГИЧНОСТИ БОЛЬШИХ АВТОМАТИЗИРОВАННЫХ СИСТЕМ

С. П. Ковалёв<sup>1</sup>

**Аннотация:** Предложен подход к снижению затрат на проектирование больших автоматизированных систем за счет привлечения современных технологий метапрограммирования. В качестве наиболее перспективных среди таких технологий рассматриваются разработка, управляемая моделями (model driven engineering, MDE), и аспектно-ориентированный подход (aspect-oriented software development). Представлены методы масштабирования этих технологий, позволяющие эффективно применять их в условиях роста размера создаваемых автоматизированных систем путем замыкания относительно системообразующих структурных отношений. В качестве примера практического применения подхода приводится проектирование математического обеспечения интеллектуальных электроэнергетических сетей. Излагаются принципы математического аппарата для построения, анализа и оптимизации процедур проектирования на базе теории категорий. Описан процесс проектирования генератора расчетных программных компонентов большой автоматизированной системы с применением теоретико-категорных методов.

**Ключевые слова:** большие автоматизированные системы; метапрограммирование; мегамодель; теория категорий; копредел; разработка, управляемая моделями; аспектно-ориентированный подход; интеллектуальная электроэнергетическая сеть

**DOI:** 10.14357/19922264160105

## 1 Введение

К числу основных тенденций развития современных автоматизированных систем относится рост масштаба до уровня больших и даже сверхбольших (ultra-large scale systems, ULS) [1].

Примером служит Smart Grid (интеллектуальная сеть) — система сквозной автоматизации крупной электроэнергетической сети [2]. Такие системы отличаются беспрецедентно большими значениями масштабных факторов, таких как количество данных, элементов, взаимосвязей, процессов, нормативов, пользователей и др. Как следствие, сверхбольшим системам присущи следующие качественные особенности, принципиально отличающие их от традиционных объектов системной инженерии [1]:

- децентрализация;
- заведомо конфликтующие, непостижимые, разноречивые требования;
- непрерывные процессы эволюции и внедрения;
- разнородные, несовместимые, изменчивые элементы;
- эрозия границ между человеком и системой;
- штатный характер сбоев;

- новые парадигмы приобретения и нормирования (policy).

Эти особенности не рассматривались в классическом программировании, поэтому применение традиционных технологий проектирования программного обеспечения при создании больших систем сопряжено со значительными затратами. Рост масштаба приводит к взрывному увеличению трудоемкости рутинных процедур, которые в ходе проектирования традиционных систем выполнялись «вручную» почти незаметно, вследствие чего сравнительно редко изучались в системной инженерии. К таким процедурам относятся согласование порядка внесения изменений в компоненты, подстройка интерфейсов программного доступа к модулям под потребности их окружения, реализация рассеянных задач (crosscutting concerns), обеспечение интегральной производительности и др. Говоря кратко, по мере роста масштаба программные изделия стремительно утрачивают *технологичность* (manufacturability).

Снижение трудозатрат на выполнение рутинных процедур возможно за счет автоматизации — перехода к метапрограммированию. Однако традиционных CASE-средств для такого перехода недостаточно: автоматизация вышеуказанных процедур

<sup>1</sup>Институт проблем управления им. В. А. Трапезникова Российской академии наук, kovalyov@nm.ru

требует специализированных инструментов. Широкие возможности по специализации инструментария заложены в технологиях разработки систем, управляемой моделями (MDE) [3]. Технологии MDE нацелены на создание средств, позволяющих описать систему набором формализованных компьютерных моделей, в том числе на специализированных языках (domain-specific languages, DSL), и далее автоматически сгенерировать из моделей тексты программного обеспечения системы. Языки моделирования строятся путем специализации на уровне метамodelей из языков общего назначения, таких как язык объектно-ориентированного моделирования UML (unified modeling language) [4], языки реального времени [5], языки логического вывода [6] и др. Сборка моделей систем из моделей компонентов описывается «мегамоделями» и выполняется при помощи разнообразных механизмов, в том числе связывания (weaving) — основного способа сборки систем в аспектно-ориентированном подходе [7]. Предоставляются инструменты для создания средств автоматического конструирования, анализа, сборки и трансформации моделей, в том числе для генерации программного кода. Большой набор таких инструментов и средств создан и опубликован в открытом исходном коде в рамках проекта Eclipse Modeling Project [8].

Разнообразные технологии MDE интенсивно развиваются в настоящее время, и для них еще не выработан общий концептуальный фундамент. Не хватает апробированных типовых решений, применение которых не влечет риска массовой генерации программ, содержащих ошибки либо потребляющих недопустимо много вычислительных ресурсов. Как следствие, не решены проблемы масштабирования технологий MDE — затраты труда на разработку, управляемую моделями, слишком быстро растут по мере увеличения размера и структурной сложности программных систем [9]. В этих условиях новые инструменты метапрограммирования больших систем должны быть специфицированы и верифицированы на максимально высоком уровне строгости — таком, который может дать только математический аппарат. Аппарат должен позволять кратко описывать механизмы масштабирования, формулировать и доказывать их основные свойства, не «потонув» в деталях описания частных моделей и языков. Ввиду этого не удается привлечь традиционные формальные методы программирования, базирующиеся на разноплановых «тяжеловесных» математических средствах, слабо совместимых друг с другом [10].

Перспективным представляется применение теории категорий [11] — раздела высшей алгебры, который позволяет прозрачно и компактно фор-

мализовать принципы системной инженерии [12]. Вводятся категории, объектами которых служат модели системных единиц (компонентов, подсистем, систем и т. д.), а морфизмы формально описывают преобразования моделей по ходу процессов разработки систем. Строятся теоретико-категорные конструкции, описывающие способы выполнения трудоемких процедур (технологические карты) на абстрактном концептуальном уровне: моделям, задействованным в процедурах, сопоставляются объекты подходящих категорий, технологическим операциям сопоставляются морфизмы, переходам между технологиями сопоставляются функторы и т. д. Путем вычислений в категориях оцениваются свойства способов выполнения процедур, выбирается такой из альтернативных способов, который доставляет экстремальное значение целевому функционалу технологичности (стоимости, надежности и т. д.). Конструкция, соответствующая самому технологичному способу, интерпретируется в терминах подходящей технологии, и выбираются либо создаются инструменты метапрограммирования для автоматизации работ по его реализации [13].

Статья построена следующим образом. В разд. 2 рассматриваются технологические проблемы проектирования больших автоматизированных систем (преимущественно в части программного обеспечения) и подходы к их решению. В разд. 3 приводится пример — повышение эффективности процесса проектирования больших расчетных задач за счет метапрограммирования. Раздел 4 посвящен принципам применения теории категорий в качестве математического аппарата метапрограммирования. В заключении подводятся итоги и намечаются перспективы представленных исследований.

## 2 Проблемы проектирования больших автоматизированных систем

Основной причиной роста значений масштабных факторов систем являются требования полноты (замкнутости) массивов сущностей, описывающих объект автоматизации, относительно тех или иных структурных отношений. Именно такие требования вызывают необходимость масштабировать технологии метапрограммирования типа MDE, поскольку сущности и отношения образуют основной предмет моделирования. К числу главных системообразующих отношений относятся топологические, каузальные, мерелогические («часть—целое»), дескриптивные («абстрактное—конкретное») и телеологические («цель—средство») [14].

В то же время результаты процессов проектирования программных систем могут вступать только в отношения «часть—целое» и «абстрактное—конкретное»: действия по созданию системы сводятся к (де)композиции составных результатов (например, детализация требований, сборка приложения из объектных модулей) и трансформации (refinement) абстрактных результатов в конкретные (например, реализация спецификации на языке программирования) [15]. Поэтому разработчики автоматизированных систем вынуждены выражать все возможные связи сущностей через мереологические и дескриптивные связи отвечающих им системных единиц. Для топологических и каузальных связей эта проблема решается путем сохранения их в базе данных: они отображаются на мереологические связи структур данных. Эффективность хранения и обработки связей достигается путем экстенсивного наращивания количества и мощности вычислительных узлов, параллельно выполняющих навигацию по ним. Здесь привлекаются технологии параллельных и распределенных вычислений типа Grid [16] или вычислительных облаков.

Однако с телеологическими связями поступить подобным образом не удастся, поскольку с ростом масштаба они неизбежно приобретают конфликтность, запутанность, изменчивость (в то время как мереологическим связям структур данных присуща низкая гибкость). Появляется большое количество рассеянных задач — таких, которые не поддаются локализации в рамках функционально замкнутых программных модулей с фиксированным интерфейсом. Обеспечение технологичности автоматизированных систем в условиях рассеяния задач является целью аспектно-ориентированного программирования (АОП) [17]. Реализация рассеянной задачи в АОП оформляется как аспект — особая программная единица, код которой автоматически вставляется в код других единиц в местах, явно задаваемых внешним образом. Процедура вставки аспектов называется связыванием (weaving). Она предоставляет аспектам полный доступ к контексту в обход ограничений модульного интерфейса. Примером технологии АОП служит AspectJ — аспектно-ориентированное расширение языка Java [18].

Однако на практике АОП применяется значительно реже, чем модульные подходы, поскольку отсутствует единое непротиворечивое понимание его методологической основы [19]. На семантическом уровне неясно, как рационально выделять и комплексировать аспекты в программах и моделях. Существующие технологии АОП предлагают лишь частные решения в рамках частных парадигм программирования. Они ограничены поддержкой программно-технических рассеянных задач, таких

как журналирование, кэширование, защита информации и т.п. Не хватает технологичных решений и инструментов для моделирования и реализации семантически богатых функциональных рассеянных задач, таких как ведение информационной модели (паспорта) объекта автоматизации, оповещение участников процессов о ходе их выполнения, оперативная оценка эффективности процессов, проверка правильности действий пользователей и компонентов системы, перевод информации на разные языки и в разные форматы.

В связи с этим в [13, 20] предложен универсальный теоретико-категорный подход к расширению «модульных» технологий проектирования систем аспектно-ориентированными приемами, пригодный к применению в технологиях MDE. В основе подхода лежит направленность на обеспечение трассирования — прослеживания воплощения абстрактно поставленных задач в конкретных фрагментах результатов процесса проектирования [21]. По существу, подход обеспечивает погружение телеологических связей задач в дескриптивные связи моделей средств их автоматизации. Такое погружение оправдано тем, что трассируемость страдает от рассеяния задач больше, чем другие показатели качества систем.

### 3 Эффективное проектирование расчетных задач

В качестве примера процесса, эффективность которого удастся повысить за счет метапрограммирования, рассмотрим проектирование базового математического обеспечения интеллектуальных электроэнергетических сетей, состоящего из задач расчета и анализа показателей распределенных энергетических объектов. Исходные данные для расчета поступают в систему от измерительных приборов, из смежных информационных систем и с автоматизированных рабочих мест пользователей. В первую очередь оперативно вычисляются режимные показатели — результаты распространения значений параметров количества и качества энергоресурсов вдоль оси времени (статистический анализ, прогнозирование), вдоль линий передачи энергии (замещение отсутствующих значений, расчет отпуска из сети по приему в сеть и др.), вдоль иерархии управления объектом (например, формирование сводного топливно-энергетического баланса города). Решение таких задач сводится к суперпозиции базовых алгоритмов, определяющих один шаг распространения (прогноз изменения потребности энергоустановки в энергоресурсе за один расчет-

ный период, объем потери энергоресурса в элементе энергетической сети и т. п.). Обычно базовые алгоритмы задаются алгебраическими формулами, разностными схемами либо задачами математического программирования. Порядок применения суперпозиции определяется каузальными, топологическими и мереологическими связями между сущностями, составляющими объект.

Расчет путем навигации по истории, топологии и иерархии объектов требуется также для показателей технического состояния, определяющих уровень износа оборудования, надежность, потребность в ремонте и/или модернизации. Массивы значений режимных и технических показателей служат входными данными для расчета необходимых управляющих воздействий на энергетический объект. Фактические значения сравниваются с целевыми, определяются причины расхождений, формируются и исполняются задания на производство работ по возврату значений в целевые рамки. В завершение цикла расчета вычисляются агрегатные показатели уровня развития объекта управления, характеризующие его способность удовлетворять потребности, новизну техники, энергетическую эффективность [22]. Эти показатели используются в процессах стратегического управления.

При программной реализации расчета показателей традиционные «ручные» процедуры составления, верификации и актуализации программ требуют затрат труда, растущих пропорционально масштабу энергетического объекта. Кроме того, для достижения необходимой производительности нужно выполнять большие расчетные задачи в среде распределенных вычислений типа Grid, развертывание программных модулей в которой требует значительных затрат труда квалифицированных системных программистов [23]. В то же время вся информация, необходимая для формирования расчетных алгоритмов, присутствует в информационной модели (паспорте) объекта управления: в нем хранятся и параметры оборудования, и всевозможные связи, и структура вычислительной среды [24]. Поэтому для снижения затрат целесообразно воспользоваться подходом типа MDE — создать автоматический генератор расчетных компонентов по наполнению паспорта. Генератор «обходит» фрагмент паспорта, отвечающий участку энергетической сети, показатели которого требуется рассчитать, и порождает тексты программ модулей, вычисляющих значения показателей. Тексты содержат расчетные формулы, обращения к специализированным инструментам расчета и анализа данных, акты межмодульного взаимодействия, системные вызовы для развертывания в среде распределен-

ных вычислений. Исполнение сгенерированных модулей позволяет обеспечить информационную поддержку принятия решений в темпе процесса оперативно-диспетчерского управления объектом.

При разработке такого генератора не удастся воспользоваться готовыми инструментами метапрограммирования наподобие созданных в рамках проекта Eclipse Modeling Project, поскольку требуется решить две серьезные проблемы.

Первая из них состоит в необходимости обеспечить не только правильность формирования расчетных формул, но и корректность и производительность их выполнения при любых входных данных с учетом особенностей архитектуры используемых вычислительных средств. Выполнение программ, составленных путем прямой реализации спецификаций алгоритмов без учета таких особенностей, может сопровождаться переполнениями, искажениями данных, коммуникационными блокировками. Например, промежуточные результаты вычисления показателей больших объектов могут выходить за рамки стандартного диапазона машинного представления чисел. Поэтому необходимо отобразить алгоритмы на архитектуру вычислительной среды [25] — подобрать типы данных, структуру управления потоком вычислений, режимы взаимодействия с другими модулями так, чтобы избежать этих проблем, затрачивая минимальное количество ресурсов машинного времени, памяти, емкости каналов связи.

При «ручной» реализации вычислительных алгоритмов программист выполняет отображение умозрительно, основываясь на документации, сопровождающей аппаратуру и средства программирования. Генератор же должен автоматически обеспечивать соответствие структур порождаемой программы характеристикам вычислительной среды, указанным в паспорте. Необходимо встроить в генератор формальную модель процедуры отображения, позволяющую адаптировать ресурсные потребности генерируемого кода к возможностям узлов вычислительной среды.

При формализации отображения каждый узел вычислительной среды представляется моделью вычислений — конечной алгеброй, состоящей из совокупности всех чисел, помещающихся в память, и вычислительных примитивов над ними. Интеграция вычислительного узла в среду заключается в кодировании чисел, поддерживаемых узлом, т. е. в сопоставлении им кодов чисел, поддерживаемых средой. Формально действия по интеграции отвечают отображения (в обычном математическом смысле) основных множеств алгебр, задающие правила кодирования так, что структура тех вычислительных операций, которые узел должен выполнять



Процесс проектирования генератора расчетных компонентов с применением теоретико-категорных методов

в составе системы, не разрушается. Все модели вычислений и все действия по их интеграции образуют категорию [26], конструкции в которой, в том числе мегамоделей, позволяют строго описывать и оптимизировать распределение ресурсов вычислительной среды под выполнение расчетных алгоритмов. Чтобы найти узлы среды, на которых можно разместить алгоритм, строится алгебра операций, выполняемых алгоритмом. Погружение этой алгебры в вычислительную среду задается вычислительным расширением — функтором специального вида, который формализует отображение задачи на среду: задача размещается в узлах, алгебраические модели которых поддерживают операции ее вычислительного расширения.

Для оптимизации размещения с точки зрения производительности применяются мощные математические методы [27]: сигнатурным операциям алгебр, отвечающих узлам среды, и связывающим узлы морфизмам приписываются оценки длительности, получаемые из паспортных характеристик среды и/или результатов прогона тестов производительности, и строится размещение, обладающее минимальной интегральной длительностью расчета. В терминах MDE процедуру вычислительного расширения можно рассматривать как частный случай перехода от платформенно-независимой модели расчетной задачи (platform independent model, PIM) к платформенно-зависимой модели развертывания задачи в вычислительной среде (platform specific model, PSM). Реализация этой процедуры и встраивается в генератор расчетных модулей.

Вторая проблема, связанная с автоматизацией порождения расчетных модулей, состоит в обеспечении их актуальности — соответствия программной реализации алгоритмов фактическому состоянию объекта управления. Нетрудно добиться соответствия алгоритмов паспорту: нужно автоматически вызывать генератор при изменениях паспорта. Значительно труднее обеспечить соответствие паспорта объекту, поскольку объект может измениться в ходе совершенно разнородных процессов: реконструкции и ремонта, смены собственника, энергетического обследования и т.д. Необходимо рассеять задачу ведения паспорта по средствам автоматизации всех таких процессов с обеспечением трассирования каждой записи паспорта к задачам, вызвавшим ее изменение. Для этого применяется процедура аспектного связывания, формально специфицированная и верифицированная при помощи конструкций теоретико-категорного подхода к АОП [20]. Благодаря связыванию достигается практически полная актуальность и достоверность паспорта, в противовес традиционным «ручным» регламентным процедурам его ведения.

Таким образом, процесс проектирования генератора расчетных компонентов с применением теоретико-категорных методов можно представить в виде схемы (см. рисунок) [28].

Такой процесс был апробирован на практике при проектировании подсистемы расчета и анализа данных программной платформы учета и управления энергообеспечением «Энергиус» [29]. Это позволило снизить стоимость и сроки внедрения

больших систем диспетчерского управления, интеллектуального учета электроэнергии, управления энергоэффективностью на базе платформы «Энергиус» по сравнению с функционально аналогичными программными продуктами.

## 4 Теория категорий как математический аппарат метапрограммирования

Рассмотрим проблему выбора математического аппарата, пригодного для строгого формального описания и анализа технологических процедур проектирования больших автоматизированных систем, в том числе с применением метапрограммирования. Известно [30], что традиционное математическое моделирование посредством дифференциальных уравнений практически не применимо в инженерии программных систем. Дело в том, что для вывода уравнений необходимы вариационные принципы, законы сохранения либо статистические закономерности, а в процессах разработки программ обнаружить их не удалось. Фактически отсутствует даже устоявшаяся система координат — числовых показателей, зависимость которых от времени характеризует все значимые аспекты процесса создания программной системы. (Очевидные показатели типа числа строк текста программы весьма слабо коррелируют с трудоемкостью написания текста, особенно в условиях применения метапрограммирования.)

Существует альтернативный прагматический подход к выбору математического аппарата, основанный на том наблюдении, что для большинства систем доступна (либо легко восстанавливаема) история сборки из составных частей. Если известны математические модели частей и сборочных операций, то можно задавать системы математическими «мегамоделями» — ориентированными графами (диаграммами), узлы которых помечены обозначениями частей, а ребра помечены обозначениями операций. Здесь требуется формировать и обрабатывать большие графы, которые целиком даже нельзя изобразить, а можно только описать структурными ограничениями. Мощный аппарат для конструирования и анализа диаграмм такого рода развит в рамках теории категорий — раздела высшей алгебры, который «начинается с наблюдения, что многие свойства математических систем можно представить просто и единообразно посредством диаграмм» [11].

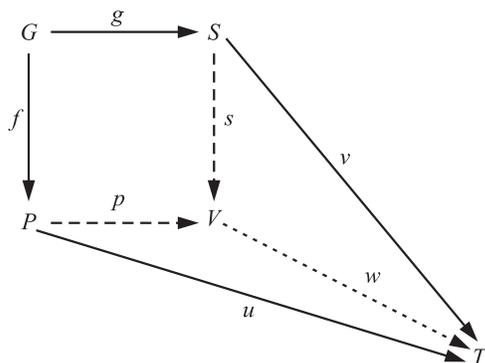
Напомним, что категория  $C$  — это класс абстрактных объектов  $\text{Ob } C$ , попарно связанных

морфизмами (абстрактными аналогами отображений) [11]: каждый морфизм  $f$  имеет область  $\text{dom } f \in \text{Ob } C$  и кообласть  $\text{codom } f \in \text{Ob } C$ . Соотношения вида  $\text{dom } f = A$  и  $\text{codom } f = B$  наглядно записываются в форме стрелки  $f : A \rightarrow B$ , а множество всех морфизмов, удовлетворяющих этим соотношениям, обозначается через  $\text{Mor}(A, B)$ . Для любой пары морфизмов  $f, g$ , такой что  $\text{codom } f = \text{dom } g$ , определена композиция — морфизм  $g \circ f : \text{dom } f \rightarrow \text{codom } g$ . Композиция ассоциативна: для любой тройки морфизмов  $f, g, h$ , если  $\text{codom } f = \text{dom } g$  и  $\text{codom } g = \text{dom } h$ , то  $h \circ (g \circ f) = (h \circ g) \circ f$ . Наконец, любой объект  $A$  обладает тождественным морфизмом  $1_A : A \rightarrow A$ , таким что для любого морфизма  $f : A \rightarrow B$  выполняется соотношение  $f \circ 1_A = 1_B \circ f = f$ .

При формализации процессов проектирования программных систем главную роль играют категории, объекты которых сопоставляются системным единицам, а морфизмы сопоставляются действиям по интеграции (мереологическим связям): каждое действие отображает компонент (область морфизма) в содержащую его систему (кообласть). Подчеркнем, что для компонента  $P$  и системы  $S$ , как правило, указывается не только возможность (или невозможность) интеграции  $P$  в  $S$ , но и совокупность всех способов интеграции (вставка, разрешение ссылок, перекомпоновка и т. д.), порождающая множество морфизмов  $\text{Mor}(P, S)$ . Композиция морфизмов отвечает многошаговым действиям (процессам), а тождественные морфизмы — «ничего неделанию». Например, объектами категорий, описывающих формальные методы программирования, служат модели программ (алгебраические спецификации, графы, термы лямбда-исчисления и т. д.), а в качестве морфизмов часто выступают гомоморфизмы и их обобщения.

Как указывалось выше, диаграммы в таких категориях задают математические «мегамоделю» систем. Процедуре сборки системы из мегамоделю отвечает копредел ( $\text{colimit}$ ) диаграммы — универсальная диаграммная конструкция [11]. Для иллюстрации этой конструкции рассмотрим соединение компонента  $P$  с системой  $S$  — прием сборки, состоящий в добавлении промежуточного компонента  $G$ , называемого «клеем» ( $\text{glue}$ ) или связкой ( $\text{connector}$ ) [31], который способен интегрироваться как с компонентом  $P$ , так и с системой  $S$ . Например, путем соединения строятся системы на базе промежуточного программного обеспечения ( $\text{middleware}$ ), которое и служит связкой. Мегамоделю соединения имеет вид пары морфизмов  $f : P \leftarrow G \rightarrow S : g$ . Ее копредел, называемый в теории категорий кодекартовым квадратом ( $\text{pushout}$ ), задается объектом-вершиной  $V$  и парой морфиз-

мов-ребер  $p : P \rightarrow V \leftarrow S : s$ , таких что  $p \circ f = s \circ g$  (т.е. соблюдается структурная корректность системного решения) и, кроме того, выполняется следующее условие универсальности: для любых объекта  $T$  и пары морфизмов  $u : P \rightarrow T \leftarrow S : v$ , если  $u \circ f = v \circ g$ , то существует единственный морфизм  $w : V \rightarrow T$ , удовлетворяющий соотношениям  $w \circ p = u$  и  $w \circ s = v$ . Тогда объект  $V$  действительно отвечает системе, которая собрана из  $S$  и  $P$  путем соединения посредством  $G$  и не содержит ничего «лишнего». Если такой объект  $V$  существует, то он определяется однозначно с точностью до изоморфизма — формального представления несущественного различия между моделями. Если же копредела не существует, то делается вывод, что клей  $G$  не способен соединить компонент  $P$  с  $S$  посредством действий  $f$  и  $g$ , т.е. что исходная мега-модель неправильно построена.



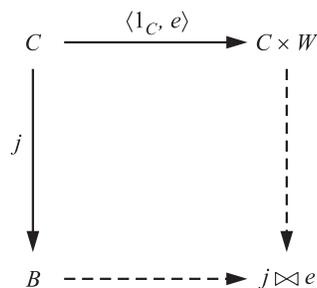
Конструкция кодекартова квадрата часто рассматривается в литературе по теоретико-категорным методам сборки программных систем (см., например, [32]). При некоторых технических ограничениях через нее можно выразить копредел любой конечной диаграммы [11] (т.е. формально собрать систему из любой мега-модели). Например, любая конечная диаграмма может быть реализована в виде системы в категории из разд. 3, формально описывающей проектирование средств распределенных вычислений [26]. Это необходимо для реализации динамического развертывания вычислительных задач в среде типа Grid: топология размещения расчетных компонентов должна диктоваться ресурсными, а не структурными ограничениями.

Путем обобщения кодекартова квадрата строится типовая формальная мега-модель основной технологической процедуры АОП — аспектного связывания [20]. В классическом АОП связывание состоит в подключении программы  $W$ , называемой советом (advice), к базовой программе (base)  $B$  в заданных местах, называемых точками соединения (join points) [33]. Каждый раз, когда при исполнении базовой программы встречается точка соеди-

нения, вызывается совет. Поэтому аспект обычно выглядит как блок программного кода, охраняемый (guarded) условием, идентифицирующим точку соединения; начало блока служит точкой вызова совета (entry point). Таким образом, инструмент связывания (weaver) принимает на вход две спецификации:

- (1) описание точек соединения в базовой программе, или срез (pointcut);
- (2) описание точек вызова совета в точках соединения.

При связывании на первом шаге (виртуально) создается достаточное количество копий совета, по одной на каждую точку соединения, с маркировкой соответствующих им точек вызова. Далее на втором шаге эти точки «склеиваются» друг с другом так, чтобы не разрушить аспектную структуру базы и совета. Для формальной записи правил связывания привлекается дополнительная модель  $C$ , называемая связкой (connector [34]), которая интегрируется с базой в точках соединения, а с советом — в точках вызова. В технологиях типа AspectJ в роли связки выступает регулярное выражение, выделяющее в тексте базовой программы синтаксические единицы, образующие срез (конструкция pointcut) [18]. Соответствие точек соединения точкам вызова задается парой морфизмов  $j : B \leftarrow C \rightarrow W : e$  (здесь наглядно проявляется отличие связывания от модульной компоновки, формализуемой одношаговым действием вида  $l : M \rightarrow S$ , где  $M$  — модуль,  $S$  — система). Первому шагу связывания отвечает построение произведения  $C \times W$  вместе с морфизмом  $\langle 1_C, e \rangle : C \rightarrow C \times W$ , который однозначно определяется условиями  $p_C \circ \langle 1_C, e \rangle = 1_C$  и  $p_W \circ \langle 1_C, e \rangle = e$ , где  $p_C : C \times W \rightarrow C$  и  $p_W : C \times W \rightarrow W$  — проекции произведения на компоненты. Второй шаг связывания, как легко видеть, формально представляет собой соединение — кодекартов квадрат над парой морфизмов  $j : B \leftarrow C \rightarrow C \times W : \langle 1_C, e \rangle$ . Результатом связывания является вершина этого квадрата, обозначаемая через  $j \bowtie e$  (при условии, что и произведение  $C \times W$ , и сам кодекартов квадрат существуют).



Путем анализа этой конструкции строго доказан ряд свойств связывания (для случаев, когда оно существует) [20]: единственность результата с точностью до изоморфизма, наличие неразрушающего вложения базы в результат, независимость результата привязывания нескольких взаимно независимых советов к одной базе от порядка связывания, сохранение аспектной структуры базы при связывании с достаточно мелкими единицами аспектной декомпозиции. В свою очередь, существование связывания доказано для ряда практически значимых частных случаев, в том числе для аспекта ведения паспорта объекта управления.

Отметим, что сама конструкция копредела, как видно из ее названия, является двойственной по отношению к конструкции предела, которая была введена первоначально — для нужд приложений теории категорий к топологии [11] (и, вообще говоря, не связана непосредственно с понятием предела из математического анализа). Предел отличается от копредела обращением направления ребер, так что в приложениях к системной инженерии предел представляет «наибольшую общую часть» составляющих мегамоделей, в то время как копредел представляет «наименьшее объемлющее целое» (систему). Например, частным случаем предела является произведение, которое содержится в мегамоделе аспектного связывания.

## 5 Заключение

Технологии метапрограммирования типа MDE и АОП обладают большим потенциалом в качестве средств повышения технологичности больших автоматизированных систем. Как показано выше, они позволяют заменять десятки программистов-кодировщиков небольшими группами, создающими инструменты для автоматического порождения массивного программного кода. Ключевую роль в проектировании, верификации и масштабировании таких инструментов играет формализация приемов метапрограммирования, которую целесообразно проводить на базе математического аппарата теории категорий.

Однако привлечение указанных технологий служит лишь первым шагом на пути к радикальному повышению эффективности процессов создания систем за счет сквозной автоматизации. На следующем шаге необходимо поднять уровень интеллектуальности инструментов программной инженерии — переложить на них построение, анализ и оптимизацию процедур проектирования. Современные CASE-средства практически не умеют этого делать — они способны только механиче-

ски выполнять команды инженеров. Значительного прогресса можно добиться, привлекая теорию категорий: можно использовать ее язык для формального внутримашинного представления процедур проектирования и свести интеллектуальные функции к распознаванию и расчету категорных конструкций. Отметим, что такой подход позволил бы повысить эффективность и в автоматизированном проектировании сложных изделий материального производства.

Уже давно существуют автоматические решатели теоретико-категорных задач, в том числе возникающих при разработке программных систем [35]. Но в дополнение к решателям для создания полноценных средств интеллектуального управления такой интеллектуальной деятельностью, как программная инженерия, требуются новые подходы, по мощности качественно превосходящие традиционные нейронные сети и численные методы оптимизации. Здесь открываются перспективные направления дальнейших исследований.

## Литература

1. Ultra-Large-Scale Systems: The software challenge of the future. — Pittsburgh: Carnegie Mellon Software Engineering Institute, 2006. 134 p.
2. Anvaari M., Cruzes D. S., Conradi R. Smart Grid software applications as an ultra-large-scale system: Challenges for evolution // Innovative Smart Grid Technologies (ISGT). — IEEE PES, 2012. P. 1–6.
3. Schmidt D. C. Model-driven engineering // IEEE Computer, 2006. Vol. 39. No. 2. P. 25–32.
4. Graaf B., van Deursen A. Visualisation of domain-specific modelling languages using UML // 14th IEEE Conference (International) on the Engineering of Computer-Based Systems ECBS'2007 Proceedings. — Tucson, 2007. P. 586–595.
5. Benveniste A., Caspi P., Edwards S. A., Halbwachs N., Le Guernic P., de Simone R. The synchronous languages 12 years later // Proc. IEEE, 2003. Vol. 91. Iss. 1. P. 64–83.
6. Васильев С. Н. Формализация знаний и управление на основе позитивно-образованных языков // Информационные технологии и вычислительные системы, 2008. № 1. С. 3–17.
7. Jouault F., Vanhooft B., Bruneliere H., Doux G., Berbers Y., Bezivin J. Inter-DSL coordination support by combining megamodeling and model weaving // 2010 ACM Symposium Applied Computing Proceedings. — Sierre, 2010. P. 2011–2018.
8. Eclipse Modeling Project. The Eclipse Foundation, 2014. <http://www.eclipse.org/modeling>.
9. Kolovos D. S., Rose L. M., Matragkas N., Paige R. F., Guerra E., Cuadrado J. S., De Lara J., Rath I., Varró D., Tisi M., Cabot J. A research roadmap towards achieving scalability

- in model driven engineering // Workshop on Scalability in Model Driven Engineering Proceedings. — Budapest, Hungary: ACM, 2013. P. 2:1–2:10.
10. Diskin Z., Maibaum T. S. E. Category theory and model-driven engineering: From formal semantics to design patterns and beyond // 7th Workshop ACCAT'2012 "Electronic Proceedings in Theoretical Computer Science," 2012. Vol. 93. P. 1–21.
  11. Маклейн С. Категории для работающего математика / Пер. с англ. — М.: Физматлит, 2004. 352 с. (*Mac Lane S. Categories for the working mathematician.* — Springer, 1978. 317 p.)
  12. Goguen J. Categorical foundations for general systems theory // Advances in cybernetics and systems research. — London: Transcripta Books, 1973. P. 121–130.
  13. Ковалёв С. П. Теоретико-категорный подход к метапрограммированию. — М.: ИПУ РАН, 2014. 112 с.
  14. Ковалёв С. П. Системный анализ жизненного цикла больших информационно-управляющих систем // Автоматика и телемеханика, 2013. № 9. С. 98–118.
  15. Sannella D. A survey of formal software development methods // Software engineering: A European prospective. — IEEE Computer Society Press, 1993. P. 281–297.
  16. Grid computing: Making the global infrastructure a reality. — New York, NY, USA: Wiley & Sons, 2003. 1060 p.
  17. Kiczales G., Lamping J., Mendhekar A., Maeda C., Lopes C. V., Loingtier J.-M., Irwin J. Aspect-oriented programming // ECOOP'97 — object-oriented programming / Eds. M. Aksit, S. Matsuoka. — Lecture notes in computer science ser. — Springer, 1997. Vol. 1241. P. 220–242.
  18. Colyer A., Clement A., Harley G., Webster M. Eclipse AspectJ: Aspect-oriented programming with AspectJ and the Eclipse AspectJ development tools. — Reading: Addison-Wesley, 2004. 504 p.
  19. Steimann F. The paradoxical success of aspect-oriented programming // Conference (International) OOPSLA'2006 Proceedings. — Portland, 2006. P. 481–497.
  20. Ковалёв С. П. Семантика аспектно-ориентированного моделирования данных и процессов // Информатика и её применения, 2013. Т. 7. Вып. 3. С. 70–80.
  21. Kannenberg A., Saiedian H. Why software requirements traceability remains a challenge // J. Defense Software Engineering, 2009. July/August. P. 14–19.
  22. Гребенюк Г. Г., Лубков Н. В., Нукишов С. М. Информационные аспекты управления муниципальным хозяйством. — М.: ЛЕНАНД, 2011. 320 с.
  23. Manset D., Verjus H., McClatchey R., Oquendo F. A formal architecture-centric model-driven approach for the automatic generation of Grid applications // 8th Conference (International) "Enterprise Information Systems: Databases and Information Systems Integration" Proceedings. — Paphos, Cyprus, 2006. P. 322–330.
  24. Uslar M., Schmedes T., Lucks A., Luhmann T., Winkels L., Appelrath H.-J. Interaction of EMS related systems by using the CIM standard // 2nd ICSC Symposium (International) on Information Technologies in Environmental Engineering ITEE 2005 Proceedings. — Magdeburg, 2005. P. 596–610.
  25. Воеводин В. В. Отображение проблем вычислительной математики на архитектуру вычислительных систем // Вычислительные методы и программирование, 2000. Т. 1. С. 37–44.
  26. Ковалёв С. П. Алгебраический подход к проектированию распределенных вычислительных систем // Сибирский журнал индустриальной математики, 2007. Т. 10. № 2. С. 70–84.
  27. Топорков В. В. Модели распределенных вычислений. — М.: Физматлит, 2004. 320 с.
  28. Ковалёв С. П. Повышение эффективности процессов проектирования больших информационно-управляющих систем // Тр. XII Всеросс. совещания по проблемам управления ВСПУ-2014. — М.: ИПУ РАН, 2014. С. 9291–9300.
  29. Ковалёв С. П., Андрюшкевич С. К., Гуськов А. Е. Интеграционная платформа учета и управления энергообеспечением «Энергиус». Свидетельство о государственной регистрации программы для ЭВМ № 2009613359 от 26.06.2009.
  30. Sommerville I. Инженерия программного обеспечения / Пер. с англ. — 6-е изд. — М.: Вильямс, 2002. 624 с. (*Sommerville I. Software engineering.* — 6th ed. — Addison Wesley, 2001. 720 p.)
  31. Allen R. J., Garland D. A formal basis for architectural connection // ACM Trans. Software Engineering Methodology, 1997. Vol. 6. No. 3. P. 213–249.
  32. Smith D. R. Composition by colimit and formal software development // Algebra, meaning, and computation / Eds. K. Futatsugi, J.-P. Jouannaud, J. Meseguer. — Lecture notes in computer science ser. — Springer, 2006. Vol. 4060. P. 317–332.
  33. Aspect-oriented software development. — Reading: Addison Wesley, 2004. 800 p.
  34. Pinto M., Fuentes L., Troya J. M. DAOP-ADL: An architecture description language for dynamic component and aspect-based development // Generative programming and component engineering / Eds. F. Pfenning, Y. Smaragkaris. — Lecture notes in computer science ser. — Springer, 2003. Vol. 2830. P. 118–137.
  35. Srinivas Y. V., Jüllig R. SPECWARE: Formal support for composing software // Mathematics of program construction / Ed. B. Möller. — Lecture notes in computer science ser. — Springer, 1995. Vol. 947. P. 399–422.

Поступила в редакцию 19.11.15

# METAPROGRAMMING TO INCREASE MANUFACTURABILITY OF LARGE-SCALE SOFTWARE-INTENSIVE SYSTEMS

S. P. Kovalyov

Institute of Control Sciences, Russian Academy of Sciences, 65 Profsoyuznaya Str., Moscow 117997, Russian Federation

**Abstract:** An approach to reduce costs of large-scale software-intensive systems design due to applying modern metaprogramming technologies is proposed. Model-driven engineering and aspect-oriented software development are considered to be the most advanced among such technologies. The methods to scale these technologies are presented in order to apply them efficiently under growth of the target system size via closure with regard to basic structural relations. Design of mathematical software for smart electric grids is considered as a case study for practical applications of the approach. Principles of mathematical device for constructing, analysis, and optimization of design technological procedures based on the category theory are described. The process to design the generator of computational software components of large-scale systems applying category-theoretical methods is drawn.

**Keywords:** large-scale software-intensive systems; metaprogramming; megamodel; category theory; colimit; model driven engineering; aspect-oriented software development; smart grid

**DOI:** 10.14357/19922264160105

## References

1. *Ultra-Large-Scale Systems: The software challenge of the future*. 2006. Pittsburgh: Carnegie Mellon Software Engineering Institute. 134 p.
2. Anvaari, M., D. S. Cruzes, and R. Conradi. 2012. Smart Grid software applications as an ultra-large-scale system: Challenges for evolution. *Innovative Smart Grid Technologies (ISGT)*. IEEE PES. 1–6.
3. Schmidt, D. C. 2006. Model-driven engineering. *IEEE Computer* 39(2):25–32.
4. Graaf, B., and A. van Deursen. 2007. Visualisation of domain-specific modelling languages using UML. *14th IEEE Conference (International) on the Engineering of Computer-Based Systems ECBS'2007 Proceedings*. Tucson. 586–595.
5. Benveniste, A., P. Caspi, S. A. Edwards, N. Halbwachs, P. Le Guernic, and R. de Simone. 2003. The synchronous languages 12 years later. *Proc. IEEE* 91(1):64–83.
6. Vasil'ev, S. N. 2008. Formalizatsiya znaniy i upravlenie na osnove pozitivno-obrazovannykh yazykov [Formalization of knowledge and control on the basis of positively-formed languages]. *Informatsionnye Tekhnologii i Vychislitel'nye Sistemy* [Information Technologies and Computer Systems] 1:3–17.
7. Jouault, F., B. Vanhooff, H. Bruneliere, G. Doux, Y. Berbers, and J. Bezin. 2010. Inter-DSL coordination support by combining megamodeling and model weaving. *ACM 2010 Symposium Applied Computing Proceedings*. Sierre. 2011–2018.
8. The Eclipse Foundation. 2014. Eclipse Modeling Project. Available at: <http://www.eclipse.org/modeling/> (accessed March 4, 2016).
9. Kolovos, D. S., L. M. Rose, N. Matragkas, R. F. Paige, E. Guerra, J. S. Cuadrado, J. De Lara, I. Ráth, D. Varró, M. Tisi, and J. Cabot. 2013. A research roadmap towards achieving scalability in model driven engineering. *Workshop on Scalability in Model Driven Engineering Proceedings*. Budapest, Hungary: ACM. 2:1–2:10.
10. Diskin, Z., and T. S. E. Maibaum. 2012. Category theory and model-driven engineering: From formal semantics to design patterns and beyond. *7th Workshop ACCAT'2012 "Electronic Proceedings in Theoretical Computer Science."* 93:1–21.
11. Mac Lane, S. 1978. *Categories for the working mathematician*. Springer. 317 p.
12. Goguen, J. 1973. Categorical foundations for general systems theory. *Advances in cybernetics and systems research*. London: Transcripta Books. 121–130.
13. Kovalyov, S. P. 2014. *Teoretiko-kategornyy podkhod k metaprogramirovaniyu* [Category-theoretical approach to metaprogramming]. Moscow: Institute of Control Sciences. 112 p.
14. Kovalev, S. P. 2013. Systems analysis of life cycle of large-scale information-control systems. *Automation Remote Control* 74(9):1510–1524.
15. Sannella, D. 1993. A survey of formal software development methods. *Software engineering: A European perspective*. IEEE Computer Society Press. 281–297.
16. *Grid computing: Making the global infrastructure a reality*. 2003. New York, NY: Wiley & Sons. 1060 p.
17. Kiczales, G., J. Lamping, A. Mendhekar, C. Maeda, C. V. Lopes, J.-M. Loingtier, and J. Irwin. 1997. Aspect-oriented programming. *ECOOP'97 — object-oriented programming*. Eds. M. Aksit and S. Matsuoka. Lecture notes in computer science ser. Springer. 1241:220–242.
18. Colyer, A., A. Clement, G. Harley, and M. Webster. 2004. *Eclipse AspectJ: Aspect-oriented programming with AspectJ and the Eclipse AspectJ development tools*. Reading: Addison-Wesley. 504 p.

19. Steimann, F. 2006. The paradoxical success of aspect-oriented programming. *Conference (International) OOPSLA'2006 Proceedings*. Portland. 481–497.
20. Kovalyov, S. P. 2013. Semantika aspektno-orientirovannogo modelirovaniya dannykh i protsessov [Semantics of aspect-oriented modeling of data and processes]. *Informatika i ee Primeneniya — Inform. Appl.* 7(3):70–80.
21. Kannenberg, A., and H. Saiedian. 2009. Why software requirements traceability remains a challenge. *J. Defense Software Engineering*. July/August:14–19.
22. Grebenjuk, G. G., N. V. Lubkov, and S. M. Nikishov. 2011. *Informatsionnye aspekty upravleniya munitsipal'nykh hozyaystvom* [Informational aspects of municipal infrastructure management]. Moscow: LENAND. 320 p.
23. Manset, D., H. Verjus, R. McClatchey, and F. Oquendo. 2006. A formal architecture-centric model-driven approach for the automatic generation of Grid applications. *8th Conference (International) "Enterprise Information Systems: Databases and Information Systems Integration" Proceedings*. Paphos, Cyprus. 322–330.
24. Uslar, M., T. Schmedes, A. Lucks, T. Luhmann, L. Winkels, and H.-J. Appelrath. 2005. Interaction of EMS related systems by using the CIM standard. *2nd ICSC Symposium (International) on Information Technologies in Environmental Engineering ITEE 2005 Proceedings*. Magdeburg. 596–610.
25. Voevodin, V. V. 2000. Otobrazhenie problem vychislitel'noy matematiki na arkhitekturu vychislitel'nykh sistem [Mapping of computational mathematics problems to computational systems architecture]. *Vychislitel'nye Metody i Programirovanie* [Numerical Methods and Programming] 1:37–44.
26. Kovalyov, S. P. 2007. Algebraicheskiy podkhod k proektirovaniyu raspredelennykh vychislitel'nykh sistem [Algebraic approach to distributed computation systems design]. *Sibirskiy Zh. Industrial'noy Matematiki* [Siberian J. Industrial Mathematics] 10(2):70–84.
27. Toporkov, V. V. 2004. *Modeli raspredelennykh vychisleniy* [Models of distributed computing]. Moscow: Fizmatlit. 320 p.
28. Kovalyov, S. P. 2014. Povyshenie effektivnosti protsessov proektirovaniya bol'shikh informatsionno-upravlyayushchikh sistem [Increasing efficiency of large-scale information-control systems design]. *Tr. XII Vseross. Soveshchaniya po Problemam Upravleniya VSPU-2014* [XII All-Russian Workshop on Control Problems VSPU-2014 Proceedings]. Moscow: Institute of Control Sciences. 9291–9300.
29. Kovalyov, S. P., S. K. Andriushkevich, and A. E. Gus'kov. June 26, 2009. Integratsionnaya platforma ucheta i upravleniya energoobespecheniem "Energius" [Integration platform for energy accounting and management "Energius"]. Certificate of state registration of computer program No. 2009613359.
30. Sommerville, I. 2001. *Software engineering*. 6th ed. Addison Wesley. 720 p.
31. Allen, R. J., and D. Garlan. 1997. A formal basis for architectural connection. *ACM Trans. Software Engineering Methodology* 6(3):213–249.
32. Smith, D. R. 2006. Composition by colimit and formal software development. *Algebra, meaning, and computation*. Eds. K. Futatsugi, J.-P. Jouannaud, and J. Meseguer. Lecture notes in computer science ser. Springer. 4060:317–332.
33. *Aspect-oriented software development*. 2004. Reading: Addison Wesley. 800 p.
34. Pinto, M., L. Fuentes, and J. M. Troya. 2003. DAOP-ADL: An architecture description language for dynamic component and aspect-based development. *Generative programming and component engineering*. Eds. F. Pfenning and Y. Smaragkaris. — Lecture notes in computer science ser. Springer. 2830:118–137.
35. Srinivas, Y. V., and R. Jüllig. 1995. SPECWARE: Formal support for composing software. *Mathematics of program construction*. Ed. B. Möller. Lecture notes in computer science ser. Springer. 947:399–422.

*Received November 19, 2015*

## Contributor

**Kovalyov Sergey P.** (b. 1972) — Doctor of Science in physics and mathematics, leading scientist, Institute of Control Problems, Russian Academy of Sciences, 65 Profsoyuznaya Str., Moscow 117997, Russian Federation; kovalyov@nm.ru

# БАЙЕСОВСКИЕ МОДЕЛИ МАССОВОГО ОБСЛУЖИВАНИЯ И НАДЕЖНОСТИ: АПРИОРНЫЕ РАСПРЕДЕЛЕНИЯ С КОМПАКТНЫМ НОСИТЕЛЕМ\*

А. А. Кудрявцев<sup>1</sup>

**Аннотация:** Данная работа является очередной в серии статей, посвященных изучению байесовских моделей массового обслуживания и надежности. В работе приводятся соотношения для функции распределения и плотности частного  $\rho$  независимых случайных величин, имеющих априорные распределения с компактным носителем, которые интерпретируются как параметр, «препятствующий» функционированию системы, и параметр, «способствующий» функционированию. Описание жизненного цикла многих реальных систем осуществляется в терминах  $\rho$ , например в теории массового обслуживания  $\rho$  называется параметром загрузки системы и входит во многие формулы, описывающие разнообразные характеристики. Рассматриваются частные случаи априорных распределений с компактным носителем, для которых плотности имеют полиномиальный или кусочно-полиномиальный вид.

**Ключевые слова:** байесовский подход; системы массового обслуживания; надежность; смешанные распределения; распределения с компактным носителем

**DOI:** 10.14357/19922264160106

## 1 Введение

Во многих областях исследования математических моделей функционирования реальных систем аналитические результаты, характеризующие жизненный цикл рассматриваемых объектов, так или иначе зависят от параметров, «способствующих» функционированию системы и «препятствующих» функционированию. Так, в моделях массового обслуживания к параметрам, «способствующим» функционированию, можно отнести интенсивность обслуживания запросов, а к параметрам, «препятствующим» функционированию, — интенсивность входящего потока требований.

Аналогично в теории надежности параметр «эффективности» средства, исправляющего ошибки в системе, «способствует» функционированию, а параметр «дефективности» — «препятствует». Очевидно, что итоговые результаты работы системы зависят не столько от значений самих параметров, влияющих на функционирование, сколько от их отношения. В общем случае такое отношение можно назвать «коэффициентом баланса системы».

Хорошо известно, что одним из основных показателей при изучении моделей массового обслуживания  $M|M|1$  является коэффициент загрузки системы  $\rho$ , равный отношению параметра входящего потока  $\lambda$  к параметру обслуживания  $\mu$ . От значения  $\rho$  зависит наличие стационарного режи-

ма у рассматриваемой системы; величина  $\rho$  входит во многие формулы, описывающие характеристики разнообразных систем массового обслуживания. По своей сути коэффициент загрузки является коэффициентом баланса, характеризующим систему: функционирование системы тем эффективнее, чем ближе к нулю значения  $\rho$ . При достаточно больших значениях  $\rho$  ( $\rho \geq 1$ ) система работает столь неэффективно, что, например, среднее число заявок в системе  $M|M|1|_{\infty}$  считается равным бесконечности.

В рекуррентных моделях роста надежности удобно рассматривать коэффициент баланса  $\rho = \lambda/\mu$ , где  $\lambda$  — параметра «эффективности» средства, исправляющего ошибки в системе, а  $\mu$  — параметр «дефективности». При этом, в отличие от теории массового обслуживания, система тем надежнее, чем больше значение  $\rho$ .

Байесовский подход к задачам массового обслуживания и надежности предполагает рандомизацию параметров  $\lambda$  и  $\mu$ .

Подробное описание предпосылок для исследования, особенностей и библиографии байесовских моделей в теории массового обслуживания и надежности можно найти в книге [1]. В основе всех результатов для байесовских моделей из [1] лежит вероятностное распределение коэффициента баланса  $\rho$ . Принципиальное отличие байесовских

\* Исследование выполнено при поддержке Российского научного фонда (проект 14-11-00397).

<sup>1</sup>Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, nubigena@mail.ru

постановок задач в указанных теориях заключается в том, что в задачах теории надежности коэффициенты «эффективности» и «дефективности» соответственно имеют смысл рандомизированных вероятностей исправления ошибки в системе и внесения новой ошибки, а следовательно, соответствующие вероятностные распределения должны быть подмножествами единичного отрезка. В теории массового обслуживания для систем  $M|M|1$  параметры входящего потока и обслуживания должны быть положительными числами, поэтому соответствующие распределения имеют лишь нижнее ограничение в нуле.

Далее приводятся результаты для распределения величины  $\rho = \lambda/\mu$  в случае, когда носителями распределений  $\lambda$  и  $\mu$  являются отрезки на положительной полупрямой. При применении изложенных ниже результатов к надежностным постановкам необходимо ограничивать правые концы носителей распределений единицей.

## 2 Основные результаты

Пусть  $\lambda$  и  $\mu$  — независимые абсолютно непрерывные случайные величины, причем  $P(\lambda \in [a_\lambda, b_\lambda]) = 1$ ,  $0 < a_\lambda < b_\lambda$ , и не существует множества  $S \subset [a_\lambda, b_\lambda]$  положительной меры Лебега такого, что  $P(\lambda \in S) = 0$ , а для случайной величины  $\mu$  выполнены аналогичные требования с параметрами  $a_\mu$  и  $b_\mu$ . Плотности случайных величин  $\lambda$  и  $\mu$  обозначим через  $f_\lambda(x)$  и  $f_\mu(x)$  соответственно. Во всех последующих выкладках будем предполагать, что  $x > 0$ .

Найдем функцию распределения  $F_\rho(x)$  и плотность  $f_\rho(x)$  случайной величины  $\rho = \lambda/\mu$ . Имеем

$$F_\rho(x) = \int_{-\infty}^{+\infty} P(\lambda < xy) dP(\mu < y) = \int_{a_\mu}^{b_\mu} \left[ \int_{a_\lambda}^{xy} f_\lambda(u) du \cdot \mathbf{1}(a_\lambda \leq xy \leq b_\lambda) + \mathbf{1}(xy > b_\lambda) \right] f_\mu(y) dy.$$

Рассмотрим всевозможные комбинации расположения точек  $a_\mu$ ,  $b_\mu$ ,  $a_\lambda/x$  и  $b_\lambda/x$  на прямой. При этом существенную роль играет взаимное расположение точек  $a_\lambda/a_\mu$  и  $b_\lambda/b_\mu$ . Так, при  $a_\mu < a_\lambda/x < b_\mu < b_\lambda/x$  и  $a_\lambda/a_\mu < b_\lambda/b_\mu$

$$F_\rho(x) = \int_{a_\lambda/x}^{b_\mu} \int_{a_\lambda}^{xy} f_\lambda(u) f_\mu(y) dudy.$$

Аналогично рассмотрев остальные случаи, убеждаемся в справедливости следующего утверждения.

**Теорема 1.** Пусть независимые абсолютно непрерывные случайные величины  $\lambda$  и  $\mu$  имеют соответственно носители распределений  $[a_\lambda, b_\lambda]$  и  $[a_\mu, b_\mu]$ ,  $0 < a_\lambda < b_\lambda$ ,  $0 < a_\mu < b_\mu$ , и плотности  $f_\lambda(x)$  и  $f_\mu(x)$ . Тогда случайная величина  $\rho = \lambda/\mu$  имеет функцию распределения:

$$F_\rho(x) = \mathbf{1}\left(\frac{a_\lambda}{b_\mu} < x \leq \min\left\{\frac{a_\lambda}{a_\mu}, \frac{b_\lambda}{b_\mu}\right\}\right) \times \int_{a_\lambda/x}^{b_\mu} \int_{a_\lambda}^{xy} f_\lambda(u) f_\mu(y) dudy + \mathbf{1}\left(\frac{a_\lambda}{a_\mu} < x \leq \frac{b_\lambda}{b_\mu}\right) \times \int_{a_\mu}^{b_\mu} \int_{a_\lambda}^{xy} f_\lambda(u) f_\mu(y) dudy + \mathbf{1}\left(\frac{b_\lambda}{b_\mu} < x \leq \frac{a_\lambda}{a_\mu}\right) \times \left[ \int_{a_\lambda/x}^{b_\lambda/x} \int_{a_\lambda}^{xy} f_\lambda(u) f_\mu(y) dudy + \int_{b_\lambda/x}^{b_\mu} f_\mu(x) dy \right] + \mathbf{1}\left(\max\left\{\frac{a_\lambda}{a_\mu}, \frac{b_\lambda}{b_\mu}\right\} < x \leq \frac{b_\lambda}{a_\mu}\right) \times \left[ \int_{a_\mu}^{b_\lambda/x} \int_{a_\lambda}^{xy} f_\lambda(u) f_\mu(y) dudy + \int_{b_\lambda/x}^{b_\mu} f_\mu(y) dy \right] + \mathbf{1}\left(x > \frac{b_\lambda}{a_\mu}\right).$$

Для нахождения плотности случайной величины  $\rho$  достаточно воспользоваться соотношением

$$f_\rho(x) = \int_{-\infty}^{+\infty} y f_\lambda(xy) f_\mu(y) dy$$

и рассуждениями, приведенными выше, для функции распределения  $F_\rho(x)$ .

**Теорема 2.** Пусть независимые абсолютно непрерывные случайные величины  $\lambda$  и  $\mu$  имеют соответственно носители распределений  $[a_\lambda, b_\lambda]$  и  $[a_\mu, b_\mu]$ ,  $0 < a_\lambda < b_\lambda$ ,  $0 < a_\mu < b_\mu$ , и плотности  $f_\lambda(x)$  и  $f_\mu(x)$ . Тогда случайная величина  $\rho = \lambda/\mu$  имеет плотность распределения

$$f_\rho(x) = \mathbf{1}\left(\frac{a_\lambda}{b_\mu} < x \leq \min\left\{\frac{a_\lambda}{a_\mu}, \frac{b_\lambda}{b_\mu}\right\}\right) \times \int_{a_\lambda/x}^{b_\mu} y f_\lambda(xy) f_\mu(y) dy +$$

$$\begin{aligned}
 & + \mathbf{1} \left( \frac{a_\lambda}{a_\mu} < x \leq \frac{b_\lambda}{b_\mu} \right) \int_{a_\mu}^{b_\mu} y f_\lambda(xy) f_\mu(y) dy + \\
 & + \mathbf{1} \left( \frac{b_\lambda}{b_\mu} < x \leq \frac{a_\lambda}{a_\mu} \right) \int_{a_\lambda/x}^{b_\lambda/x} y f_\lambda(xy) f_\mu(y) dy + \\
 & + \mathbf{1} \left( \max \left\{ \frac{a_\lambda}{a_\mu}, \frac{b_\lambda}{b_\mu} \right\} < x \leq \frac{b_\lambda}{a_\mu} \right) \int_{a_\mu}^{b_\lambda/x} y f_\lambda(xy) f_\mu(y) dy.
 \end{aligned}$$

**Замечание 1.** Применительно к байесовским моделям массового обслуживания  $M|M|1$  теоремы 1 и 2 позволяют не только находить распределение коэффициента загрузки  $\rho$ , но и распределения таких характеристик, как вероятность «непотери» вызова  $\pi = 1/(1 + \rho)$  в модели  $M|M|1|0$  или среднее число требований в очереди  $N = \rho/(1 - \rho)$  в модели  $M|M|1|\infty$  и др. Приведенные теоремы также упрощают поиск средней надежности системы, поскольку для рекуррентных моделей байесовской теории надежности априорные распределения всегда ограничены и являются подмножествами единичного отрезка.

**Замечание 2.** В формулировке теоремы 1 вместо плотности случайной величины  $\lambda$  можно использовать ее функцию распределения, учитывая

$$\int_{a_\lambda}^{xy} f_\lambda(u) du = F_\lambda(xy), \quad y \in \left[ \frac{a_\lambda}{x}, \frac{b_\lambda}{x} \right].$$

**Замечание 3.** В формулировках теорем 1 и 2 один из индикаторов  $\mathbf{1}(a_\lambda/a_\mu < x \leq b_\lambda/b_\mu)$  и  $\mathbf{1}(b_\lambda/b_\mu < x \leq a_\lambda/a_\mu)$  всегда равен нулю в зависимости от взаимного расположения точек  $a_\lambda/a_\mu$  и  $b_\lambda/b_\mu$ .

Отдельный интерес в классе распределений с компактным носителем (для некоторой случайной величины  $\xi$ ) представляют распределения, плотности которых могут быть представлены в виде полинома:

$$f_\xi(x) = \sum_{i=0}^{n_\xi} c_{\xi,i} x^i \cdot \mathbf{1}(x \in [a_\xi, b_\xi]). \quad (1)$$

К таким распределениям, в частности, относятся равномерное (при  $n_\xi = 0$ ) и параболическое (при  $n_\xi = 2$ ) распределения.

Пусть случайные величины  $\lambda$  и  $\mu$  удовлетворяют условиям теоремы 2, а их плотности — соотношению (1) с соответствующими параметрами. Для некоторых  $a$  и  $b$ , одновременно принадлежащих отрезкам  $[a_\mu, b_\mu]$  и  $[a_\lambda/x, b_\lambda/x]$ , определим

$$\begin{aligned}
 I(a, b, x) &= \int_a^b y f_\lambda(xy) f_\mu(y) dy = \\
 &= \int_a^b \sum_{i=0}^{n_\lambda} \sum_{j=0}^{n_\mu} c_{\lambda,i} c_{\mu,j} x^i y^{i+j+1} dy = \\
 &= \sum_{i=0}^{n_\lambda} \sum_{j=0}^{n_\mu} c_{\lambda,i} c_{\mu,j} \frac{b^{i+j+2} - a^{i+j+2}}{i+j+2} x^i. \quad (2)
 \end{aligned}$$

Теорема 2 дает возможность сформулировать следующее утверждение для распределений с плотностями, имеющими полиномиальный вид.

**Следствие 1.** Пусть распределения случайных величин  $\lambda$  и  $\mu$  удовлетворяют условиям теоремы 2 и соотношению (1) с соответствующими параметрами. Тогда случайная величина  $\rho = \lambda/\mu$  имеет плотность:

$$\begin{aligned}
 f_\rho(x) &= \\
 &= \mathbf{1} \left( \frac{a_\lambda}{b_\mu} < x \leq \min \left\{ \frac{a_\lambda}{a_\mu}, \frac{b_\lambda}{b_\mu} \right\} \right) I \left( \frac{a_\lambda}{x}, b_\mu, x \right) + \\
 &+ \mathbf{1} \left( \frac{a_\lambda}{a_\mu} < x \leq \frac{b_\lambda}{b_\mu} \right) I(a_\mu, b_\mu, x) + \\
 &+ \mathbf{1} \left( \frac{b_\lambda}{b_\mu} < x \leq \frac{a_\lambda}{a_\mu} \right) I \left( \frac{a_\lambda}{x}, \frac{b_\lambda}{x}, x \right) + \\
 &+ \mathbf{1} \left( \max \left\{ \frac{a_\lambda}{a_\mu}, \frac{b_\lambda}{b_\mu} \right\} < x \leq \frac{b_\lambda}{a_\mu} \right) I \left( a_\mu, \frac{b_\lambda}{x}, x \right), \quad (3)
 \end{aligned}$$

где величины  $I(a, b, x)$  определены соотношением (2).

Следствие 1 дает возможность вычислить моменты

$$\mathbb{E} \rho^k = \int x^k f_\rho(x) dx, \quad k = 1, 2, \dots,$$

для случая, когда  $f_\rho(x)$  имеет вид (3). Будем полагать для определенности, что  $c_{\mu,j} = 0$  при  $j > n_\mu$ . Введем обозначение:

$$\begin{aligned}
 J(d) &= \sum_{i=0}^{n_\lambda} \sum_{j=0}^{n_\mu} \frac{c_{\lambda,i} c_{\mu,j} d^{i+j+2}}{i+j+2} \int_{d/b_\mu}^{d/a_\mu} x^{k-j-2} dx = \\
 &= \sum_{i=0}^{n_\lambda} \frac{c_{\lambda,i} c_{\mu,k-1} d^{k+i+1}}{k+i+1} \ln \frac{b_\mu}{a_\mu} + \\
 &+ \sum_{i=0}^{n_\lambda} \sum_{j=0}^{n_\mu} \mathbf{1}(j \neq k-1) \times \\
 &\times \frac{c_{\lambda,i} c_{\mu,j} (b_\mu^{k-j-1} - a_\mu^{k-j-1}) d^{k+i+1}}{(i+j+2)(k-j-1) a_\mu^{k-j-1} b_\mu^{k-j-1}}. \quad (4)
 \end{aligned}$$

Непосредственно из следствия 1 вытекает следующее утверждение.

**Следствие 2.** Пусть распределения случайных величин  $\lambda$  и  $\mu$  удовлетворяют условиям теоремы 2 и соотношению (1) с соответствующими параметрами. Тогда моменты  $k$ -го порядка ( $k = 1, 2, \dots$ ) случайной величины  $\rho = \lambda/\mu$  имеют вид:

$$E\rho^k = \sum_{i=0}^{n_\lambda} \sum_{j=0}^{n_\mu} c_{\lambda,i} c_{\mu,j} (b_\lambda^{k+i+1} - a_\lambda^{k+i+1}) (a_\mu^{k-j-1} - b_\mu^{k-j-1}) / ((i+j+2)(k+i+1)a_\mu^{k-j-1}b_\mu^{k-j-1}) + J(b_\lambda) - J(a_\lambda),$$

где величины  $J(d)$  определены соотношением (4).

**Замечание 4.** Следствия 1 и 2 из теоремы 2 обобщают некоторые полученные ранее результаты (см., например, [1]).

**Замечание 5.** На величину моментов  $E\rho^k$  не влияет взаимное расположение точек  $a_\lambda/a_\mu$  и  $b_\lambda/b_\mu$ .

Изложенный выше метод дает возможность находить характеристики распределения частного случайных величин, плотности которых имеют полиномиальный вид не на всем носителе, а на некоторых его подмножествах. Рассмотрим следующее разбиение отрезка  $[a_\xi, b_\xi]$ :

$$0 < a_\xi = a_\xi^{(1)} < b_\xi^{(1)} = a_\xi^{(2)} < \dots < b_\xi^{(l_\xi)} = b_\xi.$$

Пусть плотность некоторой случайной величины  $\xi$ , носителем распределения которой является отрезок  $[a_\xi, b_\xi]$ , имеет кусочно-полиномиальный вид:

$$f_\xi(x) = \sum_{l=1}^{l_\xi} \sum_{i=0}^{n_\xi} c_{\xi,i,l} x^i \cdot \mathbf{1}(x \in [a_\xi^{(l)}, b_\xi^{(l)}]). \quad (5)$$

Частными случаями таких распределений являются распределение Симпсона (при  $l_\xi = 2$  и  $n_\xi = 1$ ) и трапецидальное распределение (при  $l_\xi = 3$  и  $n_\xi = 1$ ).

**Замечание 6.** Для распределений случайных величин  $\lambda$  и  $\mu$ , удовлетворяющих условиям теоремы 2 и соотношению (5) с соответствующими параметрами, при каждом конкретном наборе параметров  $n_\lambda, n_\mu, l_\lambda$  и  $l_\mu$  несложно сформулировать утверждения, аналогичные следствиям из теоремы 2, в которых величины  $I(a, b, x)$  будут определяться соотношением:

$$I(a, b, x) = \int_a^b \sum_{l=1}^{l_\lambda} \sum_{i=0}^{n_\lambda} \sum_{m=1}^{l_\mu} \sum_{j=0}^{n_\mu} c_{\lambda,i,l} c_{\mu,j,m} x^i y^{i+j+1} \times \mathbf{1}(y \in [a_\lambda^{(l)}/x, b_\lambda^{(l)}/x]) \mathbf{1}(y \in [a_\mu^{(m)}, b_\mu^{(m)}]) dy.$$

## Литература

1. Кудрявцев А. А., Шоргин С. Я. Байесовские модели в теории массового обслуживания и надежности. — М.: ФИЦ ИУ РАН, 2015. 76 с.

Поступила в редакцию 17.01.16

## BAYESIAN QUEUEING AND RELIABILITY MODELS: A PRIORI DISTRIBUTIONS WITH COMPACT SUPPORT

A. A. Kudryavtsev<sup>1,2</sup>

<sup>1</sup>Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M.V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation

<sup>2</sup>Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** This work is the latest in a series of articles devoted to the study of Bayesian queueing and reliability models. The paper presents relations for the distribution function and the density of the quotient  $\rho$  of independent random variables with *a priori* distributions with compact support, which are interpreted as a parameter “obstructing” the functioning of the system and a parameter “conducting” to the functioning of the system. Description of the life cycle of many real systems is carried out in terms of  $\rho$ ; for example, in the queueing theory, parameter  $\rho$  is called

the “system load factor” and is a part of many formulas that describe various characteristics. The paper considers particular cases of *a priori* distributions with compact support for which densities have polynomial or piecewise polynomial form.

**Keywords:** Bayesian approach; mass service theory; reliability theory; mixed distributions; distributions with compact support

**DOI:** 10.14357/19922264160106

## Acknowledgments

This work was financially supported by the Russian Science Foundation (grant No. 14-11-00397).

## References

1. Kudryavtsev, A. A., and S. Ya. Shorgin. 2015. *Bayesovskie modeli v teorii massovogo obsluzhivaniya i nadezhnosti* [Bayesian models in mass service and reliability theories]. Moscow: Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 2015. 76 p.

*Received January 17, 2016*

## Contributor

**Kudryavtsev Alexey A.** (b. 1978) — Candidate of Sciences (PhD) in physics and mathematics, associate professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; nubigena@mail.ru

# КОНЦЕПЦИЯ ОНЛАЙН-КОМПЛЕКСА ДЛЯ СТОХАСТИЧЕСКОГО МОДЕЛИРОВАНИЯ РЕАЛЬНЫХ ПРОЦЕССОВ\*

А. К. Горшенин<sup>1</sup>

**Аннотация:** Анализ информационных потоков с использованием разнообразных вероятностных моделей достаточно широко распространен в различных прикладных областях. В статье описаны основные принципы построения новой онлайн-системы стохастического моделирования реальных процессов, не имеющей прямых аналогов в силу универсальности предлагаемого набора методов, а также выбранной концепции реализации в виде интернет-ресурса, который позволит конечному пользователю не заботиться о соответствии технических характеристик своего компьютера каким-либо специальным требованиям, а сразу загружать данные на сервер и обрабатывать их.

**Ключевые слова:** смеси вероятностных распределений; метод скользящего разделения смесей; интеллектуальный анализ данных; онлайн-комплекс; матричные вычисления

**DOI:** 10.14357/19922264160107

## 1 Введение

При практическом решении задачи моделирования и анализа нестационарных информационных потоков в современных информационных, телекоммуникационных и вычислительных системах и сетях ключевым этапом является разделение смесей вероятностных распределений, т.е. статистическое оценивание неизвестных параметров. При этом в качестве базовых аппроксимирующих моделей в рамках так называемого метода скользящего разделения смесей (СРС-метод) [1] могут рассматриваться как конечные сдвиг-масштабные смеси нормальных и гамма-распределений [1], так и непрерывные распределения, например дисперсионно-сдвиговые смеси нормальных законов [2]. Основная идея СРС-метода заключается в специальном разбиении исходной выборки на подвыборки (окна) и дальнейшем анализе поведения данных на каждом окне. Это позволяет отслеживать появление и исчезновение формирующих структур в изучаемых процессах во времени.

Можно привести широкий спектр примеров успешного использования вероятностных моделей для реальных данных различной природы. Например, статьи [3–5] предлагают анализ финансовых и физических рядов, а работа [6] ориентирована на анализ биомедицинских сигналов. В статье [7] проводится анализ данных фондового индекса S&P500 на основе симметричных масштабных смесей нормальных распределений, а работа [8] посвящена

изучению курса обмена австралийского доллара с использованием представлений распределений Стьюдента в виде масштабной смеси нормальных законов с применением специальной программы WinBUGS для проведения вычислений.

Однако зачастую методы реализованы в виде программ для специализированных пакетов (в частности, MATLAB), что накладывает ряд ограничений на удобство распространения и использования. Во-первых, конечный пользователь должен иметь доступ к дорогостоящему программному обеспечению и опыт работы с ним. Во-вторых, далеко не все алгоритмы могут быть запущены на любой физической архитектуре. К таковым можно отнести, например, решения, предназначенные для выполнения вычислений с помощью CUDA (Compute Unified Device Architecture). У конечного пользователя может оказаться графическая карта от другого производителя либо его GPU (graphics processing unit) не предназначены для осуществления вычислений на должном уровне. В-третьих, могут возникнуть различные тонкости в использовании разработанных третьими лицами программ, например связанные с лицензиями.

Все это приводит к идее создания специализированного сервиса с онлайн-доступом, который поможет преодолеть указанные трудности.

В настоящей статье рассмотрим требования, которым должна удовлетворять такая система. В частности, будут описаны функциональные возможности специализированного комплекса, разра-

\* Работа выполнена при поддержке РФФИ (проект 15-37-20851 мол.а.вед).

<sup>1</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Московский технологический университет (МИРЭА), agorshenin@frccsc.ru

ботанного автором на языке программирования пакета MATLAB, который выступает прообразом функциональных возможностей, предназначенных для реализации в рамках сервиса с онлайн-доступом. Кроме того, значительное внимание уделяется получению явных формул для ряда моментных характеристик, в том числе в матричном представлении, для повышения производительности вычислений.

## 2 Матричные представления некоторых моментных характеристик конечных смесей нормальных законов

Пусть случайная величина  $Z_t$  имеет функцию распределения  $F_Z(x, t)$ , представимую в виде конечной сдвиг-масштабной смеси нормальных законов, а именно:

$$F_Z(x, t) = \sum_{i=1}^{k(t)} \frac{p_i(t)}{\sigma_i(t)\sqrt{2\pi}} \int_{-\infty}^x \exp\left\{-\frac{(y - a_i(t))^2}{2\sigma_i^2(t)}\right\} dy, \\ \forall x \in \mathbb{R}, \quad a_i(t) \in \mathbb{R}, \quad \sigma_i(t) > 0, \quad i = 1, \dots, k(t), \\ \sum_{i=1}^{k(t)} p_i(t) = 1, \quad p_i(t) \geq 0. \quad (1)$$

Параметр  $t$  здесь обозначает время (номер окна) для СРС-метода и подчеркивает тот факт, что распределение изменяется для каждого положения окна (возможно, весьма существенным образом). В частности, параметр  $k(t)$ , описывающий число компонент в смеси вида (1), при аппроксимации реальных данных может принимать различные значения с течением времени (см., например, статьи [3, 5], при этом вопрос статистической значимости дополнительных компонент обсуждается в работе [9]). Это обстоятельство значительно затрудняет прогнозирование, так как появление новых компонент в процессе зачастую объясняется новыми факторами, которые отсутствовали при построении первоначальной модели и, следовательно, не могли быть учтены. Поэтому необходимо перейти к рассмотрению некоторых «интегральных» характеристик, которые можно рассчитать для любого распределения вида (1), независимо от конкретных значений параметров.

В качестве таких величин можно использовать моменты различных порядков, а также производные от них коэффициенты асимметрии и эксцесса. Ниже получим соответствующие выражения для

случайной величины с функцией распределения (1) для первых четырех моментов.

Известно, что для начальных моментов случайной величины  $X$  с нормальным распределением с параметрами  $a$  и  $\sigma^2$  (т.е.  $X \sim N(a, \sigma^2)$ ) справедливы следующие соотношения:

$$\mathbb{E}X^m = \begin{cases} a^2 + \sigma^2, & m = 2; \\ a^3 + 3a\sigma^2, & m = 3; \\ a^4 + 6a^2\sigma^2 + 3\sigma^4, & m = 4. \end{cases} \quad (2)$$

Для начальных моментов случайной величины  $Z_t$  с функцией распределения, задаваемой формулой (1), по определению имеем:

$$\mathbb{E}Z_t^m = \int_{-\infty}^{+\infty} z^m dF_Z(z, t) = \sum_{i=1}^{k(t)} p_i(t) \frac{1}{\sigma_i(t)\sqrt{2\pi}} \times \\ \times \int_{-\infty}^{+\infty} z^m \exp\left\{-\frac{(z - a_i(t))^2}{2\sigma_i^2(t)}\right\} dz = \\ = \sum_{i=1}^{k(t)} p_i(t) \mathbb{E}X_i^m, \quad X_i \sim N(a_i(t), \sigma_i^2(t)).$$

Тогда для начальных моментов случайной величины  $Z_t$  справедлив следующий аналог выражений (2):

$$\mathbb{E}Z_t^m = \begin{cases} \sum_{i=1}^{k(t)} p_i(t) a_i(t), & m = 1; \\ \sum_{i=1}^{k(t)} p_i(t) (a_i^2(t) + \sigma_i^2(t)), & m = 2; \\ \sum_{i=1}^{k(t)} p_i(t) (a_i^3(t) + 3a_i\sigma_i^2(t)), & m = 3; \\ \sum_{i=1}^{k(t)} p_i(t) (a_i^4(t) + 6a_i^2\sigma_i^2(t) + 3\sigma_i^4(t)), & m = 4. \end{cases} \quad (3)$$

Воспользуемся этими выражениями для получения явных формул для дисперсии, а также коэффициентов асимметрии и эксцесса, зависящих только от параметров распределения, а именно: величин  $p_i(t)$ ,  $a_i(t)$  и  $\sigma_i(t)$ .

Сразу отметим, что математическое ожидание случайной величины  $Z_t$ , согласно первой строке в формуле (3), имеет вид:

$$\mathbb{E}Z_t = \sum_{i=1}^{k(t)} p_i(t) a_i(t). \quad (4)$$

## 2.1 Дисперсия

Дисперсия случайной величины  $Z_t$  с функцией распределения  $F_Z(x, t)$ , определяемой выражением (1), имеет вид:

$$\mathbb{D}Z_t = \mathbb{E}Z_t^2 - (\mathbb{E}Z_t)^2. \quad (5)$$

Пользуясь выражениями (3) и проводя необходимые преобразования, получим:

$$\begin{aligned} \mathbb{D}Z_t &= \int_{-\infty}^{+\infty} z^2 dF_Z(z, t) - \left( \sum_{i=1}^{k(t)} p_i(t) a_i(t) \right)^2 = \\ &= \sum_{i=1}^{k(t)} \frac{p_i(t)}{\sigma_i(t) \sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 \exp \left\{ -\frac{(z - a_i(t))^2}{2\sigma_i^2(t)} \right\} dz - \\ &\quad - \left( \sum_{i=1}^{k(t)} p_i(t) a_i(t) \right)^2 = \\ &= \sum_{i=1}^{k(t)} p_i(t) a_i^2(t) - \left( \sum_{i=1}^{k(t)} p_i(t) a_i(t) \right)^2 + \sum_{i=1}^{k(t)} p_i(t) \sigma_i^2(t) = \\ &= \sum_{i=1}^{k(t)} p_i(t) \left( a_i(t) - \sum_{i=1}^{k(t)} p_i(t) a_i(t) \right)^2 + \\ &\quad + \sum_{i=1}^{k(t)} p_i(t) \sigma_i^2(t). \quad (6) \end{aligned}$$

Заключительное представление в формуле (6) легко получить путем раскрытия квадрата в первой сумме и приведения подобных слагаемых (например, именно в таком виде дисперсия приводится в книге [1]).

## 2.2 Коэффициент асимметрии

Общий вид коэффициента асимметрии случайной величины  $Z_t$  с функцией распределения  $F_Z(x, t)$  (1) задается следующим выражением:

$$\gamma_{Z,t} = \frac{\mathbb{E}(Z_t - \mathbb{E}Z_t)^3}{(\mathbb{D}Z_t)^{3/2}}.$$

Выпишем отдельно выражение для числителя дроби, указанной выше:

$$\begin{aligned} \mathbb{E}(Z_t - \mathbb{E}Z_t)^3 &= \\ &= \mathbb{E}Z_t^3 - 3\mathbb{E}Z_t \cdot \mathbb{E}Z_t^2 + 3(\mathbb{E}Z_t)^3 - (\mathbb{E}Z_t)^3 = \\ &= \mathbb{E}Z_t^3 - 3\mathbb{E}Z_t \cdot (\mathbb{E}Z_t^2 - (\mathbb{E}Z_t)^2) - (\mathbb{E}Z_t)^3 = \\ &= \mathbb{E}Z_t^3 - 3\mathbb{E}Z_t \cdot \mathbb{D}Z_t - (\mathbb{E}Z_t)^3. \end{aligned}$$

Тогда коэффициент асимметрии может быть представлен в виде:

$$\gamma_{Z,t} = \frac{\mathbb{E}Z_t^3 - 3\mathbb{E}Z_t \cdot \mathbb{D}Z_t - (\mathbb{E}Z_t)^3}{(\mathbb{D}Z_t)^{3/2}}. \quad (7)$$

Воспользуемся формулами (2) и (6). Имеем:

$$\begin{aligned} \mathbb{E}(Z_t - \mathbb{E}Z_t)^3 &= \\ &= \sum_{i=1}^{k(t)} \frac{p_i(t)}{\sigma_i(t) \sqrt{2\pi}} \int_{-\infty}^{+\infty} z^3 \exp \left\{ -\frac{(z - a_i(t))^2}{2\sigma_i^2(t)} \right\} dz - \\ &\quad - 3\mathbb{E}Z_t \cdot \mathbb{D}Z_t - (\mathbb{E}Z_t)^3 = \\ &= \sum_{i=1}^{k(t)} p_i(t) (a_i^3(t) + 3a_i \sigma_i^2(t)) - 3\mathbb{E}Z_t \cdot \mathbb{D}Z_t - (\mathbb{E}Z_t)^3 = \\ &= \sum_{i=1}^{k(t)} p_i(t) (a_i^3(t) + 3a_i \sigma_i^2(t)) - \left( \sum_{i=1}^{k(t)} p_i(t) a_i(t) \right)^3 \times \\ &\quad \times \left( 3 \sum_{i=1}^{k(t)} p_i(t) \left( a_i(t) - \sum_{i=1}^{k(t)} p_i(t) a_i(t) \right)^2 + \right. \\ &\quad \left. + 3 \sum_{i=1}^{k(t)} p_i(t) \sigma_i^2(t) - \left( \sum_{i=1}^{k(t)} p_i(t) a_i(t) \right)^2 \right). \end{aligned}$$

Окончательно получим следующее выражение для коэффициента асимметрии:

$$\begin{aligned} \gamma_{Z,t} &= \\ &= \left[ \sum_{i=1}^{k(t)} p_i(t) (a_i^3(t) + 3a_i \sigma_i^2(t)) - \left( \sum_{i=1}^{k(t)} p_i(t) a_i(t) \right)^3 \times \right. \\ &\quad \times \left( 3 \sum_{i=1}^{k(t)} p_i(t) \left( a_i(t) - \sum_{i=1}^{k(t)} p_i(t) a_i(t) \right)^2 + \right. \\ &\quad \left. \left. + 3 \sum_{i=1}^{k(t)} p_i(t) \sigma_i^2(t) - \left( \sum_{i=1}^{k(t)} p_i(t) a_i(t) \right)^2 \right) \right] \times \\ &\quad \times \left[ \sum_{i=1}^{k(t)} p_i(t) \left( a_i(t) - \sum_{i=1}^{k(t)} p_i(t) a_i(t) \right)^2 + \right. \\ &\quad \left. + \sum_{i=1}^{k(t)} p_i(t) \sigma_i^2(t) \right]^{-3/2}. \quad (8) \end{aligned}$$

### 2.3 Коэффициент эксцесса

Общий вид коэффициента эксцесса случайной величины  $Z_t$  с функцией распределения  $F_Z(x, t)$ , определяемой выражением (1), задается следующим выражением:

$$\kappa_{Z, t} = \frac{\mathbb{E}(Z_t - \mathbb{E}Z_t)^4}{(\mathbb{D}Z_t)^2} - 3.$$

Выпишем выражения для числителя в указанной выше дроби через известные величины:

$$\begin{aligned} \mathbb{E}(Z_t - \mathbb{E}Z_t)^4 &= \mathbb{E}(Z_t^4 - 4Z_t^3 \cdot \mathbb{E}Z_t + \\ &+ 6Z_t^2 \cdot (\mathbb{E}Z_t)^2 - 4Z_t \cdot (\mathbb{E}Z_t)^3 + (\mathbb{E}Z_t)^4) = \\ &= \mathbb{E}Z_t^4 - 3(\mathbb{E}Z_t)^4 - 4\mathbb{E}Z_t \cdot \mathbb{E}Z_t^3 + 6(\mathbb{E}Z_t)^2 \cdot \mathbb{E}Z_t^2. \end{aligned}$$

Тогда коэффициент эксцесса может быть представлен в виде:

$$\kappa_{Z, t} = \left( \mathbb{E}Z_t^4 - 4\mathbb{E}Z_t \cdot \mathbb{E}Z_t^3 + 6(\mathbb{E}Z_t)^2 \cdot \mathbb{E}Z_t^2 - 3(\mathbb{E}Z_t)^4 \right) / (\mathbb{D}Z_t)^2 - 3. \quad (9)$$

Воспользуемся формулами (2) и (4). Имеем:

$$\begin{aligned} \mathbb{E}(Z_t - \mathbb{E}Z_t)^4 &= \\ &= \sum_{i=1}^{k(t)} \frac{p_i(t)}{\sigma_i(t)\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^4 \exp \left\{ -\frac{(z - a_i(t))^2}{2\sigma_i^2(t)} \right\} dz - \\ &- 3 \left( \sum_{i=1}^{k(t)} p_i(t)a_i(t) \right)^4 - 4 \left( \sum_{i=1}^{k(t)} p_i(t)a_i(t) \right) \times \\ &\times \left( \sum_{i=1}^{k(t)} p_i(t) (a_i^3(t) + 3a_i\sigma_i^2(t)) \right) + \\ &+ 6 \left( \sum_{i=1}^{k(t)} p_i(t)a_i(t) \right)^2 \left( \sum_{i=1}^{k(t)} p_i(t) (a_i^2(t) + \sigma_i^2(t)) \right) = \\ &= \sum_{i=1}^{k(t)} p_i(t) (a_i^4(t) + 6a_i^2\sigma_i^2(t) + 3\sigma_i^4(t)) - \\ &- 4 \left( \sum_{i=1}^{k(t)} p_i(t)a_i(t) \right) \left( \sum_{i=1}^{k(t)} p_i(t) (a_i^3(t) + 3a_i\sigma_i^2(t)) \right) + \\ &+ 6 \left( \sum_{i=1}^{k(t)} p_i(t)a_i(t) \right)^2 \left( \sum_{i=1}^{k(t)} p_i(t) (a_i^2(t) + \sigma_i^2(t)) \right) - \\ &- 3 \left( \sum_{i=1}^{k(t)} p_i(t)a_i(t) \right)^4. \end{aligned}$$

Окончательно получим следующее выражение для коэффициента эксцесса:

$$\begin{aligned} \kappa_{Z, t} &= \left[ \sum_{i=1}^{k(t)} p_i(t) (a_i^4(t) + 6a_i^2\sigma_i^2(t) + 3\sigma_i^4(t)) - \right. \\ &- 3 \left( \sum_{i=1}^{k(t)} p_i(t)a_i(t) \right)^4 - 4 \left( \sum_{i=1}^{k(t)} p_i(t)a_i(t) \right) \times \\ &\times \left( \sum_{i=1}^{k(t)} p_i(t) (a_i^3(t) + 3a_i\sigma_i^2(t)) \right) + \\ &+ 6 \left( \sum_{i=1}^{k(t)} p_i(t)a_i(t) \right)^2 \left( \sum_{i=1}^{k(t)} p_i(t) (a_i^2(t) + \sigma_i^2(t)) \right) \left. \right] \times \\ &\times \left[ \sum_{i=1}^{k(t)} p_i(t) \left( a_i(t) - \sum_{i=1}^{k(t)} p_i(t)a_i(t) \right)^2 + \right. \\ &\left. + \sum_{i=1}^{k(t)} p_i(t)\sigma_i^2(t) \right]^{-2} - 3. \quad (10) \end{aligned}$$

Стоит отметить, что формулы (4), (6), (8) и (10) могут быть использованы для непосредственного определения моментных характеристик по величинам  $p_i(t)$ ,  $a_i(t)$  и  $\sigma_i(t)$  и не требуют знания значения момента предыдущего порядка. Для упрощения вычислений, в том числе и с точки зрения программирования, при «последовательном» поиске этих характеристик проще воспользоваться формулами (5), (7) и (9). Более того, многие системы оптимизированы для выполнения матричных вычислений. Поэтому разумно воспользоваться представлением выражений (3) в виде

$$\mathbb{E}Z_t^m = \begin{cases} P_t \cdot A_t^T, & m = 1; \\ P_t (D_{A, t} \cdot A_t^T + D_{\Sigma, t} \cdot \Sigma_t^T), & m = 2; \\ P_t \cdot D_{A, t} (D_{A, t} \cdot A_t^T + 3 \cdot D_{\Sigma, t} \cdot \Sigma_t^T), & m = 3; \\ P_t (D_{A, t}^3 \cdot A_t^T + 6 \cdot D_{\Sigma, t}^2 \cdot D_{A, t} \cdot A_t^T + \\ + 3 \cdot D_{\Sigma, t}^3 \cdot \Sigma_t^T), & m = 4, \end{cases}$$

где

$$\begin{aligned} P_t &= (p_1, \dots, p_{k(t)}); \quad A_t = (a_1, \dots, a_{k(t)}); \\ \Sigma_t &= (\sigma_1, \dots, \sigma_{k(t)}); \\ D_{A, t} &= \text{diag} \{a_1, \dots, a_{k(t)}\}; \\ D_{\Sigma, t} &= \text{diag} \{\sigma_1, \dots, \sigma_{k(t)}\}, \end{aligned}$$

а обозначение  $\text{diag}\{\dots\}$  использовано для записи диагональных матриц с указанными элементами.

Получаемое преимущество в скорости в рамках подобного подхода количественно оценивается в работе [10] на примере классического EM-алгоритма (expectation-maximization).

### 3 Реализация метода скользящего разделения смесей на встроенном языке программирования пакета MATLAB

Возможности СРС-метода были реализованы в рамках единого программного комплекса на встроенном языке программирования пакета MATLAB. Для удобства работы был создан оконный пользовательский интерфейс, с помощью которого задаются параметры вычислительных методов, определяется диапазон графического вывода для сохранения и т. д. Данный комплекс объединил в себе различные варианты EM-алгоритмов, включая и реализацию матричных вычислений на GPU, для различных типов смесей вероятностных распределений.

В начале работы пользователю предлагается ввести имя (путь) для файла с данными. Может быть выбран любой диапазон внутри ряда, при этом по умолчанию предлагаются все значения от первого до последнего (с автоматическим определением длины ряда при указании параметра «End»). В случае ввода пользователем некорректных значений (например, отрицательных или превышающих объем выборки) программа возвращает настройки к предустановленным. Вид начального окна представлен на рис. 1.

Затем пользователь может выбрать отображение исходных данных или разностей, вывести гистограммы для этих рядов с автоматической аппрок-

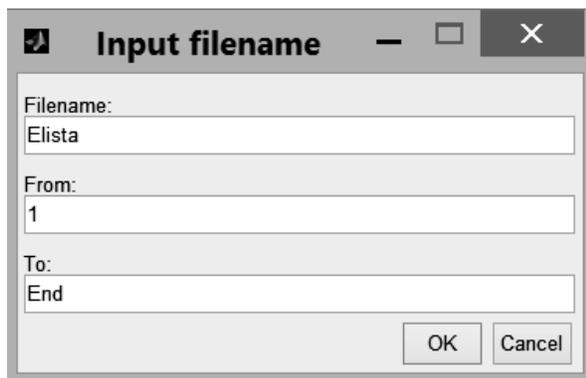


Рис. 1 Начальное окно программы

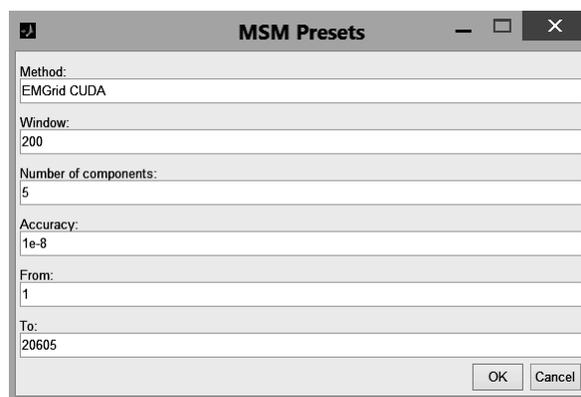


Рис. 2 Настройки параметров СРС-метода

симацией конечной смесью вероятностных распределений. После этого предлагается возможность запуска СРС-метода как для исходной выборки, так и для разностей. Настройка параметров СРС-метода осуществляется с помощью нового диалогового окна, в котором задается название вычислительного алгоритма, размер подвыборки, максимальное число компонент в аппроксимирующей смеси, точность приближений и диапазон для сдвига (можно выбрать любую часть внутри выборки; если заданные пользователем значения некорректны, расчет производится от первого элемента выборки до  $N - \text{width} + 1$ , где  $N$  — длина всей выборки для анализа, а  $\text{width}$  — ширина окна). Для каждого из полей предусмотрены значения по умолчанию. На рис. 2 приводится англоязычный вариант интерфейса с аббревиатурой MSM (от moving separation of mixtures).

После нажатия на кнопку «OK» и запуска шагов алгоритма отображается окно с индикатором выполнения и названием выбранного вычислительного метода, позволяющего следить за текущим положением окна и состоянием работы программы. После завершения вычислений появляется график с динамической и диффузионной компонентами волатильности (подробнее см. в [1]), а также диалоговое окно, с помощью которого можно настроить диапазон вывода параметров для сохранения, включая и диапазон для окон. После закрытия данного диалогового окна график автоматически сохраняется в формате PNG на диск в папку с программой. Пример графического вывода представлен на рис. 3.

Данный функционал реализуется с помощью модуля «Ядро СРС-метода» (свидетельство государственной регистрации программ для ЭВМ № 2015618673). Он используется с целью унификации для пользователя работы с любым алгоритмом, при этом с точки зрения разработчика возможно

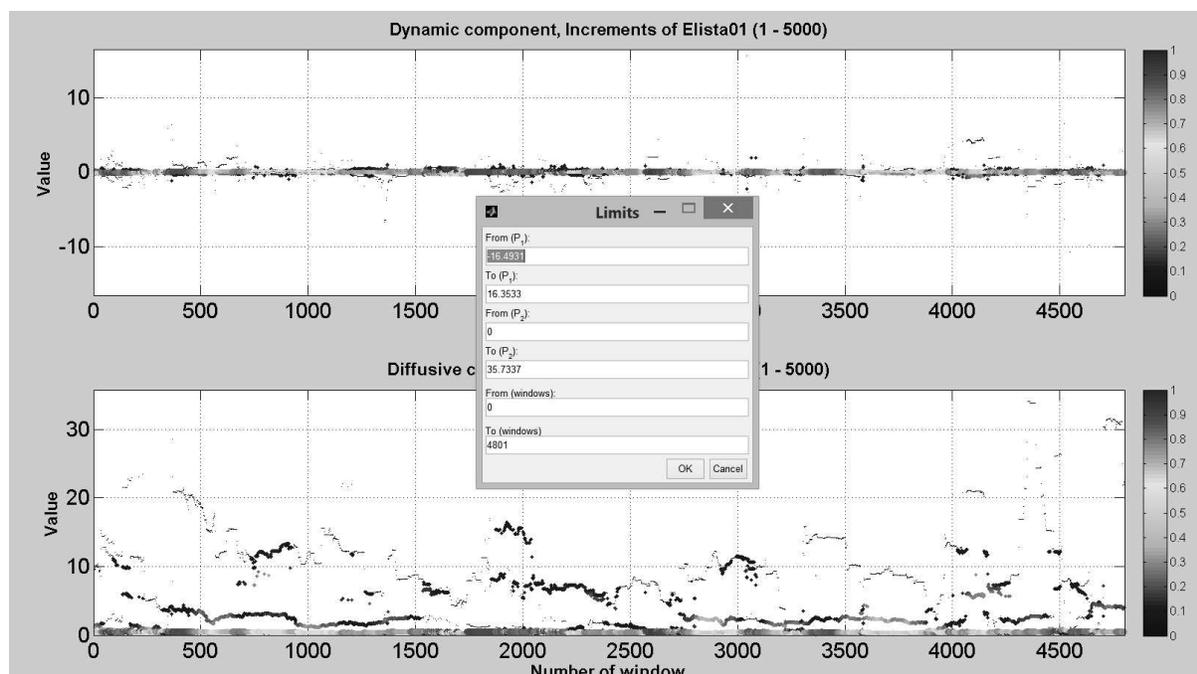


Рис. 3 Пример графического вывода с окном изменения диапазонов по каждой из осей

простое включение в состав любых новых методов без изменения пользовательского опыта. Найденные оценки сохраняются в процессе расчетов, а также по окончании работы программы, что позволяет избежать потери результатов при возникновении прерывания. Непосредственно в ядро встроены различные модификации EM-алгоритма (классический, сглаженный, основанный на повторных вычислениях на окне и т. п.).

Кроме того, в состав комплекса включен модуль визуализации моментных характеристик и квантилей (свидетельство государственной регистрации программ для ЭВМ № 2015618564), формулы для которого были рассмотрены в предыдущем разделе. Модуль содержит функции для отыскания математического ожидания, дисперсии, коэффициента асимметрии и эксцесса, а также квантилей различного уровня для конечных смесей нормаль-

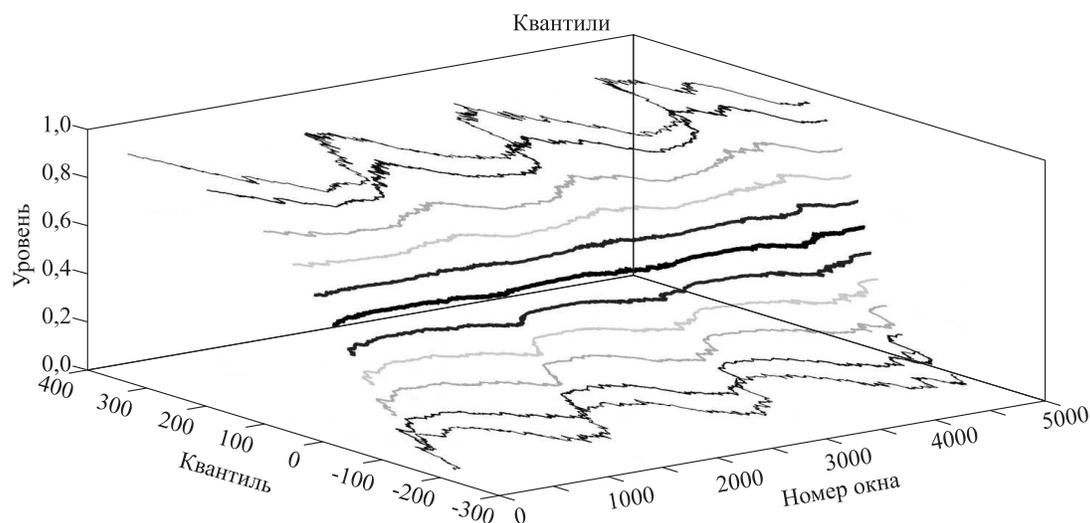


Рис. 4 Пример изображения эволюции квантилей различных уровней (от 0,05 до 0,95)

ных законов. Указанные величины вычисляются для каждого положения окна в СРС-методе, а затем выводятся на графиках: четыре моментные характеристики в разных осях на одном листе; эволюция квантилей различного уровня изображается на отдельном графике в трехмерном пространстве (рис. 4). Для заданного набора окон (параметр одной из функций) предусмотрен вывод приближения гистограммы аппроксимирующими кривыми с заданием расположения набора графиков на листе. С привлечением возможностей данного модуля были созданы графики, например в работе [11].

#### 4 Вычислительный комплекс с онлайн-доступом

В предыдущем разделе был коротко описан программный комплекс, реализованный средствами программирования пакета MATLAB, функциональные возможности которого должны стать базой для онлайн-сервиса.

Реализация широкого спектра методов интеллектуального анализа данных, основанного на идеологии смесей вероятностных распределений (не только конечных нормальных, как было рассмотрено выше) будет представлена в онлайн-системе на основе инструментов, предлагаемых в пользовательских профилях (с реализацией механизмов регистрации и авторизации). Пользователи смогут загружать на сервер данные, а полученные после обработки модели будут сохраняться в их профиле, что позволит избежать повторной загрузки одних и тех же рядов (возможно, значительного объема) и сделает возможным многократное дальнейшее использование результатов обработки.

Из-за высокой вычислительной сложности в ряде ситуаций может потребоваться значительное время для корректной аппроксимации данных выбранной моделью. В этой ситуации предполагается оповещение исследователей о статусе процесса с помощью электронных каналов связи. В случае завершения расчетов пользователь сможет в любое удобное время обратиться к результатам анализа как в числовом (с помощью экспорта оцененных параметров), так и в графическом виде. При этом будет доступна работа с несколькими рядами одновременно, в том числе в разных режимах обработки.

Основным элементом интерфейса онлайн-системы является область для отображения исходных временных рядов и визуального представления результатов СРС-метода, включая вывод нескольких графиков одновременно (динамической и диффузионной компоненты волатильности, моментных характеристик и др.).

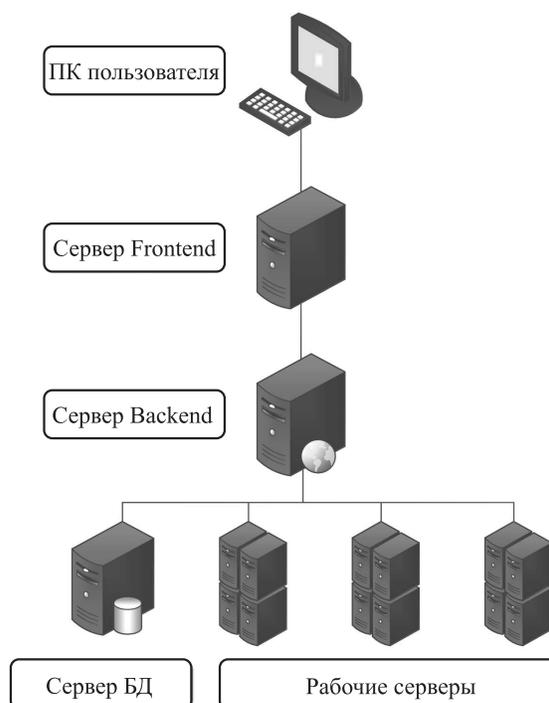


Рис. 5 Схема архитектуры комплекса

Описание архитектуры, предлагаемой для решения перечисленных задач, включая организацию потоков данных в системе, приведено в работе [12]. Рассмотрим ключевые элементы, представленные на рис. 5.

Первый уровень представляет персональный компьютер (ПК) пользователя, обеспечивающий ему доступ к интерфейсу онлайн-комплекса для загрузки данных, настройки параметров и получения результатов (в том числе в визуальной форме). Frontend-сервер предназначен для реализации интерфейса взаимодействия между пользователем и вычислительными узлами системы, при этом непосредственная обработка данных на нем не производится. Backend-сервер осуществляет взаимодействие между Frontend-сервером, рабочими серверами и серверами баз данных (БД), в частности готовит данные для обработки и распределяет задачи по вычислительным компонентам системы.

Наибольшая нагрузка по передаче данных ложится на взаимодействие Frontend- и Backend-серверов, поэтому необходима разработка механизмов ускорения их работы (в частности, за счет кэширования). Взаимодействие Backend- и рабочих серверов должно осуществляться с помощью специализированных API (application programming interface) для реализации возможностей, связанных с параллельной обработкой данных. После проведения вычислений для одного положения окна СРС-метода

данные сохраняются на сервере БД и пересылаются с помощью Frontend-сервера пользователю для возможного контроля процесса выполнения анализа с его стороны.

Подобная архитектура позволит интегрировать в систему различные методы интеллектуального анализа данных, при этом их реализация может быть основана на специальных программных и аппаратных подходах, но для конечного пользователя особенности решений будут скрыты.

## 5 Заключение

Анализ информационных потоков с использованием разнообразных вероятностных моделей широко распространен в различных прикладных областях, что подчеркивает актуальность создания описываемой в статье онлайн-системы. При этом следует отметить отсутствие непосредственных аналогов сервиса из-за универсальности предлагаемого набора методов и выбранной концепции, позволяющих конечному пользователю не заботиться о соответствии ресурсов своего компьютера каким-либо специальным техническим требованиям, а сразу загружать данные на сервер, обрабатывать их с помощью различных методов интеллектуального анализа и получать результат. Данная система сможет предложить широкие возможности для различных групп исследователей во всем мире.

Автор выражает признательность д.ф.-м.н., профессору Виктору Юрьевичу Королеву за полезные обсуждения, а также Виктору Кузьмину за плодотворное участие в разработке архитектуры комплекса.

## Литература

1. *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во МГУ, 2011. 512 с.
2. *Королев В. Ю., Корчагин А. Ю., Горшенин А. К.* Некоторые свойства дисперсионно-сдвиговых смесей нормальных законов // *Статистические методы оценивания и проверки гипотез*, 2015. Вып. 26. С. 134–153.
3. *Горшенин А. К., Королев В. Ю., Турсунбаев А. М.* Медичанские модификации EM- и SEM-алгоритмов для разделения смесей вероятностных распределений и их применение к декомпозиции волатильности финансовых временных рядов // *Информатика и её применения*, 2008. Т. 2. Вып. 4. С. 12–47.
4. *Горшенин А. К., Королев В. Ю., Малахов Д. В., Скворцова Н. Н.* Анализ тонкой стохастической структуры хаотических процессов с помощью ядерных оценок // *Математическое моделирование*, 2011. Т. 23. № 4. С. 83–89.
5. *Горшенин А. К., Королев В. Ю., Малахов Д. В., Скворцова Н. Н.* Об исследовании плазменной турбулентности на основе анализа спектров // *Компьютерные исследования и моделирование*, 2012. Т. 4. № 4. С. 793–802.
6. *Hamed M., Salleh S.-H., Chee-Ming Ting, Samdin S. B., Mohd Noor A.* Sensor space time-varying information flow analysis of multiclass motor imagery through Kalman smoother and EM algorithm // *Conference (International) on BioSignal Analysis, Processing and Systems (ICBAPS) Proceedings*, 2015. P. 118–122.
7. *Abanto-Vallea C. A., Bandyopadhyay D., Lachose V. H., Enriquez I.* Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions // *Comput. Stat. Data An.*, 2010. Vol. 54. No. 12. P. 2883–2898.
8. *Wang J. J. J., Chan J. S. K., Choy S. T. B.* Stochastic volatility models with leverage and heavytailed distributions: A Bayesian approach using scale mixtures // *Comput. Stat. Data An.*, 2011. Vol. 55. No. 1. P. 852–862.
9. *Горшенин А. К.* Проверка статистических гипотез в модели расщепления компоненты // *Вестник Московского ун-та. Сер. 15: Вычислительная математика и кибернетика*, 2011. № 4. С. 26–32.
10. *Gorshenin A. K.* On implementation of EM-type algorithms in the stochastic models for a matrix computing on GPU // *AIP Conference Proceedings*, 2015. Vol. 1648. P. 250008.
11. *Королев В. Ю., Горшенин А. К., Гулев С. К., Беляев К. П.* Статистическое моделирование турбулентных потоков тепла между океаном и атмосферой с помощью метода скользящего разделения конечных нормальных смесей // *Информатика и её применения*, 2015. Т. 9. Вып. 4. С. 3–13.
12. *Gorshenin A., Kuzmin V.* Online system for the construction of structural models of information flows // *7th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT) Proceedings*. — Piscataway, NJ, USA: IEEE, 2015. P. 216–219.

Поступила в редакцию 06.02.16

# CONCEPT OF ONLINE SERVICE FOR STOCHASTIC MODELING OF REAL PROCESSES

A. K. Gorshenin<sup>1,2</sup>

<sup>1</sup>Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

<sup>2</sup>Moscow Technological University (MIREA), 78 Vernadskogo Ave., Moscow 119454, Russian Federation

**Abstract:** Information flows analysis based on various probabilistic models is widely used in various applied fields. The article describes the basic principles of construction of the new online system for stochastic modeling of real processes, which has no analogue due to the universality of the set of methods and the concept of an Internet resource; so, the end user should not check personal computer’s specifications and could upload data to the server immediately and then process the samples.

**Keywords:** probability mixtures; moving separation of mixtures; data mining; online software; matrix computing

**DOI:** 10.14357/19922264160107

## Acknowledgments

The research was supported by the Russian Foundation for Basic Research (project 15-37-20851).

## References

1. Korolev, V. Yu. 2011. *Veroyatnostno-statisticheskie metody dekompozitsii volatil’nosti khaoticheskikh protsessov* [Probabilistic and statistical methods of decomposition of volatility of chaotic processes]. Moscow: Moscow University Publishing House. 512 p.
2. Korolev, V. Yu., A. Yu. Korchagin, and A. K. Gorshenin. 2015. Nekotorye svoystva dispersionno-sdvigovykh smesey normal’nykh zakonov [Some properties of variance-mean normal mixtures]. *Statisticheskie Metody Otsenivaniya i Proverki Gipotez* [Statistical Methods of Estimation and Hypothesis Testing] 26:134–153.
3. Gorshenin, A. K., V. Yu. Korolev, and A. M. Tursunbaev. 2008. Mediannye modifikatsii EM- i SEM-algoritmov dlya razdeleniya smesey veroyatnostnykh raspredeleniy i ikh primeneniye k dekompozitsii volatil’nosti finansovykh vremennykh ryadov [Median modification of EM- and SEM-algorithms for separation of mixtures of probability distributions and their application to the decomposition of volatility of financial time series]. *Informatika i ee Primeneniya — Inform. Appl.* 2(4):12–47.
4. Gorshenin, A. K., V. Yu. Korolev, D. V. Malakhov, and N. N. Skvortsova. 2011. Analiz tonkoy stokhasticheskoy struktury khaoticheskikh protsessov s pomoshch’yu yadernykh otsenok [Analysis of fine stochastic structure of chaotic processes by kernel estimators]. *Matematicheskoe Modelirovanie* [Mathematical Modeling] 23(4):83–89.
5. Gorshenin, A. K., V. Yu. Korolev, D. V. Malakhov, and N. N. Skvortsova. 2012. Ob issledovanii plazmennoy turbulentnosti na osnove analiza spektrov [On the investigation of plasma turbulence by the analysis of the spectra]. *Komp’yuternye Issledovaniya i Modelirovanie* [Computer Research and Modeling] 4(4):793–802.
6. Hamed, M., S.-H. Salleh, Ting Chee-Ming, S. B. Samdin, and A. Mohd Noor. 2015. Sensor space time-varying information flow analysis of multiclass motor imagery through Kalman smoother and EM algorithm. *Conference (International) on BioSignal Analysis, Processing and Systems (ICBAPS) Proceedings*. 118–122.
7. Abanto-Valle, C. A., D. Bandyopadhyay, V. H. Lachos, and I. Enriquez. 2010. Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Comput. Stat. Data An.* 54(12):2883–2898.
8. Wang, J. J. J., J. S. K. Chan, and S. T. B. Choy. 2011. Stochastic volatility models with leverage and heavytailed distributions: A Bayesian approach using scale mixtures. *Comput. Stat. Data An.* 55(1):852–862.
9. Gorshenin, A. K. 2011. Proverka statisticheskikh gipotez v modeli rasshchepeniya komponenty [Testing of statistical hypotheses in the splitting component model]. *Vestnik Moskovskogo Un-ta. Vychislitel’naya Matematika i Kibernetika* [Moscow University Computational Mathematics and Cybernetics] 4:26–32.
10. Gorshenin, A. K. 2015. On implementation of EM-type algorithms in the stochastic models for a matrix computing on GPU. *AIP Conference Proceedings*. 1648:250008
11. Korolev, V. Yu., A. K. Gorshenin, S. K. Gulev, and K. P. Belyaev. 2015. Statisticheskoe modelirovanie turbulentnykh potokov tepla mezhdru okeanom i atmosferoy

- s pomoshch'yu metoda skol'zyashchego razdeleniya konechnykh normal'nykh smesey [Statistical modeling of air–sea turbulent heat fluxes by the method of moving separation of finite normal mixtures]. *Informatika i ee Primeneniya — Inform. Appl.* 9(4):3–13.
12. Gorshenin, A., and V. Kuzmin. 2015. Online system for the construction of structural models of information flows. *7th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT) Proceedings*. Piscataway, NJ: IEEE, 2015. 216–219.

*Received February 6, 2016*

## Contributor

**Gorshenin Andrey K.** (b. 1986) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; associate professor, Moscow Technological University (MIREA), 78 Vernadskogo Ave., Moscow 119454, Russian Federation; agorshenin@frcsc.ru

# РАЗРАБОТКА АЛГОРИТМА ЧИСЛЕННОГО РЕШЕНИЯ ЗАДАЧИ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ ИНВЕСТИЦИЯМИ В ЗАКРЫТОЙ ДИНАМИЧЕСКОЙ МОДЕЛИ ТРЕХСЕКТОРНОЙ ЭКОНОМИКИ

П. В. Шнурков<sup>1</sup>, В. В. Засыпко<sup>2</sup>, В. В. Белоусов<sup>3</sup>, А. К. Горшенин<sup>4</sup>

**Аннотация:** Настоящее исследование посвящено разработке численного метода решения задачи оптимального управления инвестициями в закрытой динамической модели трехсекторной экономики. В предшествующих работах было проведено аналитическое исследование поставленной задачи оптимального управления на основе принципа максимума. В данной работе полученные аналитические представления для функций состояний и сопряженных переменных используются как основа для численного алгоритма. Предлагаемый алгоритм позволяет проанализировать класс допустимых функций управления, имеющих не более заданного конечного числа точек переключения, и найти среди них те, которые удовлетворяют необходимым условиям экстремума и ограничениям исходной задачи. Общая схема предложенного алгоритма может быть использована и при решении других задач оптимального управления, связанных с различными предметными областями. В ходе проведенного исследования разработанный алгоритм реализован в комплексе прикладных программ.

**Ключевые слова:** модель трехсекторной экономики; принцип максимума Понтрягина; численное решение задачи оптимального управления

**DOI:** 10.14357/19922264160108

## 1 Введение

В работах П. В. Шнуркова и В. В. Засыпко [1, 2] была поставлена и исследована задача оптимального управления инвестициями, сформулированная на основе закрытой динамической модели трехсекторной экономики. По форме поставленная задача представляет собой классическую задачу оптимального управления на заданном конечном интервале времени с закрепленным левым концом траектории. Теоретическое исследование задачи проводилось на основе принципа максимума Понтрягина.

Исследованию задач оптимального управления в экономических системах посвящена обширная литература. Отметим, например, фундаментальные издания [3–5]. Работы, в которых исследуются многосекторные экономические модели, встречаются достаточно редко [6]. Данное обстоятельство определяет актуальность проведенного исследования.

Следует отметить, что метод исследования задач оптимального управления, основанный на использовании принципа максимума, позволяет получить

аналитические решения только для сравнительно небольшого числа задач, и такие решения хорошо известны в научной и учебной литературе [7, 8]. Для остальных задач принцип максимума позволяет только определить структуру, т. е. аналитическое устройство функций оптимального управления.

Для продолжения исследования требуется решение весьма сложной системы соотношений, состоящей из необходимых условий экстремума и ограничений исходной задачи. Именно такая ситуация возникает и при решении рассматриваемой задачи.

Дальнейшее аналитическое исследование проводилось для функций управления, удовлетворяющих условию максимума и при этом имеющих произвольное конечное число переключений. Для таких функций выписаны аналитические решения системы сопряженных уравнений и системы уравнений дифференциальной связи. В результате получены явные представления для сопряженных переменных, которые по содержанию являются множителями Лагранжа в рассматриваемой экстремальной задаче, и функций удельного капитала в различных секторах, которые играют роль

<sup>1</sup>НИУ Высшая школа экономики, pshnurkov@hse.ru

<sup>2</sup>НИУ Высшая школа экономики, vzasypko@gmail.com

<sup>3</sup>Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, vbousov@ipiran.ru

<sup>4</sup>Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Московский технологический университет (МИРЭА), agorshenin@frccsc.ru

состояний в этой задаче [2]. Однако на этом аналитические средства исследования исчерпывают свои возможности. Дальнейшее исследование поставленной задачи оптимального управления возможно только при помощи численных методов.

В работе [2] была кратко описана общая схема алгоритма, позволяющего завершить исследование поставленной задачи численно. Для заданного набора исходных данных (параметров модели) этот алгоритм позволяет определить конкретную функцию управления и соответствующий набор функций состояний, которые удовлетворяют системе соотношений, состоящей из необходимых условий и ограничений исходной задачи, т. е. является допустимой экстремалью.

В настоящей работе результаты проведенного исследования представлены в целом. При этом основное внимание уделяется описанию и реализации указанного численного алгоритма. Приводится математическая постановка задачи оптимального управления. Формулируется утверждение о необходимых условиях экстремума в форме принципа максимума. Условие максимума позволяет определить структуру оптимального управления. Приводятся формулировки отдельных теорем об аналитических представлениях функций состояний и сопряженных переменных. Вычисление значений этих функций по найденным аналитическим формулам используется в ходе реализации основного алгоритма.

Разработанный алгоритм реализован в виде комплекса прикладных программ. В данной работе приведено описание структуры этого комплекса и его отдельных составляющих. Дано также описание исходных данных, для которых произведены численные исследования. Результатом такого исследования является конкретный управляемый процесс, удовлетворяющий необходимым условиям и ограничениям поставленной задачи.

## 2 Математическая постановка задачи оптимального управления

Рассматриваемая задача оптимального управления формулируется на основе трехсекторной модели экономики. Трехсекторная модель предложена В. А. Колемаевым [9, 10]. Дополнительные предположения, внесенные в данную модель, изложены в работе [1]. Измененную версию основной модели можно условно называть трехсекторной инвестиционной динамической моделью макроэкономической системы (национальной экономики).

В исходной задаче оптимального управления [1] состояние системы описывается трехмерным вектором, компонентами которого являются функции фондовооруженности (удельного капитала) в каждом из секторов. Параметром управления служит скалярная функция, представляющая собой долю инвестиций в первый (фондосоздающий) сектор по отношению к общему объему инвестиций в системе.

Математическая постановка задачи управления имеет следующий вид.

1. Целевой функционал смешанного типа с интегральной и терминальной частями:

$$\int_0^T e^{-\delta t} A_2 \theta_2 k_2^{\alpha_2}(t) dt + e^{-\delta T} \psi(k_0(T), k_1(T), k_2(T)) \longrightarrow \max. \quad (1)$$

2. Дифференциальная связь:

$$\left. \begin{aligned} \dot{k}_0 &= -\lambda_0 k_0 + l_0^{(1)} \rho A_1 k_1^{\alpha_1} (1 - u_1); \\ \dot{k}_1 &= -\lambda_1 k_1 + A_1 k_1^{\alpha_1} u_1; \\ \dot{k}_2 &= -\lambda_2 k_2 + l_2^{(1)} (1 - \rho) A_1 k_1^{\alpha_1} (1 - u_1). \end{aligned} \right\} \quad (2)$$

3. Начальные условия на основные параметры (состояния системы):

$$k_0(0) = k_{0,0}, \quad k_1(0) = k_{1,0}, \quad k_2(0) = k_{2,0}. \quad (3)$$

4. Ограничения на допустимое управление:

$$0 \leq u_1(t) \leq 1, \quad 0 \leq t \leq T. \quad (4)$$

Приведем описание исходных параметров и основных характеристик модели, входящих в указанную постановку задачи. Отметим, что во введенных обозначениях нижний индекс  $j$  соответствует номеру сектора рассматриваемой экономической системы,  $j = 0, 1, 2$ .

$K_j$  — объем основных производственных фондов (капитал) в  $j$ -м секторе.

$L_j$  — число занятых (объем трудовых ресурсов) в  $j$ -м секторе.

$I_j$  — объем инвестиций в  $j$ -й сектор.

$k_j = K_j/L_j$  — фондовооруженность  $j$ -го сектора экономики (удельный капитал).

$i_j = I_j/L_j$  — удельные инвестиции в  $j$ -й сектор экономики.

Указанные выше характеристики являются функциями от времени.

$\nu$  — доля прироста единицы объема трудовых ресурсов за единицу времени во всей экономической системе.

$\mu_j$  — доля выбывших за единицу времени основных производственных фондов в  $j$ -м секторе экономики.

Заметим, что в рассматриваемой модели выполняются следующие соотношения:

$L_j = L_{j,0}e^{\nu t}$ ,  $L_{j,0}$  — объем трудовых ресурсов в  $j$ -м секторе в начальный момент времени.

$f_j(k_j) = A_j k_j^{\alpha_j}$  — объем произведенного продукта на одну единицу трудовых ресурсов (производительность труда) в  $j$ -м секторе.

$A_j$  — нормирующий коэффициент  $A_j > 0$ .

$\alpha_j$  — коэффициент эластичности  $0 < \alpha_j < 1$ .

Параметры  $\nu$ ,  $\mu_j$ ,  $L_{j,0}$ ,  $A_j$  и  $\alpha_j$ ,  $j = 0, 1, 2$ , предполагаются заданными.

Введем дополнительные обозначения:

$$\lambda_j = \mu_j + \nu, \quad j = 0, 1, 2;$$

$$\theta_j = \frac{L_j}{L} = \frac{L_j}{L_0 + L_1 + L_2} = \frac{L_{j,0}}{L_{0,0} + L_{1,0} + L_{2,0}},$$

$$j = 0, 1, 2;$$

$$l_0^{(1)} = \frac{L_{1,0}}{L_{0,0}}; \quad l_2^{(1)} = \frac{L_{1,0}}{L_{2,0}}; \quad B_2 = A_2 \theta_2.$$

Функция

$$u_1(t) = \frac{I_1(t)}{I_0(t) + I_1(t) + I_2(t)} = \frac{i_1(t)}{A_1 k_1^{\alpha_1}(t)}$$

играет в данной модели роль параметра управления.

Постоянная величина  $0 \leq \rho \leq 1$  определяет распределение инвестиций между нулевым (материальным) и вторым (потребительским) секторами после задания параметра управления, который, как уже отмечалось, определяет долю инвестиций, направляемых в первый (фондосоздающий) сектор. Эта величина предполагается заданной.

Постоянная величина  $\delta > 0$  называется коэффициентом дисконтирования и также предполагается известной. Данная величина характеризует темп инфляции в рассматриваемой экономической системе.

С формальной точки зрения задача оптимизации (1)–(4) представляет собой классическую задачу оптимального управления с фиксированным интервалом времени, закрепленным левым и свободным правым концом траектории [1, 2]. Экономическое содержание этой задачи описано в работе [1]. Напомним его вкратце. Целевой функционал (1) состоит из интегральной и терминальной составляющих. Интегральная часть

представляет собой накопленный объем удельного производства (производительность труда) потребительского сектора, т. е. накопленный удельный объем потребительских благ. Терминальная часть характеризует уровень производства во всей системе в конечный момент времени  $T$  и зависит от значений удельного капитала в различных секторах. Система соотношений (2) или дифференциальная связь описывает изменение состояний системы  $k_0(t)$ ,  $k_1(t)$ ,  $k_2(t)$  при заданной функции управления  $u_1(t)$ . Данная система непосредственно выводится из исходных динамических соотношений, характеризующих трехсекторную модель экономической системы. Соотношения (3), или начальные условия, означают, что заданы фиксированные значения функций состояний в начальный момент времени  $t = 0$ . Соотношение (4) представляет собой ограничение на управление, связанное с экономическим содержанием функции  $u_1(t)$ .

### 3 Необходимые условия экстремума и их теоретический анализ

Приведем в краткой форме результаты анализа задачи (1)–(4) методом, основанным на принципе максимума Понтрягина [7, 8, 11]. Основным теоретическим утверждением является утверждение о необходимых условиях экстремума в поставленной задаче, доказанное в работе [1]. В данной работе это утверждение приводится в новой форме.

**Теорема 1.** Пусть  $(k_{0*}(t), k_{1*}(t), k_{2*}(t); u_{1*}(t))$  — оптимальный управляемый процесс, т. е. решение задачи оптимального управления (1)–(4). Тогда найдутся не равные нулю одновременно множители Лагранжа  $\lambda_0^* \in R$ ,  $(p_0(t), p_1(t), p_2(t))$ ,  $t \in [0, T]$ , такие что выполняются следующие соотношения:

1. Сопряженные уравнения (система дифференциальных уравнений относительно функций  $(p_0(t), p_1(t), p_2(t))$ ), которые в теории называются сопряженными переменными [7, 8, 11]:

$$\left. \begin{aligned} \dot{p}_0(t) &= \lambda_0 p_0(t); \\ \dot{p}_1(t) &= \lambda_1 p_1(t) - A_1 \alpha_1 k_1^{\alpha_1 - 1}(t) \left[ l_0^{(1)} \rho p_0(t) + \right. \\ &\quad \left. + l_2^{(1)} (1 - \rho) p_2(t) \right] (1 - u_1(t)) - \\ &\quad - A_1 \alpha_1 k_1^{\alpha_1 - 1}(t) u_1(t); \\ \dot{p}_2(t) &= \lambda_2 p_2(t) - B_2 e^{-\delta t} \alpha_2 k_2^{\alpha_2 - 1}(t) \end{aligned} \right\} (5)$$

$$\text{при } k_0(t) = k_{0*}(t), \quad k_1(t) = k_{1*}(t), \quad k_2(t) = k_{2*}(t) \\ \text{и } u_1(t) = u_{1*}(t).$$

2. Условия трансверсальности (система граничных условий к сопряженным уравнениям в точке  $t = T$ ):

$$\left. \begin{aligned} p_0(T) &= e^{-\delta T} \psi_{k_0^{(1)}}(k_0(T), k_1(T), k_2(T)) = \\ &= \psi_0^{(0)}(T); \\ p_1(T) &= e^{-\delta T} \psi_{k_1^{(1)}}(k_0(T), k_1(T), k_2(T)) = \\ &= \psi_1^{(0)}(T); \\ p_2(T) &= e^{-\delta T} \psi_{k_2^{(1)}}(k_0(T), k_1(T), k_2(T)) = \\ &= \psi_2^{(0)}(T) \end{aligned} \right\} (6)$$

при  $k_0(t) = k_{0*}(t)$ ,  $k_1(t) = k_{1*}(t)$  и  $k_2(t) = k_{2*}(t)$ .

3. Условие максимума функции Понтрягина:

$$\begin{aligned} \max_{u_1 \in U} H(t, k_0, k_1, k_2; u_1, p_0, p_1, p_2) &= \\ &= \max_{u_1 \in U} \left[ -[\lambda_0 k_0 p_0 + \lambda_1 k_1 p_1 + \lambda_2 k_2 p_2] + \right. \\ &+ B_2 e^{-\delta t} k_2^{\alpha_2} + A_1 k_1^{\alpha_1} [l_0^{(1)} \rho p_0 + l_0^{(2)} (1 - \rho) p_2] + \\ &\left. + A_1 k_1^{\alpha_1} [-l_0^{(1)} \rho p_0 + p_1 - l_0^{(2)} (1 - \rho) p_2] u_1 \right] = \\ &= H(t, k_0, k_1, k_2; u_{1*}, p_0, p_1, p_2) \end{aligned} \quad (7)$$

при  $k_0(t) = k_{0*}(t)$ ,  $k_1(t) = k_{1*}(t)$ ,  $k_2(t) = k_{2*}(t)$  и  $u_1(t) = u_{1*}(t)$ , где  $U = \{u : 0 \leq u \leq 1\}$  — множество допустимых значений параметра управления  $u_1 = u_1(t)$ .

Теорема 1 и аналогичные утверждения в теории оптимального управления называются теоремами о необходимых условиях экстремума в форме принципа максимума или просто принципом максимума Понтрягина. В классических задачах управления эти условия состоят из сопряженных уравнений (5), граничных условий к сопряженным уравнениям вида (6), которые обычно называются условиями трансверсальности, а также из условия максимума некоторой специальной функции, называемой функцией Понтрягина или гамильтонианом (соотношение (7)).

Дальнейшее исследование заключается в анализе общей системы соотношений, состоящей из необходимых условий (5)–(7) и ограничений исходной задачи (2)–(4). Важную роль в этом анализе играет условие максимума функции Понтрягина.

Из условия максимума непосредственно следует, что структура функции оптимального управления  $u_{1*}(t)$  зависит от значения некоторой вспомогательной функции  $Q(p_0(t), p_1(t), p_2(t))$  в каждой точке  $t \in [0, T]$  и может быть выражена следующей формулой:

$$u_{1*}(t) = \begin{cases} 1, & \text{если } Q(p_0(t), p_1(t), p_2(t)) > 0; \\ u_1^{(0)}(t), & \text{если } Q(p_0(t), p_1(t), p_2(t)) = 0; \\ 0, & \text{если } Q(p_0(t), p_1(t), p_2(t)) < 0, \end{cases} \quad (8)$$

где

$$\begin{aligned} Q(t) &= Q(p_0(t), p_1(t), p_2(t)) = \\ &= -l_0^{(1)} \rho p_0(t) + p_1(t) - l_0^{(2)} (1 - \rho) p_2(t). \end{aligned}$$

Данная функция называется функцией переключения в рассматриваемой задаче. В соотношении (8) через  $u_1^{(0)}(t)$  обозначено особое управление, возникающее при условии, когда  $Q(p_0(t), p_1(t), p_2(t)) = 0$  и функция Понтрягина не зависит явно от параметра управления  $u_1$ . Особое управление не определяется из условий максимума функции Понтрягина, и для его нахождения необходимо отдельное исследование. Аналитические представления для особых режимов управления удается определить в очень редких случаях. В связи с этим ограничимся только стандартным вариантом структуры управления без особого режима.

Функция  $Q(p_0(t), p_1(t), p_2(t))$  является непрерывно дифференцируемой функцией от аргументов  $p_0$ ,  $p_1$  и  $p_2$ . В свою очередь, сопряженные переменные  $p_0(t)$ ,  $p_1(t)$  и  $p_2(t)$  непрерывны и непрерывно дифференцируемы при всех значениях  $t \in [0, T]$ , кроме точек разрыва функции  $u_1(t)$ , определяющей значение управления. Однако, несмотря на эти хорошие аналитические свойства, поведение функции  $Q(p_0(t), p_1(t), p_2(t))$  на конечном интервале  $[0, T]$  может быть достаточно сложным. В частности, теоретически возможен вариант, когда эта функция бесконечное число раз меняет знак, что приведет к необходимости бесконечного числа переключений управления на данном конечном интервале времени. С точки зрения экономического содержания такое управление неоправданно и практически не реализуемо. В связи с этим в данной работе будет рассматриваться только вариант, когда функция  $Q(p_0(t), p_1(t), p_2(t))$  лишь конечное число раз меняет знак на интервале времени  $[0, T]$ .

Будем предполагать, что точки, в которых функция  $Q(p_0(t), p_1(t), p_2(t))$  изменяет знак, изолированы. Из соотношения (8) следует, что в указанных точках характер управления должен изменяться. В связи с этим такие точки на рассматриваемом интервале времени  $[0, T]$  будем называть точками переключения управления. Обозначим точки переключения через  $\tau_1, \tau_2, \dots, \tau_n$ ,  $0 < \tau_1 < \tau_2 < \dots < \tau_{n-1} < \tau_n < T$ ,  $\tau_0 = 0$ ,  $\tau_{n+1} = T$ . Аналитические представления решений систем сопряженных уравнений будут зависеть от того, является ли число переключений  $n$  четным или нечетным,

а также от знака функции  $Q(p_0(t), p_1(t), p_2(t))$  на начальном интервале времени  $[0, \tau_1]$ . В соответствии с этим рассмотрено четыре возможных варианта структуры оптимального управления. Для каждого из них получены аналитические решения системы сопряженных уравнений и уравнений дифференциальной связи. В данной работе приводится только часть полученных результатов.

#### 4 Аналитические решения для функций состояний и сопряженных переменных

В ходе дальнейшего изложения будем использовать для функции переключений  $Q(p_0(t), p_1(t), p_2(t))$  наряду с указанным также и краткое обозначение  $Q(t) = Q(p_0(t), p_1(t), p_2(t))$ .

Для краткости изложения выпишем решения уравнений системы сопряженных уравнений только для одного из вариантов, остальные имеют аналогичный вид. Пусть число переключений нечетно  $n = 2m - 1, m \geq 1$  — заданное целое число:

$$\left. \begin{aligned} Q(t) > 0 \text{ при } 0 \leq t < \tau_1; \\ Q(t) < 0 \text{ при } \tau_1 < t < \tau_2; \\ \dots\dots\dots \\ Q(t) < 0 \text{ при } \tau_{2m-1} < t < \tau_{2m} = T. \end{aligned} \right\} \quad (9)$$

Для данного варианта функция управления имеет вид:

$$u_{1*}(t) = \begin{cases} 1, & \tau_{2j-2} \leq t < \tau_{2j-1}, \\ & j = 1, 2, \dots, m; \\ 0, & \tau_{2j-1} \leq t < \tau_{2j}, \\ & j = 1, 2, \dots, m. \end{cases} \quad (10)$$

**Теорема 2.** Предположим, что в исходной задаче оптимального управления функция переключений  $Q(t)$  удовлетворяет условиям (9), а соответствующая функция управления  $u_{1*}(t)$  задается формулой (10). Тогда решение системы сопряженных уравнений определяется формулами:

$$\begin{aligned} p_0^{(2j-1)}(t) &= p_{0, \tau_{2j}} e^{\lambda_0(t-\tau_{2j})}; \\ p_1^{(2j-1)}(t) &= e^{\lambda_1 t} \left[ p_{1, \tau_{2j}} e^{-\lambda_1 \tau_{2j}} + \right. \\ &+ A_1 \alpha_1 \int_t^{\tau_{2j}} e^{-\lambda_1 z_3} k_1^{\alpha_1-1}(z_3) \left( l_0^{(1)} p_{p_0, \tau_{2j}} e^{\lambda_0(z_3-\tau_{2j})} + \right. \end{aligned}$$

$$\begin{aligned} &+ l_0^{(2)}(1-\rho) \left( e^{\lambda_2(z_3-\tau_{2j})} p_{2, \tau_{2j}} + \right. \\ &\left. \left. + e^{\lambda_2 z_3} B_2 \int_{z_3}^{\tau_{2j}} e^{(-\delta-\lambda_2)z_4} k_2^{\alpha_2-1}(z_4) dz_4 \right) \right] dz_3; \\ p_2^{(2j-1)}(t) &= e^{\lambda_2 t} \left[ e^{-\lambda_2 \tau_{2j}} p_{2, \tau_{2j}} + \right. \\ &\left. + B_2 \int_t^{\tau_{2j}} e^{(-\delta-\lambda_2)z_1} k_2^{\alpha_2-1}(z_1) dz_1 \right], \\ &\tau_{2j-1} \leq t \leq \tau_{2j}, \end{aligned}$$

где значения  $p_{i, \tau_{2j}}, i = 0, 1, 2$ , определяются на интервале  $[\tau_{2j}, \tau_{2j+1}]$  при  $t = \tau_{2j}, j = 1, 2, \dots, m$ , причем  $p_{i, \tau_{2m}} = \psi_i^{(0)}(T), i = 0, 1, 2$ ;

$$\begin{aligned} p_0^{(2j-2)}(t) &= p_{0, \tau_{2j-1}} e^{\lambda_0(t-\tau_{2j-1})}; \\ p_1^{(2j-2)}(t) &= p_{1, \tau_{2j-1}} \times \\ &\times e^{A_1 \alpha_1 \int_t^{\tau_{2j-1}} k_1^{\alpha_1-1}(z_2) dz_2 + \lambda_1(t-\tau_{2j-1})}; \\ p_2^{(2j-2)}(t) &= e^{\lambda_2 t} \left[ p_{2, \tau_{2j-1}} e^{-\lambda_2 \tau_{2j-1}} + \right. \\ &\left. + B_2 \int_t^{\tau_{2j-1}} e^{(-\delta-\lambda_2)z_1} k_2^{\alpha_2-1}(z_1) dz_1 \right], \\ &\tau_{2j-2} \leq t \leq \tau_{2j-1}, \end{aligned}$$

где значения  $p_{i, \tau_{2j-1}}$  определяются равенствами

$$p_{i, \tau_{2j-1}} = p_i^{(2j-1)}(\tau_{2j-1}), i = 0, 1, 2, j = 1, 2, \dots, m.$$

Аналогичные утверждения доказаны для системы уравнений дифференциальной связи. Приведем формулировку одного из них, соответствующего рассматриваемому варианту структуры функции управления.

**Теорема 3.** Предположим, что в исходной задаче оптимального управления функция переключений  $Q(t)$  удовлетворяет условиям (9), а соответствующая функция управления  $u_{1*}(t)$  задается формулой (10). Тогда решение системы уравнений дифференциальной связи определяется формулами:

$$\left. \begin{aligned} k_0^{(2j-2)}(t) &= k_{0, \tau_{2j-2}} e^{-\lambda_0(t-\tau_{2j-2})}; \\ k_1^{(2j-2)}(t) &= \left( e^{-\lambda_1(1-\alpha_1)(t-\tau_{2j-2})} \times \right. \\ &\times \left( k_{1, \tau_{2j-2}}^{1-\alpha_1} - \frac{A_1}{\lambda_1} \right) + \frac{A_1}{\lambda_1} \Big)^{1/(1-\alpha_1)}; \\ k_2^{(2j-2)}(t) &= k_{2, \tau_{2j-2}} e^{-\lambda_2(t-\tau_{2j-2})}, \\ &\tau_{2j-2} \leq t \leq \tau_{2j-1}, \end{aligned} \right\} \quad (11)$$

где значения  $k_{i,\tau_{2j-2}}$ ,  $i = 0, 1, 2$ , определяются на интервале  $[\tau_{2j-3}, \tau_{2j-2}]$  при  $t = \tau_{2j-2}$ ,  $j = 2, 3, \dots, m$ , причем  $k_{i,\tau_0} = k_{i,0}$ ,  $i = 0, 1, 2$ , при  $j = 1$ ;

$$\left. \begin{aligned} k_0^{(2j-1)}(t) &= e^{-\lambda_0 t} \left( k_{0,\tau_{2j-1}} e^{\lambda_0 \tau_{2j-1}} + \frac{l_0^{(1)} \rho A_1 k_{1,\tau_{2j-1}}^{\alpha_1} e^{\lambda_1 \alpha_1 \tau_{2j-1}}}{\lambda_0 - \lambda_1 \alpha_1} (e^{(\lambda_0 - \lambda_1 \alpha_1)t} - e^{(\lambda_0 - \lambda_1 \alpha_1)\tau_{2j-1}}) \right); \\ k_1^{(2j-1)}(t) &= k_{1,\tau_{2j-1}} e^{-\lambda_1(t - \tau_{2j-1})}; \\ k_2^{(2j-1)}(t) &= e^{-\lambda_2 t} \left( k_{2,\tau_{2j-1}} e^{\lambda_2 \tau_{2j-1}} + \frac{l_2^{(1)}(1 - \rho) A_1 k_{1,\tau_{2j-1}}^{\alpha_1} e^{\lambda_1 \alpha_1 \tau_{2j-1}}}{\lambda_2 - \lambda_1 \alpha_1} (e^{(\lambda_2 - \lambda_1 \alpha_1)t} - e^{(\lambda_2 - \lambda_1 \alpha_1)\tau_{2j-1}}) \right), \tau_{2j-1} \leq t \leq \tau_{2j}, \end{aligned} \right\} (12)$$

где значения  $k_{i,\tau_{2j-1}}$  определяются равенствами

$$k_{i,\tau_{2j-1}} = k_i^{(2j-2)}(\tau_{2j-1}), \quad i = 0, 1, 2, \quad j = 1, 2, \dots, m.$$

Таким образом, формулы (11) и (12) полностью определяют аналитические представления для функций состояний в рассматриваемом варианте поведения функции переключений.

Явные аналитические представления для сопряженных переменных  $p_0(t)$ ,  $p_1(t)$  и  $p_2(t)$  и функций состояний  $k_0(t)$ ,  $k_1(t)$  и  $k_2(t)$  получены также и для трех остальных вариантов поведения функции переключений в случае, когда число переключений  $n$  является произвольным положительным числом  $n \geq 1$ . Отметим, что соответствующие представления для сопряженных переменных и функций состояний в случае отсутствия переключений  $n = 0$  и в случае одного переключения  $n = 1$  были приведены в работе [2].

Полученные результаты позволяют последовательно вычислить значения функций состояний  $k_0(t)$ ,  $k_1(t)$  и  $k_2(t)$ , а затем сопряженных переменных  $p_0(t)$ ,  $p_1(t)$  и  $p_2(t)$  в любой точке  $t \in [0, T]$ . Но тогда и функция переключений  $Q(p_0(t), p_1(t), p_2(t))$ , определяемая через сопряженные переменные, также может быть вычислена в любой точке  $t \in [0, T]$  для любого варианта функции управления  $u_{1*}(t)$  без переключений или с произвольным конечным числом точек переключения.

## 5 Описание численного алгоритма решения задачи оптимального управления

Воспользовавшись обстоятельством, отмеченным в конце предыдущего раздела, можно предложить численный алгоритм определения функции управления  $u_{1*}(t)$ , удовлетворяющей необходимым условиям экстремума в форме принципа максимума. Идея этого алгоритма заключается в следующем.

Для каждого из рассмотренных вариантов структуры управления необходимо проанализировать поведение функции переключения  $Q(t)$  на всем интервале времени  $t \in [0, T]$ . Такой анализ можно осуществлять численными методами компьютерных программ, вычисляющих значения функции  $Q(t)$  в отдельных точках на интервале  $t \in [0, T]$ . При этом задается некоторое разбиение интервала  $[0, T]$  конечным числом точек, в которых будут вычисляться значения функции переключений  $Q(p_0(t), p_1(t), p_2(t))$ . Программа, вычисляющая значения функции  $Q(t)$ , должна использовать подпрограммы, вычисляющие значения функций  $p_0(t)$ ,  $p_1(t)$  и  $p_2(t)$ . В свою очередь, подпрограммы, в результате выполнения которых вычисляются значения сопряженных переменных, должны реализовать аналитические представления этих функций, зависящих от функций состояний  $k_0(t)$ ,  $k_1(t)$  и  $k_2(t)$ , полученные в ходе проведенного исследования.

Если выявленное поведение функции переключения  $Q(t)$  соответствует выбранной структуре функции управления  $u_1(t)$ , то рассматриваемый вариант функции управления и соответствующих функций состояний системы  $k_0(t)$ ,  $k_1(t)$  и  $k_2(t)$  можно считать управляемым процессом, который удовлетворяет системе соотношений, состоящей из необходимых условий экстремума и ограничений исходной задачи. При этом соответствие функции переключений  $Q(t)$  и выбранного варианта функции управления  $u_1(t)$  понимается в смысле выполнения соотношения (8).

Перейдем теперь к последовательному описанию предлагаемого алгоритма нахождения допустимых экстремалей в исходной задаче оптимального управления.

1. Выберем в качестве начального варианта функции управления  $u_1(t)$  один из вариантов с отсутствием переключений.
2. Для выбранного варианта вычислим значения функций  $k_0(t)$ ,  $k_1(t)$  и  $k_2(t)$  по имеющимся ана-

литическим формулам (см. формулы (19) и (21) работы [2]).

3. Используя значения выбранной функции управления  $u_{1*}(t)$  и найденные значения функций  $k_0(t)$ ,  $k_1(t)$  и  $k_2(t)$  вычислим значения сопряженных функций  $p_0(t)$ ,  $p_1(t)$  и  $p_2(t)$  по имеющимся аналитическим формулам (см. формулы (29) и (30) работы [2]).
4. Используя найденные значения сопряженных переменных  $p_0(t)$ ,  $p_1(t)$  и  $p_2(t)$ , определим численное представление функции переключений  $Q(t) = Q(p_0(t), p_1(t), p_2(t))$ .
5. Осуществим проверку соответствия характера вычисленной функции  $Q(t)$  выбранному исходному варианту функции управления. В данном случае необходимо проверить выполнение одного из условий: либо условия  $Q(t) > 0$ ,  $t \in [0, T]$ , если в п. 1 данной процедуры был выбран вариант управления  $u_1(t) = 1$ ,  $t \in [0, T]$ , либо условия  $Q(t) < 0$ ,  $t \in [0, T]$ , если в п. 1 данной процедуры был выбран вариант управления  $u_1(t) = 0$ ,  $t \in [0, T]$ . Отметим, что указанные условия на функцию  $Q(t)$  должны быть проверены для всех значений аргумента  $t$ , принадлежащих заданному изначально разбиению отрезка  $[0, T]$ .
6. Если указанное условие выполняется, то можно считать, что выбранный в п. 1 данной процедуры вариант функции управления удовлетворяет необходимым условиям экстремума в форме принципа максимума и ограничениям исходной задачи. Тогда управляемый процесс  $(u_{1*}(t), k_0(t), k_1(t), k_2(t))$  представляет собой допустимую экстремаль в рассматриваемой задаче оптимального управления. Данный вариант управляемого процесса сохраним в специально отведенном месте памяти. Перейдем к следующему варианту функции управления.
7. Предположим теперь, что проверяемое условие не выполняется. Для определенности будем считать, что при выбранном варианте управления  $u_1(t) = 1$ ,  $t \in [0, T]$  функция  $Q(t)$  меняет знак в некоторой точке  $\tau$ :  $Q(t) > 0$ ,  $t \in [0, \tau)$ ,  $Q(t) < 0$ ,  $t \in (\tau, T]$ . Такой результат означает, что при заданных исходных параметрах модели управляемый процесс, состоящий из функции  $u_1(t)$  и соответствующих функций состояний  $k_0(t)$ ,  $k_1(t)$  и  $k_2(t)$ , не удовлетворяет системе соотношений, состоящей из необходимых условий экстремума и ограничений исходной задачи. Следует перейти к анализу другого варианта функции управления.

8. В качестве следующего варианта функции управления можно выбрать тот, который наиболее близок к результату, описанному в п. 7, а именно:

$$u_1(t) = \begin{cases} 1 & \text{при } 0 \leq t \leq \tau; \\ 0 & \text{при } \tau \leq t \leq T. \end{cases} \quad (13)$$

В то же время такой выбор нельзя обосновать строго. Можно осуществить переход к любому из вариантов управления с одной точкой переключения вида (13), где  $\tau$  — произвольная внутренняя точка из интервала  $[0, T]$ , совпадающая с одной из точек заданного разбиения отрезка  $[0, T]$ .

9. Для нового варианта управления вида (13) повторяются действия, описанные в пп. 2–4 данной процедуры.
10. Проверяется условие соответствия поведения вычисленной функции  $Q(t)$  и выбранного варианта (13) функции управления.

Если поведение функции  $Q(t)$  соответствует выбранному варианту функции управления, а именно: выполняется условие  $Q(t) > 0$ ,  $0 \leq t \leq \tau$ ,  $Q(t) < 0$ ,  $\tau < t \leq T$ ,  $Q(\tau) = 0$ , то можно утверждать что выбранный вариант функции управления  $u_{1*}(t)$  вида (13) и соответствующие ему функции состояний  $k_{0*}(t)$ ,  $k_{1*}(t)$  и  $k_{2*}(t)$  удовлетворяют системе, состоящей из необходимых условий экстремума и ограничений исходной задачи. Управляемый процесс  $(u_{1*}(t); k_{0*}(t), k_{1*}(t), k_{2*}(t))$  является допустимой экстремалью в исходной задаче оптимального управления. Данный вариант управляемого процесса сохраняется в специально отведенном месте памяти, после чего происходит переход к анализу следующего варианта функции управления.

Если же поведение вычисленной функции  $Q(t)$  не соответствует характеру выбранного варианта функции управления вида (13), то набор, состоящий из функций  $u_1(t)$  и соответствующих ей функций состояний  $k_0(t)$ ,  $k_1(t)$  и  $k_2(t)$ , не удовлетворяет системе соотношений, состоящей из необходимых условий экстремума и ограничений исходной задачи. Данный набор не является допустимой экстремалью.

11. Производится переход к следующему варианту функции управления  $u_1(t)$  из рассматриваемого множества функций управления. При этом прежде всего перебираются все варианты функции управления с одной точкой переключения вида (13), в которых точка  $\tau$  пробегает заданное множество точек, составляющих разбиение отрезка  $[0, T]$ .

12. Для каждого нового варианта функции управления производятся аналогичные действия, описанные в пп. 2–4 данного алгоритма. После этого вновь проверяется соответствие поведения вычисленной функции  $Q(t)$  выбранному варианту функции управления. Выводы, которые могут быть сделаны по результатам проверки такого соответствия, аналогичны выводам, сформулированным в пп. 6, 7 и 10 данного алгоритма.
13. Указанные действия производятся для всех вариантов функции управления, входящих в некоторое множество функций, заданных на отрезке  $[0, T]$ . Более подробно данное множество будет описано ниже. Сейчас отметим лишь, что это множество конечно и данный алгоритм будет реализован за конечное число шагов.

Иллюстрацией к описанному выше алгоритму нахождения допустимых экстремалей в исходной задаче оптимального управления может служить схема, изображенная на рис. 1.

Сделаем важное замечание, относящееся к п. 13 приведенного алгоритма. Из общей теории экстремальных задач известно, что может существовать несколько допустимых экстремалей, удовлетворяющих необходимым условиям экстремума и ограничениям исходной задачи. Таким образом, процедура поиска должна продолжаться с использованием целенаправленного перебора возможных вариантов поведения функции управления. Указанный перебор должен осуществляться при помощи изменения положения точек переключения и числа этих точек. Основываясь на общих особенностях функций оптимального управления, следующих из условия максимума и соотношения (8), а также на сделанных в дальнейшем дополнительных предположениях о характере возможных

управлений, можно предложить следующее описание основных аналитических особенностей множества возможных вариантов функции управления  $u_1(t)$ , на котором реализуется описанный алгоритм. Данный объект должен представлять собой множество функций, заданных на отрезке  $[0, T]$ , принимающих два возможных значения: 0 или 1 — и имеющих не более заданного конечного числа точек разрыва первого рода (скачков). Для определенности будем предполагать также, что эти функции являются непрерывными справа.

Проведем теперь формальное описание множества возможных функций управления, на котором действует предлагаемый алгоритм. Зафиксируем целое положительное число  $N \geq 1$ . Пусть  $\Delta = T/(N + 1) > 0$ . Обозначим через  $t_i = i\Delta$ ,  $i = 0, 1, 2, \dots, N, N + 1$ , точки, принадлежащие отрезку  $[0, T]$ ,  $t_0 = 0$ ,  $t_{N+1} = (N + 1)\Delta = T$ . Совокупность точек  $\{t_1, t_2, \dots, t_N\}$  будем называть разбиением отрезка  $[0, T]$ . В каждой реализации алгоритма параметры  $T$  и  $N$ , а следовательно, и все разбиение данного отрезка времени, должны быть заданы. Будем предполагать, что заданное разбиение  $\{t_1, t_2, \dots, t_N\}$  представляет собой множество возможных точек переключения у рассматриваемых функций управления. При этом граничные точки данного отрезка  $t_0 = 0$  и  $t_{N+1} = T$  не считаются точками переключения.

Зафиксируем целое положительное число  $n$ ,  $1 \leq n \leq N$ . Выберем  $n$  различных точек из разбиения  $\{t_1, t_2, \dots, t_N\}$ . Обозначим через  $S_n^{(N)}$  множество возможных функций управления  $u_1(t)$ , обладающих следующими свойствами:

- (а) данные функции принимают только два возможных значения: 0 или 1;
- (б) данные функции являются кусочно-постоянными и меняют свое значение в точках

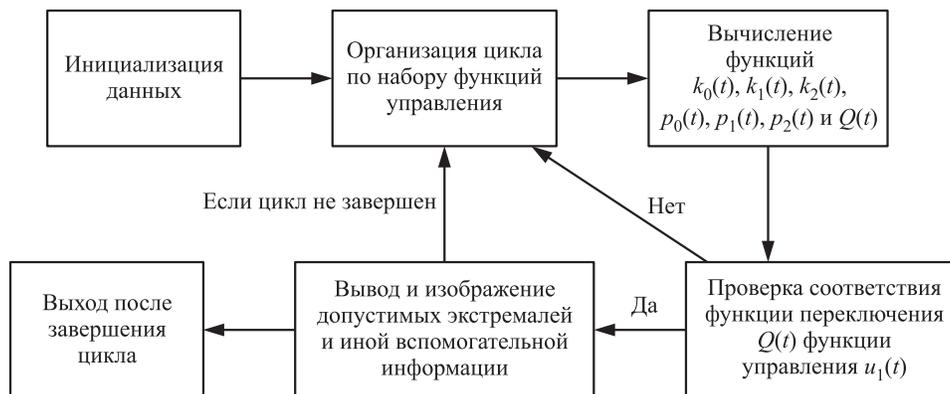


Рис. 1 Теоретическая схема реализации алгоритма

переключения. В точках переключения эти функции обладают свойством непрерывности справа;

- (в) точками переключения для функций, принадлежащих множеству  $S_n^{(N)}$ , может служить любой набор, состоящий из  $n$  различных точек, принадлежащих разбиению  $\{t_1, t_2, \dots, t_N\}$ .

Совокупность функций  $S_n^{(N)}$ , обладающих указанными свойствами, будет рассматриваться как набор возможных функций управления с  $n$  переключениями, где  $1 \leq n \leq N$ . Для удобства введем также множество  $S_0 = S_0^{(N)}$ , состоящее из двух функций:  $u_1(t) = 0, 0 \leq t \leq T$ , и  $u_1(t) = 1, 0 \leq t \leq T$ . Функции управления, входящие в множество  $S_0^{(N)}$ , не имеют переключений. Очевидно, что при  $n_1 \neq n_2$  множества  $S_{n_1}^{(N)}$  и  $S_{n_2}^{(N)}$  являются непересекающимися.

Теперь рассмотрим объединение множеств

$$\hat{S}_n^{(N)} = \bigcup_{k=0}^n S_k^{(N)}.$$

По содержанию множество  $\hat{S}_n^{(N)}$  представляет собой совокупность функций, обладающих свойствами (а) и (б), указанными выше, для которых точками переключения может служить любой набор, состоящий не более чем из  $n$  различных точек, принадлежащих множеству  $\{t_1, t_2, \dots, t_N\}$ . Именно такие наборы функций будут рассматриваться в данной алгоритмической части настоящего исследования как множества возможных функций управления. Для каждой реализации изложенного выше алгоритма задаются параметры  $N$  и  $n$ , т. е. шаг разбиения  $\Delta$  и максимально возможное число точек переключения  $n \leq N$ .

Докажем утверждение о числе элементов множества  $\hat{S}_n^{(N)}$ .

**Утверждение 1.** Число элементов множества  $\hat{S}_n^{(N)}$  определяется следующей формулой:

$$\hat{N}_n = 2 \sum_{k=0}^n C_N^k.$$

**Доказательство.** Определим число элементов в каждом из множеств, входящих в множество  $\hat{S}_n^{(N)}$ . Множество  $S_0^{(N)}$  состоит из двух элементов. Рассмотрим множество  $S_k^{(N)}$  с фиксированным номером  $k$ . Число возможных вариантов точек переключения для функций из данного множества совпадает с числом возможных вариантов выбора  $k$  различных элементов из  $N$  без возвращения и без учета порядка элементов. Указанное число вариантов переключения равно числу сочетаний

$C_N^k = N!/(k!(N-k)!)$ . Каждому фиксированному набору точек переключения соответствуют два варианта возможного вида функции управления, которые отличаются значением функции на интервале от  $t = 0$  до первой точки переключения. Таким образом, множество  $S_k^{(N)}$  содержит  $2C_N^k$  элементов. Поскольку множества  $S_k^{(N)}$  не пересекаются при различных  $k$ , число элементов в объединении множеств равно сумме числа элементов в каждом множестве. Отсюда следует, что

$$\hat{N}_n = 2 + 2C_N^1 + 2C_N^2 + \dots + 2C_N^n = 2 \sum_{k=0}^n C_N^k.$$

Утверждение 1 доказано.

Доказанное утверждение позволяет точно определить число вариантов функции управления, которые необходимо исследовать в ходе реализации предложенного алгоритма. Если удастся каким-либо образом, например при помощи численного эксперимента, оценить время, необходимое для анализа каждого варианта, то полученная формула позволит оценить общее время работы программы, зависящее от параметров  $N$  и  $n$ .

## 6 Программная реализация алгоритма и результаты вычислений

Предложенный алгоритм численного решения задачи оптимального управления реализован в виде программного комплекса. Комплекс состоит из нескольких функциональных частей, которые можно назвать модулями. Перечислим эти модули в той последовательности, в которой происходит реализация всего алгоритма.

1. Модуль, непосредственно реализующий аналитические формулы для функций состояний  $k_0(t)$ ,  $k_1(t)$  и  $k_2(t)$ .
2. Модуль, непосредственно реализующий аналитические формулы для сопряженных переменных  $p_0(t)$ ,  $p_1(t)$  и  $p_2(t)$ . В этом модуле используются результаты вычислений по программным продуктам первого модуля.
3. Модуль, реализующий аналитическое представление для функции переключения  $Q(t) = Q(p_0(t), p_1(t), p_2(t))$ . В этом модуле используются результаты вычислений из модуля 1 и модуля 2.
4. Модуль, реализующий перебор некоторого заданного множества возможных функций управ-

ления и проверку соответствия каждого варианта функции управления характеру функции переключения, вычисляемой для этого варианта. В этом модуле используется результаты вычислений модуля 3.

В результате использования данного комплекса определяются все возможные варианты управляемых процессов, состоящих из некоторой функции управления  $u_{1*}(t)$  и соответствующих ей функций состояний  $(k_{0*}(t), k_{1*}(t), k_{2*}(t))$ , которые удовлетворяют системе, состоящей из необходимых условий экстремума (5)–(7) и ограничений исходной задачи (2)–(4). Каждый из таких управляемых процессов  $(u_{1*}(t); k_{0*}(t), k_{1*}(t), k_{2*}(t))$  представляет собой допустимую экстремаль для заданного набора исходных параметров математической модели.

Отметим еще одну важную особенность программного комплекса, реализующего разработанный алгоритм. Данная особенность связана с организацией выбора точек переключения и всей функции управления  $u_{1*}(t)$ . В программном комплексе предусмотрены три возможности такого выбора и вычисления функций состояний  $(k_0(t), k_1(t), k_2(t))$ , сопряженных переменных  $(p_0(t), p_1(t), p_2(t))$  и функции переключений  $Q(t) = Q(p_0(t), p_1(t), p_2(t))$ , соответствующих выбранной функции управления:

1. Случайный выбор точек переключения  $(t_1, t_2, \dots, t_n)$ . При этом задается число  $n \geq 1$  и значение функции управления на начальном интервале  $[0, t_1]$ . Такой выбор условно называется стохастическим моделированием. Полученные результаты могут быть использованы для качественного анализа исходной задачи управления.
2. Задание конкретного набора точек переключения  $(t_1, t_2, \dots, t_n)$ . При этом конкретный вид функции управления  $u_1(t)$  определяется при помощи задания ее значения на начальном интервале времени  $[0, t_1]$ . Такая функция позволяет исследовать возможные варианты управляемых процессов по отдельности и изучать влияние отдельных параметров на вид управляемого процесса.
3. Организация непосредственного перебора возможных вариантов функций управления  $u_1(t)$ , входящих в определенный класс  $S_k^{(N)}$ ,  $k = 0, 1, \dots, n$ . При этом должно быть задано число точек переключения  $k$  и значение функции управления  $u_1(t)$  на начальном интервале времени  $[0, t_1]$ . Все наборы точек переключения функции управления, входящих в указанный класс, перебираются в ходе реализации

программы. Пользователю необходимо лишь ввести информацию о числе точек переключения, а также о значении функции управления на начальном интервале, которое может быть равно 0 или 1. В результате действия этой функции программного комплекса определяются все допустимые экстремали  $u_{1*}(t)$ , входящие в рассматриваемый класс  $S_k^{(N)}$ , и соответствующие им функции состояний  $(k_{0*}(t), k_{1*}(t), k_{2*}(t))$ . Данная функция предназначена для непосредственного численного решения поставленной задачи оптимального управления, т. е. нахождения всех допустимых экстремалей.

Теперь приведем краткое описание основного набора числовых значений исходных параметров рассматриваемой математической модели. Указанные числовые значения частично взяты из работы Колемаева [12], в которой они были определены на основе анализа реальных данных. Указанные значения определены на основе анализа данных о советской экономике за период с 1960 по 1991 гг. (в ценах 1983 г.). Для описания функционирования национальной экономики СССР за этот период использовалась трехсекторная модель и соответствующие значения параметров были определены для каждого сектора. В частности, в работе [12] было установлено, что численные значения параметров, входящих в функции  $f_j(k_j) = A_j k_j^{\alpha_j}$ ,  $j = 1, 2$ , принимают следующие числовые значения:

$$A_1 = 1,35; \quad A_2 = 2,71; \quad \alpha_1 = 0,68; \quad \alpha_2 = 0,72.$$

Коэффициенты  $\mu_j$ ,  $j = 0, 1, 2$ , характеризующие скорость выбывания основных фондов в различных секторах, принимают следующие значения:

$$\mu_0 = 0,1; \quad \mu_1 = 0,3; \quad \mu_2 = 0,2.$$

Коэффициент  $\nu$  характеризует скорость прироста объема трудовых ресурсов. Поскольку для различных национальных экономических систем такой параметр может быть очень различным, выберем условное расчетное значения для  $\nu = 0,01$  исходя из предположения о достаточно медленном темпе прироста объема трудовых ресурсов.

Поскольку  $\lambda_j = \mu_j + \nu$ ,  $j = 0, 1, 2$ , получаем:

$$\lambda_0 = 0,11; \quad \lambda_1 = 0,31; \quad \lambda_2 = 0,21.$$

Параметр  $\delta > 0$ , характеризующий скорость уменьшения реального содержания в единице денежных ресурсов (показатель инфляции), выберем равным  $\delta = 0,06$ , что соответствует среднему темпу инфляции.

Величина  $T$  — длительность интервала управления, которая иногда также называется горизонтом

планирования. Рассмотрим краткосрочный вариант, когда горизонт планирования представляет собой величину  $T = 1$  условных единиц времени (лет).

В рассматриваемой модели величины  $\theta_0, \theta_1$  и  $\theta_2$  представляют собой доли трудовых ресурсов  $j$ -го сектора в общем объеме трудовых ресурсов системы. Отметим, что в упомянутой работе В. А. Колмаева [12] эти параметры также были определены на основе данных об экономике СССР за 1960–1991 гг., а именно:

$$\theta_0 = 0,22; \quad \theta_1 = 0,16; \quad \theta_2 = 0,62.$$

Учитывая, что в рассматриваемой задаче параметры  $l_0^{(1)}$  и  $l_2^{(1)}$  выражаются через  $\theta_0, \theta_1$  и  $\theta_2$  по формулам

$$l_0^{(1)} = \frac{L_{1,0}}{L_{0,0}} = \frac{\theta_1}{\theta_0}; \quad l_2^{(1)} = \frac{L_{1,0}}{L_{2,0}} = \frac{\theta_1}{\theta_2},$$

получаем

$$l_0^{(1)} = \frac{L_{1,0}}{L_{0,0}} = \frac{\theta_1}{\theta_0} = \frac{16}{22} = \frac{8}{11};$$

$$l_2^{(1)} = \frac{L_{1,0}}{L_{2,0}} = \frac{\theta_1}{\theta_2} = \frac{16}{62} = \frac{8}{31}.$$

Зададим теперь начальные значения функций фондовооруженности (удельного капитала):

$$k_{0,0} = 1000; \quad k_{1,0} = 2000; \quad k_{2,0} = 1500.$$

Предположим, что в рассматриваемой задаче терминальный член целевого функционала линейным образом зависит от значений удельного капитала  $k_0(T), k_1(T), k_2(T)$ :

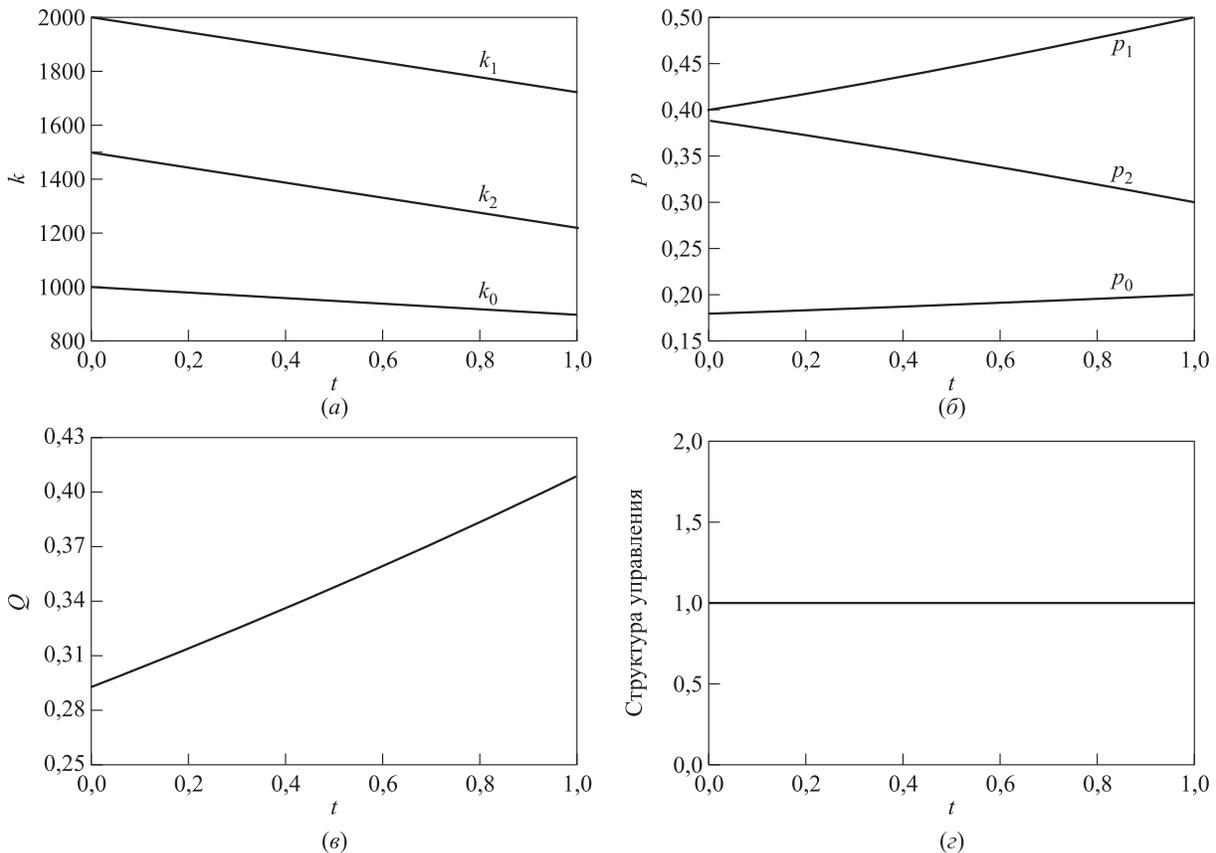
$$\psi(k_0(T), k_1(T), k_2(T)) = a_0 k_0(T) + a_1 k_1(T) + a_2 k_2(T),$$

где  $a_0, a_1$  и  $a_2$  — заданные коэффициенты, характеризующие веса (вклады) величин  $k_0(T), k_1(T)$  и  $k_2(T)$  в терминальную часть целевого функционала. Данные величины могут быть определены экспертами-экономистами. В этом случае производные терминального члена целевого функционала имеют вид:

$$\psi_{k_0}(k_0(T), k_1(T), k_2(T)) = \psi_0^{(0)}(T) = a_0;$$

$$\psi_{k_1}(k_0(T), k_1(T), k_2(T)) = \psi_1^{(0)}(T) = a_1;$$

$$\psi_{k_2}(k_0(T), k_1(T), k_2(T)) = \psi_2^{(0)}(T) = a_2.$$



**Рис. 2** Графическое представление функций фондовооруженности  $k_i(t)$  (а), сопряженных переменных  $p_i(t)$  (б), функции  $Q(t)$  (в) и структура управления (г)

Величины  $a_0$ ,  $a_1$  и  $a_2$  определяют граничные условия для системы сопряженных уравнений относительно функций  $p_0(t)$ ,  $p_1(t)$  и  $p_2(t)$  в точке  $t_{N+1} = T$ , т. е. условия трансверсальности. В рассматриваемом случае

$$p_0(T) = \psi_0^{(0)}(T) = a_0;$$

$$p_1(T) = \psi_1^{(0)}(T) = a_1; \quad p_2(T) = \psi_2^{(0)}(T) = a_2.$$

Зададим численные значения коэффициентов  $a_0$ ,  $a_1$  и  $a_2$  следующим образом:

$$a_0 = 0,2; \quad a_1 = 0,5; \quad a_2 = 0,3.$$

Наконец, зададим числовое значение величины  $\rho$ , характеризующей распределение инвестиций между материальным и потребительским секторами. Предположим, что  $\rho = 0,2$ . Таким образом, все необходимые значения параметров трехсекторной модели экономики заданы.

Разбиение отрезка времени  $[0, T]$  задается параметром  $N = 99$ , откуда  $\Delta = T/(N + 1) = 0,01$ . Рассмотрено множество возможных функций управления  $\hat{S}_3^{(N)}$ , состоящее из управлений, имеющих не более трех точек переключения на заданном интервале времени. В результате установлено, что в данном множестве имеется единственная функция управления  $u_{1*}(t) = 1$ ,  $t \in [0, T]$ , которая вместе с соответствующими ей функциями состояний  $(k_{0*}(t), k_{1*}(t), k_{2*}(t))$  является решением системы соотношений, содержащей необходимые условия и ограничения исходной задачи. Графические представления для функции  $u_{1*}(t)$ ,  $t \in [0, T]$ , а также соответствующих ей функций состояний  $k_{0*}(t)$ ,  $k_{1*}(t)$  и  $k_{2*}(t)$ , сопряженных переменных  $p_0(t)$ ,  $p_1(t)$  и  $p_2(t)$  и функции переключений  $Q(t) = Q(p_0(t), p_1(t), p_2(t))$  приведены на рис. 2.

Итак, для рассматриваемого набора численных значений исходных параметров в поставленной задаче оптимального управления имеется единственная допустимая экстремаль, задаваемая функциями  $(u_{1*}(t); k_{0*}(t), k_{1*}(t), k_{2*}(t))$ , изображенными на рис. 2, а и 2, г. В этом управляемом процессе переключения отсутствуют.

Отметим, что аналогичные результаты были получены еще для нескольких вариантов набора значений исходных параметров, частично отличающихся от основного.

## 7 Заключительные замечания

Численный алгоритм, изложению и анализу которого посвящена данная работа, позволяет завершить решение поставленной задачи оптимального управления. В ходе реализации данного алгоритма

используются программы, вычисляющие значения функции состояний и сопряженных переменных. Алгоритм осуществляет перебор возможных вариантов функции управления и выбирает те из них, которые являются допустимыми экстремальми. Таким образом, решение исходной задачи оптимального управления находится при помощи сочетания аналитических и численных методов.

Предлагаемый алгоритм может быть использован не только в рассмотренной задаче оптимального управления. Его можно использовать при решении различных задач, которые имеют следующие общие свойства:

- исследование задачи производится на основе принципа максимума;
- структура оптимального управления определяется некоторой функцией переключений, зависящей от сопряженных переменных;
- система сопряженных уравнений может зависеть, вообще говоря, от функций состояний и управлений.

Такие общие свойства имеет весьма широкий класс задач оптимального управления, возникающих при анализе экономических и технических систем.

## Литература

1. Шнурков П. В., Засыпко В. В. Оптимальное управление инвестициями в закрытой динамической модели трехсекторной экономики: математическая постановка задачи и общий анализ на основе принципа максимума // Вестник МГТУ им. Н. Э. Баумана. Сер. Естественные науки, 2014. Вып. 2. С. 101–115.
2. Шнурков П. В., Засыпко В. В. Аналитическое исследование задачи оптимального управления инвестициями в закрытой динамической модели трехсекторной экономики // Вестник МГТУ им. Н. Э. Баумана. Сер. Естественные науки, 2014. Вып. 4. С. 101–120.
3. *Intriligator M.* Математические методы оптимизации и экономическая теория / Пер. с англ. — М.: Айрис-Пресс, 2002. 553 с. (*Intriligator M. D.* Mathematical optimization and economic theory. — SIAM, 1971. 508 p.)
4. *Arrow A., Intriligator M. D., Hildenbrand W., Sonnenschein H.* Handbook of mathematical economics. — Amsterdam: North Holland, 1991. 750 p.
5. *Barro R., Sala-i-Martin X.* Economic growth. — Boston: MIT Press, 2003. 672 p.
6. *Ngai L. R., Pissarides C. A.* Structural change in a multi-sector model of growth. — London: Centre for Economic Policy Research, 2004. Discussion Paper No. 4763. 25 p.
7. *Алексеев В. М., Галеев Э. М., Тухомиров В. М.* Сборник задач по оптимизации. Теория. Примеры. Задачи. — М.: Физматлит, 2005. 256 с.

8. Арутюнов А. А., Магарил-Ильяев Г. Г., Тихомиров В. М. Принцип максимума Понтрягина. — М.: Факториал, 2006. 144 с.
9. Колемаев В. А. Трёхсекторная модель экономики // Сб. науч. тр. Междунар. академии информатизации. — М.: Копия-Принт, 1997. С. 335–345.
10. Колемаев В. А. Математическая экономика. — М.: Юнити-Дана, 2002. 399 с.
11. Алексеев В. М., Тихомиров В. М., Фомин С. В. Оптимальное управление. — М.: Физматлит, 2007. 408 с.
12. Колемаев В. А. Оптимальный сбалансированный рост открытой трёхсекторной экономики // Прикладная эконометрика, 2008. Вып. 3. С. 14–42.

Поступила в редакцию 08.11.15

## DEVELOPMENT OF THE ALGORITHM OF NUMERICAL SOLUTION OF THE OPTIMAL INVESTMENT CONTROL PROBLEM IN THE CLOSED DYNAMICAL MODEL OF THREE-SECTOR ECONOMY

P. V. Shnurkov<sup>1</sup>, V. V. Zasytko<sup>1</sup>, V. V. Belousov<sup>2</sup>, and A. K. Gorshenin<sup>2,3</sup>

<sup>1</sup>National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow 123458, Russian Federation

<sup>2</sup>Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

<sup>3</sup>Moscow Technological University (MIREA), 78 Vernadskogo Ave., Moscow 119454, Russian Federation

**Abstract:** The paper develops the numerical method of solution of the optimal investment control problem in the closed dynamical model of three-sector economy. The preceding papers described an analytical research of this problem by the method based on the Pontryagin maximum principle. In the present paper, the authors obtained analytical representations for state functions. Conjugate variables are used as the foundation of the numerical algorithm. The developed algorithm makes it possible to analyze the class of admissible control functions, having not more than the given finite number of points of switch, and to find among them those that satisfy the necessary optimality conditions and restrictions of the original task. The general scheme of the proposed algorithm can be used to investigate another optimal control tasks, connected with different subject areas. The developed algorithm is realized in a system of applied programs.

**Keywords:** model of three-sector economy; Pontryagin maximum principle; numerical method of solution of the optimal control problem

**DOI:** 10.14357/19922264160108

## References

1. Shnurkov, P. V., and V. V. Zasytko. 2014. Optimal'noe upravlenie investitsiyami v zakrytoy dinamicheskoy modeli trekhsektornoy ekonomiki: Matematicheskaya postanovka zadachi i obshchiy analiz na osnove printsipa maksimuma [Optimal control of investments in the closed-form dynamic model of three-sector economy: Mathematical statement of the problem and general analysis based on the maximum principle]. *Vestnik MGTU im. N. E. Baumana. Ser. Estestvennye nauki* [Herald of the Bauman Moscow State Technical University. Ser. Natural Sciences] 2:101–115.
2. Shnurkov, P. V., and V. V. Zasytko. 2014. Analiticheskoe issledovanie zadachi optimal'nogo upravleniya investitsiyami v zakrytoy dinamicheskoy modeli trekhsektornoy ekonomiki [Analytical study of optimal investments control problem in closed-form dynamic model of three-sector economics]. *Vestnik MGTU im. N. E. Baumana. Ser. Estestvennye nauki* [Herald of the Bauman Moscow State Technical University. Ser. Natural Sciences] 4:101–120.
3. Intriligator, M. D. 1971. *Mathematical optimization and economic theory*. SIAM. 508 p.
4. Arrow, A., M. D. Intriligator, W. Hildenbrand, and H. Sonnenschein. 1991. *Handbook of mathematical economics*. Amsterdam: North Holland. 750 p.
5. Barro, R., and X. Sala-i-Martin. 2003. *Economic growth*. MIT Press. 672 p.
6. Ngai, L. R., and C. A. Pissarides. 2004. Structural change in a multi-sector model of growth. London: Centre for Economic Policy Research. Discussion Paper No. 4763. 25 p.
7. Alekseev, V. M., E. M. Galeev, and V. M. Tikhomirov. 2005. *Sbornik zadach po optimizatsii. Teoriya. Primery*.

- Zadachi* [Collection of tasks on optimization. Theory. Examples. Tasks]. Moscow: Fizmatlit. 256 p.
8. Arutyunov, A. A., G. G. Magaril-II'yaev, and V. M. Tikhomirov. 2006. *Printsip maksimuma Pontryagina* [The Pontryagin maximum principle]. Moscow: Faktorial. 144 p.
  9. Kolemaev, V. A. 1997. Trekhsektornaya model' ekonomiki [Three-sector model]. *Sb. nauch. tr. Mezhdunar. Akademii Informatizatsii* [International Academy of Informatization Proceedings]. Moscow: Kopya-Print. P. 335–345.
  10. Kolemaev, V. A. 2002. *Matematicheskaya ekonomika* [Mathematical economics]. Moscow: Yuniti-Dana. 399 p.
  11. Alekseev, V. M., V. M. Tikhomirov, and S. V. Fomin. 2007. *Optimal'noe upravlenie* [Optimal control]. Moscow: Fizmatlit. 408 p.
  12. Kolemaev, V. A. 2008. Optimal'nyy sbalansirovanny rost otkrytoy trekhsektornoy ekonomiki [The optimal balanced growth of the open three-sector economy]. *Prikladnaya Ekonometrika* [Applied Econometrics] 3:14–42.

Received November 8, 2015

## Contributors

**Shnurkov Peter V.** (b. 1953) — Candidate of Science (PhD) in physics and mathematics, associate professor, National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow 123458, Russian Federation; pshnurkov@hse.ru

**Zasytko Veronika V.** (b. 1988) — PhD student, National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow, 123458, Russian Federation; vzasytko@gmail.com

**Belousov Vasilii V.** (b. 1977) — Candidate of Science (PhD) in technology, Head of Laboratory, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; VBelousov@ipiran.ru

**Gorshenin Andrey K.** (b. 1986) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; associate professor, Moscow Technological University (MIREA), 78 Vernadskogo Ave., Moscow 119454, Russian Federation; agorshenin@frcsc.ru

## МЕЛКОЗЕРНИСТЫЕ ГИБРИДНЫЕ ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ. ЧАСТЬ 2: ДВУНАПРАВЛЕННАЯ ГИБРИДИЗАЦИЯ

И. А. Кириков<sup>1</sup>, А. В. Колесников<sup>2</sup>, С. В. Листопад<sup>3</sup>, С. Б. Румовская<sup>4</sup>

**Аннотация:** Рассматривается проблематика междисциплинарных инструментариев и делается вывод об актуальности исследований свойства «зернистости» гибридов в информатике. Исследованы свойства функциональной и инструментальной неоднородности сложных задач и приведены результаты моделирования мелкозернистых гибридов в теории схем ролевых концептуальных моделей (РКМ). Результаты исследований показаны в рамках лингвистического подхода, суть которого состоит в трансформации вербализованной информации об объектах-оригиналах (сложных задачах) и объектах-прототипах (методах моделирования), имеющейся в полиязыках профессиональной деятельности, в объекты-результаты (функциональные гибридные интеллектуальные системы). Трансформация направляется эвристиками — схемами ролевых концептуальных моделей в неформальной аксиоматической теории. Категориальное ядро теории — «ресурс—свойство—действие—отношение». Введено понятие двунаправленной гибридации. Рассмотрены ее преимущества и приведены первые результаты.

**Ключевые слова:** логико-математический интеллект; гибридные интеллектуальные системы; лингвистический подход; теория ролевых концептуальных моделей; мелкозернистые гибриды; двунаправленная гибридация

**DOI:** 10.14357/19922264160109

### 1 Введение

В первой половине 1990-х гг. L. Medsker (Вашингтон, США) были анонсированы гибридные интеллектуальные системы (ГиИС), по существу совпадающие с интеллектуальными гибридными системами [1], гибридными интегрированными системами [2, 3], гибридными информационными системами [4] и гибридными интеллектуальными адаптивными системами [5]. В России сложилось несколько школ в этой области междисциплинарных знаний: Д. А. Поспелова—В. Б. Тарасова; В. Н. Вагина—А. П. Еремеева; Г. В. Рыбиной в Москве [3]; В. Ф. Пономарева—А. В. Колесникова в Калининграде [6]; Н. Г. Ярушкиной в Ульяновске [7].

Вследствие противоречий между свойствами объективной реальности, процессами, явлениями и событиями окружающего мира и научной картиной мира, а также из-за разнообразия информации, перерабатываемой при решении задач, и разнообразия знаний об объекте и окружающем мире перспективна выработка принципов, переводящих создание гибридов из уникальной ремесленной

мастерской в проектно-конструкторскую деятельность, в частности мелкозернистых функциональных ГиИС (ФГиИС).

В сочетании с лингвистическим подходом к гибридации [6, 8], развитым из [9, 10], в частности аксиоматической теорией ролевых концептуальных моделей и построенной в ней многоуровневой моделью внешнего мира как взаимоуязванной совокупностью логико-лингвистических, знаковых представлений «мира измерений», «мира ресурсов, свойств, процессов», «мира состояний и поведения», «мира субъектов-решателей (мира задач)», «мира решения задач», «мира методов моделирования задач», «мира субъектов-разработчиков (мира моделирования)» с использованием тетрады концептов «ресурс—свойство—действие—отношение», мелкозернистые ФГиИС позволят объединить естественный и визуальный языки в логико-математических и визуально-пространственных рассуждениях над проблемами, что релевантно феномену человеческого мышления [11].

Настоящая работа является продолжением статьи [12] и призвана сформулировать понятие класса мелкозернистых ГиИС в терминах РКМ как

<sup>1</sup>Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, baltbipiran@mail.ru

<sup>2</sup>Балтийский федеральный университет им. И. Канта; Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, avkolesnikov@yandex.ru

<sup>3</sup>Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, ser-list-post@yandex.ru

<sup>4</sup>Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, sophiyabr@gmail.com

цели двухнаправленной гибридизации и рассмотрим ее методологические и технологические стороны.

## 2 Объекты-оригиналы гибридизации

Решение задач системами управления шло в условиях изменяющегося мирового рынка, их эволюции от тейлоровских к посттейлоровским организациям и смены научной картины мира [8]. В то время как стандартизация и практика автоматизированного управления определяли термины «задача» (технично-экономическая сущность, выходная информация, входная информация, алгоритм решения) и «комплекс задач», наука (математика, исследование операций, теория управления, теория принятия решений, искусственный интеллект, системный анализ, лингвистика), исследуя методы решения в мире искусственно упрощенных, игровых ситуаций, выработала широкий спектр терминов и мнений — от «простой задачи» до «сложной задачи», «проблемы».

В целях упорядочения терминологии в [12] введена качественная шкала и мера сложности моделирования задач. Тем не менее в [13] высказывалось мнение, что, несмотря на одинаковое содержание терминов «проблема» и «задача», между ними есть ситуативная разница. Первый применяется в ситуациях, когда «у нас есть проблема», а второй применялся опосредованно в ситуациях, когда «необходимо решить задачу». Измерения по шкале сложности задач («простая»—«сложная») относили проблему ближе к правой границе и переводили в сферу коллективного интеллекта. Такое же мнение высказал и В. Ф. Спиридонов [14], исследуя психологию решения задач и проблем. Сравни-

тельный анализ свойств проблем и сложных задач (см. таблицу) показывает их содержательное, смысловое совпадение.

Исследования информации о сложных задачах [12, 17] показали ее разнообразие: функциональную и инструментальную неоднородность.

*Функциональная неоднородность* сложных задач проявляется в наличии взаимосвязанных областей однородных параметров-переменных: детерминированных, стохастических, лингвистических, генетических, — на которых аналитическими, статистическими, экспертными, нечеткими, нейросетевыми, генетическими представлениями заданы причинно-следственные связи в рассуждениях экспертов. Таким образом, функциональная неоднородность связана с проблемой «целое—части» или «сложная задача — подзадачи» (рис. 1). Это явление слабо изучено в системном анализе и инженерии знаний. Качество ее исследования определяет глубину понимания задачи.

На рис. 1 модель задачи имеет двухуровневое представление: на макроуровне — задача как целое и ее свойства ( $prb^u$ ); на микроуровне — система из подзадач  $prb^h$  (светлые кружки) и координирующей задачи  $prb^k$  (темный кружок).

Зададим множество языков профессиональной деятельности (ЯПД)

$$LANG^p = \{LANG_1^p, \dots, LANG_{N_{LANG}}^p\}$$

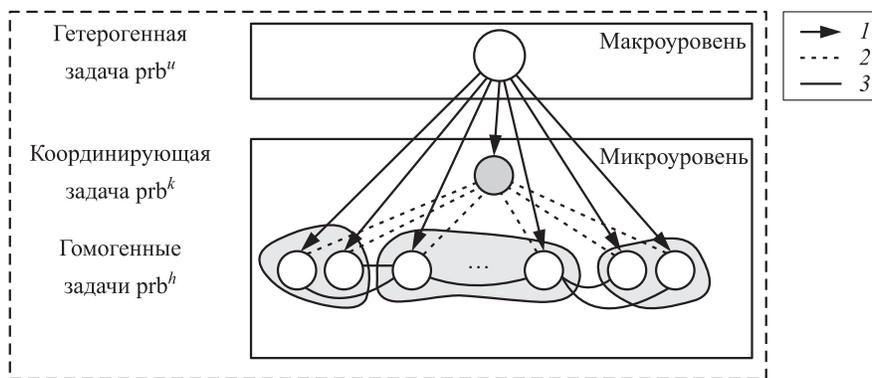
в системе  $S$  и соответствие

$$\Psi_1 \subseteq LANG^p \times S | \Psi_1 \neq \emptyset,$$

где  $S = \{STR_1, STR_2, STR_3\}$  для модели  $mod_1^S$  [11, 17] заочных консультаций и  $S = \{STR_1, \dots, STR_4\}$  для модели  $mod_2^S$  [11, 17] очных консультаций. Одной страте может соответствовать более одного

Определения и свойства проблем и сложных задач

<b>Определение проблемы по В. Ф. Спиридонову</b>	<b>Свойства проблемы</b> (психология, В. Ф. Спиридонов [14])
Проблема (англ. complex problem, ill-structured problem) — это затруднительные условия без явно сформулированной цели или четкая цель, не связанная со сложившимися неблагоприятными условиями	Отсутствие и необходимость поиска исходной формулировки, комплексное строение, «навязчивый» и сетевой характер, непрозрачность, собственная динамика, «человеческое» измерение, межпредметное содержание
Формула идентичности терминов «проблема» $\cong$ «сложная задача»	
<b>Определение сложной задачи</b>	<b>Свойства сложной задачи</b>
<b>Сложная задача</b> (англ. complex problem) — взаимодействие в форме гетерогенной системы с неопределенностью	(искусственный интеллект, А. В. Колесников, И. А. Кириков [8, 15–17]) Субъективность, системность, неоднородность, динамичность, неопределенность ситуации решения, полиязыковой характер, внутренняя несогласованность



**Рис. 1** Двухуровневое представление сложной задачи: 1 — отношения включения; 2 — отношения координации; 3 — отношения декомпозиции

$LANG_1^p \in LANG^p | q = 1, \dots, N_{LANG}$ . Тогда *гетерогенной предметной областью* назовем

$$E^{LANG} = \langle S, LANG^p, \Psi_1 \rangle.$$

Пусть в  $E^{LANG}$  есть гетерогенные задачи  $PRB^u = \{prb_1^u, \dots, prb_{N_{PRBU}}^u\}$  и  $\forall prb_l^u \exists PRB_l^h = \{prb_1^h, \dots, prb_{N_{lh}}^h\}$ , где  $l = 1, \dots, N_{PRBU}$ ,  $\forall l (N_{lh} = \text{var } y)$ ,  $prb_l^u \in PRB^u$ . Допустим, что  $prb_l^u$  могут возникать только на  $STR_j | j = 2, 3$  и  $STR_4$  для  $\text{mod}_1^S$ - и  $\text{mod}_2^S$ -стратификаций соответственно. Зададим соответствия  $\Psi_2 \subseteq PRB^u \times S$  и  $\Psi_3 \subseteq PRB^h \times S$ , где  $PRB^h = \prod_l^{N_{PRBU}} PRB_l^h$ , причем более одной  $prb^h$  есть на  $STR_j | j = 1, \dots, 3$  и для  $\text{mod}_2^S$ -стратификации  $\forall (prb_4^u STR_4) \in \Psi_2 \exists (PRB_1^h STR_1), \exists (PRB_2^h STR_2), \exists (PRB_3^h STR_3) \in \Psi_3$ , где  $PRB_1^h, \dots, PRB_3^h \subseteq PRB^h$ , а для  $\text{mod}_1^S$ -стратификации  $\forall (prb_3^u STR_3) \in \Psi_2 \exists (PRB_1^h STR_1), \exists (PRB_2^h STR_2) \in \Psi_3$  и  $\forall (prb_2^u STR_2) \in \Psi_2 \exists (PRB_1^h STR_1) \in \Psi_3$ .

Тогда *гетерогенной проблемной средой* назовем

$$E^u = \langle E^{LANG}, PRB^u, PRB^h, \Psi_2, \Psi_3 \rangle. \quad (1)$$

Сформулируем свойства  $E^u$ .

1. В силу свойств  $E^{LANG}$  это редуцированное, многоуровневое представление системы  $S$ , на стратах которого есть свои ЯПД и задачи.
2. Задачи на страте  $STR_4$  по определению гетерогенные, а на стратах  $STR_1$ – $STR_3$  по определению гомогенные (однородные).
3. По определению методы решения  $prb^h$  известны, и на стратах  $STR_1$ – $STR_3$  могут строиться модели решения подзадач.
4. По определению методы решения  $prb^u$  на страте  $STR_4$  неизвестны, и нет априори заданных моделей.

5. В силу свойств  $S$  между стратами существуют двунаправленные потоки информации, координирующие решение однородных задач в составе гетерогенной.

### 3 Объекты-прототипы гибридации

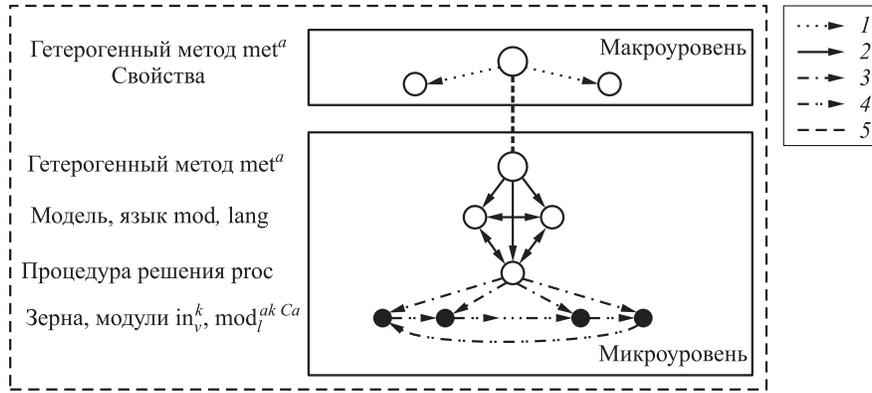
*Инструментальная неоднородность* сложных задач проявляется в разнообразии методов решения ее частей — подзадач — как следствие редукции сложной задачи. Это явление из мира субъекта моделирования, оно также еще до конца не изучено.

В гетерогенной проблемной среде (1) проявляются ограниченность, достоинства и недостатки метода. Он модифицируется настолько, что превращается в новый метод. Иногда метод-первооснова перестает существовать, впитывая «дополнения» и «изменения», заимствуемые из других методов. Такой процесс можно рассматривать как эволюцию популяции методов и распространить на него подходы генетики. Для этого представим методы-сущности в макро- и микроуровнях (рис. 2).

На макроуровне метод — ресурс разработчика для решения задачи с отличительными свойствами и представляется схемой РКМ [11, 17]:

$$\begin{aligned} \text{met}^a &= R^{\text{met met}}(\text{met}, \text{MET}) \circ R^{\text{met pr}}(\text{met}, \text{ch}_1) \circ \\ &\circ R^{\text{met pr}}(\text{met}, \text{ch}_2) \circ R^{\text{met pr}}(\text{met}, \text{ch}_3) \circ \\ &\circ R^{\text{met act}}(\text{met}, \text{ACT}_1) \circ R^{\text{met act}}(\text{met}, \text{ACT}_2) \circ \\ &\circ R^{\text{met prb}}(\text{met}, \text{prb}^h) \circ R^{\text{met pr}}(\text{met}, \text{SPC}^m), \end{aligned}$$

где  $\text{ch}_i \in \text{CH} \subseteq \text{PR}$  — классификатор (аналитический, статистический, символичный, коннекционистский, эволюционный);  $\text{ch}_2 \in \text{CH} \subseteq \subseteq \text{PR}$  — модель (черный ящик, система массового



**Рис. 2** Двухуровневое представление метода: 1 — отношения определения, предназначения; 2 — отношения ресурсов, ресурсов и действий; 3 — отношения включения; 4 — отношения следования; 5 — «соответствовать»

обслуживания, система автоматического управления, условие-действие, серый ящик, эволюция, ситуация-решение);  $ch_3 \in CH \subseteq PR$  — язык описания (уравнения, алгоритм, продукции, матрицы и др.);  $ACT_1 \subseteq ACT$  — процедура получения решения (прямой, обратный вывод в экспертных системах, прямое распространение в нейросетях, нечеткий вывод, методы решения уравнений, машинные эксперименты с генетическим или моделирующим алгоритмом; методы установления соответствия на множествах прецедентов, поиска аналогов и сохранения единиц опыта в памяти и др.);  $ACT_2 \subseteq ACT$  — процедура обучения (обратное распространение, алгоритм Кохонена, непараметрическое обучение и др.);  $SPC^m$  — спецификатор (погрешность, гибридные возможности [11, 17], знания о преимуществах и недостатках, работа с шумом, адаптивность и др. [11, 17]);  $R^{met\ met}$ ,  $R^{met\ act}$  — отношения определения;  $R^{met\ prb}$  — отношения предназначения.

Микроуровневое представление  $met^a$  содержит генетическую информацию и трактует метод как отношения модели, языка, процедуры, специфицируемых схемой РКМ:

$$\begin{aligned}
 met^a = & R^{met\ res}(met, RES) \circ R^{res\ res}(RES, mod) \circ \\
 & \circ R^{res\ res}(RES, lang) \circ R^{res\ act}(RES, proc) \circ \\
 & \circ R^{res\ res}(lang, mod) \circ R^{res\ act}(mod, proc) \circ \\
 & \circ R^{res\ act}(lang, proc),
 \end{aligned}$$

где  $met$  — метод;  $mod \in MOD \subseteq RES$ ,  $lang \in LANG \subseteq RES$ ,  $proc \in ACT$  — схемы РКМ модели, языка и процедуры;  $R^{res\ res}$  — отношения ресурсов («метод–модель», «метод–язык», «язык–модель»);  $R^{res\ act}$  — отношения ресурсов и действий («метод–процедура», «модель–процедура», «язык–процедура»).

Составные части «модель», «язык» и «процедура» на микроуровне ( $mod$ ,  $lang$  и  $proc$ ) определяются схемами РКМ:

$$\begin{aligned}
 mod = & R^{res\ res}(mod, RES) \circ R^{res\ res}(RES, RES) \circ \\
 & \circ R^{pr\ pr}(PR, PR) \circ R^{act\ act}(ACT, ACT) \circ \\
 & \circ R^{res\ pr}(RES, PR) \circ R^{res\ act}(RES, ACT) \circ \\
 & \circ R^{act\ pr}(ACT, PR); \\
 lang = & R^{res\ res}(lang, RES) \circ R^{res\ res}(RES, RES) \circ \\
 & \circ R^{pr\ pr}(PR, PR) \circ R^{act\ act}(ACT, ACT) \circ \\
 & \circ R^{res\ pr}(RES, PR) \circ R^{res\ act}(RES, ACT) \circ \\
 & \circ R^{act\ pr}(ACT, PR); \\
 proc = & R^{act\ act}(proc, ACT) \circ R^{res\ res}(RES, RES) \circ \\
 & \circ R^{pr\ pr}(PR, PR) \circ R^{act\ act}(ACT, ACT) \circ \\
 & \circ R^{res\ pr}(RES, PR) \circ R^{act\ res}(ACT, RES) \circ \\
 & \circ R^{act\ pr}(ACT, PR).
 \end{aligned}$$

Таким образом, «зернистость» возникает на микроуровне применительно к структурным особенностям метода.

Пусть  $Iv^m = \{Iv_1^m, \dots, Iv_{N_{Iv^m}}^m\}$  — множество экспертов в коллективе разработчиков. Введем множество их ЯПД  $LANG^m = \{LANG_1^m, \dots, LANG_{N_{LANG^m}}^m\}$  и установим невязимо однозначное соответствие  $\Psi^{LANG, Iv} : LANG^m \rightarrow Iv^m$ : не всюду определено (есть языки, которыми модельеры не владеют), сюръективно (некомпетентные модельеры к гибридизации не привлекаются), нефункционально (одним и тем же языком владеют несколько экспертов).

Введем множества макроуровневых и микроуровневых представлений автономных методов:  $MET^a$ ,  $MET^a$ . Зададим взаимно однозначные соответствия  $\Psi_i^{met\ met}$  при  $i \in \overline{1, 5}$  и  $(met_{dj}^a, met_{ql}^a)$ ,

$j = l; d, q \in \{An, St, Lg, Li, Ep\}; d = q: \Psi_i^{\text{met met}} = \text{MET}_d^a \rightarrow \text{MET}_q^a$ .

Зададим множества кортежей  $((\text{met}_{dj}^a, \text{met}_{ql}^a), \text{LANG}_w^m)$ , где  $w \in \overline{1, N_{\text{LANG}}}$ , как соответствия  $\Psi_i^{\text{MET LANG}}: \Psi_i^{\text{met met}} \rightarrow \text{LANG}^m$ : не всюду определены (не каждый метод «знаком» модельерам), сюръективны (модельер владеет минимум одним методом), не функциональны (метод выражается на разных языках) и не взаимно однозначны.

Неоднородной исследовательской средой назовем

$$E^M = \langle \text{LANG}^m, \text{MET}^a, \text{MET}^a, \Psi_i^{\text{met met}}, \Psi^{\text{MET LANG}} \rangle.$$

Данная модель отображает в ходе гибридизации основные особенности коллективной разработки: двухуровневое представление метода моделирования; многообразие методов работы с различными видами знаний; разнообразие языков профессиональной деятельности модельеров.

## 4 Объекты-результаты гибридизации

Актуальность применения в системах управления мелкозернистых ФГиИС связана с проблемными ситуациями, для которых нет ни знаний, ни опыта и которые могли бы быть отображены в гетерогенном модельном поле. Причина таких ситуаций — уникальность проблемы в сочетании с усугубляющим трудность ситуации положением, когда известные методы не могут быть применены к решению однородных(ой) задач(и).

Например, пусть для решения некоторой однородной задачи подходит экспертная система. Однако поскольку предположительно ее база знаний имеет большой размер, то известные алгоритмы ее интерпретации будут работать медленно. Такая ситуация не позволяет применить при гибридизации крупнозернистый элемент, построенный как экспертная система, и требует вмешательства в микроуровневое представление. Цель — заменить отдельные процедуры интерпретатора — мелкие зерна — на, чаще всего, полнофункциональный, т. е. крупнозернистый элемент, построенный по другому методу. Иное объяснение следует из необходимости использовать не только классы трансформационных и ФГиИС, как в случае крупнозернистой гибридизации, но и полиморфические ГиИС, которые сконструировать из крупного зерна невозможно.

Наконец, в рамках мелкозернистых гибридов проще решать технологические задачи, сопутствующие решению подзадач, например: извлечения

знаний для экспертной системы из декларативной информации других элементов ФГиИС; корректировки знаний элементов ГиИС после того, как в базу знаний одного из элементов были внесены изменения; пополнения исходной информации, поступающей на вход элемента как за счет его собственной базы знаний и данных, так и баз данных и баз знаний других элементов. Все они требуют разработки механизма вмешательства в декларативную и(или) процедурную составляющие элемента ФГиИС априори, еще до синтеза ГиИС. На рис. 3 дана иерархия схем РКМ для представления знаний о мелкозернистых ФГиИС. Для этого в направлении сверху вниз изображены схемы РКМ мелкозернистой ФГиИС ( $\alpha^u$ ), крупнозернистой ФГиИС ( $\alpha^u(t)$ ), двух вариантов элементов ФГиИС ( $\alpha_j^h(t)$ ), метода на макро- ( $\text{met}^a$ ) и микроуровне ( $\text{met}^a$ ), модели ( $\text{mod}$ ), языка ( $\text{lang}$ ) и процедуры ( $\text{proc}$ ) как частей метода.

Движение по иерархии сверху вниз — детализация, а снизу вверх — агрегирование. В месте, где указано отношение «соответствовать», возникает «разрыв» включения одних знаний в другие.

На рис. 3 использованы обозначения:  $\text{RES}^{n-1}$  — множество элементов-зерен функциональных  $\alpha^h \in \in A^h$  и технологических  $\alpha^\tau \in A^\tau$  элементов;  $R^{\text{res res}}$  и  $\dot{R}^{\text{res res}}$  — отношения включения и интеграции зерен из  $\text{RES}^{n-1}$  в знак ФГиИС  $\text{res}^n$ ;  $\text{pr}_1^n$  — вектор исходных данных  $\text{DAT}^u$  задачи  $\text{prb}^u$ ;  $\text{pr}_2^n$  — вектор выходных данных элементов, цель  $\text{GL}^u$  задачи  $\text{prb}^u$ ;  $\text{pr}_3^n$  — вектор состояния ФГиИС;  $\text{RES}^n$  — множество знаков информационного языка;  $\text{PR}_1^n$  и  $\text{PR}_2^n$  — множества свойств «вход» и «выход» элементов из  $\text{RES}^n$  соответственно;  $R_6^{\text{pr pr}}, R_7^{\text{pr pr}}$  и  $R_8^{\text{pr pr}}$  — отношения функционирования ФГиИС;  $\dot{R}^{\text{res res}}$  — отношения интеграции [11, 17];  $R_9^{\text{pr pr}}$  и  $R_{10}^{\text{pr pr}}$  — отношения «входа» ФГиИС и «входов» элементов, а также «выходов» элементов и «выхода» ФГиИС соответственно;  $t$  — автоматное время.

Возникает задача комбинирования на микроуровне. Пусть имеем множества:  $\text{In}^k = \{in_1^k, in_2^k, \dots, in_{N_{\text{in}}}^k\}$  — интерпретаторов моделей вычислений автономных методов [11, 17]);  $A^\tau = \{\alpha_{j1}^\tau, \alpha_{j2}^\tau, \dots, \alpha_{jN_{\alpha^\tau}}^\tau\}$  — технологических компонентов;  $\text{MOD}^{ak} \text{Ca} = \{\text{mod}_1^{ak} \text{Ca}, \text{mod}_2^{ak} \text{Ca}, \dots, \text{mod}_{N_{\text{met Ca}}}^{ak} \text{Ca}\}$  — декларативных составляющих автономных методов;  $\text{ACT}^k = \{\text{act}_1^k, \text{act}_2^k, \dots, \text{act}_{N_{\text{act k}}}^k\}$  — действий разработчика ФГиИС. Тогда имеют смысл следующие соответствия:

$$\Psi^{\text{in } \alpha} = \left\{ (in_v^k, \alpha_{j\omega}^\tau) \mid in_v^k \in \text{IN}^k, \alpha_{j\omega}^\tau \in A^\tau, \right. \\ \left. k \in \{An, St, Lg, Li, Ep\}, v, j \in \{1, \dots, 7\}, \right. \\ \left. \omega = 1, \dots, N_{\alpha, \tau} \right\},$$

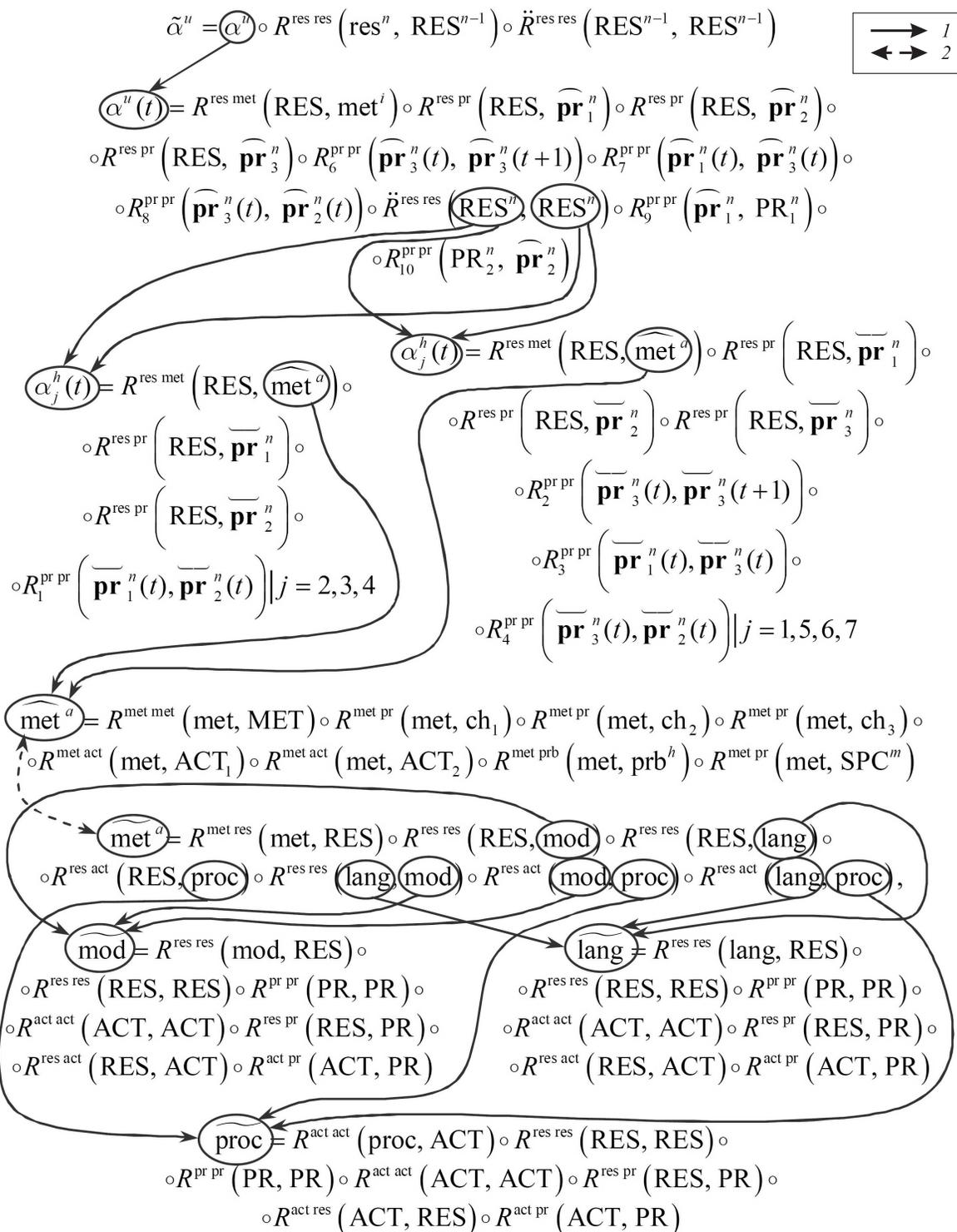


Рис. 3 Иерархия схем ролевых концептуальных моделей для моделирования знаний о мелкозернистых ФГИИС: 1 — отношение замены («что» и «на что» заменяется); 2 — «соответствовать»

$$\Psi^{\text{met Ca } \alpha} = \left\{ \left( \text{mod}_l^{\text{ak Ca}}, \alpha_{j\omega}^{\tau} \right) \mid \text{mod}_l^{\text{ak Ca}} \in \text{MOD}^{\text{ak Ca}}, l \in \{1, \dots, 7\} \right\},$$

$$\Psi^{\text{actk } \alpha} = \left\{ \left( \text{act}_v^k, \alpha_{j\omega}^{\tau} \right) \mid \text{act}_v^k \in \text{ACT}_k \right\},$$

где  $\text{in}_v^k$ ,  $\text{mod}_l^{\text{ak Ca}}$  и  $\alpha_{j\omega}^{\tau}$  — части интерпретатора в модели вычислений автономного метода (назовем их *зернами*), декларативной составляющей модели вычислений автономного метода и полнофункциональный технологический элемент (назовем их *модулями*) соответственно.

### 5 Двухнаправленная гибридизация

В зависимости от особенностей объектов-результатов будем различать и виды гибридизации в информатике (рис. 4). Исследования гетерогенной проблемной среды на искусственных гетерогенных системах класса «крупнозернистые ФГиИС» [11, 17] названы *однонаправленной гибридизацией* [8]. Исследования гетерогенной проблемной среды и гетерогенной исследовательской среды на искусственных гетерогенных системах класса «мелкозернистые ФГиИС» [8, 11, 17] названы *двухнаправленной гибридизацией* [8].

Первый вариант связан с трансформацией информации так, что гибриды-потомки как интегрированный метод решения сложной задачи наследуют родительские признаки — функциональную неоднородность сложной задачи как объекта-оригинала, а родительские признаки методов опосредованно (как есть), через гетерогенное модельное поле. Это направление показано номерами 1–4. Ему соответствует однонаправленная гибридизация. Ее смысл состоит в том, что сложная задача

должна быть рассмотрена на макро- и микроуровне. Макроуровень (фенотип задачи) — вся задача в целом как сложная сущность, система. Микроуровень (генотип сложной задачи) — совокупность подзадач, связанная в декомпозициях классами отношений. Макро- и микроуровневые отображения задачи взаимосвязаны и должны рассматриваться в единстве. Однонаправленная гибридизация требует исследований и извлечения знаний о макро- и микроуровневом представлении задачи и их взаимозависимостях. Эти исследования должны проводить системные аналитики совместно с экспертами из коллектива, принимающего решения.

Второй вариант связан с трансформацией информации и наследованием родительских признаков как сложной задачи (ее функциональной неоднородности), так и непосредственно (как надо) объектов-прототипов. Это направление показано стрелками 1, 5–9. Этому направлению соответствует двухнаправленная гибридизация. Ее смысл состоит в том, что каждый метод из ограниченной совокупности должен быть рассмотрен на макро- и микроуровне. Макроуровень (фенотип метода) — метод в целом как сложная сущность, система. Микроуровень (генотип метода) — совокупность зерен «модель», «язык описания», «процедура получения решения» или зерен более детального уровня как составных частей процедуры решения. Макро- и микроуровневые отображения метода взаимосвязаны и должны рассматриваться в единстве. Двухнаправленная гибридизация помимо работ направления 1–4 требует: исследования и извлечения знаний о возможностях методов; исследования и извлечения знаний о макро- и микроуровневых представлениях методов и взаимозависимостях. Эти исследования должны проводить системные аналитики совместно с экспертами, владеющими методами моделирования (модельерами).

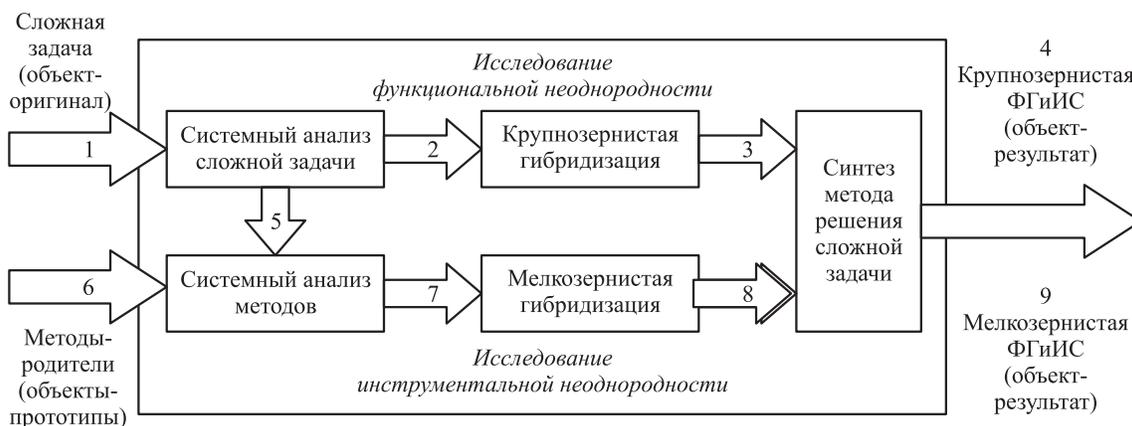


Рис. 4 Одно- и двухнаправленная гибридизация в информатике

Каждая из ФГиИС, полученных двухнаправленной гибридизацией, должна иметь «сходство» со сложной задачей, а для каждой ее части конструируется «хороший» метод-потомок.

Преимущества двухнаправленной гибридизации:

- (1) модельная архитектура мелкозернистых ФГиИС отображает на компьютере функциональную структуру задачи, области ее однородных параметров (детерминированных, стохастических, логических, лингвистических и др.), представленных множествами подзадач, что позволяет вести несколько линий рассуждений;
- (2) для решения подзадач можно использовать не только известные методы моделирования (методы-родители), но и конструировать из набора инструментальных средств (зерен) методы-потомки, лишенные родительских недостатков;
- (3) мелкозернистые ФГиИС разрабатываются методами целенаправленной гибридизации, управляемой знаниями о функциональной и инструментальной неоднородности сложной задачи, позволяющей выявить плюсы и минусы и путем декомпозиции процедурной составляющей построить наборы типовых инструментальных средств-зерен;
- (4) компьютерная интерпретация модели «мелкозернистая ФГиИС» может рассматриваться как синтез метода решения задачи, адаптируемого к изменениям во внешней среде, условиям задачи, релевантного неоднородности подобных задач, наследующего сильные стороны методов-родителей, приближающегося к возможностям естественного коллективного интеллекта.

Для выполнения однонаправленной гибридизации разработана и апробирована проблемно-структурная методология, соответствующая технология и инструментальные средства поддержки [6, 8]. Для выполнения двухнаправленной гибридизации разработана и частично апробирована проблемно-инструментальная методология [12]. Технология и инструментарий апробированы для узкого класса дигридов искусственных нейронных сетей и генетических алгоритмов.

## 6 Заключение

В рамках лингвистического подхода к гибридизации рассмотрено моделирование мелкозернистых ФГиИС в неформальной аксиоматической теории схем ролевых концептуальных моделей. Такие системы обладают значительным потенциалом

имитации рассуждений специалистов, решающих сложные задачи, и разрабатываются методами целенаправленной гибридизации, управляемой двумя видами знаний. Прежде всего, это знания о функциональной неоднородности сложной практической задачи, полученные в ходе ее системного анализа. Второй вид знаний — это информация о микроуровневых представлениях методов, полученная в ходе их системного анализа с использованием триады «язык—модель—процедура» и позволяющая не только выявить и систематизировать плюсы (силы) и минусы (слабости), но и путем декомпозиции процедурной составляющей построить наборы типовых инструментальных средств-зерен. В итоге компьютерная интерпретация модели «мелкозернистая ФГиИС» рассматривается как синтез метода решения сложной задачи, легко адаптируемого к изменениям во внешней среде, условиям решения задачи, релевантного неоднородности подобных задач и наследующего сильные стороны методов-родителей.

## Литература

1. *Goonatilake S., Khebbal S.* Intelligent hybrid systems // 1st Singapore Conference (International) on Intelligent Systems Proceedings, 1992. P. 356–364.
2. *Поспелов Г. С.* Искусственный интеллект — основа новой информационной технологии. — М.: Наука, 1988. 280 с.
3. *Рыбина Г. В.* Интегрированные экспертные системы: современное состояние, проблемы и тенденции // Известия РАН. Теория и системы управления, 2002. № 5. С. 111–126.
4. *Medsker L. R.* Hybrid intelligent systems. — Kluwer Academic Publ., 1995. 295 p.
5. *Kasabov N., Kozma R.* Hybrid intelligent adaptive systems: A framework and a case study on speech recognition // Int. J. Intell. Syst., 1998. Vol. 13. P. 455–466.
6. *Колесников А. В.* Гибридные интеллектуальные системы. Теория и технология разработки / Под ред. А. М. Яшина. — СПб.: СПбГТУ, 2001. 711 с.
7. *Ярушкина Н. Г.* Основы теории нечетких и гибридных систем. — М.: Финансы и статистика, 2004. 320 с.
8. *Колесников А. В., Кириков И. А.* Методология и технология решения сложных задач методами функциональных гибридных интеллектуальных систем. — М.: ИПИ РАН, 2007. 387 с.
9. *Уемов А. И.* Вещи, свойства, отношения. — М.: Институт философии АН СССР, 1963. 184 с.
10. *Поспелов Д. А.* Ситуационное управление: теория и практика. — М.: Наука, 1986. 288 с.
11. *Колесников А. В.* Моделирование естественных гетерогенных систем коллективного принятия решений // Системный анализ и информационные техно-

- логии (САИТ-2015): Тр. 6-й Междунар. конф. — М.: ИСА РАН, 2015. Т. 1. С. 7–16.
12. Колесников А. В., Кириков И. А., Листопад С. В., Румовская С. Б., Доманицкий А. А. Решение сложных задач коммивояжера методами функциональных гибридных интеллектуальных систем / Под ред. А. В. Колесникова. — М.: ИПИ РАН, 2011. 295 с.
  13. Самсонова М. В., Ефимов В. В. Технология и методы коллективного решения проблем. — Ульяновск: УлГТУ, 2003. 152 с.
  14. Спиридонов В. Ф. Психология мышления: Решение задач и проблем. — М.: Генезис, 2006. 319 с.
  15. Колесников А. В., Кириков И. А. Концептуальная модель двунаправленной гибридизации при разработке компьютерных систем поддержки принятия решений // Системы и средства информатики. — М.: Наука, 2008. Доп. вып. С. 21–53.
  16. Кириков И. А., Колесников А. В., Листопад С. В. Моделирование самоорганизации групп интеллектуальных агентов в зависимости от степени согласованности их взаимодействия // Информатика и ее применения, 2009. Т. 3. Вып. 4. С. 76–86.
  17. Колесников А. В. Принципы и методология разработки информационных гетерогенных систем двунаправленной гибридизации // Интегрированные модели и мягкие вычисления в искусственном интеллекте: Сб. науч. тр. VIII Междунар. науч.-практич. конф. — М.: Физматлит, 2015. Т. 1. С. 36–53.

Поступила в редакцию 20.09.15

## FINE-GRAINED HYBRID INTELLIGENT SYSTEMS. PART 2: BIDIRECTIONAL HYBRIDIZATION

I. A. Kirikov<sup>1</sup>, A. V. Kolesnikov<sup>1,2</sup>, S. V. Listopad<sup>1</sup>, and S. B. Rumovskaya<sup>1</sup>

<sup>1</sup>Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation

<sup>2</sup>Immanuel Kant Baltic Federal University, 14 Nevskogo Str., Kaliningrad 236041, Russian Federation

**Abstract:** The problematic of interdisciplinary tools is considered and the conclusion about the relevance of research of the “grain” property of hybrids in informatics is made. Properties of functional and instrumental heterogeneity of complex tasks are investigated and the results of fine-grained hybrids modeling within the theory of the schemes of role conceptual models are presented. The results are presented within the linguistic approach, the core of which is in transformation of the verbalized information about objects-originals (complex subjects) and objects-prototypes (modeling approaches) to objects-results (functional hybrid intelligent systems). It exists in poly-languages of professional activity. The transformation is directed by heuristics which are the schemes of the conceptual role models in the informal axiomatic theory. The categorical core of the theory is “resource–property–operation–relation.” The notion of bidirectional hybridization, its benefits, and the first results are represented.

**Keywords:** logical-mathematical intelligence; hybrid intelligent systems; linguistic approach; theory of role conceptual models; fine-grained hybrids; bidirectional hybridization

**DOI:** 10.14357/19922264160109

## References

1. Goonatilake, S., and S. Khebbal. 1992. Intelligent hybrid systems. *1st Singapore Conference (International) on Intelligent Systems Proceedings*. 356–364.
2. Pospelov, G. S. 1988. *Iskusstvennyy intellekt — osnova novoy informatsionnoy tekhnologii* [Artificial intelligence — the base of the new information processing technology]. Moscow: Nauka. 280 p.
3. Rybina, G. V. 2002. Integrated expert systems: State of the art, problems, and trends]. *J. Comput. Sys. Sc. Int.* 41(5):780–793.
4. Medsker, L. R. 1995. *Hybrid intelligent systems*. Kluwer Academic Publ. 295 p.
5. Kasabov N., and R. Kozma. 1998. Hybrid intelligent adaptive systems: A framework and a case study on speech recognition. *Int. J. Intell. Syst.* 13:455–466.
6. Kolesnikov, A. V. 2001. *Gibridnye intellektual'nye sistemy. Teoriya i tekhnologiya razrabotki* [Hybrid artificial systems. Theory and development technology]. St. Petersburg: SPbGTU. 711 p.
7. Jarushkina, N. G. 2004. *Osnovy teorii nechetkikh i gibridnykh sistem* [Foundations of theory of fuzzy and hybrid systems]. Moscow: Finance and Statistics. 320 p.
8. Kolesnikov, A. V., and I. A. Kirikov. 2007. *Metodologiya i tekhnologiya resheniya slozhnykh zadach metodami funktsional'nykh gibridnykh intellektual'nykh sistem* [Methodology and technology for solving of complex problems

- using the methodology of functional hybrid artificial systems]. Moscow: IPI RAN. 387 p.
9. Uemov, A. I. 1963. *Veshchi, svoystva, otnosheniya* [Items, properties, relations]. Moscow: Institute of Philosophy of the Academy of Sciences of the USSR. 184 p.
  10. Pospelov, D. A. 1986. *Situatsionnoe upravlenie: Teoriya i praktika* [Situation control: Theory and practice]. Moscow: Nauka. 288 p.
  11. Kolesnikov, A. V. 2015. Modelirovanie estestvennykh geterogennykh sistem kollektivnogo prinyatiya resheniy [Modeling of the natural heterogeneous systems of collective decision-making]. *6th Conference (International) "Systems Analysis and Information Technology" Proceedings*. Moscow: ISA RAS. 1:7–16.
  12. Kolesnikov, A. V., I. A. Kirikov, S. V. Listopad, S. B. Rumovskaya, and A. A. Domanitskiy. 2010. *Reshenie slozhnykh zadach kommvoyazhera metodami funktsional'nykh gibridnykh intellektual'nykh sistem* [Solving of the complex traveling salesman problem using the methodology of functional hybrid artificial systems]. Ed. A. V. Kolesnikov. Moscow: IPI RAN. 295 p.
  13. Samsonova, M. V., and V. V. Efimov. 2003. *Tehnologiya i metody kollektivnogo resheniya problem* [The technology and methods of collective decision-making]. Ulyanovsk: UISTU. 152 p.
  14. Spiridonov, V. F. 2006. *Psikhologiya myshleniya: Reshenie zadach i problem* [The psychology of mind: Solving of tasks and problems]. Moscow: Genesis. 319 p.
  15. Kolesnikov, A. V., and I. A. Kirikov. 2008. Kontseptual'naya model' dvunapravlennoy gibridizatsii pri razrabotke komp'yuternykh sistem podderzhki prinyatiya resheniy [Substructured model of bidirectional hybridization within the development of computer-based systems for decision-making support]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics*. Add. Issue:21–53.
  16. Kirikov, I. A., A. V. Kolesnikov, and S. V. Listopad. 2009. Modelirovanie samoorganizatsii grupp intellektual'nykh agentov v zavisimosti ot stepeni soglasovannosti ikh vzaimodeystviya [Imitation of self-organization of intelligent agents teams in case of agreement dimension of their interaction]. *Informatika i ee Primneniya — Inform. Appl.* 3(4):76–86.
  17. Kolesnikov, A. V. 2015. Printsipy i metodologiya razrabotki informatsionnykh geterogennykh sistem dvunapravlennoy gibridizatsii [Approach and methodology of the development of informational heterogeneous systems of bidirectional hybridization]. *Integrirovannye Modeli i Myagkie Vychisleniya v Iskusstvennom Intellekte: Sb. nauch. tr. VIII Mezhdunar. Nauch.-Praktich. Konf.* [8th Research and Practice Conference (International) "Integrated Models and Soft Computing in Artificial Intelligence" Proceedings]. Moscow: Fizmatlit. 1:36–53.

Received September 20, 2015

## Contributors

**Kirikov Igor A.** (b. 1955) — Candidate of Sciences (PhD) in technology; director, Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation; baltbipiran@mail.ru

**Kolesnikov Alexander V.** (b. 1948) — Doctor of Sciences in technology; professor, Department of Telecommunications, Immanuel Kant Baltic Federal University, 14 Nevskogo Str., Kaliningrad 236041, Russian Federation; senior scientist, Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation; avkolesnikov@yandex.ru

**Listopad Sergey V.** (b. 1984) — Candidate of Sciences (PhD) in technology; scientist, Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation; ser-list-post@yandex.ru

**Rumovskaya Sophiya B.** (b. 1985) — programmer I, Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation; sophiyabr@gmail.com

## ПРЕДСТАВЛЕНИЕ КРОССЯЗЫКОВЫХ ЗНАНИЙ О КОННЕКТОРАХ В НАДКОРПУСНЫХ БАЗАХ ДАННЫХ\*

И. М. Зацман<sup>1</sup>, О. Ю. Инькова<sup>2</sup>, М. Г. Кружков<sup>3</sup>, Н. А. Попкова<sup>4</sup>

**Аннотация:** Рассматриваются используемые в контрастивных лингвистических исследованиях базы данных (БД), получившие название «надкорпусных». Они формируются в результате обработки текстов, хранящихся в двуязычных параллельных подкорпусах Национального корпуса русского языка. В них каждому тексту на русском языке соответствует один или несколько его переводов на другой язык, а каждому тексту на иностранном языке — один его перевод на русский язык. Каждый текст на языке оригинала и его переводы выровнены по предложениям. Надкорпусные БД (НБД) представляют собой новый вид лингвистических ресурсов, которые предназначены для целенаправленного извлечения новых знаний о широком спектре языковых единиц (ЯЕ). Эти знания необходимы для повышения качества машинного перевода, актуализации моно- и двуязычных грамматик, а также для обновления многообразных образовательных курсов по лингвистике, теории и практике перевода. В статье дается описание концепции формирования НБД и примера реализации такой базы для представления знаний о коннекторах русского языка и об их переводах на французский язык.

**Ключевые слова:** кроссыязыковые знания; коннекторы русского языка; представление знаний о коннекторах; надкорпусные базы данных

**DOI:** 10.14357/19922264160110

### 1 Введение

Создание электронных корпусов текстов позволяет значительно ускорить сбор исходных данных для проведения лингвистических исследований, включая выполнение запросов на поиск исследуемых языковых единиц [1]. При решении задач контрастивной лингвистики используются параллельные корпуса, в которых хранятся тексты на одном языке и их переводы на другой язык, выровненные, как правило, по предложениям. Параллельные корпуса могут объединяться в рамках более крупных хранилищ текстов.

Примером такого хранилища является Национальный корпус русского языка (<http://ruscorpora.ru>). На данный момент он содержит параллельные подкорпуса для английского, армянского, белорусского, болгарского, бурятского, испанского, итальянского, латышского, немецкого, польского, украинского, французского и эстонского языков.

Когда исследователь находит в корпусе интересные его объекты, у него нередко возникает потребность в создании для этих объектов формализованных описаний, которые могли бы значительно расширить данные, содержащиеся в корпусе. При этом создание таких описаний иногда подразумевает формирование и использование авторских систем классификации, которые отсутствуют в корпусе. В некоторых корпусах существует возможность экспорта данных, найденных в результате лексико-грамматического поиска исследуемых языковых единиц, но их последующее хранение, авторское описание и классификация не входят в перечень функций корпусов. Более того, сама идея создания электронных корпусов текстов не предполагает реализации этих функций.

В работах [2–4] была описана технология формирования БД, которая существенно дополняет функциональность параллельных корпусов при решении задач контрастивной лингвистики. В пе-

\* Работа выполнена в Институте проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук при поддержке РФФИ (проект 14-07-00785), РГНФ (проект 16-24-41002) и ШННФ (проект IZLRZ1\_164059).

<sup>1</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, [izatsman@yandex.ru](mailto:izatsman@yandex.ru)

<sup>2</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Женевский университет, [Olga.Inkova@unige.ch](mailto:Olga.Inkova@unige.ch)

<sup>3</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, [magnit75@yandex.ru](mailto:magnit75@yandex.ru)

<sup>4</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, [natasha\\_popkova@mail.ru](mailto:natasha_popkova@mail.ru)

риод 2013–2015 гг. эта БД использовалась для исследования глагольных форм русского языка в зеркале их переводов на французский язык [5]. В ней реализован двуязычный лексико-грамматический поиск, который позволяет находить интересные пользователя глагольные формы русского языка, задавая для них сочетания лексических и грамматических признаков и одновременно соответствующие им функционально эквивалентные фрагменты в текстах на французском языке, а также получать информацию о частотности моделей перевода этих глагольных форм ([http://a179.ipi.ac.ru/corpora\\_dynasty/main.aspx](http://a179.ipi.ac.ru/corpora_dynasty/main.aspx)).

В стадии формирования находятся еще две БД [6], предназначенные для контрастивного исследования лингвоспецифичных единиц (ЛСЕ) и коннекторов русского языка; описание второй из них будет дано в разд. 3. Опыт их применения показал, что, с одной стороны, расширение функциональности параллельных корпусов необходимо для решения современных задач контрастивной лингвистики, а с другой стороны, три БД (глагольных форм, ЛСЕ и коннекторов) представляют собой новый вид лингвистических информационных ресурсов, которые получили название «надкорпусные БД» [7, 8].

Основным логическим элементом НБД является установленное лингвистом переводное соответствие, т.е. двухместный кортеж, включающий языковую единицу в оригинальном тексте и ее функционально эквивалентный фрагмент<sup>1</sup> в тексте перевода в привязке к контексту их употребления (такие кортежи далее будут называться двуязычными кортежами). Двуязычные кортежи формируются профессиональными лингвистами, которые анализируют и сопоставляют выровненные параллельные тексты, загруженные в НБД (см. подробнее [3]). В процессе их формирования нередко выявляются лакуны в существующих сопоставительных грамматических описаниях. Например, лингвист может обнаружить некоторые переводные соответствия, которые отсутствуют в этих описаниях, но которые дают важную дополнительную информацию о семантике и функционировании изучаемых языковых явлений. Возможность специфицирования таких новых для контрастивной лингвистики переводных соответствий с помощью двуязычных кортежей является характерной чертой НБД, отличающей ее от других видов лингвистических информационных ресурсов (параллельных корпусов, типологических БД и т.д. [7, 8]).

Основная цель этой статьи состоит в описании концепции формирования НБД как нового

вида лингвистических информационных ресурсов, а также примера ее реализации для представления в НБД кроссязыковых знаний о коннекторах русского языка и об их переводах на французский язык.

## 2 Основные термины и концептуальные положения

В этом разделе приведены определения терминов из работ [3, 4], предложенные Анной Зализняк и адаптированные к описанию концепции формирования НБД. В этих работах для типологизации глагольных форм применяются лингвистические понятия «базовый вид» и «дополнительный признак» формы. В терминах информатики как информационно-компьютерной науки [11, 12] применение этих понятий для типологизации представляет собой одновременное и независимое использование двух классификаций по разным основаниям.

Таким образом, по сути, применяется фасетная классификация глагольных форм, в которой выделены одно главное и одно дополнительное основания. Рубрика, присвоенная языковой единице по главному основанию, в НБД глагольных конструкций названа базовым видом, а в НБД коннекторов — речевой реализацией. По дополнительному основанию ЯЕ могут быть присвоены ноль, одна или несколько рубрик, значения которых были названы ее дополнительными признаками во всех НБД. Отметим, что названия рубрик фасетной классификации (ФК) выбираются лингвистами. От этого выбора не зависит функционирование НБД. Использование ФК с двумя основаниями иллюстрируется в статье на примере формирования НБД коннекторов.

Отметим три важных свойства ФК, которая применяется для категоризации языковых единиц в процессе проведения контрастивных лингвистических исследований. Во-первых, по главному основанию может быть присвоена только одна рубрика. Во-вторых, по дополнительному основанию может быть присвоено любое число рубрик. В-третьих, эта классификация является темпоральной, т.е. число уровней классификации по каждому основанию, рубрик каждого уровня, их названия и дефиниции могут изменяться во времени. Это, с одной стороны, позволяет удовлетворить требование адаптируемости НБД к видам исследуемых ЕЯ и специфике решаемых лингвистических задач, с другой стороны, решать задачи представления в НБД системы кроссязыковых знаний, которые формируются и пополняются в процессе проведения контрастивных исследований.

<sup>1</sup>Термин «функционально эквивалентный фрагмент» был введен Д. О. Добровольским в его работах с соавторами [9, 10].

Принимая во внимание требование адаптируемости НБД в качестве исходного, сформулируем первое положение концепции формирования НБД.

**Положение 1.** Должна быть обеспечена возможность формировать и редактировать ФК, используемую в НБД для описания исследуемых ЯЕ, которая включает одно главное основание и одно дополнительное.

**Таблица 1** Названия девяти рубрик первого уровня по главному основанию и число рубрик на втором уровне

№	Рубрика первого уровня по главному основанию	Число рубрик второго уровня
1	при	5
2	иначе	2
3	в.связи	3
4	прочие	21
5	сочетания	6
6	только	3
7	то CNT part*	2
8	кстати	4
9	не то чтобы	6

\* Рубрика «то CNT part» используется в тех случаях, когда слово *то* является элементом многокомпонентного коннектора, например «*Не то что к Соне, а к Дуне пойдешь!*» (Ф. М. Достоевский. «Преступление и наказание»).

Для иллюстрации этого положения приведем краткое описание ФК с одним главным и одним дополнительным основаниями, используемой в НБД коннекторов русского языка. Для ее главного основания определены два уровня классификации. На первом уровне используется 9 рубрик по состоянию на 1 января 2016 г. (табл. 1).

В табл. 2 приведены названия шести рубрик второго уровня для 9-й рубрики первого уровня «не то чтобы» и число двуязычных кортежей, сформированных по каждой из этих рубрик на конец каждого квартала 2015 г.

Рубрика «прочие» (см. строку № 4 в табл. 1) включает те коннекторы, для которых на 1 января 2016 г. не была проведена классификация по восьми другим рубрикам первого уровня главного основания. В результате анализа содержания этой рубрики в дальнейшем могут добавляться новые рубрики первого уровня или редактироваться существующие. Начиная с 2016 г. помимо изменения числа рубрик могут редактироваться их названия и дефиниции.

Приведем пример предложения с коннектором, который относится к рубрике второго уровня «не то что (расстояние) а» и его перевод на французский язык, кроме текста в квадратных скобках.

<sup>1</sup> В этом примере ФЭФ — это слова *non que... mais* в тексте перевода, которые, по мнению лингвиста, являются переводным соответствием для коннектора *не то что... а*.

**Пример 1** [И. А. Гончаров. «Обломов».]:

«— Как: ты бы и мне не поверил?»

— Ни за что; **не то что** тебе, а все может случиться: [ну, как лопнет, вот я и без гроша];

«Pour rien au monde **non que** je ne te fasse confiance **mais** tout peut arriver».

В этом примере в процессе построения двуязычного кортежа для коннектора *не то что... а* лингвист сначала определяет в переводе его функционально эквивалентный фрагмент (ФЭФ)<sup>1</sup> и контекст ФЭФ. Затем путем объединения коннектора *не то что... а* и его ФЭФ *non que... mais* в НБД строится двуязычный кортеж типа *не то что... а; non que... mais*.

Кроме главного основания ФК для ЯЕ (ФК ЯЕ) в НБД коннекторов есть еще дополнительное. По состоянию на 1 января 2016 г. для него также определены два уровня классификации, включая 6 рубрик первого уровня (табл. 3).

В табл. 4 приведены названия 10 рубрик второго уровня для рубрики первого уровня «Отношения», включая название особой рубрики «отношение подлежит определению», а также число двуязычных кортежей, сформированных по каждой из этих рубрик на конец каждого квартала 2015 г.

Для построения двуязычных кортежей на основе параллельных текстов лингвисты должны проводить анализ исследуемой ЯЕ оригинального текста, идентифицировать и классифицировать соответствующий ей ФЭФ с использованием текста перевода. При этом для классификации ФЭФ создается еще одна темпоральная ФК для ФЭФ (ФК ФЭФ). Анализ приведенного примера позволяет сформулировать следующее положение концепции формирования НБД.

**Положение 2.** Должна быть обеспечена возможность загружать и использовать параллельные тексты, содержащие исследуемые ЯЕ и их ФЭФ в тексте перевода, которые идентифицируются и классифицируются с использованием ФК ФЭФ.

Для иллюстрации этого положения приведем описание ФК ФЭФ, которая по состоянию на 1 января 2016 г. включала два уровня по главному основанию. Первый уровень включает три рубрики:

- (1) «CNT» для ФЭФ, которые являются коннекторами французского языка;
- (2) «TVD (прочие)» для ФЭФ, рубрика которых временно не определена;
- (3) «Zego» для отсутствующих ФЭФ (к этой рубрике относятся те ситуации, когда отсутствует перевод исследуемой ЯЕ).

**Таблица 2** Названия шести рубрик второго уровня для рубрики первого уровня «не то чтобы» и число двуязычных кортежей на конец каждого квартала 2015 г.

№	Рубрика второго уровня	Число кортежей в НБД			
		на 31.03.2015	на 30.06.2015	на 30.09.2015	на 31.12.2015
1	не то что	0	0	4	24
2	не то что (расстояние) а	0	0	1	11
3	не то что (расстояние) но	0	0	0	5
4	не то чтобы	0	0	7	8
5	не то чтобы (расстояние) а	0	0	5	5
6	не то чтобы (расстояние) но	0	1	1	1

**Таблица 3** Названия шести рубрик первого уровня по дополнительному основанию и число рубрик на втором уровне

№	Рубрика первого уровня по дополнительному основанию	Число рубрик второго уровня
1	Отношения	10
2	Структура	7
3	Позиция	3
4	Порядок	3
5	Статус	6
6	Расположение	2

В НБД коннекторов ФК ЯЕ и ФК ФЭФ сейчас полностью совпадают по дополнительному основанию. Однако в общем случае такое совпадение может отсутствовать — лингвисты при необходимости могут формировать разные ФК ЯЕ и ФК ФЭФ по дополнительному основанию. Например, в НБД глагольных форм классификации по дополнительному основанию в ФК ЯЕ и ФК ФЭФ не совпадают (см. табл. 3 и 4 в работе [3]).

Возможность формировать и редактировать ФК ЯЕ зафиксирована в первом концептуальном положении. Приведем аналогичное положение для ФК ФЭФ.

**Положение 3.** Должна быть обеспечена возможность формировать и редактировать ФК ФЭФ, используя

емую в НБД для описания и классификации ФЭФ, которая включает одно главное основание и одно дополнительное. В общем случае ФК ЯЕ и ФК ФЭФ по дополнительному основанию не совпадают и могут создаваться независимо друг от друга.

Число двуязычных кортежей, построенных в НБД глагольных форм, сейчас превышает десять тысяч. Формирование ФК ЯЕ и ФК ФЭФ дает возможность реализовать не только лексический, но и грамматический поиск кортежей, используя их рубрики, причем для решения задач контрастивной лингвистики часто необходим двуязычный лексико-грамматический поиск. Он дает возможность одновременно задать в запросе поисковые признаки для исследуемой ЯЕ на одном языке и для ее ФЭФ на другом. Такая поисковая возможность редко присутствует в параллельных корпусах. Опишем эту возможность в виде отдельного концептуального положения.

**Положение 4.** В НБД должна быть предусмотрена возможность формировать двуязычные кортежи и задавать лексико-грамматические критерии их поиска, используя одновременно слова, а также рубрики ФК для исследуемой ЯЕ и ФК ФЭФ для ее перевода, включая случаи его отсутствия.

В НБД каждому тексту на русском языке может соответствовать несколько его переводов на другой

**Таблица 4** Названия десяти рубрик второго уровня для рубрики «Отношения» и число двуязычных кортежей

№	Рубрика второго уровня по дополнительному основанию	Число кортежей в НБД			
		на 31.03.2015	на 30.06.2015	на 30.09.2015	на 31.12.2015
1	Временные отношения	0	0	0	0
2	Отношение подлжит определению	38	78	365	381
3	Отношения замещения	0	3	15	35
4	Присоединительные отношения	0	0	0	0
5	Противительные отношения	4	8	8	8
6	Соединительные отношения	38	38	38	38
7	Сопоставительные отношения	4	4	4	4
8	Сравнительные отношения	0	0	0	0
9	Условные отношения	0	0	0	0
10	Уступительные отношения	8	8	8	8

язык, а каждому тексту на иностранном языке — несколько его переводов на русский язык. Вернемся к примеру 1 с предложением, содержащим коннектор *ne то что... , а*. Для этого предложения в НБД есть еще один перевод, так как в НБД коннекторов есть доступ к текстам двух переводов этого произведения.

**Пример 2** со вторым переводом текста на русском языке из примера 1 на французский язык [И. А. Гончаров. «Обломов»]:

«...Aucunement. Et **ce n'est pas**, je t'assure, **parce que** c'est toi, **mais** je pense à tout ce qui peut se passer...» (ФЭФ *ce n'est pas... parce que... , mais*).

Для второго перевода был построен еще один двуязычный кортеж типа *(не то что... , а; ce n'est pas... parce que... , mais)*. В этом случае недостаточно иметь возможность лексико-грамматического поиска для одного варианта перевода. Нужно также учесть случаи, когда один текст имеет несколько переводных вариантов.

**Положение 5.** При использовании НБД должна быть обеспечена возможность формировать описания и классифицировать несколько вариантов перевода одной и той же ЯЕ; должен быть предусмотрен лексико-грамматический поиск с использованием одновременно лексики, рубрик ФК исследуемой ЯЕ и ФК ФЭФ для нескольких вариантов ее перевода.

Приведенные в этом разделе концептуальные положения формирования НБД являются основой для разработки ее концептуальной и логической схем данных, а также информационной технологии создания НБД и их применения для решения задач контрастивной лингвистики по исследованию широкого спектра ЯЕ, в частности коннекторов русского языка.

### 3 Надкорпусные базы данных коннекторов и категории двуязычных кортежей

Выровненные параллельные тексты являются потенциальным источником новых, но труднодоступных кроссязыковых знаний. Тексты переводов, выровненные относительно оригинальных текстов, являются уникальным и постоянно пополняемым источником таких знаний. Однако для их извлечения необходимо существенно дополнить функциональность параллельных корпусов, что и предлагается делать с помощью НБД, используя методы

и модели теории генерации знаний [5, 13–16]. Согласно концепции формирования НБД, основные положения которой были рассмотрены в предыдущем разделе, для описания новых знаний предназначены следующие четыре компонента НБД: ФК исследуемых ЯЕ, ФК их ФЭФ, двуязычные кортежи и система связей между первыми тремя компонентами. В данном разделе дано описание этой системы и предлагается подход к категоризации двуязычных кортежей на примере НБД коннекторов, структура которых и некоторые аспекты функционирования на данный момент малоизучены.

Коннектор — это «языковая единица, функция которой состоит в выражении определенного типа отношений — логико-семантических, иллюкутивных<sup>1</sup>, структурных, — существующих между двумя соединенными с ее помощью компонентами, имеющими предикативный характер, выраженными имплицитно или эксплицитно» [17]. Лексико-грамматическая природа коннекторов и их структура разнообразны. Коннектор может состоять из одного (*хотя, притом, кстати*) и большего числа слов (*не только... , но и; скорее... , чем...* ); первые являются по своему составу однокомпонентными (пример 3), вторые — многокомпонентными (примеры 1 и 4).

**Пример 3.** [И. А. Гончаров. «Обломов»]:

Нет, у него чернильница полна чернил, на столе лежат письма, бумага, даже гербовая, **притом** испи-санная его рукой.

**Пример 4.** [И. А. Гончаров. «Обломов»]:

Со времени смерти стариков хозяйственные дела в деревне **не только** не улучшились, **но**, как видно из письма старосты, становились хуже.

С лингвистической точки зрения интерес представляют как многокомпонентные коннекторы в целом, так и их составные части; этот последний аспект особенно важен для изучения структуры коннекторов [18]. Поэтому основным принципом формирования НБД коннекторов является следующий: многокомпонентный коннектор как ЯЕ описывается с помощью двуязычных кортежей сначала в целом, а затем описываются по отдельности составляющие его компоненты. На всех этапах описания у исследователя есть возможность приписать левому и правому элементам сформированного двуязычного кортежа:

- по одной рубрике ФК ЯЕ и ФК ФЭФ по главному основанию;

<sup>1</sup>То есть устанавливаемых на уровне речевых актов. Так, в высказывании *Он не пришел, потому что заболел* коннектор *потому что* устанавливает логико-семантическое отношение причины на уровне двух ситуаций, описанных в главном и в придаточном предложениях. А в высказывании *Позвони ему, пожалуйста, потому что он может обидеться* тот же коннектор устанавливает то же отношение, но на уровне речевого акта, совершенного в главном предложении (я тебя **прошу** позвонить), и вводит причину этой просьбы (иначе он обидится).

— ноль, одну или несколько рубрик ФК ЯЕ и ФК ФЭФ по дополнительному основанию этих классификаций.

Таким образом, традиционные и новые знания о структуре коннекторов и их функциях в тексте представлены в НБД в виде двуязычных кортежей, т. е. упорядоченных пар, состоящих из исследуемой ЯЕ и ее ФЭФ, рубрик эволюционирующих ФК ЯЕ и ФК ФЭФ, а также системы связей между элементами кортежей и рубриками. Для иллюстрации приведем еще один пример двуязычного кортежа, построенного в НБД для коннектора *не только*... , *но* из примера 4 и следующего его перевода на французский язык:

**Пример 5** с переводом текста на русском языке из примера 4 на французский язык [И. А. Гончаров. «Обломов»]:

Depuis la mort des parents, les affaires du domaine **non seulement** ne s'amélioraient pas, **mais**, à en croire la lettre du régisseur, empiraient.

Для этого коннектора и его ФЭФ в предложении из примера 5 построен двуязычный кортеж типа  $\langle \text{не только} \dots, \text{но}; \text{non seulement} \dots \text{mais} \rangle$ . В НБД коннекторов этот кортеж сопровождается рубриками ФК ЯЕ и ФК ФЭФ, которые будут описаны в следующем разделе.

В завершение этого раздела опишем кратко 9 категорий двуязычных кортежей, формируемых в НБД коннекторов. В процессе сопоставительного анализа параллельных русских текстов и их французских переводов, полученных из Национального корпуса русского языка (НКРЯ), в настоящее время лингвистами формируются русско-французские соответствия в виде трех из девяти теоретически определенных категорий. Следующие три категории используются в тех случаях, когда некоторому тексту соответствует только один перевод:

- (1)  $\langle c, Fc \rangle$ , где  $c$  — коннектор русского языка;  $Fc$  — его ФЭФ на французском языке;
- (2)  $\langle b, Fb \rangle$ , где  $b$  — блок коннектора (его неэлементарная составная часть);  $Fb$  — его ФЭФ;
- (3)  $\langle e, Fe \rangle$ , где  $e$  — элемент коннектора (его элементарная составная часть);  $Fe$  — его ФЭФ.

В процессе описания каждого многокомпонентного коннектора могут использоваться двуязычные кортежи всех трех перечисленных категорий, но необходимость их построения и количество кортежей определяются лингвистом. Обязательным является только построение кортежа первой категории, т. е. для всего многокомпонентного коннектора.

Приведем пример коннектора и его ФЭФ, для описания структуры которого может понадобиться

использование двуязычных кортежей нескольких категорий.

**Пример 6** [И. А. Гончаров. «Обломов»]:

...мягкость, которая была господствующим и основным выражением, **не** лица **только**, а всей души;

...l'expression de douceur, qui dominait **non seulement** le visage **mais aussi** l'âme.

Для многокомпонентного коннектора из этого примера *не*... *только*, а могут быть построены кортежи всех трех категорий, поскольку он состоит из разложимого блока *не только*, для которого может быть построен кортеж второй категории, и из элемента коннектора *а*, для которого может быть построен кортеж третьей категории (также при необходимости могут быть построены кортежи для элементов, составляющих блок *не только*).

Для текстов, имеющих несколько вариантов перевода, после построения лингвистом двуязычных кортежей первых трех категорий автоматически, т. е. программным способом, без участия лингвистов, могут быть сформированы русско-французские кортежи еще трех категорий, в которых второй компонент является множеством ФЭФ из разных переводов одного и того же текста:

- (4)  $\langle c, \{Fc_1, \dots, Fc_n\} \rangle$ , где  $c$  — коннектор русского языка;  $Fc_1$  — его ФЭФ в первом переводе; ...;  $Fc_n$  — его ФЭФ в  $n$ -м переводе;
- (5)  $\langle b, \{Fb_1, \dots, Fb_n\} \rangle$ , где  $b$  — блок коннектора;  $Fb_1$  — его ФЭФ в первом переводе; ...;  $Fb_n$  — его ФЭФ в  $n$ -м переводе;
- (6)  $\langle e, \{Fe_1, \dots, Fe_n\} \rangle$ , где  $e$  — неделимый элемент коннектора;  $Fe_1$  — его ФЭФ в первом переводе; ...;  $Fe_n$  — его ФЭФ в  $n$ -м переводе.

В перспективе также планируется провести сопоставительный анализ параллельных текстов, состоящих из выровненных по предложениям оригинальных произведений на французском языке и их переводов на русский, полученных на основе соответствующего подкорпуса НКРЯ. Это даст возможность сформировать в НБД коннекторов французско-русские кортежи еще трех категорий:

- (7)  $\langle Sc, c \rangle$ , где  $c$  — многокомпонентный коннектор русского языка;  $Sc$  — «стимул» французского языка, обусловивший появление этого коннектора в переводе на русский язык;
- (8)  $\langle Sb, b \rangle$ , где  $b$  — блок коннектора русского языка;  $Sb$  — «стимул» французского языка, обусловивший появление этого блока коннектора в переводе на русский язык;
- (9)  $\langle Se, e \rangle$ , где  $e$  — неделимый элемент коннектора русского языка;  $Se$  — «стимул» французского

языка, обусловивший появление этого элемента коннектора в переводе на русский язык.

Одновременное формализованное описание в НБД не только всего многокомпонентного коннектора как единой ЯЕ, но и его составляющих позволяет, с одной стороны, фиксировать все возможные блоки и элементы, составляющие коннекторы, характеризующиеся высокой степенью вариативности, а с другой стороны, даст возможность отслеживать, в какие многокомпонентные коннекторы может входить тот или иной блок или его элементы.

#### 4 Рубрицирование коннекторов и их функционально эквивалентных фрагментов

Приведенные ранее примеры построения двуязычных кортежей предполагают, что предварительно был выполнен поиск в НБД по шаблонам, которые формируются для каждого исследуемого коннектора. В результате поиска по сформированному шаблону из заданного массива параллельных текстов отбираются все пары предложений, удовлетворяющие критерию поиска (см. табл. 5 с первыми

четырьмя парами выровненных предложений, найденных в НБД коннекторов по сочетанию «не только»; для четвертой пары с использованием текстов примеров 4 и 5 был построен двуязычный кортеж типа *⟨не только... , но; non seulement... mais⟩*).

Лингвист анализирует каждую найденную пару предложений в отдельности, отмечает все составляющие коннектора и его контекст, а также ФЭФ и его контекст. Коннектор русского языка и его ФЭФ во французском тексте — это левый и правый элементы формируемого двуязычного кортежа соответственно.

Как отмечалось ранее, для рубрицирования коннекторов (левый элемент двуязычных кортежей) и их ФЭФ (правый элемент) используются две в общем случае независимые системы классификации: ФК ЯЕ и ФК ФЭФ. Таблица 6 отражает результаты рубрицирования в НБД коннекторов левого и правого элементов (а также их контекстов) двуязычного кортежа типа *⟨не только... , но; non seulement... mais⟩* в виде рубрик по двум основаниям ФК ЯЕ и ФК ФЭФ.

Первый ее столбец содержит коннектор *не только... , но* в контексте всего предложения на русском языке. Второй столбец содержит рубрики ФК ЯЕ для этого коннектора. По главному основанию

**Таблица 5** Первые четыре пары выровненных предложений, найденных в НБД коннекторов по сочетанию слов «не только»

Оригинальный текст	Перевод
Захар не старался изменить <b>не только</b> данного ему Богом образа, но и своего костюма, в котором ходил в деревне.	Zakhar n'avait rien fait pour changer l'apparence que Dieu lui avait donnée ni le costume qu'il avait porté à la campagne.
Он его представлял себе чем-то вроде второго отца, который только и дышит тем, как бы за дело и не за дело, сплошь да рядом, награждать своих подчиненных и заботиться <b>не только</b> о их нуждах, но и об удовольствиях.	Il se l'imaginait comme une sorte de second père qui ne pensait qu' à distribuer des primes à ses employés, qu'ils le méritent ou non, à tort et à travers, et qu' à veiller non seulement à leurs besoins mais aussi à leur bien-être.
Это происходило, как заметил Обломов впоследствии, оттого, что есть такие начальники, которые в испуганном до одурения лице подчиненного, выскочившего к ним навстречу, видят <b>не только</b> почтение к себе, но даже ревность, а иногда и способности к службе.	Comme Oblomov le remarqua plus tard, la cause en était que certains supérieurs voyaient dans la mine effrayée d'un employé qui s'empressait à leur rencontre, non seulement une preuve de respect pour eux, mais aussi un signe de zèle et même d'aptitude au service.
Со времени смерти стариков хозяйственные дела в деревне <b>не только</b> не улучшились, но, как видно из письма старосты, становились хуже.	Depuis la mort des parents, les affaires du domaine non seulement ne s'amélioraient pas, mais, à en croire la lettre du régisseur, empiraient.

**Таблица 6** Результаты рубрицирования левого и правого элементов двуязычного кортежа типа *⟨не только... , но; non seulement... mais⟩*

Оригинальный текст	Перевод
Со времени смерти стариков хозяйственные дела в деревне <b>не только</b> не улучшились, <b>но</b> , как видно из письма старосты, становились хуже.	Depuis la mort des parents, les affaires du domaine <b>non seulement</b> ne s'amélioraient pas, <b>mais</b> , à en croire la lettre du régisseur, empiraient.
<b>не</b>    <b>только</b>    <b>но</b> ⟨ TBD ⟩ ⟨ CNT p CNT q ⟩ ⟨ CNT ⟩ ⟨ Дистант ⟩	<b>non seulement</b>    <b>mais</b> ⟨ TBD ⟩ ⟨ CNT p CNT q ⟩ ⟨ CNT ⟩ ⟨ Дистант ⟩

Таблица 7 Дополнительные признаки двуязычных кортежей

№	Код дополнительного признака кортежа	Описание дополнительного признака кортежа
1	Exp	Требуется экспертиза кортежа экспертом
2	Ext_Up	Для построения кортежа необходимо учесть контекст предыдущей пары предложений
3	Ext_Down	Для построения кортежа необходимо учесть контекст следующей пары предложений
4	Up_Down	Для построения кортежа необходимо учесть контекст предыдущей и следующей пар предложений
5	NB	Интересный лингвистический пример (для использования в публикациях)
6	Cngrn	Коннектор русского языка переведен коннектором французского языка
7	Dvrg	Коннектор русского языка переведен ЯЕ или конструкцией другой категории (не коннектором)

**Замечание:** Определения дополнительных признаков Cngrn (конгруэнтность) и Dvrg (дивергентность) заимствованы из работы [1] и адаптированы для случая контрастного исследования коннекторов. В основу использования признака Cngrn положено функциональное соответствие исследуемой языковой единицы и ее ФЭФ, т.е. в данном случае принадлежность ФЭФ к функциональному классу коннекторов.

рубрика второго уровня «не||только||но» для этого коннектора принадлежит рубрике первого уровня «только» (см. строку 6 в табл. 1). По дополнительному основанию этому коннектору присвоено 4 рубрики:

- (1) ⟨TBD⟩ относится к рубрике первого уровня «Отношения» (см. строку 1 в табл. 3) и говорит о том, что отношение, выражаемое этим коннектором, будет определено позже;
- (2) ⟨CNT р CNT q⟩ относится к рубрике первого уровня «Порядок» (см. строку 4 в табл. 3) и говорит о том, что элементы многокомпонентного коннектора находятся в каждом из соединяемых фрагментов текста р и q;
- (3) ⟨CNT⟩ относится к рубрике первого уровня «Статус» (см. строку 5 в табл. 3) и говорит о том, что кортеж построен для всего коннектора, а не для отдельных его блоков или элементов;
- (4) ⟨Дистант⟩ относится к рубрике первого уровня «Расположение» (см. строку 6 в табл. 3) и говорит о том, что части коннектора разделены текстом.

Третий столбец таблицы содержит ФЭФ *non seulement... mais* в контексте перевода всего предложения на французский язык. Четвертый столбец содержит рубрики ФК ФЭФ. По главному основанию рубрика второго уровня «non seulement... mais» (коннектор и имя его рубрики посимвольно могут совпадать, поэтому имя ставится в кавычки) говорит о том, что ФЭФ *non seulement... mais* относится к категории коннекторов французского языка, т.е. принадлежит рубрике первого уровня

«CNT». По дополнительному основанию этому коннектору присвоены те же 4 рубрики, что и коннектору *не только... но* в тексте оригинала. Однако эти рубрики могут, в принципе, не совпадать.

После завершения рубрицирования левого и правого элементов каждого кортежа лингвист может присвоить несколько дополнительных признаков всему кортежу в целом, используя табл. 7.

## 5 Двуязычный поиск в надкорпусной базе данных коннекторов

Построенные двуязычные кортежи могут быть найдены по лексике обеих своих составляющих, рубрикам ФК ЯЕ, ФК ФЭФ и дополнительным признакам кортежей. Приведем три примера, иллюстрирующих отдельно лексические и грамматические двуязычные поисковые возможности, а также лексико-грамматический двуязычный запрос на поиск. Для первого примера в шаблоне на поиск был задан тип кортежа ⟨*притом; au surplus*⟩. В результате лексического поиска были найдены<sup>1</sup> три двуязычных кортежа, соответствующих заданному типу, с разными наборами рубрик ФК ЯЕ и ФК ФЭФ, так как контексты коннектора *притом* и его ФЭФ *au surplus* в этих трех случаях различаются (см. табл. 8 с контекстами и наборами рубрик).

В табл. 8 рубрики ⟨TBD⟩, ⟨р CNT q⟩ и ⟨CNT⟩ были описаны выше (с небольшой разницей: в отличие от ⟨CNT р CNT q⟩ рубрика ⟨р CNT q⟩ означает, что коннектор находится между соединяемыми

<sup>1</sup> В этом и следующих двух примерах результаты поиска даны по состоянию НБД коннекторов на 23.01.2016.

**Таблица 8** Результат двуязычного лексического поиска — три двуязычных кортежа, соответствующих заданному типу (*притом; au surplus*), с контекстами и наборами рубрик ФК ЯЕ и ФК ФЭФ

Оригинальный текст		Перевод	
Притом этот человек не любил неизвестности,	<b>притом</b> < TBD > < повествовательное > < начальная > < p CNT q > < CNT >	<b>Au surplus</b> , cet homme n'aimait pas l'incertitude	<b>au surplus</b> < TBD > < повествовательное > < начальная > < p CNT q > < CNT >
Притом она ничего бы и не поняла.	<b>притом</b> < TBD > < повествовательное > < начальная > < p CNT q > < CNT >	<b>Au surplus</b> , elle n'aurait rien compris!	<b>au surplus</b> < TBD > < восклицательное > < начальная > < p CNT q > < CNT >
Не противоречу вам и <b>притом</b> не мастер я философствовать.	<b>притом</b> < Part >	Je ne vous contredis pas et <b>au surplus</b> je ne suis pas fort pour faire de la philosophie.	<b>au surplus</b> < Part >

фрагментами текста или во втором из них). Рубрика <начальная> относится к рубрике «Позиция» (см. строку 3 в табл. 3) и говорит о том, что коннектор находится в начале маркируемого им фрагмента текста. Рубрика <Part> относится к рубрике первого уровня «Статус» и означает, что в данном случае *притом* входит в состав многокомпонентного коннектора (*и притом*), являясь его частью. Здесь для него построен кортеж третьей категории (см. разд. 3 с описанием категорий). Наконец, рубрики <повествовательное> и <восклицательное> по состоянию на 1 января 2016 г. относятся к рубрике первого уровня «Структура» (см. строку 2 в табл. 3), но в будущем они, возможно, будут выделены в отдельную рубрику первого уровня в процессе развития ФК ЯЕ и ФК ФЭФ. Эти две рубрики используются ниже для иллюстрации двуязычных грамматических поисковых возможностей.

Во втором примере надо было найти любые двуязычные кортежи, коннекторы которых маркируют повествовательные предложения (в запросе была задана рубрика <повествовательное> для контекста коннектора в русском языке), а их ФЭФ — восклицательные (т. е. одновременно задан поиск и с использованием рубрики <восклицательное> для контекста ФЭФ). В этом запросе на поиск присутствуют только эти две рубрики и отсутствуют лексемы. В результате поиска были найдены два двуязычных кортежа: один типа <*притом; au surplus*> и второй типа <*притом же; et enfin*> (см. табл. 9 с двумя кортежами, контекстами и наборами рубрик).

В третьем примере был задан одновременно поиск кортежа типа <*притом; zero*>, чтобы найти случаи, когда коннектор *притом* не переведен на французский язык, и задана рубрика второго уров-

**Таблица 9** Результат двуязычного грамматического поиска — два двуязычных кортежа, контексты и рубрики

Оригинальный текст		Перевод	
Притом она ничего бы и не поняла.	<b>притом</b> < TBD > < повествовательное > < начальная > < p CNT q > < CNT >	<b>Au surplus</b> , elle n'aurait rien compris!	<b>au surplus</b> < TBD > < восклицательное > < начальная > < p CNT q > < CNT >
Притом же в деревне одному очень скучно.	<b>притом же</b> < TBD > < повествовательное > < начальная > < p CNT q > < CNT > < Контакт >	<b>Et enfin</b> , vivre à la campagne tout seul, c'est tellement ennuyeux!	<b>et enfin</b> < TBD > < восклицательное > < начальная > < p CNT q > < CNT > < Контакт >

**Таблица 10** Результат двуязычного лексико-грамматического поиска — три кортежа, соответствующих типу *⟨притом; zero⟩* в комбинации с рубрикой *⟨начальная⟩*

Оригинальный текст		Перевод	
<b>Притом</b> их связывало детство и школа —	<b>притом</b> ⟨ TBD ⟩ ⟨ повествовательное ⟩ ⟨ начальная ⟩ ⟨ p CNT q ⟩ ⟨ CNT ⟩	Les deux hommes étaient unis par les souvenirs de l'école et de l'enfance,	<b>zero</b>
<b>Притом</b> карет несло такое множество взад и вперед и с такою быстротою, что трудно было даже приметить;	<b>притом</b> ⟨ TBD ⟩ ⟨ повествовательное ⟩ ⟨ начальная ⟩ ⟨ p CNT q ⟩ ⟨ CNT ⟩	Les équipages se croisaient si nombreux et roulaient à si belle allure qu'il était difficile d'en distinguer un parmi les autres;	<b>zero</b>
<b>Притом</b> история о танцующих стульях в Колюшенной улице была еще свежа,	<b>притом</b> ⟨ TBD ⟩ ⟨ повествовательное ⟩ ⟨ начальная ⟩ ⟨ p CNT q ⟩ ⟨ CNT ⟩	L'histoire des chaises tournantes de la rue des Grandes-Écuries était encore présente à toutes les mémoires.	<b>zero</b>

ня *⟨начальная⟩*, т. е. этот коннектор должен быть в начале предложения на русском языке. В результате лексико-грамматического поиска были найдены три кортежа, соответствующих заданному типу, в комбинации с рубрикой *⟨начальная⟩*.

Таблицы 8–10 содержат примеры реализации в НБД тех функций двуязычного поиска, которые отсутствуют в корпусах параллельных текстов, но необходимы для сопоставительных исследований и получения новых кроссязыковых знаний.

## 6 Заключение

Необходимость разработки и применения НБД для описания разных видов ЯЕ проявляется наиболее наглядно в случае машинного перевода, качество которого на сегодняшний день остается мало удовлетворительным. Для повышения качества машинного перевода требуется существенное развитие двуязычных грамматических описаний, благодаря полученным на основе сопоставления параллельных текстов новым кроссязыковым знаниям. Сопоставление выровненных предложений двух (а в перспективе и большего числа) языков с помощью НБД позволяет выявлять и заполнять лакуны в системе лингвистических знаний о разных видах языковых единиц и явлений [5, 16]. В статье были рассмотрены примеры из НБД коннекторов, грамматическое описание которых (как в одном языке, так и в сопоставительном аспекте [17]) должно быть приведено в соответствие

с современным уровнем развития лингвистической науки.

Рассмотренные примеры построения двуязычных кортежей позволяют сделать вывод о том, что в общем случае рубрицирование коннекторов и их ФЭФ дает возможность исследователям, применяющим НБД, выполнять следующие задачи:

- (1) разрабатывать методики формирования новых лингвистических знаний, ориентированные на применение современных информационно-компьютерных технологий локального и/или сетевого «мозгового штурма» для сопоставительного анализа параллельных текстов;
- (2) используя новые методики, выявлять лакуны в системах моно- и кроссязыковых знаний;
- (3) специфицировать лакуны, маркируя их особыми рубриками фасетных классификаций НБД (см., например, рубрику *⟨отношение подлжит определению⟩* в табл. 4);
- (4) формулировать цели лингвистических исследований в интересах генерации новых лингвистических знаний, направленной на заполнение выявленных лакун;
- (5) пополнять моно- и кроссязыковые знания, заполняя лакуны в процессе развития фасетных классификаций НБД, включая создание их новых оснований, уровней, названий и дефиниций рубрик, используемых для описания исследуемых ЯЕ, явлений и их контекстов.

Первые две и четвертая функции являются общими для параллельных корпусов и НБД. Однако третья и пятая функции реализуются в настоящее время только в НБД для разных видов ЯЕ.

В рамках российско-швейцарского проекта «Контрастивное корпусное исследование коннекторов русского языка» в НБД сейчас проектируются еще две функции, которые позволят:

- (6) выявлять лакуны в системах моно- и кросс-языковых знаний, используя одновременно тексты, созданные профессиональными переводчиками, и машинные переводы;
- (7) специфицировать отдельно лакуны в системе лингвистических знаний и отдельно ошибки систем машинного перевода.

Иначе говоря, эти новые функции НБД ориентированы на прикладное применение и будут непосредственно способствовать повышению качества машинного перевода.

## Литература

1. Johansson S. Seeing through Multilingual Corpora: On the use of corpora in contrastive studies. — Amsterdam: John Benjamins, 2007. 377 p.
2. Loiseau S., Sitchinava D. V., Zalizniak Anna A., Zatsman I. M. Information technologies for creating the database of equivalent verbal forms in the Russian–French multivariant parallel corpus // Информатика и её применения, 2013. Т. 7. Вып. 2. С. 100–109.
3. Kruzchkov M. G., Buntman N. V., Loshchilova E. Ju., Sitchinava D. V., Zalizniak Anna A., Zatsman I. M. A database of Russian verbal forms and their French translation equivalents // Компьютерная лингвистика и интеллектуальные технологии: По мат-лам ежегодной Междунар. конф. «Диалог». — М.: РГГУ, 2014. Вып. 13(20). С. 284–296.
4. Бунтман Н. В., Зализняк Анна А., Зацман И. М., Кружков М. Г., Лоцилова Е. Ю., Сичинава Д. В. Информационные технологии корпусных исследований: принципы построения кросс-лингвистических баз данных // Информатика и её применения, 2014. Т. 8. Вып. 2. С. 98–110.
5. Zatsman I., Buntman N. Outlining goals for discovering new knowledge and computerised tracing of emerging meanings discovery // 16th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2015. P. 851–860.
6. Зализняк Анна А., Зацман И. М., Инькова О. Ю., Кружков М. Г. Надкорпусные базы данных как лингвистический ресурс // Корпусная лингвистика-2015: Тр. 7-й Междунар. конф. — СПб.: СПбГУ, 2015. С. 211–218.
7. Кружков М. Г. Информационные ресурсы контрастивных лингвистических исследований: типологические базы данных // Системы и средства информатики, 2015. Т. 25. № 1. С. 198–212.
8. Кружков М. Г. Информационные ресурсы контрастивных лингвистических исследований: электронные корпуса текстов // Системы и средства информатики, 2015. Т. 25. № 2. С. 140–159.
9. Добровольский Д. О., Кретов А. А., Шаров С. А. Корпус параллельных текстов: архитектура и возможности использования // Национальный корпус русского языка: 2003–2005. — М.: Индрик, 2005. С. 263–296.
10. Добровольский Д. О., Кретов А. А., Шаров С. А. Корпус параллельных текстов // Научная и техническая информатика. Сер. 2: Информационные процессы и системы, 2005. № 6. С. 16–27.
11. Zatsman I. Computer and information science: Background of formation // Scientific Technical Information Processing, 2013. Vol. 40. No. 3. P. 119–130.
12. Zatsman I. Table of interfaces of informatics as computer and information science // Scientific Technical Information Processing, 2014. Vol. 41. No. 4. P. 233–246.
13. Зацман И. М., Косарик В. В., Курчавова О. А. Задачи представления личностных и коллективных концептов в цифровой среде // Информатика и её применения, 2008. Т. 2. Вып. 3. С. 54–69.
14. Зацман И. М. Семиотическая модель взаимосвязей концептов, информационных объектов и компьютерных кодов // Информатика и её применения, 2009. Т. 3. Вып. 2. С. 65–81.
15. Зацман И. М. Нестационарная семиотическая модель компьютерного кодирования концептов, информационных объектов и денотатов // Информатика и её применения, 2009. Т. 3. Вып. 4. С. 87–101.
16. Zatsman I., Buntman N., Kruzchkov M., Nuriev V., Zalizniak Anna A. Conceptual framework for development of computer technology supporting cross-linguistic knowledge discovery // 15th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2014. Vol. 3. P. 1063–1071.
17. Инькова-Манзотти О. Ю. Коннекторы противопоставления во французском и русском языках: сопоставительное исследование. — М.: Информэлектро, 2001. 434 с.
18. Инькова О. Ю. К проблеме описания многокомпонентных коннекторов русского языка: не только... но и // Вопросы языкознания, 2016. № 2. С. 37–60.

Поступила в редакцию 28.01.16

# REPRESENTATION OF CROSS-LINGUAL KNOWLEDGE ABOUT CONNECTORS IN SUPRACORPORA DATABASES

I. M. Zatsman<sup>1</sup>, O. Yu. Inkova<sup>1,2</sup>, M. G. KruzHKov<sup>1</sup>, and N. A. Popkova<sup>1</sup>

<sup>1</sup>Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

<sup>2</sup>University of Geneva, 22 Bd des Philosophes, CH-1205 Geneva 4, Switzerland

**Аннотация:** The article considers “supracorpora databases,” which are used in contrastive linguistic studies. Such databases result from processing of parallel texts from bilingual parallel subcorpora within the Russian National Corpus. Each of these parallel texts contains either one original Russian text with one or more translations into a foreign language, or one original text in a foreign language with one translation into Russian. Every source text is aligned with its translation(s) at the level of sentences. Supracorpora databases are a new type of linguistic resources designed for goal-oriented discovery of new knowledge about various linguistic units. This knowledge is needed to improve the quality of machine translation, to update monolingual and bilingual grammars, and to modernize a wide range of academic courses in such fields as linguistics and translation studies. The article describes the underlying conceptual foundations of the database and gives an example of how it can be implemented to represent knowledge about Russian connectors and their French translation correspondences.

**Keywords:** cross-lingual studies; Russian connectors; representation of knowledge about connectors; supracorpora databases

**DOI:** 10.14357/19922264160110

## Acknowledgments

This research was performed at the Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, and supported by the Russian Foundation for Basic Research (grant No. 14-07-00785), by the Russian Foundation for Humanities (grant No. 16-24-41002), by Swiss National Science Foundation (grant No. IZLRZ1\_164059).

## References

- Johansson, S. 2007. *Seeing through Multilingual Corpora: On the use of corpora in contrastive studies*. Amsterdam: John Benjamins. 377 p.
- Loiseau, S., D. V. Sitchinava, Anna A. Zalizniak, and I. M. Zatsman. 2013. Information technologies for creating the database of equivalent verbal forms in the Russian–French multivariant parallel corpus. *Informatika i ee Primeneniya — Inform. Appl.* 7(2):100–109.
- KruzHKov, M. G., N. V. Buntman, E. Ju. Loshchilova, D. V. Sitchinava, Anna A. Zalizniak, and I. M. Zatsman. 2014. A database of Russian verbal forms and their French translation equivalents. *Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii. Po mat-lam Ezhegodnoy Mezhdunar. Konf. “Dialog-2014”* [Computational Linguistics and Intellectual Technologies: Conference (International) “Dialog-2014” Proceedings]. Moscow. 13(20):284–297.
- Buntman, N. V., Anna A. Zaliznyak, I. M. Zatsman, M. G. KruzHKov, E. Yu. Loshchilova, and D. V. Sitchinava. 2014. Informatsionnye tekhnologii korpusnykh issledovaniy: printsipy postroeniya kross-lingvisticheskikh baz dannykh [Information technologies for corpus studies: Underpinnings for cross-linguistic database creation]. *Informatika i ee Primeneniya — Inform. Appl.* 8(2):98–110.
- Zatsman, I., and N. Buntman. 2015. Outlining goals for discovering new knowledge and computerised tracing of emerging meanings discovery. *16th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 851–860.
- Zaliznyak, Anna A., I. M. Zatsman, O. Yu. Inkova, and M. G. KruzHKov. 2015. Nadkorpusnye bazy dannykh kak lingvisticheskiy resurs [Subcorpora databases as linguistic resource]. *Korpusnaya Lingvistika: Tr. 7-y Mezhdunar. Konf. [7th Conference (International) on Corpus Linguistics Proceedings]*. St. Petersburg: St. Petersburg State University. 211–218.
- KruzHKov, M. G. 2015. Informatsionnye resursy kontrastivnykh lingvisticheskikh issledovaniy: Tipologicheskie bazy dannykh [Information resources for contrastive studies: Typological databases]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(1):198–212.
- KruzHKov, M. G. 2015. Informatsionnye resursy kontrastivnykh lingvisticheskikh issledovaniy: Elektronnye korpusa tekstov [Information resources for contrastive studies: Electronic text corpora]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(2):140–159.
- Dobrovol'skiy, D. O., A. A. Kretov, and S. A. Sharov. 2005. Korpus parallel'nykh tekstov: Arkhitektura i vozmozhno-

- sti ispol'zovaniya [Corpus of parallel texts: Architecture and applications]. *Natsional'nyy korpus russkogo yazyka: 2003–2005* [Russian National Corpus: 2003–2005]. Moscow: Indrik. 263–296.
10. Dobrovol'skiy, D. O., A. A. Kretov, and S. A. Sharov. 2005. Korpus parallel'nykh tekstov [Corpus of parallel texts]. *Nauchnaya i Tekhnicheskaya Informatsiya* [Scientific and Technical Information]. Ser. 2: Informatsionnye protsessy i sistemy [Informational processes and systems]. 6:16–27.
  11. Zatsman, I. 2013. Computer and information science: Background of formation. *Scientific Technical Information Processing* 40(3):119–130.
  12. Zatsman, I. 2014. Table of interfaces of informatics as computer and information science. *Scientific Technical Information Processing* 41(4):233–246.
  13. Zatsman, I. M., V. V. Kosarik, and O. A. Kurchavova. 2008. Zadachi predstavleniya lichnostnykh i kollektivnykh kontseptov v tsifrovoy srede [Representation of individual and collective concepts in digital medium]. *Informatika i ee Primeneniya — Inform. Appl.* 2(3):54–69.
  14. Zatsman, I. 2009. Semioticheskaya model' vzaimosvyazey kontseptov, informatsionnykh ob"ektov i komp'yuternykh kodov [Semiotic model of relationships of concepts, information objects, and computer codes]. *Informatika i ee Primeneniya — Inform. Appl.* 3(2):65–81.
  15. Zatsman, I. 2009. Nestatsionarnaya semioticheskaya model' komp'yuternogo kodirovaniya kontseptov, informatsionnykh ob"ektov i denotatov [Nonstationary semiotic model of computer coding of concepts, information objects, and denotata]. *Informatika i ee Primeneniya — Inform. Appl.* 3(4):87–101.
  16. Zatsman, I., N. Buntman, M. Kruzhkov, V. Nuriev, and Anna A. Zalizniak. 2014. Conceptual framework for development of computer technology supporting cross-linguistic knowledge discovery. *15th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 3:1063–1071.
  17. Inkova-Manzotti, O. Yu. 2001. *Konnektory protivopostavleniya vo frantsuzskom i russkom yazykakh: Sopostavitel'noe issledovanie* [Connectors of opposition in French and Russian: A comparative study]. Moscow: Informelektro. 434 p.
  18. Inkova, O. Yu. 2016. K probleme opisaniya mnogokomponentnykh konnektorov russkogo yazyka: Ne tol'ko... no i [Towards the description of multiword connectives in Russian: Ne tol'ko... no i (non only... but also)]. *Voprosy Jazykoznanija* [Topics in the Study of Language] 2:37–60.

Received January 28, 2016

## Contributors

**Zatsman Igor M.** (b. 1952) — Doctor of Science in technology, Head of Department, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; izatsman@yandex.ru

**Inkova Olga Yu.** (b. 1965) — Doctor of Science in philology, Faculty member, University of Geneva, 22 Bd des Philosophes, CH-1205 Geneva 4, Switzerland; senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; Olga.Inkova@unige.ch

**Kruzhkov Mikhail G.** (b. 1975) — leading programmer, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; magnit75@yandex.ru

**Popkova Natalia A.** (b. 1992) — junior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; natasha\_\_popkova@mail.ru

# ПРИМЕНЕНИЕ КОНТЕКСТНО-СВОБОДНЫХ ГРАММАТИК ДЛЯ ИЗВЛЕЧЕНИЯ ОНТОЛОГИИ ИЗ ТЕКСТОВ КОРОТКИХ ОПИСАНИЙ СТАТЕЙ БИОЛОГИЧЕСКОЙ ТЕМАТИКИ\*

Д. А. Алексеевский<sup>1</sup>

**Аннотация:** Обработка текстов биологической и медицинской тематики представляет интерес как с точки зрения биологии, для которой она предоставляет ценные результаты, так и в качестве источника более сложных задач для обработки текстов. Одной из важных задач автоматической обработки текстов является построение онтологий. Предложен метод построения онтологий промежуточного уровня по корпусу текстов на ограниченном подмножестве английского языка. Онтологии промежуточного уровня служат одним из инструментов решения задачи установления соответствия между фактами в априорных онтологиях и фрагментами текста. Предложен новый подход, основанный на расширенном определении контекстно-свободных (КС) грамматик, позволяющий порождать онтологии, обладающие указанным свойством. Показаны преимущества использования корпусов на ограниченном подмножестве естественного языка для построения таких онтологий.

**Ключевые слова:** КС-грамматики; построение онтологий; биомедицинские тексты

**DOI:** 10.14357/19922264160111

## 1 Введение

За последние десятилетия биология, а следом за ней и медицина претерпели несколько научных переворотов, каждый из которых приводил к бурному росту числа публикаций, а также и прочих текстов этих тематик. Многие полученные данные были собраны в базы данных, которые играют большую роль в этих науках. В то же время с ростом объема опубликованных текстов обнаруживаются новые виды данных, доступные в текстовом виде, требующие структурирования и верификации. Этим объясняется растущая актуальность темы извлечения фактов из текстов биологической и медицинской тематики. Следует заметить, что эта тема имеет существенные отличия от автоматической обработки текстов в целом, что обуславливает выделение ее в отдельную область.

Задачам автоматической обработки текстов медицинской и биологической тематики посвящено много работ. Среди современных направлений исследований: извлечение и нормализация именованных сущностей [1], извлечение событий и составных отношений [2], анализ дискурса и ко-референции [3], построение и пополнение онтологий и баз данных [4]. Среди наиболее широко используемых биологических баз данных встречаются ресурсы, совмещающие структурированные данные (ссылки на другие базы данных, чис-

ловые характеристики объектов, номенклатурные названия объектов и т. п.), неструктурированные текстовые данные (текстовые описания, цитаты из статей и энциклопедий) и частично формализованные текстовые данные (описания на ограниченном подмножестве языка с использованием контролируемых словарей) [5, 6].

Наряду с задачей извлечения фактов, соответствующих заранее заданной онтологии, для некоторых областей актуальна задача определения онтологической структуры и извлечения самих онтологических элементов. В настоящей статье предложен метод преобразования частично структурированных текстовых описаний в онтологии, основанный на использовании гетерогенных частотных списков и семантически ориентированных КС (СОКС) грамматик.

Для иллюстрации работы метода выбраны краткие аннотации статей, используемые в одной из баз данных (см. подразд. 2.4). Приведена последовательность действий по преобразованию аннотаций в онтологическое представление, дана оценка применимости метода в выбранном примере. В настоящее время указанные краткие аннотации заполняются кураторами вручную, но затем автоматически посредством простых шаблонов по ним определяется уровень доверия к записи в базе данных. Приведение таких аннотаций к онтологическому представлению является необходимым первым ша-

\* Работа выполнена при частичной поддержке РФФИ (проект 15-07-09306).

<sup>1</sup> НИУ Высшая школа экономики, dalexeyevsky@hse.ru

гом для последующего автоматического построения аннотаций по тексту статьи.

## 2 Контекст работы

### 2.1 Специфика обработки биологических текстов

Медицинская и биологическая тематика текстов привносит особенности во многие этапы их обработки. В значительной мере именно это и обуславливает выделение *bionlp* как отдельной предметной области.

Один из часто используемых шагов обработки текстов — идентификация фрагментов текста, соответствующих известным сущностям в базах данных, так называемое извлечение именованных сущностей.

В биологических текстах этот шаг обработки усложнен несколькими обстоятельствами:

- именованная сущность может являться лишь частью слова, например в предложении «The acid-promoted expression of the PmrD protein was *phoPQ*-dependent, which is in agreement with the fact that PhoP is the only known direct transcriptional activator of *pmrD* (Kox *et al.*, 2000)» в слове *phoPQ*-dependent выделяют двухбелковый комплекс «*phoPQ*», состоящий из белка «*phoP*» и белка «*Q*»;
- некоторые сущности, такие как белки или химические соединения, имеют множество синонимичных названий, при этом в текстах могут использоваться не полные названия, а их сокращения, смысл которых возможно восстановить лишь из контекста. Например, белок<sup>1</sup>, имеющий в базе данных названия «Sweet protein *mabinlin-2*», «*Mabinlin II*», «*MAB II*», «Sweet protein *mabinlin-2 chain A*», «Sweet protein *mabinlin-2 chain B*», может встречаться в статьях как «heat-stable sweet protein, *mabinlin-II*», «*mabinlin*» (в пределах текста одной статьи это название может в разных контекстах обозначать как название класса белков, так и конкретный белок), «*Sm-MaIIA*» (обозначение одной цепочки модифицированного белка, введенное в статье) [7];
- для некоторых сущностей после определения их названия необходимо точнее идентифицировать сущность, о которой идет речь: например, одно и то же название белка может иметь несколько аллелей в одном организме, белок может различаться или не различаться в зависимости от ткани, для которой проводился

эксперимент, одно имя могут иметь схожие, но различные белки из разных организмов. Для каждого белка, имеющего то же имя, имеется отдельная запись в базе данных, и необходимо определить, о какой именно записи идет речь.

### 2.2 Задача построения онтологий

В литературе встречаются разнообразные определения понятия онтологии в зависимости от темы и специфики выбранной задачи. Встречающиеся определения этого понятия в контексте извлечения фактов описывают способы представления знаний, как правило, состоящие из описаний сущностей, их свойств, классификации, связей между ними и логических правил пополнения их свойств и связей [8].

Выделяют онтологии, построенные априори путем логической классификации и лексические, в которых отражаются семантические связи между языковыми единицами [9]. Они обладают разными свойствами: априорная точнее отражает предметную область и позволяет применять богатые механизмы логического вывода, в то время как сущности лексических онтологий, как правило, проще выделять в тексте. В связи с этим одна из часто возникающих задач состоит в установлении соответствия между сущностями лексической и априорной онтологии [10].

Подход, предлагаемый в настоящей статье, позволяет построить онтологию, занимающую промежуточное положение. Такая онтология строится частично по базам данных как априорная, частично по корпусу текстов как лексическая. Это определяет ее главное достоинство: она содержит как ссылки на конкретные сущности из базы данных, так и их текстовое представление.

### 2.3 Семантически-ориентированные контекстно-свободные грамматики

Для настоящей работы в качестве формализма для описания синтаксической структуры предложения были выбраны КС-грамматики. Контекстно-свободная грамматика — это способ описания структуры предложения в виде иерархии составляющих частей [11]. Дадим ей формальное определение.

**Определение 1.** Контекстно-свободной грамматикой называется четверка  $G = (V, \Sigma, R, S)$ , где  $V$  — конечное множество нетерминальных символов;  $\Sigma$  — конечное множество терминальных символов;  $R \subset \{V \times (V \cup \Sigma)^*\}$  — множество правил вывода вида  $v \rightarrow a_1 a_2 \dots$ , где  $v \in V$ ,  $a_i \in V \cup \Sigma$ ;  $S \in V$  — начальный символ.

<sup>1</sup>UniProt AC P30233, <http://www.uniprot.org/uniprot/P30233>.

Формальным определением для описания онтологии в настоящей работе было выбрано следующее: онтология — это ориентированный граф, в котором каждая вершина и каждое ребро сопровождаются пометой. С помощью пометы множество вершин делится на вершины-классы и вершины-экземпляры. Пометы на ребрах устанавливают тип отношений, в которых находятся две выбранные вершины. Приведем более формализованное определение.

**Определение 2.** Онтологией называется пара  $O = (G_O, L_O)$  из ориентированного графа и меток к нему. В свою очередь, граф  $G_O = (E_O, R_O)$  состоит из множества вершин  $E_O$ , называемого множеством сущностей, и множества ребер  $R_O \subset E_O \times E_O$ , называемого множеством отношений; метки  $L_O = (T_E, T_R, L_E, L_R)$  задаются алфавитом возможных меток для вершин  $T_E$ , алфавитом возможных меток для ребер  $T_R$ , отображением  $L_E : E_O \rightarrow T_E$  вершины на ее метку и отображением  $L_R : R_O \rightarrow T_R$  ребра на его метку.

Одно из свойств предлагаемого в настоящей работе алгоритма состоит в простоте выделения онтологических фактов из деревьев синтаксического разбора. Такой алгоритм требует введения нового понятия: семантически ориентированной КС-грамматики. Контекстно-свободная грамматика является семантически ориентированной для данной онтологии, если часть ее правил описывает сущности и отношения в онтологии. Предлагается следующее определение.

**Определение 3.** Семантически ориентированной КС-грамматикой называется тройка  $S = (G, O, M)$  из КС-грамматики  $G = (V, \Sigma, \dots)$ , онтологии  $O = ((E - O, R_O), (T_E, T_R, L_E, L_R))$  и отображения  $M$  между ними. Отображение  $M = (M_E, M_R)$  состоит из отображения  $M_E \subset (\Sigma \cup V, E_O)$ , где  $\forall (v, e), (v', e') \in M_E : v = v' \Leftrightarrow e = e'$ ; символов грамматики на вершины онтологии и отображения  $M_R \subset (V, L_R)$ , где  $\forall (v, r), (v', r') \in M_R : v = v' \Leftrightarrow r = r'$ .

Терминальный символ грамматики может быть отображен на вершину-класс или вершину-экземпляр либо не использоваться в онтологии. В последнем случае терминальный символ будем называть синтаксическим по последнему этапу обработки текста, в котором он используется. Нетерминальный символ может быть отображен на вершину-класс, метку ребра (тип отношения), в том числе одновременно, либо не использоваться в онтологии. Как и в случае с терминальными вершинами, в последнем случае такой нетерминал будет называться синтаксическим.

## 2.4 База данных UniProt

Материалом для разработки и тестирования предлагаемой процедуры построения онтологий послужила свободно распространяемая база UniProt [6].

UniProt является хранилищем аминокислотных последовательностей белков наряду с их краткими описаниями. База содержит ссылки на другие базы данных, посвященные исследованиям белков специфическими методами. Кроме того, частью описания белка в базе является список литературы, описывающей белок.

Для каждого белка база содержит:

- описание его аминокислотной последовательности (поле « » — два пробела);
- обозначения белка согласно различным номенклатурам (поля «DE» и «GN»);
- идентификаторы в различных биологических базах данных самого белка (поля «ID», «AC» и «DR») и его носителя («OC» и «OX»);
- биологический контекст белка (поля «OS», «OG» и «OH»);
- библиографическую информацию (поля «RN», «RP», «RC», «RX», «RG», «RA», «RT» и «RL»);
- описания известных свойств белка: текстовые (поле «CC»), на ограниченном подмножестве английского языка (поля «RP» и «KW»), формализованные (поле «FT»);
- уровень доверия данной записи (поле «PE»);
- прочую служебную информацию (поля «DT» и «SQ»).

Значение «PE» уровня достоверности записи базы данных определяется тем, какими экспериментальными средствами установлен факт существования белка и его соответствия представленным данным. Описания того, какие экспериментальные средства применялись к белку, хранятся в базе в полях «CC», «RP» и «KW», для некоторых методов факт их применения можно опознать по свойствам в поле «FT». В базе заданы формальные правила выставления значения уровня доверия («PE») в зависимости от наличия некоторых шаблонных выражений в этих полях [12].

База данных UniProt состоит из двух частей: UniProt/TrEMBL, пополняемой полностью автоматически, и UniProt/Swiss-Prot, пополняемой кураторами вручную на основе материалов UniProt/TrEMBL, существующих публикаций и материалов других баз данных. Поля «KW» и «FT» получают начальные значения автоматически в базе

данных UniProt/TrEMBL, хотя затем могут быть изменены в процессе курирования. Поля «CC» и «RP» заполняются только кураторами вручную.

Этими обстоятельствами обусловлено то, что в качестве материала для настоящей работы был собран корпус предложений в поле «RP» из базы данных UniProt/Swiss-Prot.

### 3 Материалы и методы

Материалом исследований послужили данные из базы UniProt/Swiss-Prot версии 2015\_01. Собранный корпус уникальных атомарных причин цитирования в поле «RP» имеет размер 173 212 предложений.

Всего база UniProt/Swiss-Prot 2015\_1 содержит:

- 547 357 записей (одна запись описывает один белок);
- 1 092 817 ссылок на литературу и, соответственно, всего предложений в поле «RP», включая повторяющиеся, среди них;
- 179 616 уникальных предложений в поле «RP», состоящих из одного или нескольких атомарных описаний (в свою очередь также включающих повторения);
- 173 212 уникальных атомарных описаний.

Для дальнейшей работы использовался описанный корпус уникальных атомарных описаний, с тем чтобы наиболее полно покрыть максимально возможное количество особых случаев в языке.

#### 3.1 Особенности предложений в поле «RP»

Предложения в поле «RP» являются полуструктурированными, так как несут признаки как структурированных, так и естественных языковых данных. Предложения порождаются кураторами. Для них не существует формализованного описания структуры или инструмента для валидации. Существует находящаяся на данный момент в стадии разработки инициатива по унификации представления названий различных классов сущностей в таких предложениях с помощью внедрения контролируемых словарей [13]. Наряду с этим для кураторов существует инструкция по заполнению, включающая в себя примеры представления большого числа типов фактов [14].

Каждое атомарное описание является именной группой. Важно заметить, что для краткости описания не содержат упоминаний описываемого объекта. Объект описания устанавливается из факта принадлежности описания записи в базе данных (рис. 1).

#### 3.2 Словники

Для извлечения именованных сущностей и насыщения списка примитивных фактов были использованы словники.

Словник имен белков был построен по значениям в поле «DE», подразделам RecName и AltName

```
X-RAY CRYSTALLOGRAPHY (1.80 ANGSTROMS) OF 44-480 OF WILD-TYPE AND MUTANTS TYR-118; ARG-168 AND ALA-309 IN ACTIVE AND RESTING STATES AND IN COMPLEX WITH PEPTIDE SUBSTRATE, FUNCTION, CATALYTIC ACTIVITY, ENZYME REGULATION, SUBSTRATE SPECIFICITY, SUBUNIT, DOMAIN, PROTEOLYTIC AUTO-CLEAVAGE, ACTIVE SITES, SITES, DISRUPTION PHENOTYPE, MUTAGENESIS OF VAL-118; ARG-168; SER-309 AND GLN-338, AND PDZ DOMAIN DELETION MUTANT.
```

(а)

```
> X-RAY CRYSTALLOGRAPHY (1.80 ANGSTROMS) OF 44-480 OF WILD-TYPE AND MUTANTS TYR-118; ARG-168 AND ALA-309 IN ACTIVE AND RESTING STATES AND IN COMPLEX WITH PEPTIDE SUBSTRATE
> FUNCTION
> CATALYTIC ACTIVITY
> ENZYME REGULATION
> SUBSTRATE SPECIFICITY
> SUBUNIT
> DOMAIN
> PROTEOLYTIC AUTO-CLEAVAGE
> ACTIVE SITES
> SITES
> DISRUPTION PHENOTYPE
> MUTAGENESIS OF VAL-118; ARG-168; SER-309
```

(б)

Рис. 1 Примеры описаний в поле «RP»: (а) полное описание, (б) атомарные факты

базы данных UniProt и полям Full, Short, Name, Synonyms в них. Суммарный объем словаря составил 308 370 словосочетаний.

Некоторые названия белков совпадают с общезначимыми словами английского языка. Для того чтобы исключить ошибки второго рода в таких случаях, из словаря имен белков были удалены все слова, являющиеся словами английского языка. Для этой фильтрации был использован словарь общеупотребительной лексики американского английского языка [15] объемом 99 171 словоформа, содержащий все падежные формы слов.

### 3.3 Методы

Для сегментации текста на слова был использован токенизатор, сохраняющий все знаки пунктуации, включая дефисы, как отдельные токены. Токенизатор был разработан на основе пакета языка Python [16].

Для построения СОКС-грамматик был использован парсер Эрли с проходом снизу вверх из пакета nltk [17] для языка Python.

Для построения частотных списков использовались средства shell script и сопутствующие программы текстовой обработки из базового комплекта операционной системы GNU: cat, sort, uniq, grep, sed, head, tail, less.

## 4 Алгоритм разработки онтологии с помощью контекстно-свободных грамматик

Задача алгоритма состоит в том, чтобы за наименьшее время преобразовать наибольшую часть заранее заданного корпуса фактов, представленного в виде полуструктурированных текстовых данных, в онтологическое представление.

Основная идея алгоритма состоит в итерационном применении и пополнении КС-грамматики. После каждого применения грамматики предложения корпуса преобразуются в гетерогенную последовательность из токенов и нетерминальных символов грамматики. Полученный корпус гетерогенных последовательностей используется для того, чтобы определить, какое правило нужно добавить в корпус для получения наибольшего прироста количества предложений, разбор которых доведен до нетерминала-вершины.

При построении КС-грамматики терминальными символами грамматики являются токены из корпуса, множество нетерминальных символов является объединением из множества типов сущностей

в онтологии и множества вспомогательных нетерминальных символов.

Входными данными для построения онтологии являются:

- корпус разбираемых текстов;
- базы данных и словники, позволяющие выделять в тексте релевантные именованные сущности.

Алгоритм состоит из пяти шагов:

1. Подготовить начальную грамматику.
2. Применить к корпусу текстов правила грамматики, заменив покрытые правилами фрагменты текста соответствующими нетерминалами.
3. Оценить покрытие корпуса текстов нетерминалами и выбрать метод пополнения грамматики (см. ниже).
4. Пополнить грамматику новым правилом (см. ниже).
5. Перейти на шаг 2.

Начальная грамматика содержит заранее определенный нетерминал-вершину; множество нетерминальных символов, состоящее только из нетерминала-вершины; множество терминальных символов, совпадающее с множеством токенов корпуса; множество правил, являющееся пустым.

Оценка покрытия может производиться одним из двух способов.

1. Выбрать из корпуса случайным образом 100 предложений, среди них найти наиболее частую синтаксическую конструкцию или тип именованной сущности, который еще не покрыт правилами грамматики.
2. Построить частотный список предложений, выбрать из них наиболее частое, для которого может быть написано правило СОКС-грамматики, не имеющее ложных срабатываний.

В результате оценки должно быть порождено правило одного из трех видов:

- (1) синтаксическое упрощение;
- (2) создание или пополнение газетера;
- (3) семантическое правило.

*Синтаксическими упрощениями* называются правила грамматики, которые не отображаются в результирующей онтологии, но обобщают однородные конструкции и упрощают последующее расширение грамматики.

К этому типу правил относятся, например,  
`and -> 'AND' | ',' | ',,' | ',,' 'AND' | ',,' | ',,' 'AND'`  
**и**  
`det -> 'A' | 'AN' | 'THE'`

Необходимость создания или пополнения газетера возникает в тех случаях, когда наиболее частым

не покрытым нетерминалами явлением в корпусе оказываются названия именованных сущностей, принадлежащие к одному классу.

Например, в предложении

PALMITOYLATION AT CYS-11, AND MUTAGENESIS OF SER-2; ARG-6 AND CYS-11.

четыре раза встречаются названия конкретных аминокислотных остатков в белке, представленные как название аминокислоты и номер ее позиции, записанные через дефис. В тот момент, когда в корпусе такие случаи становятся самыми частотными из неразобранных, необходимо пополнить газетир списком названий аминокислот.

Третий вариант действий состоит в том, чтобы пополнить СОКС-грамматику *семантическим правилом*. Для этого необходимо выявить самую частотную конструкцию, такую что в ней нет токенов, которые могли бы войти в именованную сущность; в ней нет лексики, играющей исключительно синтаксическую роль; она не сведена к нетерминалу, являющемуся вершиной онтологии.

Такая конструкция может являться предложением целиком, в этом случае из нее будет образовано новое правило для СОКС-грамматики, в левой части которого будет находиться вершина онтологии:

```
feature -> 'STRUCTURE' 'BY' method
feature -> modification 'AT' range
```

Пример предложений, использующих приведенный фрагмент грамматики:

STRUCTURE BY ELECTRON MICROSCOPY  
(9.4 ANGSTROMS).  
PHOSPHOPANTHETHEINYLATION AT SER-37.

Такая конструкция может одновременно быть предложением и сводиться к нетерминалу, который при этом не является вершиной онтологии, например:

```
feature -> method
feature -> interaction
```

Пример предложений, использующих приведенный фрагмент грамматики:

IDENTIFICATION BY MASS SPECTROMETRY.  
CALMODULIN-BINDING.

Такая конструкция может являться частью предложения, в этом случае нетерминал в левой части правила не будет являться вершиной онтологии, например:

```
interaction -> interaction 'WITH' protein
```

Пример предложений, использующих приведенный фрагмент грамматики:

INTERACTION WITH MPK6

#### 4.1 Преобразование деревьев синтаксического разбора в онтологическое представление данных

В результате работы СОКС-парсера предложения исходного текста преобразуются в деревья синтаксического разбора. Например, предложение

FUNCTION, AND INTERACTION WITH RBM8A; NXF1 AND THE EXON JUNCTION COMPLEX.

после разбора преобразуется в следующее дерево:

```
(description
(feature
(feature FUNCTION)
(and , AND)
(feature
(interaction
(interaction INTERACTION)
WITH
(protein
(protein (protein RBM8A) (and ;)
(protein NXF1))
(and AND)
(protein (det THE)
(protein (words EXON JUNCTION
COMPLEX))))))
.)
```

Такое дерево содержит набор связей, которые в точности соответствуют онтологическим. Помимо таких связей в дереве имеются связи и узлы, имеющие синтаксическую роль (сочетание и детерминанты). Кроме того, связи, отвечающие за сочетание, представлены здесь не как однородные связи внутри одного объекта, а как вложенная рекурсивная цепочка связей.

Для преобразования деревьев такого вида в онтологические факты необходимо:

- заменить текстовое описание именованных сущностей на идентификатор базы данных (например, заменить RBM8A на Q9Y5S9; RBM8A является названием белка, общего для многих видов, база данных UniProt содержит 64 белка с идентичным названием, текст данного предложения получен из описания белка, извлеченного из h.sapiens; следовательно, нас интересуют и белки RBM8A только из h.sapiens, такой белок только один);

- нормализовать числовые значения (например, заменить на 4.2 поддерево

```
(float (digits 4) . (digits 2));
```

- раскрыть случаи сочетания необходимым для данного онтологического класса способом;
- удалить нетерминалы, играющие синтаксическую роль (например, поддерево: (det THE));
- в случаях, когда несколько аргументов обозначаются одним и тем же нетерминалом, дать аргументам различные имена;
- преобразовать правила грамматики в объявление онтологических классов, отношений класс–подкласс и объявлений свойств;
- преобразовать газетеры в объявление онтологических индивидов и отношений класс–индивид;
- преобразовать дерево разбора в объявление набора онтологических индивидов, объявление их отношения к соответствующим онтологическим классам и отношений часть–целое и атрибут для этих индивидов.

Для приведенного примера фрагмент грамматики (вместе с вставленными в него для наглядности фрагментами необходимых газетеров) выглядит следующим образом:

```
description -> feature '.'
feature -> feature and feature
feature -> interaction
feature -> 'FUNCTION'
interaction -> interaction 'WITH' protein
interaction -> 'INTERACTION'
protein -> protein and protein
protein -> words 'COMPLEX'
protein -> det protein
protein -> Q9Y5S9 | Q9UBU9
and -> 'AND' | ',' | ';' | 'AND' | ';' |
';' 'AND'
```

Он однозначным образом преобразуется в набор определений (здесь авторы используют OWL2 functional notation [18]:

```
Declaration(Class(:Description))
Declaration(Class(:Feature))
Declaration(Class(:Function))
Declaration(Class(:Interaction))
Declaration(Class(:Protein))
Declaration(ObjectProperty
(:InteractionWith))
ObjectPropertyDomain(:InteractionWith
:Protein)

SubClassOf(:Feature :Description)
SubClassOf(:Interaction :Feature)
SubClassOf(:Function :Feature)
```

```
Declaration(NamedIndividual(:Q9Y5S9))
ClassAssertion(:Protein :Q9Y5S9)
Declaration(NamedIndividual(:Q9UBU9))
ClassAssertion(:Protein :Q9UBU9)
```

При этом приведенное описание трансформируется в набор онтологических объектов:

```
Declaration(NamedIndividual(:function1))
ClassAssertion(:Function :function1)

Declaration(NamedIndividual(:interaction1))
ClassAssertion(:Interaction :interaction1)

ObjectPropertyAssertion(:InteractionWith
:interaction1 :Q9Y5S9)
ObjectPropertyAssertion(:InteractionWith
:interaction1 :Q9UBU9)
```

```
> [X - RAY CRYSTALLOGRAPHY [1 . 80 ANGSTROMS]resolution OF [44 - 480]range
OF [WILD - TYPE AND MUTANTS [TYR - 118 ; ARG - 168 AND ALA - 309]range]variant
IN [ACTIVE AND RESTING STATES]form AND IN [COMPLEX WITH [PEPTIDE SUBSTRATE]chemical]chemenv]feature
> [FUNCTION]feature
> [CATALYTIC ACTIVITY]feature
> [ENZYME REGULATION]feature
> [SUBSTRATE SPECIFICITY]feature
> [SUBUNIT]feature
> [DOMAIN]feature
> [PROTEOLYTIC AUTO - CLEAVAGE]feature
> [ACTIVE SITES]feature
> [SITES]feature
> [DISRUPTION PHENOTYPE]feature
> [MUTAGENESIS OF [VAL - 118 ; ARG - 168 ; SER - 309 AND GLN - 338]range]feature AND
```

Рис. 2 Разбор описания

```
Declaration(NamedIndividual(:protein1))
AnnotationAssertion( rdfs:comment
:protein1 "EXON JUNCTION COMPLEX" )
ClassAssertion(:Protein :protein1)
```

## 5 Результаты и обсуждение

В ходе работы была построена СОКС-грамматика, содержащая 179 правил (рис. 2).

### 5.1 Оценка покрытия

Для оценки была выбрана случайным образом тестовая выборка из 100 предложений, 96 из них уникальные. Тестовая выборка содержит 205 атомарных причин цитирования, 135 из них уникальные.

Задача построения газетированных предложений находится за пределами настоящей работы, поэтому в тестовой выборке перед тестированием сущности, входящие в газетированные предложения, были вручную заменены на соответствующие им нетерминалы. Дополнительно в грамматику были добавлены правила, позволяющие обрабатывать такие преобразованные входные данные.

В тех случаях, где в тестирующей выборке одна и та же сущность могла быть описана более длинной или более короткой цепочкой, использовалась более короткая цепочка. Таким образом вручную были размечены классы: белок, вещество, болезнь, лекарство, химическая модификация.

Полученные в результате тестирования оценки покрытия представлены в таблице.

Результаты тестирования покрытия

Тестирование	Доля
Все атомарные причины цитирования	73%
Уникальные атомарные причины цитирования	43%
Все предложения	54%
Уникальные предложения	52%

Следует обратить внимание на значительный (в 1,7 раза) прирост покрытия при отключении процедуры удаления дубликатов из корпуса атомарных причин цитирования. Это является косвенным следствием большого числа дубликатов, которые, в свою очередь, являются следствием ограниченности выбранного языка (он использует только именные группы) и его лексической ограниченности (кураторы следуют инструкции, регламентирующей используемую лексику). Такие ограничения приводят к значительному объему дублирования в корпусе. Это дает возможность при меньшем числе правил в грамматике добиваться более высокого покрытия корпуса, что и предложено в настоящей статье.

Очевидно, что более сложные конструкции обладают большим разнообразием и, следовательно, меньшей степенью дублирования, что и продемонстрировано на оценке покрытия полных предложений. Таким образом, для более сложных или менее ограниченных языков кажется осмысленным в качестве предобработки выделять наиболее узко лишь такие конструкции, которые имеют сущности, значимые для составляемой онтологии. Для построения онтологий, описывающих объекты и их свойства, такой предобработкой может служить выделение именных групп.

## 6 Заключение

В работе поставлена актуальная задача разработки новых онтологий на основе корпусных данных и предложен подход к ее решению. Для составления онтологий в работе дано определение и представлен алгоритм составления семантически ориентированных КС-грамматик. Важным аспектом подхода является использование в качестве материала для построения онтологии корпуса предложений на ограниченном подмножестве естественного языка.

Алгоритм опробован для текстов именных групп ограниченного языка, используемого в базе UniProt для описания причин цитирования статьи, в результате чего составлена грамматика и разработан синтаксический анализатор таких причин цитирования.

## Литература

1. Doğan R. I., Leaman R., Lu Zh. Ncbi disease corpus: A resource for disease name recognition and concept normalization // J. Biomed. Inform., 2014. Vol. 47. P. 1–10. doi: 10.1016/j.jbi.2013.12.006.
2. Li Ch., Song R., Liakata M., Vlachos A., Seneff S., Zhang X. Using word embedding for bio-event extraction // 2015 Workshop on Biomedical Natural Language Processing (BioNLP 2015) Proceedings. — Beijing, China: ACL, 2015. P. 121–126.
3. Kim J., Nguyen N., Wang Yu., Tsujii J., Takagi T., Yonezawa A. The genia event and protein coreference tasks of the BioNLP shared task 2011 // BMC Bioinformatics, 2012. Vol. 13. Suppl. 11:S1. doi:10.1186/1471-2105-13-S11-S1.
4. Nédellec C., Bossy R., Kim Ji., Kim Ju., Ohta To., Pyysalo S., Zweigenbaum P. Overview of BioNLP shared task 2013 // BioNLP Shared Task 2013 Workshop (BioNLP-ST 2013) Proceedings. — ACL, 2013. P. 1–7.
5. Tanabe M., Kanehisa M. Unit 1–12 using the KEGG database resource // Current protocols in bioinformatics. — John Wiley & Sons, Inc., 2012. P. 1.12.1–1.12.43. doi: 10.1002/0471250953.bi0112s38.

6. The UniProt Consortium. UniProt: A hub for protein information // *Nucleic Acids Res.*, 2015. Vol. 43. P. D204–D212. doi: 10.1093/nar/gku989.
7. Tonkon M. J., Miller R. R., DeMaria A. N., Vismara L. A., Amsterdam E. A., Mason D. T. Multifactor evaluation of the determinants of ischemic electrocardiographic response to maximal treadmill testing in coronary disease // *Am. J. Med.*, 1977. Vol. 62. Iss. 3. P. 339–346. doi: 10.1016/0002-9343(77)90830-0.
8. Giaretta P., Guarino N. Ontologies and knowledge bases towards a terminological clarification // *Towards very large knowledge bases.* — Amsterdam: IOS Press. P. 25–32.
9. Jones D., Bench-Capon T., Visser P. Methodologies for ontology development // *IT&KNOWS Conference, XV IFIP World Computer Congress Proceedings.* — Budapest, 1998.
10. Reed S. L., Lenat D. B. Mapping ontologies into Cyc // *AAAI 2002 Conference Workshop on Ontologies For The Semantic Web*, 2002. P. 1–6.
11. Хомский Н. Аспекты теории синтаксиса / Пер. В. А. Звегинцева. — М.: Изд-во Моск. ун-та, 1972. 258 с. (*Chomsky N. Aspects of the theory of Syntax.* — MIT Press, 1969. 261 p.)
12. Criteria used to assign the pe level of entries. [http://www.uniprot.org/docs/pe\\_criteria](http://www.uniprot.org/docs/pe_criteria).
13. Controlled vocabulary. [http://www.uniprot.org/help/controlled\\_vocabulary](http://www.uniprot.org/help/controlled_vocabulary).
14. *UniProt Consortium*. UniProt manual curation sop. [http://www.uniprot.org/docs/sop\\_manual\\_curation.pdf](http://www.uniprot.org/docs/sop_manual_curation.pdf).
15. Beale A. Spell checker oriented word lists. 1999–2015. <http://wordlist.aspell.net/12dicts-readme>.
16. Van Rossum G. Python programming language // *USENIX Annual Technical Conference*, 2007.
17. Bird S., Klein E., Loper E. *Natural language processing with Python.* — O'Reilly Media, 2009. 512 p.
18. Horridge M., Patel-Schneider P. F. *OWL 2 Web Ontology Language Manchester Syntax.* — 2nd ed. — W3C Working Group Note, 2009. <http://www.w3.org/TR/owl2-manchester-syntax>.

Поступила в редакцию 23.09.15

## BioNLP ONTOLOGY EXTRACTION FROM A RESTRICTED LANGUAGE CORPUS WITH CONTEXT-FREE GRAMMARS

D. A. Alexeyevsky

National Research University Higher School of Economics; 20 Myasnitskaya Str., Moscow 101000, Russian Federation

**Abstract:** BioNLP is an emerging area of NLP that brings new challenging objects for language processing and new valuable resources for bioinformatics and medicine. One notable task in BioNLP is creating de-novo ontologies. This is generally a tedious process; however, in some cases, it is possible to automate it to some extent. One such case is when a corpus of texts in a restricted subset of natural language is available. This paper presents a simple approach to automate ontology creation in such cases. The approach is aimed to simplify mapping of entities in natural texts to predefined ontologies wherever possible. The paper discusses which properties of the corpus enable the approach presented.

**Keywords:** BioNLP; ontology creation; context-free grammar

**DOI:** 10.14357/19922264160111

## Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (project 15-07-09306).

## References

1. Doğan, R. I., R. Leaman, and Zh. Lu. 2014. Ncbi disease corpus: A resource for disease name and concept normalization. *J. Biomed. Inform.* 47:1–10. doi: 10.1016/j.jbi.2013.12.006.
2. Li, Ch., R. Song, M. Liakata, A. Vlachos, S. Senef, and X. Zhang. 2015. Using word embedding for bio-event extraction. *2015 Workshop on Biomedical Natural Language Processing (BioNLP 2015) Proceedings*. Beijing, China: ACL. 121–126.
3. Kim, J., N. Nguyen, Yu. Wang, J. Tsujii, T. Takagi, and A. Yonezawa. 2012. The genia event and protein coreference tasks of the BioNLP shared task 2011. *BMC Bioinformatics* 13(Suppl. 11:S1). doi: 10.1186/1471-2105-13-S11-S1.
4. Nédellec, C., R. Bossy, Ji. Kim, Ju. Kim, To. Ohta, S. Pyysalo, and P. Zweigenbaum. 2013. Overview of

- BioNLP shared task 2013. *BioNLP Shared Task 2013 Workshop (BioNLP-ST 2013) Proceedings*. ACL. 1–7.
5. Tanabe, M., and M. Kanehisa. 2012. Unit 1–12 using the KEGG database resource. *Current protocols in bioinformatics*. John Wiley & Sons, Inc. 1.12.1–1.12.43. doi: 10.1002/0471250953.bi0112s38.
  6. The UniProt Consortium. 2015. Uniprot: A hub for protein information. *Nucleic Acids Res.* 43:D204–D212. doi:10.1093/nar/gku989.
  7. Tonkon, M. J., R. R. Miller, A. N. DeMaria, L. A. Vis-mara, E. A. Amsterdam, and D. T. Mason. 1977. Multi-factor evaluation of the determinants of ischemic electrocardiographic response to maximal treadmill testing in coronary disease. *Am. J. Med.* 62(3):339–346. doi: 10.1016/0002-9343(77)90830-0.
  8. Giaretta, P., and N. Guarino. 1995. Ontologies and knowledge bases towards a terminological clarification. *Towards very large knowledge bases*. Amsterdam: IOS Press. 25–32.
  9. Jones, D., T. Bench-Capon, and P. Visser. 1998. Methodologies for ontology development. *IT&KNOWS Conference, XV IFIP World Computer Congress Proceedings*. Budapest. 62–75.
  10. Reed, S. L., and D. B. Lenat. 2002. Mapping ontologies into Cyc. *AAAI 2002 Conference Workshop on Ontologies For The Semantic Web*. 1–6.
  11. Chomsky, N. 1969. *Aspects of the theory of Syntax*. MIT Press. 261 p.
  12. Criteria used to assign the pe level of entries. Available at: [http://www.uniprot.org/docs/pe\\_criteria](http://www.uniprot.org/docs/pe_criteria) (accessed January 21, 2016).
  13. Controlled vocabulary. Available at: [http://www.uniprot.org/help/controlled\\_vocabulary](http://www.uniprot.org/help/controlled_vocabulary) (accessed January 21, 2016).
  14. UniProt Consortium. Uniprot manual curation sop. Available at: [http://www.uniprot.org/docs/sop\\_manual\\_curation.pdf](http://www.uniprot.org/docs/sop_manual_curation.pdf) (accessed January 21, 2016).
  15. Beale, A. 1999–2015. *Spell checker oriented word lists*. Available at: <http://wordlist.aspell.net/12dicts-readme> (accessed January 21, 2016).
  16. Van Rossum, G. 2007. Python programming language. *USENIX Annual Technical Conference*.
  17. Bird, S., E. Klein, and E. Loper. 2009. *Natural language processing with Python*. O’Reilly Media. 512 p.
  18. Horridge, M., and P. F. Patel-Schneider. 2009. OWL 2 Web Ontology Language Manchester Syntax. 2nd ed. W3C Working Group Note. Available at: <http://www.w3.org/TR/owl2-manchester-syntax> (accessed January 21, 2016).

Received September 23, 2015

## Contributor

**Alexeyevsky Daniil A.** (b. 1983) — PhD student, Faculty of Humanities, National Research University Higher School of Economics; 20 Myasnitskaya Str., Moscow 101000, Russian Federation; [dalexeyevsky@hse.ru](mailto:dalexeyevsky@hse.ru)

## COMPLEXITY AND ITS INFORMATION CONTENT

N. Callaos<sup>1</sup> and R. Seyful-Mulyukov<sup>2</sup>

**Abstract:** The word ‘information’ has been used in many senses and its related concepts have been defined in different ways. One of the senses in which the word is used relates to a concept which is considered one of the main properties of matter. The definition of this conception of information supports the expression of concepts such as Complexity and Self-Organization. In this paper, Complexity and Self-Organization concepts are applied to systems at the macro- and microlevels. Their similarities and differences are analyzed and information content is considered. The regularities of Complexity and Self-Organization are applied to petroleum as a complex natural thermodynamic system. Petroleum reflects all of the main and widely understood features of Complexity and Self-Organization but demonstrates additional properties which were not considered earlier. Complexity and Self-Organization can help to deepen our understanding of the origin of hydrocarbon molecules, their age, and behavior in the process of petroleum generation in general.

**Keywords:** complexity; complexity properties; complex system; self-organization; artificial complexity; natural complexity; petroleum origin; hydrocarbon molecule complexity; petroleum information content

**DOI:** 10.14357/19922264160112

### Introduction

Among the phenomena most disputed by scientists during the last three decades are Information and Complexity. Various scientists’ understanding and description of the phenomena are rather different. At present, there is no generally accepted definition. This is not unusual and is quite understandable because the significance of one’s conception of Complexity and Information corresponds to the philosophy and world-view in a given discipline. For example, in Cybernetics, Complexity refers to the structure and control of a Complex System. In biology, Complexity is the basis from which the appearance and evolution of living species are studied. In mathematics, Complexity is not a phenomenon but the rate of difficulty, confusion or entanglement of any discipline with problems solved by the abstract method of mathematical calculation. Mathematical theories of Complexity help our understanding of philosophical problems dealing with intelligence and cognition.

The cognition of Information as a phenomenon has about two-thousand-year history. Its initial meaning was simply data, language and knowledge that people transmitted to each other. It became clear only at the beginning of the XX century that one of the several concepts of Information is related to fundamental properties of matter<sup>3</sup>.

The present authors are not attempting to create a new definition of Information or Complexity. The aim of this study is to analyze Complexity as a phenomenon, the expression of which depends on the level of matter organization. It aims to show that self-organization is the main feature of Complexity. It is considered to be inherent in objects, especially at the microlevel. On that level, any characteristic of a natural system is an expression of uncertainty, which is subordinate to informatics laws [3]. The unit of measurement of uncertainty is information, in the sense being used in this article [4, 5].

### Complexity and Complex Systems

In a general sense, Complexity as a phenomenon has two definitions supplementing each other. According to the first, Complexity is an aggregate of interdependent parts which comprise a **Whole**. The properties and behavior of a **Whole** are not to be found in any of its parts. The parts possess their own individual behavior and unique properties. They could be the results of biotic or abiotic phenomena, both of which were named agents by Heylighen [6]. Usually Complexity as the **Whole** is associated with the notion of Complex Systems.

A second definition considers *Complexity* to be a state between order and chaos, or disorder. A state of order

<sup>1</sup>International Institute of Systemic, Cybernetics and Informatics, USA-Venezuela, 2206 Tillman Av., Winter Garden, FL 34787, USA

<sup>2</sup>Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

<sup>3</sup>For a detailed analysis regarding the different senses in the meaning of ‘information’ and the different concepts and definitions related to this word, see [1]; see [2] for an extended version.

could be structures, signals, data, images, and many other objects [7].

In order to distinguish between the definitions of *Complexity* and Complex System, *Complexity* considered as phenomena will be assigned *italics* and Complex System considered as concrete (three-dimensional (3D)) manifestations of *Complexity* assigned in ordinary letters for the remainder of this paper.

In both definitions, *Complexity* means the **Whole** has been organized by its parts from Chaos to Order. The key idea in both is organization, in the sense of a transition from one state of matter to another. Transition is the process of development of the parts which make up the **Whole**. Development is the second property of *Complexity*. Without development, there could be no organization and *Complexity* could not be created. At the same time, development is one of the principal properties of matter. Any complex system tends to the state of minimal free energy and entropy. The change of state is the third property of *Complexity*.

Leading from such assumptions and independently from different views on *Complexity* as phenomena, one can see three main features of *Complexity*: *state*, *development*, and *properties*. Each feature is evaluated by its own independent measurements in units expressing time, mass, velocity, equilibrium, temperature, and some others.

The properties and character of the manifestation of *Complexity* depend upon the level of matter organization. Matter on the microlevel — atoms, molecules, crystals, and cells are organized into *Complexity* by one type of parts, while matter on the macrolevel — materials, substances, machine components and others, are organized into *Complexity* by other kind of parts.

## Complex Systems on the Macrolevel

Various parts of any macrolevel Complex System can be observed, measured by instruments, investigated, and analyzed. Some are created consciously by humans or instinctively by living creatures (insects, marine and terrestrial animals and birds, etc.). Humans create Complex Systems consciously by means of their intelligence and work in order to construct something, ordering it according to what was conceived and planned in advance. This includes all human-made technical facilities, mechanisms, and hugely Complex Systems such as national and international rail way networks with rails, carriages, locomotives, and other parts. In these complex systems, parts of the Whole work together in an organized mechanism.

Complex Systems like termite mounds, beehives, cocoons, nests, beaver dams, coral reefs, etc. are instinctively made by living creatures. Their constructions are guided by instinct programmed at the genetic level.

The basis of this instinct is the motivation or impulse to strive for survival in the environment and to reproduce a complex system similar to itself.

Complex Systems which are created artificially, consciously, and independently from the genetic level and brought into being by humans through manufacturing have common features as well as differences. Such forms contain the common attribute of utility or applicability. The technical nature of these Complex Systems are planned and realized by humans, according to their purposes. Aspects of these Complex Systems such as construction, operation, and development are planned in advance.

The commands which regulate the sequence of actions and interactions of various parts can be made by software algorithms and/or manual operations. On the other hand, living creatures construct Complex Systems instinctively via conditioned and unconditioned reflexes.

All types of *complexity* and complex systems created artificially (at the macrolevel) have a common feature: they are not the result of self-organization which is an internal natural property at the microlevel. Artificial *complexity* is manufactured by men or living creatures by means of ordering and relating macrolevel substances and components. Their process of manufacturing can be perceived as linear, open for observation, planned in advance consciously or instinctively in the case of nonhuman living creatures.

These systems are constructed and developed with the support of physics, chemistry, geophysics, and other general scientific laws to overcome Earth's gravity, air mass turbulence, entropy decrease, survival under temperature changes, as well as by creating microconditions in the environment, using magnetic and gravitation fields to navigate and resolve other problems, etc.

Any Complex System is not *Complexity* or a **Whole** until the function or purpose for which it has been created has been fulfilled. When the purpose for creating it ceases to exist, then it can be considered a 'dead mechanism.' A Complex System which has lost its basic components, for example, an abandoned seashell or bird nest is not considered a **Whole** and, therefore, not considered as *Complexity*. At such a point in its existence, it is considered as chaos or disorder. For example, discarded piles of plastic, wood, metal, sea shells or dead coral, and many other forms can be conceived as just chaos at the macrolevel.

There are large formations of Complex Systems at the macrolevel in the human social sphere and complex natural phenomena like the matter and energy cycle within the Earth—ocean—atmosphere system. Complex natural processes of the organic matter cycle, complex trophic chains among fauna and flora communities, etc. are not considered in this study.

## The Expression of *Complexity* at the Microlevel

The *Complexity* of Natural Complex Systems is the result of Self-Organization of the parts which can be elementary particles, atoms, molecules, crystals, or cells. Natural *Complexity* and corresponding Complex Systems are created without human intervention and without any effort by living creatures. Their constructions are subordinated to general laws of Mechanics and Physics and depend upon the mechanisms and modes by which their parts are organized on an atomic, molecular, or cellular levels.

The main feature of a microlevel natural Complex System is uncertainty. Uncertainty is determined by many properties of *Complexity* and the self-organization process of each level. The Self-Organization of a Complex System, i. e., to create order out of chaos, is the result of internal energy flow, nonlinear interactions among the parts, unpredictability, uncontrollability, and multilevel balancing [6].

The simplest Complex System is the hydrogen atom. It consists of a nucleon carrying one proton and one orbit with one rotating electron. According to Quantum Mechanics, the state, properties, and development of the parts, being elementary particles, are affected by quantum processes and possess their own impulses and momentum. In our case when one considers *Complexity* of macrolevel objects, the laws of Newtonian Mechanics play the main role. When one considers the *Complexity* of microlevel objects, both Newtonian and Quantum Mechanics Laws need to be taken into consideration. One of the main quantum characteristics is uncertainty which might be measured using Shannon's definition of information [8, 9], which was made in the context of electronic communication systems. This definition happened to be *mathematically isomorphic* with the definition of thermodynamic entropy in statistical mechanics. But this does not imply that they have the same meaning. Several authors alerted with regards to this issue.

Myron Tribus asked Shannon: "What he had thought about when he had finally confirmed his famous measure." Shannon replied [10]: "My greatest concern was what to call it. I thought of calling it 'information,' but the word was overly used, so I decided to call it 'uncertainty.' When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one knows what entropy really is, so in a debate you will always have the advantage.'"

With regarding to this issue, Denbigh affirmed [11] that "in my view Von Neumann did Science a disservice! . . . There are, of course, good mathematical reasons why information theory and statistical mechanics because require functions having the same formal structure. They have a common origin in probability theory and they also need to satisfy certain common requirements such as additivity. Yet this formal similarity [mathematical isomorphism] does not imply the functions are necessarily signifies or represents the same concepts. The term 'entropy' had already been given a well established physical meaning in thermodynamics. And it remains to be seen under what conditions, if any, thermodynamic entropy and information are mutually inconvertible."

Several authors (see, for example, [12]) agree with Denbigh [11] regarding the disservice Von Neumann made to Science. Since then, conceptual confusion has generated more confusions and some time implicit and unnoticed nonsense. This is why, several authors recommend using the respective units in order to avoid confusions. Schneider and Lewis [13], for example, affirm that "if you are making computations from symbols, *always* use the term uncertainty, with recommended units of bits per symbol. If you mean the entropy of a physical system, then use the term entropy, which has units of joules per kelvin (energy per temperature)."

In other very detailed articles, Callaos and Callaos affirmed the following [1, 2]: "The words selected by Shannon to refer to his mathematical definition and the identification of different concepts by the same mathematical definition has created semantic and conceptual confusions, and generated significant controversies. Thomas D. Schneider (from the National Institutes of Health), for example, affirms, referring to the words used by Shannon in his communications theory, that "Information Is Not Entropy, Information Is Not Uncertainty!"

Stonier affirmed that "a result of Von Neumann advice, the communications engineers and information theorists all became the victims of a bad joke: that the potential indeterminacy of a message is the same thing as entropy. The confusion still reigns today . . . Shannon's sleight of hand has been attacked by a number of authorities," among of whom are the authors mentioned above. For example, Hubert P. Yockey, physicist and information theorist, who worked at the University of California, Berkeley, and under Robert Oppenheimer on the Manhattan Project, after carefully examining this issue concluded that ". . .there is, therefore, no relation between Maxwell–Boltzmann–Gibbs entropy of statistical mechanics and Shannon's entropy of communications systems." Should we then differentiate between Shannon's entropy and Maxwell–Boltzmann–Gibbs's entropy, i. e., between informational and ther-

modynamic entropies? These kinds of confusions and contradictions are even found among the same communications engineers and researchers working on the same kind of engineering problems at the same time. Norbert Wiener, for example, affirms explicitly that “the notion of the amount of information attaches itself very naturally to a classical notion in statistical mechanics: that of entropy. Just as the amount of information in a system is a measure of its degree of organization, so the entropy of a system is a measure of its degree of disorganization. . . . The amount of information, being the negative logarithm of a quantity which we may consider as a probability, is essentially a negative entropy.” So, does “information equal entropy,” as it has been identified in Shannon’s Theory? Or is it essentially “negative entropy” as Wiener affirmed? We conceive information, at least from a subjective perspective and, more generally, in a biological context, as Wiener’s conceived it. It is almost common sense. Entropy is related to disorder and information to order. Entropy is related to disorganization and information is associated with organization. Entropy is related to uncertainty and information is associated with certainty, or a decrease in the level of uncertainty. In 1956, about eight years after Shannon’s and Wiener’s opposite conceptions regarding the relationship between information and entropy, Leon Brillouin published his book titled “Science and Information Theory,” where he affirmed that “we prove that information must be considered as a negative term in the entropy of a system; in short information is negentropy. . . . Entropy measures the lack of information.” So, it seems that Brillouin takes Wiener’s side, in conceiving information as negative entropy.

We recommend keeping in mind what has been quoted above in order to avoid conceptual confusions which might generate implicit and unnoticed nonsense. Many more details regarding this potential conceptual confusion and how to avoid it can be found in [2]. To help keeping in mind the different concepts named with the same word we proposed, a nominal distinction and we also suggested to flown other authors proposals regarding this issue. Accordingly we affirmed that Peters [14] proposed to use the term “*spread*” or “measure of spread of the probability distribution function” to refer to what he called “superior expression” [the genre] of both kinds of entropies [species]. Peters affirms that “only when this spread refers to the distribution function of microphysical state properties will be called entropy, when it refers to a quantity or a set of symbols not representing thermodynamical state properties it will be called information theoretical entropy, in short **intropy**.” In order to “generalize the notion of information,” Peters affirms that “information is created by an act, by an event which reduces the *a priori* spread of any quantity. The numerical difference (*a priori* spread minus *a poste-*

*riori* spread) is called information. In the limit case, the spread is totally removed. This happens when an individual symbol is selected out of a set of *a priori* possible symbols.” Consequently, information is produced when an act reduces the *a priori* **intropy** or the information theoretical entropy. Reiterating what was said above, but using Peter’s terms and conceptual perspective, one can say that in a two-symbol, or two-state, system, with maximum informational entropy, or **intropy** (i. e., where the two states are equiprobables,  $p_1 = p_2 = 0.5$ ), has by definition 1 bit of information capacity. This means that its spread, or intropy, is 1 bit. The respective limit case (where the spread is completely removed) is when one of the two states is selected, i. e., when the intropy is reduced to zero. Consequently,

$$\begin{aligned} A \text{ priori intropy} - a \text{ posteriori intropy} \\ = 1 \text{ bit} - 0 \text{ bit} = 1 \text{ bit.} \end{aligned}$$

So, as it can be easily noticed, although the measure we are using for *a priori* intropy, *a posteriori* intropy, and information is the same, the concepts are different. Intropy (or information theoretical entropy) and information (*a priori* intropy – *a posteriori* intropy) are not the same concepts. The concept of “information capacity” is not the same as the “information delivered” by an action, or selection. The “quantity of information capacity” is equal the “quantity of information delivered” just in the limit case where *a posteriori* intropy is equal to zero. In nonlimit cases (i. e., *a posteriori* intropy > 0), both measures are not the same. We cannot deliver more information than the source capacity, but we can deliver less (or equal, in the limit case) quantity of information. Consequently, the information measure of the source should be differentiated from the information measure of the delivered information, if one wants to avoid confusing ambiguities.

As it was noted above, the uncertainty, which is one of the main quantum characteristics, could be measured by Shannon’s definition of information or “*information entropy*” if one can keep in mind the distinction referred to above in order to avoid the confusions. Broglie [15] established that elementary particles behave as a particle and as a wave possessing by energy and impulse. The cinematic and dynamics of elementary particle movement in the quantum field was described by Heisenberg [4]. Zeilinger proved that the **bit** (binary unit), introduced by Shannon in 1948 as information unit, is, in fact, the information capability (*a priori* intropy) [5]. Rashovsky [16] and Trucco [17] proved that the 3D volume of a molecule has topological information that can be measured in **bits**.

Natural Complex Systems at the molecular level are of two types. One type is associated to living organisms composed of very complex cells consisting of billions

of molecules. Petroleum is another type composed of thousands of hydrocarbon molecules. According to the molecular organization of petroleum, it is considered to be in a state between organic and inorganic.

On the other hand, Self-Organization of *Complexity* and its information content can be understood if one considers two fundamental discoveries of science. First is the discovery of amino acids or macromolecules DNA and RNA in biology. The second is the discovery of the quantum behavior of elementary particles in all chemical elements including carbon and hydrogen.

DNA and RNA are the self-organizing molecules into a macroform. The self-organized parts of the macroform molecule are genes. These genes form genetic code or the matrix carrying genetic information. The  $10^{14}$  cells of the most complex natural system — human beings — are reproduced according to the gene-matrix. Ashby established the relationship between Variety, Organization, Complexity, and Information of microlevel objects [18]. He showed their role in the appearance and development of biotic complex natural systems.

Self-Organization on the atomic level objects is generated by elementary particles (electrons and protons). Their quantum behavior, uncertainty, and corresponding information content have close connection to quantum mechanics principles.

## Petroleum *Complexity* is Identified by Its Composition and Properties

Petroleum has three features showing its Self-Organization into a Complex System.

1. The parts of petroleum are hydrocarbon molecules; which compose 95% of its content. These parts are self-organized into three types of hydrocarbon molecules: (i) saturated (paraffin's) molecules which are a chain of carbon and hydrogen atoms; (ii) unsaturated hydrocarbons (naphthenes) molecules which are the closed cycles; and (iii) aromatic in which molecules are hexahedron in structure with a carbon atom located in each of its corners. All these parts self-organize to form petroleum.
2. During the petroleum generation process, carbon and hydrogen atoms are not drawn or extracted from any rocks, composed from geological media (matter) containing these elements. The initial quantity of carbon and hydrogen atoms in organic sediments (according to an organic model) or where such atoms were supplied by the Earth's mantle substance (according to inorganic model) remains constant during the entire process of petroleum generation. The three main types of hydrocarbon

molecules mentioned above are contained in the final petroleum in their original respective quantities. This occurs despite the differences in the structure, molecular mass, geological media, and time of generation. This phenomenon shows that petroleum complexity is a Self-Organizing process in which its parts are the hydrocarbon molecules. Self-Organization in that case is the transformation of one type of molecule into another. In the case of petroleum, this process is triggered by catalytic reactions between hydrocarbon molecules and the elements composing the surrounding rocks of geological media.

3. Hydrocarbon molecules are generated during different stages of the process in geological media where thermodynamic conditions are constantly changing. Therefore, the hydrocarbon molecules as the parts of the Complex System preserve their composition and structure. The fact of self-organization has been proved as the main mechanism of petroleum generation. Petroleum in fields located on The Arabian Peninsula, North Africa, Oklahoma, Western Siberia, and other areas are an aggregate of paraffin, naphthenic, and aromatic hydrocarbons molecules. There are direct indices of the presence of discrete and nonlinear mechanisms that account for the preservation of structural and chemical individuality in each type of hydrocarbon molecule. Seyful-Mulyukov showed this mechanism is the quantum matrix of the hydrocarbon molecule [19]. Via analogical thinking, one can say that it is some sort of '*hydrocarbon genetic code*' what guarantees the development of the given types of molecules during all stages of petroleum generation. The matrix is a subatomic level phenomenon in which the elementary particles of atoms interact to generate hydrocarbon molecules.

The generation of any hydrocarbon molecule depends upon the electron orbitals of carbon and hydrogen atoms hybridization forming a molecule. New molecule generation results from a change from one type of hybridization to another. This change is the reaction of a Complex System to changes in the surrounding thermodynamic and geological media. Therefore, during creation of a new type of molecules, all previously generated molecules maintain their spin or wave fields of elementary particles. The Quantum matrix provides for the preservation of each type of molecules appearing during the different stages of petroleum genesis.

The petroleum quantum matrix is comparable to the macromolecule DNC and DNA in living species. In spite of the differences at the organization level, both fulfill the same function — *Complex System Reproduction*. For example, the reproduction of a human being,

comprising some  $10^{14}$  cells, occurs as the DNA of the older organism is replicated within the new organism by use of the same quantity and quality of cells. The petroleum quantum matrix contains coded information that insures the parallel existence of the specific types of molecules within the system. DNC and DNA macromolecules are a higher level of Organization compared to petroleum's quantum matrix.

Natural Complex Systems such as petroleum demonstrate such features as *Existence, Development, and Cognoscibility* besides various physical properties. They expose the regularities and mechanism of *Complexity* development as well as *Composition, Structure, and Duration of Existence*. These features are the key elements to understand petroleum genesis.

*Existence* determines the behavior and properties of *Complexity* as the **Whole** which is different from any of its parts. By existence, we mean such Complex System characteristics as *Invariance, Synergetics, Uniqueness, and Unpredictability*. Petroleum completely demonstrates these features. As the **Whole**, petroleum features invariance as it maintains its composition and properties in different geological structures, stratigraphic levels, and geophysical media. Petroleum is also synergetic because it exists as the **Whole** in which properties and characteristics are different from any of its parts. Petroleum is unique due to its hydrocarbon composition, structure, and molecular mass being inimitable and able to be generated and exist only in a specific geological period of Earth's development. Petroleum generation requires specific thermodynamic, geological, geochemical conditions which appeared in the Earth's crust not long ago. The genesis process remains unpredictable as long as data on the trajectory of geophysical matter development at any moment of geological time cannot be established and expressed mathematically.

Development is the totality of such dynamic petroleum features as *Openness, Nonstationary, and Movement Permanency*. Petroleum openness is a property that allows petroleum to exchange energy and information with the environment. This leads to changes in the hydrocarbon molecules structure and composition. This situation triggers a change in the Complex System as a **Whole**. Nonstationary is another property of petroleum which permits the ability to change structure and phase state depending on the surrounding geological media. If change leads to the loss of the main types of hydrocarbon molecules, the remainder is hydrocarbons but not petroleum as Complex System.

Movement Permanency refers to constant changes in petroleum's internal structure and composition caused by changes in its entropy. Movement Permanency explains the impossibility of petroleum in maintaining its native composition and structure during millions of years. Entropy change refers to the difference between

any initial and finite moment of petroleum development. That is why, Devonian, Carboniferous, Jurassic, or Cretaceous age petroleum could not exist now.

*Cognoscibility* has a direct relationship with petroleum origin, age, and *Complexity*. Cognoscibility is the process of interaction between two systems: (i) the *subject* of cognition, i. e., the researcher in our case, and (ii) the *object* of cognition, which, in our case, is the petroleum. Researchers are studying the properties and composition of petroleum which has recently been extracted. Two opinions could be expressed on the age of that petroleum — either it was generated recently or it was generated hundreds of millions of years ago.

The first opinion means that the researcher examined and fixed the data, facts and conditions characterizing the petroleum as a Complex System, created under recently existing thermodynamic and geological conditions. The second opinion means that the petroleum is a Complex System generated hundreds of millions of years ago. If so, that system maintained its composition and structure over hundreds of millions of years. Also, it means that the specific thermodynamic and geological conditions required for petroleum generation over those hundreds of millions of years were the same as they are today. This absolutely contradicts the postulates and laws of geochemistry and geology. Petroleum could not exist in its native state for hundreds of millions years.

*Cognoscibility* is the interaction between the subject and objects of cognition within a determined time and space. Any natural complex system development can be comprehended if one takes into consideration that interaction. For example, a paleontologist while modeling Cambrian fauna, namely, the trilobite (650 million years old) had to deal with a simpler type (dummy) of natural stone that is not considered as a Complex System but is rather a chaotic mixture of minerals replicating the system by simple repetition of its form but not its content or natural *Complexity*. Cognoscibility provides proof for the idea that any Complex System generated long ago and remaining unchanged for millennia could not exist. All *Complexity* considered as a Whole has a life circle, i. e., creation, development, existence, and decay. Creating and developing petroleum Complexity as a phenomenon is characterized by the Laws of *Simplicity, Uncertainty, and Requisite Variety*.

Simplicity is Natural or Universal Law. Nature generates any Complex System by selecting the simplest option among all existing options since it is the most efficient use of energy and resources. Petroleum *Complexity* is the simplest aggregate of hydrocarbon molecules generated in a given geochemical, geological, thermodynamic, and other circumstances relevant to specific stages in the development of the Earth. The Law of Simplicity provides researchers with an understanding of the age of any Complex System (petroleum, animals, etc.).

This Law is the embodiment of the simplest optimal form of *Complexity* possible in a given, specific period of the Earth evolution.

Uncertainty is a notion used in many sciences including Informatics, Mathematics, Philosophy, Cybernetics, Physics, and others. In these contexts, Informatics might be conceived as a science of Uncertainty, i. e., science using mathematical and cybernetic methods to transform Uncertainty (potential information) into Information (actual or delivered information). Informatics, by means of applying a range of technologies, allows the use, transmission, storage, retrieval, and many other ways of handling, managing, and utilizing Information.

Uncertainty is important for understanding the interactions between elementary particles of the carbon and hydrogen atoms during the process of generating hydrocarbon molecules. The key feature of that process is the conjugate variable states of the interacting objects subordinate to the Laws of Quantum Mechanics. Uncertainty is applied to express the passage of the complex system from one state to other. Uncertainty and Information share causal reciprocity insofar that a change in Uncertainty leads to change in Information, potential information might be transformed in delivered or actual information.

As it is known, Requisite Variety Law was formulated first by Ashby [20]. The Law expresses the interconnection of two systems: (i) the one which is controlling and (ii) that which is under control. According to this Law, the larger the variety of reactions that a control system can generate (in its interaction with the controlled system), the larger the variety of perturbations it can compensate. Metaphorically expressed, the Requisite Variety Law states that *only the internal variety of a system can 'destroy' (deal with) its external variety*. A control system of two states, for example, cannot control a system of three possible states or more. Consequently, the potential adaptability of a system to the uncertainties of its environment depends on having more internal variety and the external ones. Regulative control prevents the transference of variety from the environment to the systems; so, the system can be less deviated from its 'goal.' This can be conceived as the opposite to information transmission where the purpose is to maximize the conservation of variety.

In general, if one has a set  $D$  of disturbances (in the environment or the controlled system), a set  $R$  of responses (of the system or of the regulative control), and a set  $O$  of outcomes (of the system in its interaction with its environment or of the control systems interacting with the control system), then, Ashby affirmed [20, p. 47] that "if the varieties are measured logarithmically, this means that if the varieties of  $D$ ,  $R$ , and actual outcomes are, respectively,  $V_d$ ,  $V_r$ , and  $V_o$ , then the minimal value of  $V_o$  is  $V_d - V_r$ . If now  $V_d$  is given,  $V_o$  minimum

can be lessened only by a corresponding increase in  $V_r$ . This is the law of requisite variety. What it means is that restriction of the outcomes to the subset that is valued as Good demands a certain variety in  $R$ ."

The Requisite Variety Law might be applied to Complex Systems such as society, industrial and agricultural production systems, economic systems and nature, including petroleum as a complex whole. The survival of the petroleum for 1–2 hundreds million years, in spite of all changes in successive geological environments with its respective high level of disturbances, shows that it has a high internal variety which might be conceived as a measure of its complexity, which increased though process in which simpler molecules formed part of larger, or more complex, ones as a result of thermodynamic changes and catalytic reactions in the continuously changing environment.

*Complexity* creation is not free-running, autonomous, and independent from environmental processes. Dependence on the state and properties of the environment demonstrates the creation of petroleum *Complexity*. Petroleum's composition, structure, and development depend directly on the totality of specific geological media. If one considers that petroleum could be a Complex System of one hundred or even two hundred million years of age, then one should also consider that the geological environment today must be the same as it was many millions of years ago. This, however, is implausible as everything is changing constantly, including the geological media and hydrocarbon molecules. This may account for the current failure to find petroleum which is not comparable to modern petroleum.

## Petroleum Complex System and Its Information Content

Information is one of the main objective characteristics of the elementary particles in any atom including carbon, hydrogen, and hydrocarbon molecules [3–5]. At the same time, Information volume (stored information or potential information) is one of the indices of a Natural Complex Systems development process. Petroleum is a perfect example for demonstrating this concept.

Petroleum hydrocarbon molecules are 99% composed of atoms of carbon (C), hydrogen (H), oxygen (O), nitrogen (N), and sulfur (S). Their information content in **bits** is H — 10, C — 109, N — 138, O — 149, and S — 317 [3]. Petroleum is generated from hydrocarbon molecules which combine in consecutive order as they change form. This process can be expressed by Information content in **bits** as molecules appear and more complex molecules are structured.

Petroleum generation results in the appearance of carbon and hydrogen atoms in the normal nucleon-

orbital configuration. Fomin showed that it occurs due to mantle plasma matter decompression [21]. The bonding of these atoms is realized through several stages, the first being hydrocarbon protomolecule formation. The following stages are consecutive processes of gaseous and liquid hydrocarbon molecule generation accruing in deep lithosphere strata of the Earth. The main stage of petroleum hydrocarbon molecule generation is, as was briefly mentioned above, the transformation of simpler molecules into more complex molecules, a process triggered by thermodynamic changes and catalytic reactions between hydrocarbon molecules and the crystal structure of surrounding geological media rocks.

Following the Heylighen idea on the *Complexity* development [6], petroleum generation completely corresponds to the process of *Complexity* formation. The process is accompanied with a corresponding increase in Information content [22]. Hydrogen and carbon atoms being formed in their normal configuration in upper mantle have the following bits. Information content H — **10 bit** and C — **109 bit**, respectively. Methane (CH<sub>4</sub>) is the first hydrocarbon molecule generated with **154 bit** of Information. Next, gases such as butane (C<sub>4</sub>H<sub>10</sub>) form hydrocarbon molecules which are possessing **547 bit** of Information.

The main types of hydrocarbon molecules including saturated (paraffin's), unsaturated (naphthenes), and aromatic molecules make up 90% of petroleum composition. The total Information content of the pure hydrocarbon part of petroleum, which consists only of carbon and hydrogen atoms, is **4148 bit** in petroleum conventional molecule.

Petroleum completes its formation process and is accumulated in pools or within fields where finally all hydrocarbon molecules join. Petroleum's final composition consists of a mixture of pure hydrocarbons (paraffin's, naphthenes, and aromatics), heteroatom molecules, and dash (sulfur, nitrogen, oxygen, and some others). Conventional molecules of light petroleum have an empirical formula of C<sub>32</sub>H<sub>66</sub>SN and a corresponding Information content of **16 224 bit**.

In specific conditions at the depth interval of 50–200 m, petroleum loses its light components and transforms into bitumen. The empirical formula of conventional molecule of bitumen is C<sub>45</sub>H<sub>51</sub>O<sub>2</sub>SN and an Information content of **17 789 bit**.

In a general sense, petroleum genesis is a process taking place in the upper lithosphere. From a chemical point of view, it is natural liquid substance formation made up of hydrocarbons molecules. From a physical point of view, the formation of petroleum *Complexity* is the natural ordered process of Self-Organization: elementary particle form atoms, atoms form hydrocarbon molecules, and these transform into three types which finally compose petroleum. The consecutive expansion of

*Complexity* is accompanied by a corresponding increase in Information Content.

## Concluding Remarks

*Complexity* and Complex Systems are considered as a **Whole** composed of the parts (also referred to as objects or agents). The types, scales and modes of organization are different. On the macrolevel, *Complexity* and Complex Systems are created artificially by human beings or other living species. They are not the result of Self-Organization and their *Complexity* is created consciously or instinctively. Their main feature is utility because they are a construct generated by a determined aim or directed by instinct. Artificial complex systems that have been created by macrolevel objects are not the self-organizing systems.

*Complexity* and Natural Complex Systems of microlevel objects are created by Nature alone. Such microlevel objects are subdivided on the lowest atomic and highest molecular levels. The main feature of microlevel Complex Systems is *Complexity* and Self-Organization. Its main feature is Uncertainty of the distribution and interaction of the parts in the context of a **Whole**. That is why, one can use the concepts of information as defined by Shannon in expressing petroleum's information content *Complexity* and Complex Systems at a molecular level composed by the cells of living species or the petroleum hydrocarbon molecules, the latter according to a scale of organization positioned between organic and inorganic.

Self-Organization of *Complexity* and their Information Content on the atomic level are conditioned by the quantum behavior of the atom's elementary particles. It is the main factor of Complex System's parts reproduction and interaction. On the molecular level, *Complexity* is determined by the properties, structure, and interaction of DNC and DNA macromolecules which contain the genes and the related biological information. Petroleum is a Natural Complex Thermodynamic System which not only shows the main and well-known features of *Complexity* but reveals some new features. Petroleum is a Self-Organizing Natural Complex System structured by its parts (hydrocarbon molecules) via nonlinear discrete processes. The structuring and preservation of these parts in the Complex System is provided by quantum matrixes. Accordingly, quantum matrixes create three main types of hydrocarbon molecules which remain constant in petroleum in its native form. On the molecular level, hydrocarbons transformation into 500 molecular structures, which make up the petroleum, is realized by their catalytic reactions within the rocks of geological media. Petroleum demonstrates additional features of *Complexity* including

existence, development, cognoscibility, simplicity, uncertainty, and requisite variety. Complexity and Natural Complex Systems at the microlevel possess Information Content corresponding to the quantity and the quality of atoms and molecules composing the parts of the Complex System.

## Acknowledgments

Originally, this article was written in Russian and English. Authors appreciate Scott Barbur for his efforts to integrate both and editing the text in English allowed a more adequate understanding of the main idea of the article to the Western scientific community.

## References

- Callaos, N., and B. Callaos. 2002. Toward a systemic notion of information: Practical consequences. *Informing Sci. J.* 5(1):1–11.
- Callaos, B., and B. Callaos. 2011. Toward a systemic notion of information: Practical consequences (extended version). Academia. 99 p. Available at: [http://www.academia.edu/4434476/Toward\\_a\\_Systemic\\_Notion\\_of\\_Information\\_Practical\\_Consequences\\_Extended\\_Version\\_](http://www.academia.edu/4434476/Toward_a_Systemic_Notion_of_Information_Practical_Consequences_Extended_Version_) (accessed September 12, 2015).
- Gurevich, I. M. 2007. *Zakony informatiki — osnovna stroeniya i pozniyani slozhnykh sistem* [Informatics laws — the basis of structure and knowledge of complex systems]. Moscow: TORUS PRESS. 399 p.
- Heisenberg, W. 1957. Quantum theory of fields and elementary particles. *Rev. Mod. Phys.* 29(3):269–278.
- Zeilinger, A. A. 1999. Foundation principle for Quantum Mechanics. *Found. Phys.* 29(4):631–643.
- Heylighen, F. 2008. Complexity and self-organization. *Encyclopedia of library and information science*. Eds. M. J. Bates and M. N. Maack. New York, NY: Taylor and Francis. 20 p. Available at: <http://pespmc1.vub.ac.be/papers/elis-complexity.pdf> (accessed September 19, 2015).
- Prigogine, I., and I. Stengers. 1984. *Order out of Chaos: Man's new dialogue with nature*. New York, NY: Bantam Books. 381 p.
- Shannon, C. E. 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27:379–423; 623–656.
- Shannon, C. E., and W. Weaver. 1963. *The mathematical theory of communication*. Urbana and Chicago: University of Illinois Press. 125 p.
- Tribus, M., and C. McIrvine. 1971. Energy and information. *Sci. Am.* 225(3):179–188.
- Denbigh, K. 1981. How subjective is entropy? *Chem. Brit.* 17(4):168–185.
- Ben-Naim, A. 2008. *A farewell to entropy: Statistical thermodynamics based on information:  $S = \log W$* . Hackensack, NJ: World Scientific. 409 p.
- Schneider, T., and K. Lewis. 2011. A glossary for biological information theory and the Delila system. Available at: <https://schneider.ncifcrf.gov/glossaryframes.html> (accessed September 27, 2015).
- Peters, J. 1975. Entropy and information: Conformities and controversies. *Entropy and information in science and philosophy*. Eds. L. Kubát and J. Zeman. Amsterdam: Elsevier Scientific Publ. Co. 61–81.
- Brogie, L. 1927. Wave mechanics and the atomic structure of matter and radiation. *J. Phys. Paris* 8(5):225–241.
- Rashevsky, N. 1955. Life, information theory and topology. *B. Math. Biophys.* 17(3):229–235.
- Trucco, E. 1956. A note on the information content of graphs. *B. Math. Biophys.* 18(2):129–135.
- Ashby, W. R. 1956. *An introduction to cybernetics*. London: Chapman & Hall. 283 p.
- Seyful-Mulyukov, R. B. 2014. Quantum matrix of hydrocarbon molecules is the key element of petroleum genesis. *14th Multidisciplinary Scientific Geoconference (International) and Expo SGEM Proceedings*. Varna, Bulgaria. 1:759–765.
- Ashby, W. R. 1958. Requisite variety and its implications for the control of complex systems. *Cybernetica* 1-2:83–89. Available at: <http://pespmc1.vub.ac.be/books/ashbyreqvar.pdf> (accessed October 11, 2015).
- Fomin, Yu. M. 2005. Verkhnyaya astenosfera — perekhodnaya zona mezhdu veshchestvom mantii i litosfery [The upper asthenosphere — the transition zone between the mantle and lithosphere]. *Problemy evolyutsii* [Evolution problems]. Available at: <http://www.evolbiol.ru/fomin.htm> (accessed March 11, 2016).
- Seyful-Mulyukov, R., and M. K. Hlava. 2012. The nature of information as a fundamental property of matter: A case study using petroleum and hydrocarbon gases. *3rd Multi-Conference (International) on Complexity, Informatics and Cybernetics Proceedings*. Orlando, FL. 173–178.

Received October 15, 2015

## Contributors

**Callaos Nagib C.** (b. 1943) — PhD in mathematics; President, International Institute of Informatics and Systemic, 14269 Lord Barclay Dr., Orlando, FL 32837, USA; [n.c.callaos@callaos.com](mailto:n.c.callaos@callaos.com)

**Seyful-Mulyukov Rustem B.** (b. 1928) — Doctor of Science in geology, professor, Head of Laboratory, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str, Moscow 119333, Russian Federation; [rust@ipiran.ru](mailto:rust@ipiran.ru)

## СЛОЖНОСТЬ И ЕЕ ИНФОРМАЦИОННОЕ СОДЕРЖАНИЕ

Н. К. Каллаос<sup>1</sup>, Р. Б. Сейфуль-Мулюков<sup>2</sup>

<sup>1</sup>Международный институт информатики и системных исследований, Орландо, Флорида, США, n.c.callaos@callaos.com

<sup>2</sup>Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, gust@ipiran.ru

**Аннотация:** Термин «информация» имеет разные толкования и определения. Авторы статьи разделяют мнение исследователей, рассматривающих информацию как одно из свойств материи. Такое представление информации использовано для анализа понятий «сложность» и «самоорганизация». Их связь с информацией анализируется на примере закономерностей создания и строения объектов субатомного, микро- и макроуровня. Понятия «сложность», «самоорганизация» и «информационное содержание» применены к нефти, на примере которой доказывается связь этих трех понятий. Анализ нефти как сложной термодинамической природной системы дал возможность расширить понятие «сложность» свойствами и особенностями, которые ранее не рассматривались. Сложность и самоорганизация помогают представить глубинное происхождение и строение углеводородных молекул, возраст, а также образование и возраст нефти в целом.

**Ключевые слова:** сложность; свойства сложности; сложные системы; самоорганизация; искусственная сложность; природная сложность; сложность углеводородных молекул; происхождение нефти; информационное содержание нефти

**DOI:** 10.14357/19922264160112

### Литература

1. Callaos N., Callaos B. Toward a systemic notion of information: Practical consequences // *Informing Sci. J.*, 2002. Vol. 5. No. 1. P. 1–11.
2. Callaos N., Callaos B. Toward a systemic notion of information: Practical consequences (extended version). — *Academia*, 2011. 99 p. [http://www.academia.edu/4434476/Toward\\_a\\_Systemic\\_Notion\\_of\\_Information\\_Practical\\_Consequences\\_Extended\\_Version](http://www.academia.edu/4434476/Toward_a_Systemic_Notion_of_Information_Practical_Consequences_Extended_Version).
3. Гуревич И. М. Законы информатики — основа строения и познания сложных систем. — М.: ТОРУС ПРЕСС, 2007. 399 с.
4. Heisenberg W. Quantum theory of fields and elementary particles // *Rev. Mod. Phys.*, 1957. Vol. 29. No. 3. P. 269–278.
5. Zeilinger A. A. Foundation principle for quantum mechanics // *Found. Phys.*, 1999. Vol. 29. No. 4. P. 631–643.
6. Heylighen F. Complexity and Self-Organization // *Encyclopedia of library and information sciences* / Eds. M. J. Bates, M. N. Maack. — New York, NY, USA: Taylor and Francis, 2008. 20 p. <http://pespmc1.vub.ac.be/papers/elis-complexity.pdf>.
7. Prigogine I., Stengers I. Order out of Chaos: Man's new dialogue with nature. — New York, NY, USA: Bantam Books, 1984. 381 p.
8. Shannon C. E. A mathematical theory of communication // *Bell Syst. Tech. J.*, 1948. Vol. 27. P. 379–423; 623–656.
9. Shannon C. E., Weaver W. The mathematical theory of communication. — Urbana and Chicago: University of Illinois Press, 1963. 125 p.
10. Tribus M., McIrvine C. Energy and information // *Sci. Am.*, 1971. Vol. 225. No. 3. P. 179–188.
11. Denbigh K. How subjective is entropy? // *Chem. Brit.*, 1981. Vol. 17. No. 4. P. 168–185.
12. Ben-Naim A. A farewell to entropy: Statistical thermodynamics based on information:  $S = \log W$ . — Hackensack, NJ, USA: World Scientific, 2008. 409 p.
13. Schneider T., Lewis K. A glossary for biological information theory and the Delila system. <https://schneider.ncifcrf.gov/glossaryframes.html>.
14. Peters J. Entropy and information: Conformities and controversies // *Entropy and information in science and philosophy* / Eds. L. Kubát, J. Zeman. — Amsterdam: Elsevier Scientific Publ. Co., 1975. P. 61–81.
15. Broglie L. Wave mechanics and the atomic structure of matter and radiation // *J. Phys. Paris*, 1927. Vol. 8. No. 5. P. 225–241.
16. Rashevsky N. Life, information theory and topology // *B. Math. Biophys.*, 1955. Vol. 17. No. 3. P. 229–235.
17. Trucco E. A note on the information content of graphs // *B. Math. Biophys.*, 1956. Vol. 18. No. 2. P. 129–135.
18. Ashby W. R. An introduction to cybernetics. — London, U.K.: Chapman & Hall, 1956. 283 p.

19. *Seyful-Mulyukov R.* Quantum matrix of hydrocarbon molecules is the key element of petroleum genesis // 14th Multidisciplinary Scientific Geoconference (International) and Expo SGEM Proceedings. — Varna, Bulgaria, 2014. Vol. 1. P. 759–765.
20. *Ashby W.R.* Requisite variety and its implications for the control of complex systems // *Cybernetica*, 1958. Vol. 1. No. 2. P. 83–89. <http://pespmc1.vub.ac.be/books/ashbyreqvar.pdf>.
21. *Фомин Ю. М.* Верхняя астеносфера — переходная зона между веществом мантии и литосферы // Проблемы эволюции, 2005. <http://www.evolbiol.ru/fomin.htm>.
22. *Seyfoul-Mulyukov R., Hlava M. M. K.* The nature of information as a fundamental property of matter: A case study using petroleum and hydrocarbon gases // 3rd Multi-Conference (International) on Complexity, Informatics and Cybernetics Proceedings. — Orlando, FL, USA, 2012. P. 173–178.

*Поступила в редакцию 15.10.2015*

**Алексеевский Даниил Андреевич** (р. 1983) — аспирант факультета гуманитарных наук Национального исследовательского университета Высшая школа экономики

**Белоусов Василий Владимирович** (р. 1977) — кандидат технических наук, заведующий сектором Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Брюхов Дмитрий Олегович** (р. 1971) — кандидат технических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Вольнова Алина Александровна** (р. 1986) — научный сотрудник Института космических исследований Российской академии наук

**Гордов Евгений Петрович** (р. 1946) — доктор физико-математических наук, руководитель Международного исследовательского центра климато-экологических исследований Института мониторинга климатических и экологических систем Сибирского отделения Российской академии наук

**Горшенин Андрей Константинович** (р. 1986) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; доцент Московского технологического университета (МИРЭА)

**Засыпко Вероника Владимировна** (р. 1988) — аспирант Национального исследовательского университета Высшая школа экономики

**Зацман Игорь Моисеевич** (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Инькова Ольга Юрьевна** (р. 1965) — доктор филологических наук, сотрудник Университета Женева; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Калиниченко Леонид Андреевич** (р. 1937) — доктор физико-математических наук, профессор, заведующий лабораторией Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; профессор факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

**Каллаос Наджиб Чарли** (р. 1943) — PhD (математика), президент Международного института информатики и системных исследований, Орlando, Флорида, США

**Кириков Игорь Александрович** (р. 1955) — кандидат технических наук, директор Калининградского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Киселева Надежда Николаевна** (р. 1949) — доктор химических наук, заведующая лабораторией Института металлургии и материаловедения им. А. А. Байкова Российской академии наук

**Ковалева Дана Александровна** (р. 1973) — кандидат физико-математических наук, научный сотрудник Института астрономии Российской академии наук

**Ковалёв Сергей Протасович** (р. 1972) — доктор физико-математических наук, ведущий научный сотрудник Института проблем управления им. В. А. Трапезникова Российской академии наук

**Колесников Александр Васильевич** (р. 1948) — доктор технических наук, профессор кафедры телекоммуникаций Российского государственного университета имени Иммануила Канта; старший научный сотрудник Калининградского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Кружков Михаил Григорьевич** (р. 1975) — ведущий программист Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Кудрявцев Алексей Андреевич** (р. 1978) — кандидат физико-математических наук, доцент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова;

Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Листопад Сергей Викторович** (р. 1984) — кандидат технических наук, научный сотрудник Калининградского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Малков Олег Юрьевич** (р. 1961) — доктор физико-математических наук, заведующий отделом Института астрономии Российской академии наук

**Окладников Игорь Георгиевич** (р. 1978) — кандидат технических наук, старший научный сотрудник Международного исследовательского центра климато-экологических исследований Института мониторинга климатических и экологических систем Сибирского отделения Российской академии наук

**Подколотный Николай Леонтьевич** (р. 1952) — заведующий Центром коллективного пользования «Биоинформатика» Федерального исследовательского центра Института цитологии и генетики Сибирского отделения Российской академии наук

**Позаненко Алексей Степанович** (р. 1962) — кандидат физико-математических наук, заведующий сектором Института космических исследований Российской академии наук

**Пономарева Наталия Васильевна** (р. 1956) — доктор медицинских наук, руководитель группы, ведущий научный сотрудник Научного центра неврологии

**Попкова Наталия Александровна** (р. 1992) — младший научный сотрудник, Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Румовская София Борисовна** (р. 1985) — программист 1-й категории Калининградского филиала Фе-

дерального исследовательского центра «Информатика и управление» Российской академии наук

**Сейфуль-Мулюков Рустем Бадриевич** (р. 1928) — доктор геолого-минералогических наук, профессор, заведующий лабораторией Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Синицын Владимир Игоревич** (р. 1968) — доктор физико-математических наук, доцент, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Синицын Игорь Николаевич** (р. 1940) — доктор технических наук, профессор, заслуженный деятель науки РФ, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Скворцов Николай Алексеевич** (р. 1973) — научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Ступников Сергей Александрович** (р. 1978) — кандидат технических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Фазлиев Александр Зарипович** (р. 1953) — кандидат физико-математических наук, заведующий Центром интегрированных информационных систем Института оптики атмосферы Сибирского отделения Российской академии наук

**Шнурков Петр Викторович** (р. 1953) — кандидат физико-математических наук, доцент Национального исследовательского университета Высшая школа экономики

---

## Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

---

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация не должна нарушать закон об авторских правах.

Направляя рукопись в редакцию, авторы сохраняют все права собственников данной рукописи и при этом передают учредителям и редколлегии неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на ее распространение в России и за рубежом. Авторы должны представить в редакцию письмо в следующей форме:

**Соглашение о передаче права на публикацию:**

*«Мы, нижеподписавшиеся, авторы рукописи «. . .», передаем учредителям и редколлегии журнала «Информатика и её применения» неисключительное право опубликовать данную рукопись статьи на русском языке как в печатной, так и в электронной версиях журнала. Мы подтверждаем, что данная публикация не нарушает авторского права других лиц или организаций, а также не содержит сведений, запрещенных к опубликованию в открытой печати.*

*Подписи авторов: (ф. и. о., дата, адрес)».*

Это соглашение может быть представлено в бумажном виде или в виде отсканированной копии (с подписями авторов).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации представленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.

3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам.

Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.

5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.

6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или ЛАТЭХ (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.

7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 20 страниц указанного формата.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления представлены в Интернете: <http://www.ipiran.ru/journal/template.doc>

8. Статья должна содержать следующую информацию на *русском и английском языках*:

- название статьи;
- Ф.И.О. авторов, на английском можно только имя и фамилию;
- место работы, с указанием почтового адреса организации и электронного адреса каждого автора;
- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:  
[http://www.ipiran.ru/journal/issues/2013\\_07\\_01\\_rus/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01_rus/authors.asp) и  
[http://www.ipiran.ru/journal/issues/2013\\_07\\_01\\_eng/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp);
- аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул;
- ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами;
- источники финансирования работы (ссылки на гранты, проекты, поддерживающие организации и т. п.).

9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала;
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://translit.ru/> бесплатной программой транслитерации русского текста в латиницу, при этом в закладке «варианты. . . » следует выбрать опцию BGN.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Ниже приведены примеры ссылок на различные виды публикаций в списке “References”.

**Описание статьи из журнала:**

Zagurenko, A. G., V. A. Korotovskikh, A. A. Kolesnikov, A. V. Timonov, and D. V. Kardymon. 2008. Tekhniko-ekonomicheskaya optimizatsiya dizayna gidrorazryva plasta [Technical and economic optimization of the design of hydraulic fracturing]. *Neftyanoe hozyaystvo [Oil Industry]* 11:54–57.

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

**Описание статьи из электронного журнала:**

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

**Описание статьи из продолжающегося издания (сборника трудов):**

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal’—kompozit” [Experimental study of the strength of joints “steel—composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

**Описание материалов конференций:**

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

**Описание книги (монографии, сборники):**

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. 1: Friksionnye protsessy pri rezanii metallov* [Tribology of cutting. Vol. 1: Frictional processes in metal cutting]. Ivanovo: Ivanovskii State Univ. 108 p.

**Описание переводной книги** (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W. Vibration problems in engineering. — 4th ed. — N.Y.: Wiley, 1974. 521 p.*)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. N.Y.: Wiley. 521 p.

**Описание неопубликованного документа:**

Latypov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

**Описание интернет-ресурса:**

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

**Описание диссертации или автореферата диссертации:**

Semenov, V. I. 2003. *Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor* [Mathematical modeling of the plasma in the compact torus]. D.Sc. Diss. Moscow. 272 p.

Kozhunova, O. S. 2009. *Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa* [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

**Описание ГОСТа:**

GOST 8.586.5-2005. 2007. *Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv* [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publs. 10 p.

**Описание патента:**

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. *Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya* [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.
11. При отправке файлов по электронной почте просим придерживаться следующих правил:
  - указывать в поле subject (тема) название журнала и фамилию автора;
  - использовать attach (присоединение);
  - в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.
12. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

**Адрес редакции журнала «Информатика и её применения»:**  
Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН  
Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05  
e-mail: [rust@ipiran.ru](mailto:rust@ipiran.ru) (Сейфуль-Мулюков Рустем Бадриевич)  
<http://www.ipiran.ru/journal/issues/>

---

## Requirements for manuscripts submitted to Journal “Informatics and Applications”

---

Journal “Informatics and Applications” (Inform. Appl.) publishes theoretical, review, and discussion articles on the research and development in the field of informatics and its applications.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

The topics covered include the following areas:

- theoretical fundamentals of informatics;
  - mathematical methods for studying complex systems and processes;
  - information systems and networks;
  - information technologies; and
  - architecture and software of computational complexes and networks.
1. The Journal publishes original articles which have not been published before and are not intended for simultaneous publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

***Agreement on the transfer of rights to publish:***

*“We, the undersigned authors of the manuscript “. . .”, pass to the Founder and the Editorial Board of the Journal “Informatics and Applications” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.*

*Author(s) signature(s): (name(s), address(es), date).*

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If the authors receive their article for correction after reviewing, it does not mean that the article is approved for publication. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or  $\text{\LaTeX}$  (.tex) formats, and additionally in the .pdf format. All documents may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.
7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 20 pages of the specified format.

Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

All pages of the manuscript should be numbered.

The templates for the manuscript typesetting are presented on site: <http://www.ipiran.ru/journal/template.doc>.

8. The articles should enclose data both in **Russian and English:**

- title;
- author’s name and surname;
- affiliation — organization, its address with ZIP code, city, country, and official e-mail address;
- data on authors according to the format: (see site)

[http://www.ipiran.ru/journal/issues/2013\\_07\\_01/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp) and

[http://www.ipiran.ru/journal/issues/2013\\_07\\_01\\_eng/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp);

- abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae;
  - indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.  
Important! Keywords must not be sentences;
  - Acknowledgments.
9. References. Russian references have to be presented both in English translation and Latin transliteration (refer <http://www.translit.ru>, option BGN).  
Please take into account the following examples of Russian references appearance:
- Article in journal:**  
Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.
- Journal article in electronic format:**  
Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).
- Article from the continuing publication (collection of works, proceedings):**  
Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental’noe issledovanie prochnosti soedineniy “stal’–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.
- Conference proceedings:**  
Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma “Novye resursoberegayushchie tekhnologii nedropol’zovaniya i povysheniya neftegazootdachi” [6th Symposium (International) “New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact” Proceedings]*. Moscow. 267–272.
- Books and other monographs:**  
Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem [Operation of turbine generators with direct cooling]*. Moscow: Energy Publs. 352 p.
- Dissertation and Thesis:**  
Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.
- State standards and patents:**  
GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch’yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publs. 10 p.  
Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel’nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.
- References in Latin transcription are presented in the original language.  
References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.
10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.
11. Submissions of files by e-mail must include:
- the journal title and author’s name in the “Subject” field;
  - an article and additional materials have to be attached using the “attach” function;
  - an electronic version of the article should contain the file with the text and a separate file with figures.
12. “Informatics and Applications” journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

**Editorial Board address:**

FRC CSC RAS, 44, block 2, Vavilov Str., Moscow 119333, Russia  
Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05  
e-mail: [rust@ipiran.ru](mailto:rust@ipiran.ru) (to Prof. Rustem Seyful-Mulyukov)  
<http://www.ipiran.ru/english/journal.asp>