

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ
ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 10

Москва 2018

ОБЩИЙ РАЗДЕЛ

УДК 37.016: [004:33]

Е.В. Васильева

Развитие креативных способностей и компетенций профессионалов цифрового будущего

Описана разработка рекомендаций по применению подхода проектного мышления в подготовке и обучении в университетах будущих специалистов для цифровой экономики. Представлен обзор основных методов дизайн-мышления и даются рекомендации по внедрению подхода дизайн-мышления в учебные дисциплины на примере поэтапного описания авторского кейса. Подход дизайн-мышление (Design Thinking) направлен на развитие креативных способностей человека через эмпирические правила и опыт, эмоциональный интеллект и признание ценности мнений других людей. Описываются методики дизайн-мышления: инструменты визуализации (карта эмпатии, пользовательского пути, заинтересованных сторон), инструменты Customer Development, партизанской этнографии, POV-фокусировки, горячего прототипирования.

Приводятся основные инструменты, которые участники дизайн-семинаров используют для разработки инновационных идей, решения адаптивных проблем. Подробно описан кейс в постановке дизайн-мышления «Как улучшить впечатление слушателей при обучении на курсах повышения квалификации по профессиональной программе?». Даются комментарии по результатам решения кейса участниками дизайн-семинаров.

Ключевые слова: цифровая экономика, информационные технологии, цифровые компетенции, компетентностный подход, мягкие навыки, дизайн-мышление, ИТ-образование, инновации

ВВЕДЕНИЕ

Внедрение новых методов преподавания учебных дисциплин важно в условиях преобразования высшей школы России для поддержки реализации Программы «Цифровая экономика Российской Федерации». Внедрение в деятельность университетов творческих методологий, применение их для обсуждения рабочих вопросов и формирования стратегии развития, могут повысить удовлетворенность и вовлеченность в процесс сотрудников, повысить их ответственность за принимаемые решения [1]. К таким технологиям относятся новые производственные технологии и промышленный Интернет, анализ Больших Данных (*Big Data*), нейронные сети, блокчейн, искусственный интеллект (*AI*), компоненты робототехники и сенсорики, системы моделирования полного жизненного цикла продукта, Интернет вещей (*IoT*), дополненная виртуальная реальность (*AR/VR*), квантовые и облачные вычисления. Согласно плану мероприятий по направлению «Формирование исследовательских компетенций и научно-технологических заделов» [2] программы «Цифровая экономика РФ» формируются системы поддержки исследований, создаются центры компетенций, которые аккумулируют информацию о возможностях реализации проектов в приоритетных направлениях международного научно-технического сотрудничества в области цифровой экономики.

Основная причина поворота к тренду развития креативных способностей людей кроется в появлении новых возможностей и вызовов, которые связаны со взрывным характером развития информационно-коммуникационных технологий (ИКТ) и ориентирами на цифровую трансформацию деятельности. Для поддержки цифровых проектов важны компетенции, причем это не только технические знания и навыки в области ИКТ. Для решения проблем в условиях неопределенности, которую создает цифровое пространство, персоналу важны развитые мягкие навыки (*soft skills*) коммуникаций, командной работы, эмоционального интеллекта, критического мышления, креативности и готовности к инновациям.

В Программе «Цифровая экономика Российской Федерации» есть позиции, в которых отражена важность шагов по реализации стратегии обучения в течение всей жизни персональных траекторий развития кадров. К 2019 г. уже планируется сформировать список требований к описаниям Топ-100 компетенций, необходимых для реализации цифровых преобразований экономики [3].

Проблемы развития креативных способностей людей, возможности раскрытия эмпатии, обучения выстраивать мыслительные процессы и воспринимать чужие идеи как основу своих изучались многими учеными. Одним из основоположников философии проектного мышления или дизайн-мышления (*Design Thinking*) называют Герберта Саймона, Нобелевского лауреата, чьи идеи в настоящее время востребованы инженерами, системотехниками и программистами в проектах создания искусственного интеллекта, обработки информации, сложных систем. В 1969 г. Герберт Саймон впервые представил размышления о природе организованной сложности, важности развития мышления человека через эмпи-

рические правила, опыт, умение адаптироваться к условиям высокой неопределенности окружающей среды [4]. Примерно в это же время Генрих Альтшуллер, советский ученый и изобретатель, опубликовал первые работы, где представил к обсуждению Теорию решения изобретательских задач (ТРИЗ). В этой теории её методологию и сейчас активно используют инженерные лаборатории для «творческого решения задачи» [5], содержится алгоритм экспериментальной обработки новаторских идей, основанный на «психических процессах технического творчества» и выстраивании мыслительных процессов. Особенности развития креативных способностей человека посвящены исследования Г.П. Щедровицкого, Н.П. Бехтерева, Н.Ю. Хрящевой, М. Микалко, Э. де Боно [6–10] и др.

В 2004 г. Дэвид Келли, основатель дизайн-агентства *IDEO*, и Хассо Платтнер, сооснователь ИТ-компании *SAP*, представили подход дизайн-мышление, который вобрал в себя лучшие наработки в области развития креативных навыков человека, изучения клиентского поведения, генерации идей, визуализации [11, 12].

Согласно В. Папанеку [13], «дизайн – это сознательные и интуитивные усилия по созданию значимого порядка». Мышление – это познавательная деятельность человека. Таким образом, в определении дизайн-мышления соединяется наше восприятие творчества от термина «дизайн» и системный подход – от слова «мышление».

Для того чтобы заниматься творческим мышлением, необходимо преодолеть барьеры очевидного и традиционного, бросая вызов тому, что установлено. Нужно быть готовым преодолеть любые вызовы, в том числе ментальные, на пути создания нового и неустанно искать идеи, слова [14]. Дизайнеры в процессе работы создают множество различных прототипов, моделируя и оценивая образ нового продукта, находя через творчество и итерационность революционные идеи.

Именно так построен подход *Design Thinking* – это последовательность шагов познания проблемы от ее восприятия, анализа, поиска паттернов (закономерностей и связей) до визуализации нашего представления о решении. Неоднозначность и неопределенность – условия, когда дизайн-мышление необходимо. Именно тогда рациональное мышление может уступить нашим интуитивным способностям и воображению.

Подход помогает получить навыки адаптации к условиям высокой неопределенности окружающей среды. Мы развиваемся в опыте, принимая и участвуя в своих ошибках. А дизайн-мышление позволяет проиграть этот опыт, применив игровые механики и работу в кросс-функциональной команде.

Как помогает философия дизайн-мышления в организации коллективного творчества? Креативность в эпоху прогресса цифровых технологий и робототехники, которая создает все новые вызовы для человечества, – это, пожалуй, одна из главных черт, которая нужна сегодня, чтобы стать успешным, выдержать в гонке за лидерство в любой профессиональной деятельности. В перечне востребованных компетенций

к 2020 г., опубликованных на Всемирном экономическом Форуме в 2016 г. [15], на первом месте находятся: когнитивные способность и креативность, активное обучение, эмоциональный интеллект и умение общаться с другими, способности комплексно решать проблемы и критическое мышление [16].

Дизайн-мышление представляет собой интегрированный дизайнерский итерационный и исследовательский подход, который направлен на создание прорывных идей, учит структурировать информацию, «быть более гибкими в применении своих знаний» [17], видеть паттерны и связи между объектами в сложных системах. Ключевые ступени процесса – эмпатия, фокусировка (*Poin-of-Vew, POV*), генерация, выбор, прототипирование и тестирование – поддерживают процесс генерации идей через комплекс всевозможных инструментов и методик. Этот комплекс постоянно обновляют и дополняют дизайн-исследователи по всему миру.

Подход учит через командную работу, игровые механики, визуализацию выстраивать процессы поиска ответов и ценить мнение других. Этот подход необходим, когда нужно разобраться в противоречивых требованиях и ожиданиях клиентов (карта эмпатии, партизанская этнография, *POV*-вопрос). Или когда ваши конкуренты сделали прорыв, а вам нечем быстро ответить на их вызов (метод триад, матрица положительного и отрицательного клиентского опыта).

Эрик Рис называет дизайн-мышление одним из основных принципов подхода «Бережливый стартап» (*Lean Start-up*) [18]. Дизайн-мышление заставляет «выйти из офиса» [19, с. 10], помогает научиться сотрудничать с другими людьми и вдохновляться их идеями, открывать для себя новые точки зрения и неожиданные возможности, уметь направлять свои творческие способности на решение проблем.

Руководитель найдет в дизайн-мышлении основу для изменения корпоративной культуры организации. В компании дизайн-мышление можно применять при проведении совещаний, стратегических сессий, мозговых штурмов, в тех случаях, когда сотрудники из разных отделов безрезультатно проводят много совещаний, или когда они не могут разобраться в запутанных и постоянно меняющихся требованиях начальства. Как правило, эти инструменты полезны при решении адаптивных проблем, для которых не существуют заранее определенные и однозначно правильные решения, а единственный способ их преодолеть – перестроиться самому, изменить свой взгляд на жизнь, свои убеждения и предпочтения [20]. Дизайн-мышление помогает в этом случае максимально использовать потенциал и идеи персонала для развития бизнеса, вовлечь его в проблему и повышать ответственность за изменения.

Организациям будущего нужны новаторы, умеющие нестандартно мыслить и находить альтернативные решения проблем в условиях неопределенности, обладающие не узкой специализацией, а междисциплинарными знаниями и широкими компетенциями в различных отраслях экономики, уметь договариваться, работать в открытых командах, а главное – постоянно развиваться. Выработка стратегии на компетенции, гибкость, предикативность требует новые форматы

обучения: компетентностное обучение, обучение по-разному (*Learning in different ways*), которое предполагает формальное и неформальное образование [21], и новые форматы коммуникаций. С целью повышения эффективности совместной работы и создания благоприятной атмосферы внутри коллектива в периодических научных подразделениях организуются регулярные семинары с использованием инструментов дизайн-мышления [22].

УСЛОВИЯ ПРОВЕДЕНИЯ ДИЗАЙН-СЕМИНАРА

Продолжительность командной работы при выполнении заданий обычного кейса составляет от 2-х до 5-ти часов. Важным фактором является четкое следование запланированному графику. Ограничение во времени позволяет спровоцировать на творчество и запустить процесс генерации большего количества перспективных идей.

Проводя дизайн-семинары, мы предлагаем выполнить задания в команде [23]. Участники исследуют клиентский опыт взаимодействия с продуктом, услугой или процессом, разрабатывают бизнес-идеи, создают прототипы продуктов или услуг. Тем самым, через проектный опыт, они получают навыки эмпатии, развивают системное и креативное мышление, изучают инструменты поддержки творчества, структурирования информации, приобретают навыки участия в кросс-функциональной команде. Так выстраиваются коммуникации и сотрудничество, возникает готовность к инновациям, проактивность, развиваются креативность и социальный интеллект. Создаваемые идеи изучаются в пересечении трех пространств инноваций: идея должна быть осуществима в обозримом будущем с точки зрения технологии, желательна пользователем, рентабельна для бизнеса. Это и есть самое главное ограничение дизайн-мышления – пространство инноваций.

Кейс «Как мотивировать студента университета прийти на вечернюю пару?»

Шаг 1. Построение карты заинтересованных сторон для ключевых участников процесса. Выявить целевых потребителей разрабатываемой идеи можно, построив *карту заинтересованных сторон (Stakeholder Map)*. На ней фиксируются позиции всех, кто соприкасается с продуктом или услугой, или будет это делать в будущем.

Шаг 2. Построение карты эмпатии. Составление *карты эмпатии* поможет раскрыть потребности целевой аудитории. В экономике впечатлений (*experience economy*), в которую мы, по мнению цифровых визионеров, перешли после экономики знаний, и ожидаемой в ближайшем времени экономике воображения (*imagination economy*) [24] лидером станет тот, кто понимает своего потребителя, его мысли и желания, кто проявляет к нему эмпатию [25]. Эмпатия, которая является основой большинства методик дизайн-мышления, помогает научиться анализировать контекст, опыт потребителя, его эмоции и ощущения, исследовать экосистему с целью выявления скрытых запросов, что открывает новые возможности для бизнеса. Исследователи, описывая протрет

потребителя, сопоставляют свое мнение о нем с мнением коллег, дополняют данными, о которых раньше никогда не задумывались. Но в целом, это позволяет приблизить понимание о потребителе к реальности.

Карта эмпатии состоит из 4-х основных блоков. В классической версии информация и гипотезы о том, что из себя представляет потенциальный потребитель разрабатываемого решения, заносятся в разделы карты: «Думает и чувствует», «Слышит», «Видит», «Говорит и делает». Пример карты, которая была построена участниками дизайн-семинара при решении кейса: «Как мотивировать студента университета прийти на вечернюю пару?», представлен на рис. 1. Дизайн-семинар проводился для преподавателей Финансового университета при Правительстве РФ. В командах над заданиями кейсов работали профессора и доценты различных направлений подготовки студентов: «Управление персоналом», «Бизнес-информатика», «Корпоративные финансы» и пр. Кейсы содержали вопросы из образовательной сферы, проблемы которой были известны всем участникам. Однако обсуждение тех или иных вопросов заставило участников дизайн-семинаров взглянуть на проблемы с других точек зрения, поделиться своим опытом с коллегами, получить подтверждение своих гипотез о возможных решениях, как оказалось, общих проблем.

Шаг 3. Исследование проблемы. Гипотезы, сделанные на этапе предварительного кабинетного анализа и при построении различных карт, должны быть обязательно проверены через проведение *глубинного интервью* тех, кто непосредственно будет заинтересован в решении. Этот этап, который в концепции «Бережливый стартап» (*Lean Start-up*), сформулиро-

ванной и популяризированной Э. Рисом [18], отводится на *Customer Discovery* (первый шаг методологии клиентского развития С. Бланка (англ. – *Customer Development*) [19]).

Основная задача проведения глубинного интервью заключается в том, чтобы как можно лучше понять опыт потребителя продукта или услуги, уточнить детали его взаимодействия с продуктом или услугой, а в процессе анализа – выйти на уровень инсайта.

Оценить гипотезы и выбрать лучшую идею для дальнейшей разработки и тестирования можно с помощью различных инструментов. С помощью метода «Включенное наблюдение» можно узнать больше об анализируемом объекте, а также протестировать идею с участием пользователя. Например, метод «Мокасины» предполагает, что исследователь примеряет на себя роль будущего потребителя продукта, воплощая метафору «носить чужие мокасины». Этим методом воспользовался в декабре 2016 г. глава Сбербанка РФ Герман Греф, чтобы оценить удобство офисов Сбербанка для людей с ограниченными возможностями здоровья [26].

Для систематизации результатов предыдущего шага можно построить *Матрицу положительного и отрицательного клиентского опыта*, в которой по осям или квадрантам матрицы изучаются Топ-5 "Действий", совершаемых пользователями при взаимодействии с изучаемым продуктом или процессом, и Топ-5 "Действий", где существуют возможности для улучшения; Топ-5 "Потребностей пользователей" (для чего нужен в реальности продукт?) и Топ-5 "Неудовлетворенных потребностей использования" (что продукт делает плохо или не делает вовсе?).

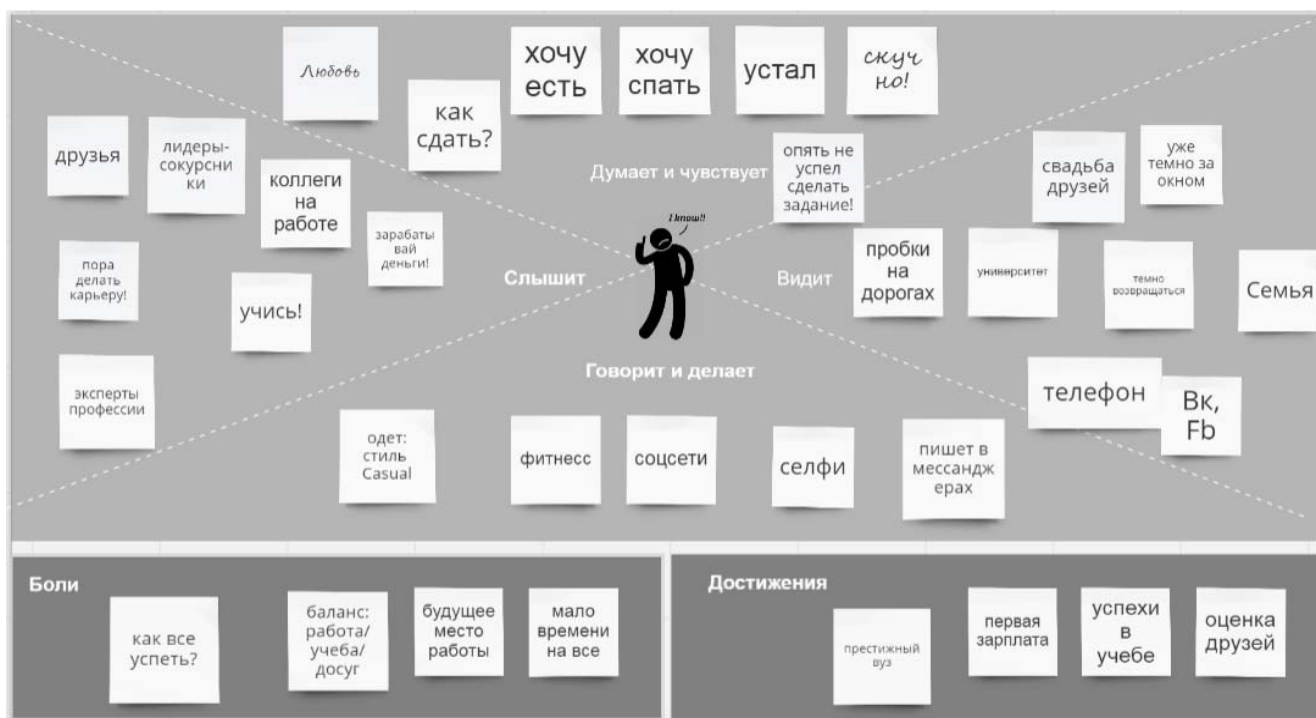


Рис. 1. Пример карты эмпатии: студент университета, каким его видят преподаватели (рисунок построен в сервисе *realtimeboard.com*).

Источник: результаты дизайн-семинара)

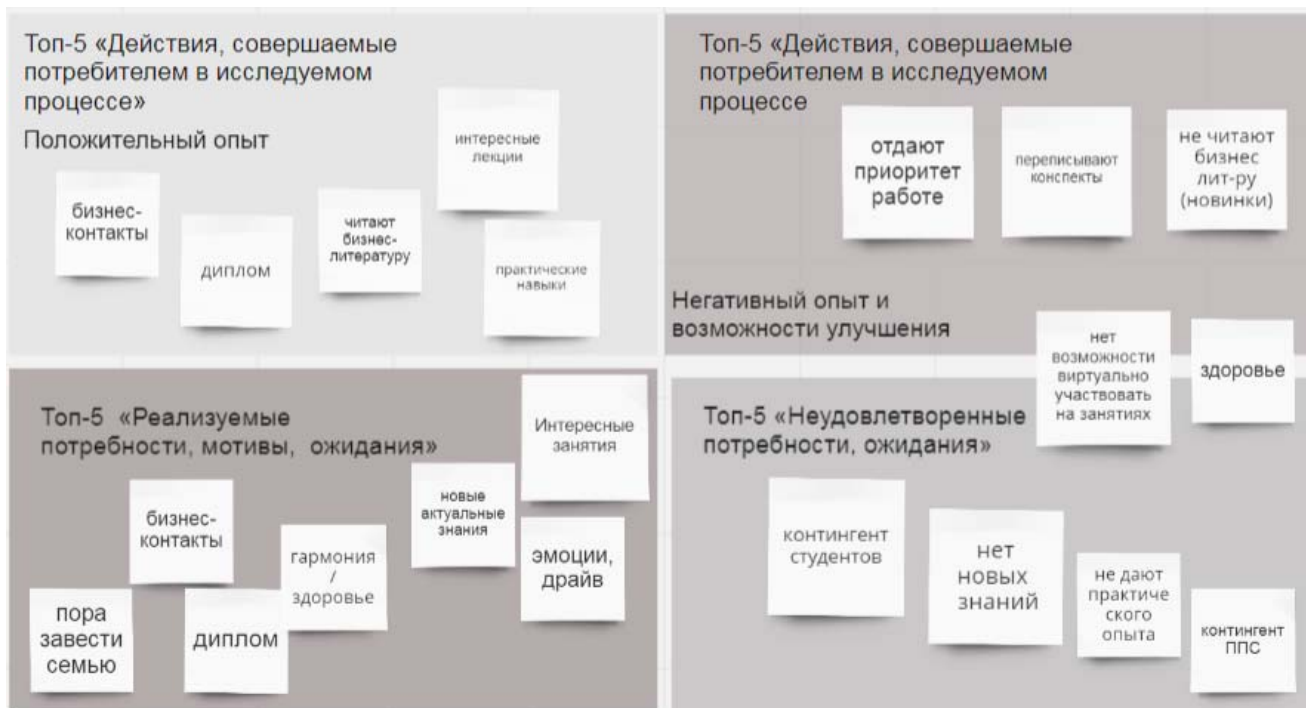


Рис. 2. Пример построения матрицы положительного и отрицательного клиентского опыта в кейсе: «Как мотивировать студента университета прийти на вечернюю пару?».

Источник: результаты дизайн-семинара

Матрица, представленная на рис. 2, создана в он-лайн сервисе *Realtimеboard*. Не все стикеры, которые участники командного обсуждения сохраняют на гранях матрицы, станут основой для дальнейшего исследования, но в дизайн-семинарах есть правило: все базовые гипотезы могут запустить прорывные идеи. Поэтому все идеи записываются и сохраняются, отбираются наиболее перспективные, но не упускается шанс вернуться и снова пересмотреть резервные.

Customer Journey Map (CJM) – это технология визуализации результатов анализа взаимодействия потребителя с продуктом, услугой или процессом по этапам жизненного цикла.

Шаг 4. Фокусировка на проблеме. Дизайн-мышление – это человеко-центрированный подход к разработке инноваций. Таким образом, правильная формулировка постановки проблемы: «Как мы можем помочь улучшить опыт, впечатление от взаимодействия человека с продуктом, процессом?»

В формулировке *POV*-вопроса (англ. *Point Of View*): «Как мы можем помочь?» применяется конструкция: **Пользователь – Потребность – Идея**. Например: «Как мы можем помочь [*пользователю*], решить [*проблему*] таким образом: [*идея*]». Вопрос служит хорошим стимулом для поиска идей, запуска мозгового штурма. В такой постановке может быть сформулировано изучение путей совершенствования процесса, продукта, но правильнее сфокусировать проблему вокруг потребностей человека.

Шаг 5. Генерация идей (Мозговой штурм). Метод *триад* направлен на анализ аналогичных продуктов или решений, которые используются в других предметных областях. Изучается положительный опыт этих решений, описываются различия или схожие черты, выявляются эмоции потребителей. Цель –

определить возможности переноса успешных сторон известных решений в плоскость новой проблемы.

При применении метода *триад* [23, 27] сначала пользователю предлагается выбрать три вида продукта, а затем – описать отличия одного от двух других. Мы предлагаем фиксировать ход анализа с помощью таблицы (табл. 1), в которой удобно выделить отличительные черты новой разработки от двух аналогичных решений. Строки могут быть заданы произвольно.

Пример сравнения положительных и отрицательных сторон опыта (впечатлений) студентов от посещения ими молодежных популярных рабочих и развлекательных площадок (марафон, хакатон, коворкинг или дискотека, ночной клуб, бал) и обучения в университете представлен в табл. 2.

Решение другого кейса «Как улучшить впечатление слушателей при обучении на курсах повышения квалификации по профессиональной программе?» дало интересный опыт интеграции карты пользовательского пути (*CJM*) с методом *триад*. В качестве аналога исследовательская группа выбрала практику организации туристского путешествия. Как создает удовольствие и приятные моменты туроператор? Встреча в аэропорту и трансфер до отеля вызывают у путешественника чувство защищенности в чужой стране и уже настраивают его на хорошие впечатления. Буклет с рекомендацией тура обеспечивает полезной информацией и создает уверенность, что отдых будет интересным. Комфортные условия и атмосфера укрепят лояльность клиента, который с удовольствием будет потом рекомендовать этот маршрут своим знакомым. Следующим этапом исследовательской командой создавался пользовательский путь слушателя программы повышения квалификации уже с учетом ключевых принципов организации маршрута путешественника.

Метод триад (шаблон для выполнения исследования)*

Характеристики	Описание аналогов		Отличия разрабатываемого продукта	
	Аналог 1	Аналог 2	Аналог 1	Аналог 2
Уникальность Минусы Плюсы Мотивы покупки Эмоции потребителя Каналы продвижения Ресурсы Эффекты				

*Источник: разработано автором

Таблица 2

Метод триад: опыт посещения молодыми людьми популярных площадок
(кейс: «Как мотивировать студента университета прийти на вечернюю пару?»)*

Аналог 1: хакатон, тусовка, коворкинг	Аналог 2: совместные баллы	Вечерние пары в вузе
<p>Плюсы:</p> <ul style="list-style-type: none"> • расширяет кругозор • умная тусовка • новые знакомства • анонимность <p>Минусы:</p> <ul style="list-style-type: none"> • не найти друзей • постоянно меняющаяся тусовка 	<ul style="list-style-type: none"> • обучение <i>pick-up</i> • взаимоотношения полов 	<ul style="list-style-type: none"> • любовь и семья • умная тусовка, расширение кругозора, знаниях из разных областей • совместные мероприятия и занятия между гуманитарными техническими факультетами • обучение – метод «смешанные группы» • кросс-дисциплинарные курсовые проекты

* Источник: результаты дизайн-семинара

Шаг 6. Тестирование. Техника *World café* («Мировое кафе» [23, 27, 28]) предполагает обсуждение выдвигаемых гипотез представителями всех команд, участвующих в дизайн-семинаре. Начиная работать над одной проблемой, через 20 минут каждая проектная группа переходит к проблеме другой проектной группе (меняются столами). Капитаны команды (хозяин стола), принимая новых участников за своим столом, обязаны быстро объяснить им суть проблемы, сформулированные ранее гипотезы, зафиксировать их сомнения и новые идеи. Смена столов проектной командой происходит в три этапа – каждые 20, 15, 10 минут соответственно.

Шаг 7. Оценка идей. Принципы выбора на основе метода *группировки идей* заключаются в рациональности и возможности реализации сгенерированной идеи. Идеи отбираются в три группы: лучшие, резервные и худшие. В качестве других критериев отбора можно выбрать: юмор, глубина проработки, количество ошибок (ответов); самые впечатляющие, рациональные, неожиданные, захватывающие, милые, рискованные; «что вы знаете?»; «что вы не знаете и не можете узнать?»; «что вы не знаете, но могли бы узнать?».

Идеи, которые были предложены участниками семинара для мотивирования студентов и преподавателей, чтобы они занимались вечером представлены в табл. 3.

Среди итогов обсуждения: организовать места для отдыха тем, кто приезжает в университет после работы задолго до начала занятий; сделать возможным дистанционное подключение к учебным классам (эффект присутствия) с рабочих мест, но с четким регламентом на возможности такого теле-посещения, созданием «буферов времени» в расписании; ввести меры по стимулированию преподавателей, разрабатывающих видео-материалы или проводящих видео-консультации; ввести должность ассистента из старшекурсников для поддержки деятельности ведущих профессоров; использовать больше активных форм и игровых механик, чтобы поддерживать в тоне мышление уставших за день студентов и преподавателей; создать досуговые центры (кафе, фитнес-залы, бассейн) для возможности неформального обмена знаниями; ввести проектное обучение; сократить количество аудиторных занятий. Некоторые из этих идей попали в резервные, поскольку команда считала их трудно реализуемыми в ближайшее время.

Метод группировки и наполнения*

Лучшие идеи	Резерв	Худшие идеи
<ul style="list-style-type: none"> • мастер-классы разных специалистов • активные методы обучения, игрофикация для поддержания драйва вечером • проектное обучение и совместные семинары для разных факультетов (гуманитариев и технарей) • план посещения и дистанционного подключения • место для отдыха до пары (релакс) • дистанционное подключение к занятиям • буфер времени в расписании для вечерних пар • сократить продолжительность пар вечером 	<ul style="list-style-type: none"> • автозаполнение журнала (технология распознавания лиц) • изменение регламента для дистанционного подключения • развитие досуговой инфраструктуры (фитнесс, бассейн) • решение проблемы с питанием через кофе-брейки на переменах • разгрузка ППС, который участвует в дистанционном обучении (ассистент из студентов старшекурсников для проведения семинаров) • создание регламента дистанционного обучения 	<ul style="list-style-type: none"> • места хранения личных вещей • дисциплины на развитие <i>soft skills</i>

* Источник: построено участниками дизайн-семинаров при решении кейса

Шаг 8. Прототипирование. *Активное прототипирование* помогает прочувствовать концепцию на ранних этапах ее создания. Процесс прототипирования предполагает создание эскиза, макета, разработанного в процессе генерации идеи решения сфокусированной проблемы пользователя. Это фаза проверки идеи на ее востребованность конечным потребителем. Прототип – это еще одна возможность доработать идею, «домыслить» ее руками (известное выражение Дэвида Кэлли [11, 12, 29]). Прототипирование может проводиться в несколько итераций, в результате которых отбрасываются не оправдавшие себя гипотезы, но появляются новые, которые тоже надо проверить.

Еще одна задача прототипирования – создание объекта или среды, которые могут быть протестированы пользователями. В этом случае преследуется цель – получение обратной связи от своей целевой аудитории.

Одно из главных правил: прототип должен быть разработан быстро и дешево («за несколько минут и пару центов» [29, с. 105]). Самый доступный и быстрый способ прототипирования – создание макета, эскиза из картона или бумаги. Однако в качестве прототипов могут выступать не только экспонаты из подручных средств, но и ролевая игра или сценарий. Можно применять такие методы, как: *Bodystorming*, чтобы изобразить сервис с помощью людей, сыграть сценку (пользователь – сервис); *Stop-Motion* – ролик для передачи атмосферы, условий, в которых будет работать пока несуществующий новый продукт; *Storybord* – нарисованный кадровый сценарий использования, презентация или стена с заметками и схемами; *Storytelling* – презентация идеи в виде рассказа и др.

Бизнес-оригами [27, с. 153] – это созданный из бумаги прототип, его можно использовать для описания взаимодействия объектов (в этом случае связи

обозначают стрелками, над стрелками могут быть надписаны сценарии или особенности отношений), отображения артефактов при изучении целевой аудитории, моделирования ситуации, презентации идеи. Для разработки прототипа можно применять метод игрового моделирования, который предполагает использование конструктора *Lego Serious Play* [27, с. 50]. Концепция этого метода была предложена профессорами швейцарского Международного института управленческого развития Йоханом Русом и Бартом в середине 1990-х гг. В основе этого метода лежат принципы игры, конструкционизма и воображения: большие участки человеческого мозга отвечают за моторику рук – во время моделирования и разговора мозговая активность повышается и улучшается циркуляция кислорода (70-80% нервных окончаний идут от мозга к рукам). Серьезная игра должна вызывать интерес к реализации разработанных идей. Так, в приведенном кейсе участниками семинара был разработан прототип класса будущего, где проводятся с теле-присутствием студентов. В этом классе есть учебные места, которые могут трансформироваться для проектной работы в один большой стол; есть множество экранов, которые транслируют занятие в любую точку подключения. Каждое из рабочих мест в классе оснащено зарядками для гаджетов, необходимыми в цифровой среде.

ВЫВОДЫ

Взрывной характер технологий, цифровизация процессов производства и услуг, общие цифровые преобразования организаций не только вносят изменения на рынке труда, влияя на сокращение спроса на работников, но и ставят под сомнение дальнейшее сохранение некоторых профессий. В докладе *Boston Consulting Group* озвучено, что в ближайшие 10 лет исчезнет до 50% профессий [30]. *Boston Consulting*

Group выявила 60 основных тенденций, стимулирующих приливную волну («A tidal wave of change» [31]), которая вскоре сделает работу почти неузнаваемой для сегодняшних лидеров бизнеса.

С точки зрения возможных изменений в спросе на таланты выделены такие две группы сил:

1) технологическая и цифровая производительность. Автоматизация, включая робототехнику и искусственный интеллект (AI), заменяет рабочие места. Большие Данные и расширенная аналитика открывают обширные клиентские, операционные и кадровые идеи. Расширенный доступ к информации и идеям краудсорсинг и экономика совместного использования (*sharing economy*) размывает границы традиционных организаций. Эти тенденции обуславливают развитие таких компетенций, как знание и навыки интеллектуального анализа данных (включая сбор, очистку, обработку, агрегирование), разработки приложений и дизайна пользовательского опыта;

2) изменения в способах создания стоимости бизнеса: умная простота или *Smart Simplicity* [32]; гибкость и инновации: внедрение гибких подходов и творческих методологий; новые стратегии клиентов. Новые, гибкие рабочие процессы требуют от сотрудников совершенно новых навыков, установок и знаний (*agile, scrum, kanban*, дизайн-мышление) таких, как креативность, умение работать в команде, быть готовым к сотрудничеству и инновациям. Они должны уметь самостоятельно принимать правильные решения.

Перемены в предложении талантов существенным образом отразятся на процессах управления организацией – это:

- сдвиги в распределении ресурсов: новая демографическая структура, старение населения; дисбаланс навыков, нехватка квалифицированных работников на одних рынках и избыток менее квалифицированных работников на других; изменение геополитической и экономической мощи. Востребованы специализированные цифровые возможности, опыт работы во временных и виртуальных командах, умение использовать преимущества облачных вычислений;

- изменение культуры и ценностей рабочей силы: разнообразие, инклюзивность; индивидуализм и предпринимательство; культура, основанная на общих ценностях. Специалист будущего должен обладать социально-экологической ответственностью, толерантностью и уважением, уметь строить межличностные коммуникации, быть предприимчивым и креативным, иметь стремление к карьерному росту и саморазвитию, способным работать с индивидуалами, понимать и воспринимать личные стремления и ценности других.

Эти глобальные тренды произведут революцию в работе организаций и заставят руководителей пересмотреть их стратегии, поскольку понадобятся новые способы управления, новые подходы к набору, развитию и привлечению кадров. Фактически, почти каждая организационная роль в конечном счете потребует использования современных технологий, а компании не останутся иного выбора, кроме как инвестировать в масштабные, продолжающиеся программы развития навыков. Профессионалам скорого будущего нужен расширенный набор компетенций, в

которых знания и умения в определенной сфере деятельности должны быть обогащены владением современными цифровыми технологиями обработки информации, знанием робототехники, навыками машинного обучения, облачных вычислений и других так называемых «сквозных» технологий. По словам Германа Грефа: «Сейчас гигантская проблема с кадрами... Экономика еще не развернулась полностью лицом к "цифре"... Все профессии будут дополняться через типе *digital*. Начиная от менеджеров и заканчивая самыми простыми специальностями» [33]. Поэтому так активно в настоящий момент внедряются проекты стратегического развития, программы цифровизации социальной сферы и др.

Ориентация на потребителя и реактивные, подчас неожиданные, изменения их потребностей, вкусов, предпочтений заставляет меняться бизнес, искать новые бизнес-моменты, что в терминологии *Gartner* означает постоянный поиск кратковременных возможностей, используемых динамически: «бизнес-момент может взяться ниоткуда, и всё же их становится всё больше» [34]. Еще недавно человеческий ресурс воспринимался как основной участник бизнес-процессов на предприятиях, как неотъемлемая часть организационного капитала компании. Однако роботы, боты, дроны и другие технологии заменяют человека, освобождая его от рутинного труда и предоставляя возможность больше заниматься творчеством. Будущие бизнес-задачи компании в условиях постоянных изменений будут требовать от её сотрудников умения работать в команде, эффективно использовать свои компетенции, быть сосредоточенными и успешными, нестандартно мыслить и находить оригинальные решения, использовать накопленные знания, что сделает применение различных техник дизайн-мышления неотъемлемой частью деятельности человека.

Нужны ли такие техники дизайн-мышления? Даже если забыть о некоторой несерьезности игровых моментов, то однозначно – да. Можно научить рационально мыслить, но навыки принятия решений в нестандартных ситуациях развиваются только исходя из собственного опыта. Лучше если этот опыт появляется (проигрывается) в игре. Современная экономика и перспективы ее развития диктуют совершенно иные требования к специалисту будущего. Он должен не только владеть технологиями, но и обладать творческими способностями, воображением, интуицией, гибким и образным мышлением, уметь быстро ориентироваться в смене обстановки и легко адаптироваться к постоянно меняющимся условиям. В лаборатории *IDEO* своих специалистов называют «Т-людьми», где горизонтальная линия от буквы «Т» – это ширина их знаний в различных сферах, а вертикальная – специализация [29, с. 14; 35].

Одна из фундаментальных задач высшего учебного заведения – предоставление студентам возможности стать ориентированными на знания, их важно не просто учить, а научить учиться *to learn how to learn* [14]. Для этого требуется высокая степень доверия и неконкурентная среда, в которой происходит обмен знаниями. Это требует глубоких изменений в образовательном процессе, расширении практики применения новых форм обучения и коммуникаций. Дизайн-

мышление может предложить тесно взаимосвязанные инструменты, которые в совокупности позволят высшему образованию реагировать на меняющиеся в условиях цифровой трансформации потребности бизнеса и общества.

СПИСОК ЛИТЕРАТУРЫ

1. Программа «Цифровая экономика Российской Федерации» от 28.07.2017 г. № 1632-р. – URL: <http://static.government.ru/media/files/9gFM4FHj4PsB7915v7yLVuPgu4bvR7M0.pdf> (дата обращения 25.04.2018).
2. План мероприятий по направлению "Формирование исследовательских компетенций и технологических заделов" программы "Цифровая экономика Российской Федерации". – URL: <http://static.government.ru/media/files/1P5evO23war1woLA0q8aJ2DtAqsydInS.pdf> (дата обращения 25.12.2017).
3. Поляк Ю.Е. О влиянии цифровизации экономики на рынок труда и подготовку кадров // Сб. трудов конференции «Преподавание информационных технологий в Российской Федерации – 2018». – URL: <http://it-education.ru/conf2018/thesis/2708/> (дата обращения 14.05.2018)
4. Саймон Г.А. Науки об искусственном. – М.: Мир, 1972. – 148 с.
5. Альшуллер Г.С., Шапиро Р.Б. О психологии изобретательского творчества // Вопросы психологии. – 1956. – № 6. – С. 37-49.
6. Путеводитель по методологии организации, руководства и управления. Хрестоматия по работам Г.П. Щедровицкого. – М.: Альпина Паблишер, 2012. – 197 с.
7. Бехтерева Н.П. Магия мозга и лабиринты жизни. – М.: АСТ; СПб, 2007. – 400 с.
8. Психогимнастика в тренинге / под ред. Н.Ю. Хрящевой. – СПб: Речь, 2001. – 250 с.
9. Michalko M. Cracking Creativity: The Secrets of Creative Genius. – Berkeley, CA: Ten Speed Press, 1998. – 352 p.
10. de Bono E. Lateral Thinking: a textbook of creativity. – London: Pelican Books, 1970. – 211 p.
11. Kelley T., Kelley D. Creative Confidence Unleashing the Creative Potential Within Us All. – NY: Crown Business, 2013. – 288 p.
12. Liedtka J., Ogilvie T. Designing for Growth: A Design Thinking Toolkit for Managers. – NY: Columbia University Press, 2011. – 256 p.
13. Папанек В. Дизайн для реального мира. – М.: Издатель Д. Аронов, 2004. – 253 с.
14. Competence-based Learning. A proposal for the assessment of generic competences / A.V. Sanchez, M.P. Ruiz, other. – Bilbao: University of Deusto, 2008. – 334 p.
15. Сайт ВЭФ. – URL: http://www.weforum.org/agenda/2016/01/11-experts-at-davos-on-the-future-of-work?utm_content=buffer552fe&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer (дата обращения 02/10/2017)
16. Отчет ВЭФ. – URL: <http://reports.weforum.org/future-of-jobs-2016/chapter-1-the-future-of-jobs-and-skills/> (дата обращения 02/10/2017)
17. Из интервью профессора Стэнфорда Ларри Лейфера: «Design Thinking — это талант использовать чужие идеи». – URL: <https://theoryandpractice.ru/posts/2213-intervyu-s-professorom-stenforda-larri-leiferom-design-thinking--eto-talant-ispolzovat-chuzhie-idei> (дата обращения 05.05.2018).
18. Ries E. The Lean Start-up: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses. – NY: Crown Business, 2013. – 328 p.
19. Blank S. The Four Steps to the Epiphany: Successful Strategies for Products that Win. – Wisconsin: Quad/Graphics, 2005. – 282 p.
20. Ertel C., Solomon L.K. Moments of Impact: How to Design Strategic Conversations That Accelerate Change. – NY: H. Simon & Schuster, 2014. – 272 p.
21. Human Resources and Security. – URL: https://ec.europa.eu/info/departments/human-resources-and-security_en (дата обращения 01.05.2018)
22. Управление изменениями: новые форматы обучения // Доклад HRD КРОК П. Хабаровой / SAP Forum 2018 Москва. – URL: <http://sap.ru/15248221810000038166/11&offset=3> (дата обращения 27.04.2018).
23. Васильева Е.В. Дизайн-мышление: немного о подходе и много об инструментах развития креативного мышления, изучения клиентских запросов и создания идей. – М.: РУСАЙНС, 2018. – 204 с.
24. Bidshahri R. These Are the Most Exciting Industries and Jobs of the Future. – URL: <https://singularityhub.com/2018/01/29/these-are-the-most-exciting-industries-and-jobs-of-the-future/> (accessed: 29.01.2018).
25. Nussbaum B. The Empathy Economy, Business Week. – URL: www.businessweek.com/bwdaily/dnflash/mar2005/nf2005037_4086.htm/ (access date: 16.12.2009).
26. РИА Новости. – URL: <https://ria.ru/economy/20161202/1482749707.html> (дата обращения 02.12.2016)
27. Мартин Б., Ханнингтон Б. Универсальные методы дизайна. – СПб: Питер, 2014. – 208 с.
28. Крогерус М., Чеппелер Р. Книга решений. 50 моделей стратегического мышления. – М.: ЗАО «Олимп – Бизнес», 2012.
29. Браун Т. Дизайн-мышление: от разработки новых продуктов до проектирования бизнес-моделей. – М.: Изд-во Манн, Иванов и Фербер, 2012. – 256 с.
30. The Global Workforce Crisis: \$10 Trillion at Risk, BCG report, July 2014. – URL: https://www.bcgperspectives.com/content/articles/management_two_speed_economy_public_sector_global_workforce_crisis/ (дата обращения 05.05.2018).
31. Bhalla V., Dyrchs S., Strack R. Twelve forces that will radically change how organizations work. – URL: <https://www.bcg.com/publications/2017/people-organization-strategy-twelve-forces-radically>

- change-organizations-work.aspx?linkId=35939610 (дата обращения: 27.03.2017).
32. Morieux Y. Smart Rules: Six Ways to Get People to Solve Problems Without You // Harvard Business Review, September 2011. – URL: <https://hbr.org/2011/09/smart-rules-six-ways-to-get-people-to-solve-problems-without-you/> (accessed: 29.01.2018).
33. Из выступления Г.О. Грефа на XI съезде организации «Опора России». – URL: <http://www.vestifinance.ru/articles/94114> (дата обращения 22.11.2017).
34. Видение будущего цифрового мира на конференции Gartner Symposium/Itxpo – 2013. – URL: <http://www.crn.ru/news/detail.php?ID=84923> (дата обращения 10.04.2013).
35. Из интервью Бёртон Раст, арт-директор из дизайн-агентства IDEO. – URL: <http://www.lookatme.ru/mag/people/experience/217099-ideo-thoughts> (дата обращения 05/10/2017).

Материал поступил в редакцию 01.06.18.

Сведения об авторе

ВАСИЛЬЕВА Елена Викторовна – доктор экономических наук, доцент, профессор кафедры «Бизнес-информатика», Финансового университета при Правительстве Российской Федерации, Москва
e-mail: evvasileva@fa.ru

Интеграция информационных ресурсов с данными по свойствам веществ и материалов. Практическая реализация и доступные средства

Проведен анализ новых методов интеграции научных данных о свойствах веществ и материалов с использованием современных информационных технологий, для чего рассмотрены возможности систем, созданных в рамках Semantic Web и Big Data. Оба подхода, несмотря на различия решаемых с их помощью задач, ориентированы на работу с большими массивами распределенных в сети и разнородных по структуре данных.

Проведено исследование ряда платформ интеграции данных, разработанных в последние годы, включая систему, предложенную авторами. Сделан вывод об отсутствии в настоящее время систем интеграции, полностью удовлетворяющих характерным особенностям научных данных, хотя ряд программных элементов и технических решений достойны внимания и могут составить основу готового к реализации проекта интеграции научных данных о свойствах веществ и материалов.

Ключевые слова: базы данных, онтология, интеграция данных, свойства веществ и материалов

ВВЕДЕНИЕ

Проведение исследований и разработка перспективных технологий и материалов связаны с генерацией и распространением новых знаний о свойствах веществ и материалов: теплофизических, механических, электронных и др. Научно-техническому прогрессу всегда сопутствовало развитие средств и методов сбора, хранения, систематизации и обмена научными данными и знаниями. И если ранее эти процессы были напрямую связаны с особенностями печатного производства, то современные технологии хранения и переработки научной информации, в основном, определяются тенденциями развития программно-технической инфраструктуры.

В последнее десятилетие обозначились тенденции активного использования новых информационных технологий *Semantic Web* [1] и *Big Data* [2] в области работы с научными данными [3–6]. Эти технологии призваны решить фундаментальные проблемы, обусловленные безграничным многообразием ресурсов (базы данных, веб-страницы и т.д.), форматов и моделей данных (реляционная, объектно-ориентированная и др.), а также их структурной и семантической неоднородностью. Их возможности применительно к интеграции, т.е. единообразному представлению данных из множества источников о свойствах вещества, в частности о теплофизических свойствах, были рассмотрены нами в работах [7–10].

Технология *Semantic Web* обеспечивает интеграцию за счет записи данных в едином формате *RDF* (*Resource Definition Framework*), состоящем из пригодных для машинной обработки утверждений в виде **триплета**, имеющего формальный вид «**субъект – предикат – объект**». В результате определенных процедур вся совокупность данных и знаний, содержащихся в источнике данных, оказывается представленной в сети в виде набора связанных триплетов. При этом *RDF* обеспечивает построение формальной модели данных, не касаясь семантики, т.е. смысла отдельных понятий, и отсылая за его интерпретацией к имеющимся в сети онтологиям и словарям.

Что касается технологии *Big Data*, с которой связывают ожидание революционных трансформаций во многих сферах общественной жизни [2], то её применимость к интеграции научных данных определяется тем, что изначально эта технология была ориентирована на работу не просто с «большими», но и качественно разнородными ресурсами, начиная от традиционных БД и до неструктурированных документов в виде текстов, изображений, медиа-файлов и т.п. При этом инструментарий *Big Data* игнорирует тонкие детали, связанные с терминологией и понятийным аппаратом, т.е. семантикой или смысловым содержанием данных. Для целей интеграции в новый инструментарий, который сам по себе решает лишь ограниченную задачу (сводит воедино разнотипные

источники и форматы данных), требуется внедрить онтологический подход, ранее успешно применявшийся в комбинации с обычными базами данных. Именно он способен обеспечить необходимое внимание к точности определений, аксиоматике, соответствию данных накопленным знаниям о закономерностях и т.п.

Таким образом, в обеих технологиях (*Semantic Web* и *Big Data*) только онтологии и контролируемые словари могут обеспечить семантическую интеграцию данных, придавая им унифицированный и принятый в научном сообществе смысл. Богатый потенциал, присущий онтологиям в решении проблем интеграции научных данных, привлек внимание исследователей, предложивших различные методы, применимые в науках о Земле, медико-биологических науках, химии и других разделах естествознания [6–11]. На сегодняшний день многие из этих методов вышли на стадию практической реализации с наличием общедоступных программных средств и детальной проработкой сценариев использования онтологий.

В настоящей статье рассмотрено несколько реализаций онтологического подхода к научным данным, «встроенного» в открытые технологические платформы для активного использования современных информационных технологий. При этом достоинства и недостатки каждой из версий будут рассмотрены с определенной точки зрения – насколько удачно принципы и конкретные решения соответствуют специфике данных о свойствах вещества и требованиям научного сообщества.

ОСОБЕННОСТИ ДАННЫХ О СВОЙСТВАХ ВЕЩЕСТВ И МАТЕРИАЛОВ

Научные данные, характеризующие различные свойства вещества, прежде всего теплофизические, физико-химические, механические и др., представлены в мировых БД для широчайшего круга веществ: чистых и растворов, органических и неорганических, конструкционных материалов и наноструктур [12, 13]. Для каждого из указанных классов существуют свои физические модели и связанные с ними словари понятий и логические структуры данных. Понятие структуры данных о свойствах в общем случае состоит из трех элементов: 1) способы идентификации объектов (атомы, молекулы, вещества, растворы, и т.п.); 2) состояние объекта (параметры состояния, фаза и/или фазовые границы, характеристики среды и т.п.); 3) физические характеристики объекта (энергетические, транспортные, оптические, механические свойства и т.п.), в том числе данные по неопределенности свойств. Структура данных включает также сведения о размерности, типах и форматах представления данных и т.п. Научная практика выработала значительный багаж систематизаций и классификаций в этой области.

Но сущностное развитие содержания научных данных приводит к изменчивости их логических структур, что связано с появлением новых знаний о свойствах веществ, созданием новых материалов и технологий, расширением понятий и классификаций [5, 7, 8]. При этом происходит увеличение объема

данных о свойствах, вызванное постоянно растущим спросом на данные по свойствам вещества со стороны множества отраслей науки и промышленности, включая энергетику, химию, металлургию и т.д. Развитие информационно-телекоммуникационной среды привело к широкому распространению БД по свойствам веществ и материалов в научной и образовательной сферах, что вызывает необходимость в проверке достоверности и согласованности данных, в их перманентном сравнении и обмене между различными источниками.

Основываясь на анализе, проведенном в работах [7–10], выделим основные особенности данных о свойствах с точки зрения:

- **структуры данных** – постоянная необходимость коррекции указанных выше трех элементов логической структуры для адекватного описания физических характеристик объектов исследования;
- **типа данных** – это, в первую очередь, числовые массивы данных (константы, таблицы, функции), во вторую очередь, данные в различной графической форме (спектрограммы, фотоматериал и т.п.) и, наконец, текстовые данные (термины, классификаторы, описания, библиография и т.п.).

Рассмотренные далее реализации современных информационных технологий (*Semantic Web* и *Big Data*) изучены на предмет их соответствия указанной специфике научных данных и возможности построения на их основе эффективно действующей платформы интеграции.

АВТОРСКИЙ ПРОЕКТ РАБОТЫ С ДАННЫМИ ПО ФИЗИЧЕСКИМ СВОЙСТВАМ ВЕЩЕСТВА

В работе [14] была предложена технология работы с данными, основанная на совместном применении онтологических моделей разного уровня и инструментария *Big Data*. Была представлена система для оперативной работы и интеграции разнородных по типу и изменяющихся по структуре данных. В качестве базовой единицы хранения данных рекомендовано использовать файлы в *JSON* (*JavaScript Object Notation*) формате [15]; оперирование и описание данных (метаданные), возложено на методы семантического моделирования и хранения. Основная идея проекта заключается в связывании атрибутов (ключей), а также значений (объектов) с концептами онтологий. Важно, что эта операция отражается как на содержании литеральных данных онтологической модели, так и в изменении наименования атрибутов/ключей наборов данных в *JSON*-формате. При этом возникают две связанные системы хранения данных:

- репозиторий онтологий и моделей метаданных для оперирования данными, хранимый в БД *графового* типа¹;
- локальные или распределенные хранилища файлов данных как в *JSON*-формате, так и других форматах (текстовых, графических и т.п.).

¹ Графовая БД реализует сетевую модель в виде графа, рассматривается как обобщение *RDF*-модели, используемой при кодировании онтологии (подробнее: Техопедия. – URL: www.techopedia.com/definition/30577).

Предлагаемое решение учитывает две важные особенности данных, отмеченные в предыдущем разделе. Использование онтологических моделей дает значительный простор для решения вопросов, связанных с изменчивостью структуры данных, в отличие от изменения структуры традиционных БД, что позволяет достаточно безболезненно модернизировать текущую структуру данных. Связь с онтологией предоставляет также возможность организовать интеграцию совершенно различных типов данных приведением их атрибутов к единым концептам предметно-ориентированной (*domain*) онтологической модели. Применение файлов в *JSON*-формате позволяет организовать «выгрузку» данных с совершенно разной исходной структурой и практически из всех типов БД, не подключаясь к выбранному ресурсу. Его владелец сам принимает решение о допустимости экспорта в объеме экспортируемого набора данных, которые он считает нужным сделать публичными. Вся дальнейшая работа по управлению данными происходит с *JSON*-файлами, ничуть не затрагивая исходные ресурсы. Наконец использование одной из наиболее популярных платформ высокопроизводи-

тельных вычислений *Apache Spark* [16] для работы с большими массивами разнородных данных позволяет организовать весь процесс хранения, обработки и поиска данных среди *JSON*-файлов, в общем случае размещенных на разных серверах сети. В то же время, наличие веб-репозитория онтологий дает возможность проводить поиск на специальном языке запросов *SPARQL*², совершая навигацию по классам онтологии, то есть переходя на верхний и нижний уровни иерархии классов [6, 10]. Зафиксировав требуемое понятие, пользователь применяет возможности платформы для запроса к содержательной части хранилища (то есть к *JSON*-файлам) на традиционном для реляционных БД языке *SQL* (*Structured Query Language*)³.

Все функциональное управление комплексом данных базируется на свободно-распространяемом программном обеспечении – пакетах программ и программных языках (*JAVA, Python, Scala, PHP*) в веб-среде. В настоящее время проходит процесс проверки выбранных решений, разработка и апробация части модулей будущей системы. Схема управления и работы платформы [14] представлена на рис. 1.



Рис. 1. Схема веб-среды интеграции разнородных и распределенных данных по свойствам веществ

² SPARQL (рекурсивный акроним от англ. SPARQL Protocol and RDF Query Language) – язык запросов к данным, представленным по модели, предусмотренной технологией Semantic Web (подробнее: Техопедия. – URL: www.techopedia.com/definition/4184/)

³ Язык структурированных запросов (подробнее: Техопедия. – URL: www.techopedia.com/definition/1245)

Отметим один из ключевых моментов предложенной технологии – стандартизацию хранения данных, полученных из множества источников, различающихся структурой, форматом и семантикой – в структурированном текстовом документе, записанном в *JSON*-формате, принятом в современных информационных технологиях для обмена данными и метаданными. Преимущество текстового документа – возможность простого чтения и редактирования, доступность для восприятия человеком, удобная форма хранения и обмена произвольной структурированной информацией. Адекватность предложенного стандарта поставленной цели (интеграции данных о свойствах) подтверждается прежним опытом использования структурированного текста (вместо традиционных БД), доказавшего свою эффективность как средства хранения теплофизических данных в проекте *ThermoML* [17] и данных по свойствам конструкционных материалов в проекте *MatML* [18]. Здесь в качестве основной единицы хранения предложен текстовый документ, записанный в *JSON*-формате, который менее перегружен деталями и упрощает представление структуры данных, сокращая их размер и время обработки документа.

АНАЛИЗ СУЩЕСТВУЮЩИХ РЕШЕНИЙ ПРОБЛЕМЫ ИНТЕГРАЦИИ ДАННЫХ

Рассмотрим три технологические платформы [19–21] с открытым программным кодом, способные решать задачи интеграции разнородных данных и базирующиеся на применении онтологий (*Semantic Web*), а также с возможным (хотя не обязательным) обращением к методам «больших данных» (*Big Data*). В работе [22] изложены идеи совместного применения методов *Semantic Web* и *Big Data* в области управления предприятием и бизнес процессами, но без конкретных программных решений. Кроме того, эта публикация сильно ориентирована на интеграцию данных, представленных в коммерческих программных продуктах, что сильно отличается от поставленной здесь проблемы. Ограничимся анализом продуктов, которые отвечают признакам:

- разработаны как интегрированные платформы, доступные в сети Интернет, предоставляют необходимые возможности управления данными и знаниями;
- имеют многоуровневую архитектуру, которая объединяет оригинальные программы с дополнительными внешними компонентами, причем используется программное обеспечение с открытым исходным кодом или ограниченные варианты коммерческого программного обеспечения со свободным доступом;
- доступны для скачивания и свободного применения.

Платформа *OSF (Open Semantic Framework)* [19]. Авторы *OSF* позиционируют ее как интегрированный программный комплекс с использованием семантических технологий для управления знаниями. *OSF* имеет множество приложений для интеграции корпоративной информации в сети Интернет, связанной с поддержкой социальных функций. Система обеспечивает:

- интеграцию данных для всех типов контента;
- управление знаниями;

- семантический поиск по всей сети Интернет;
- распределенный и дифференцированный доступ к данным;
- публикацию и управление информацией, предоставляемой пользователем.

Авторы [19] заявляют, что *OSF* может интегрировать и управлять всеми типами контента – неструктурированными документами, полуструктурированными файлами, электронными таблицами и структурированными базами данных – с использованием лучших в ИТ средств индексирования и управления данными. Весь внешний контент преобразуется в каноническую модель данных *RDF (Resource Definition Framework)* с предоставлением общих инструментов для разметки и управления всем контентом. Система реализована как пакет программ на *JAVA* и *PHP*, в качестве хранилищ *RDF* используется графовая БД, включенная в платформу [23]. Часть системных параметров и, собственно, данные содержатся в БД *MySQL*⁴. *OSF* предполагает встроенным редактором онтологий. Фрагмент работы данной опции представлен на рис. 2.

Анализ применения *OSF* показал наличие практических средств работы с онтологическими моделями, обеспечивающими ввод, модернизацию и публикацию данных. К недостаткам, с точки зрения работы с научными данными, следует отнести принцип хранения данных. В системе предлагается два варианта: хранилище *RDF* на основе *OpenLink Virtuoso* [23] и реляционная БД *MySQL*. Первый из них не является приемлемым вариантом для оперирования с численными типами данных, хотя дает неограниченные удобства при вариации структуры метаданных. Вторым вариантом показал, что хранение иерархически выстроенных таксономий реализовано в *MySQL* крайне примитивным образом, а именно с использованием метода *parent id*. Это означает, что вся иерархия атрибутов описывается в одной таблице с применением дополнительного атрибута *parent_id* – уникального номера записи «родителя» в этой же таблице. В этом случае организация хранения сложных ветвящихся структур в реляционных БД приводит к росту числа таблиц и снижению производительности при выполнении запросов, требуя выполнения операторов соединения таблиц.

Следует отметить, что эта проблема в последние годы вызвала необходимость ввода в ряде реляционных БД нового типа данных *json* – аналогичного *JSON*-формату (в *MySQL*, например, начиная с версии 5.7). Однако его полная реализация пока далека от идеала.

Система интеграции данных *MOMIS (Mediator environment for Multiple Information Source)* [20]. Как заявляют авторы, система *MOMIS* является структурой, способной интегрировать в полуавтоматическом режиме данные, поступающие из распределенных и гетерогенных источников (то есть, структурированные и полуструктурированные), чтобы выявить новую информацию из явно несвязанных существующих данных.

⁴ Свободно-распространяемая реляционная БД, успешно конкурирующая с коммерческими продуктами (подробнее: Техопедия. – URL: www.techopedia.com/definition/3498).

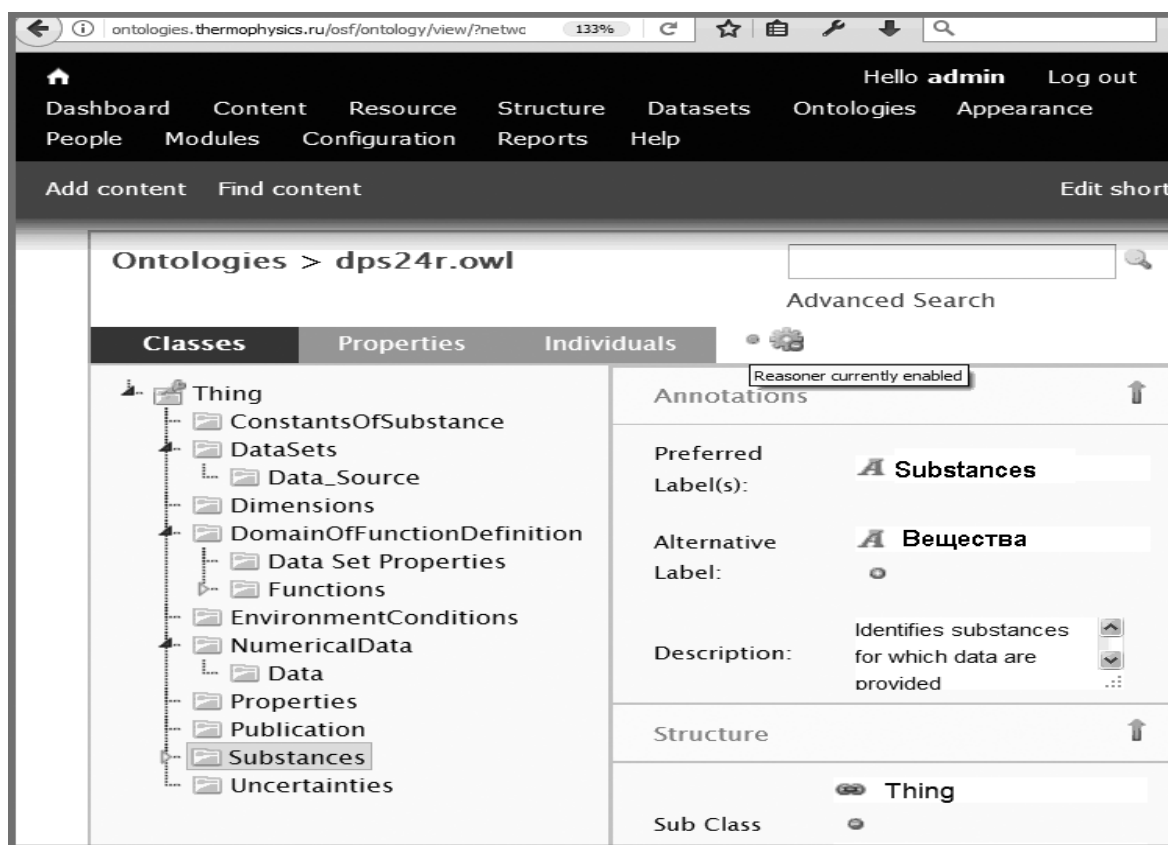


Рис. 2. Демонстрация встроенного редактора онтологий OSF

Платформа опирается на технологию ODMG (*Object Data Management Group*), задача которой – разработка принципов хранения объектов в БД, то есть поддержка объектно-ориентированных БД. В основе технологии лежит применение объектно-ориентированного языка ODL3, который служит для анализа и извлечения информации о характеристиках интегрируемых наборов данных. При этом выполняется обнаружение связей между схемами источников информации, используется семантика источников данных, методы кластеризации и выводы дескриптивной логики.

Целью системы MOMIS является минимизация затрат при выполнении процесса интегрирования. Интеграция данных в системе – это проблема объединения данных, находящихся в разных локальных источниках, и предоставления пользователю единой информации из этих данных. Такой унифицированный вид информации называется в MOMIS глобальной схемой – *Global Schema (GS)*. Проект интеграции предполагает набор локальных источников и набор глобальных схем, а GS – это набор глобальных классов и семантических отношений между ними. Таким образом, MOMIS является классической архитектурой «обертка/посредник» (*wrapper/mediator*): локальные источники данных содержат реальные данные, в то время как глобальная GS предоставляет согласованное, интегрированное, доступное только для чтения представление базовых источников (ме-

таданные). Схема процесса интеграции данных MOMIS представлена на рис. 3.

Вся информация, связанная с глобальной схемой GS, хранится в файлах формата XML. Рабочая область проекта интеграции – это каталог, в котором хранятся все папки проекта. Поддерживаемые типы БД: MySQL, Microsoft SQL Server, Oracle, DB2, PostgreSQL, а также файлы в форматах Microsoft Excel, CSV и источники данных, доступные через веб-сервис.

В качестве базы семантического аннотирования исходных данных служит система лексических словарей WordNet [24]. Подробнее с принципами работы системы MOMIS можно ознакомиться в работах [25, 26].

Система MOMIS не предлагает автономной системы хранения данных и является «чистой» системой интеграции. Ее непосредственное применение к задаче интеграции данных по свойствам, как и других научных данных, сильно ограничены опорой на семантику, предлагаемую WordNet [24]. Лексическая БД позволяет выявить лишь семантические отношения между конструктами языка со своими филологическими особенностями, но не дефиниции и строгие логические связи между научными понятиями. Кроме того, опора на WordNet автоматически исключает работу с многоязычным контентом.

Ограничивающим фактором в работе с множеством ресурсов является и необходимость постоянного доступа и привязки к источникам данных, особенно к различным веб-ресурсам.

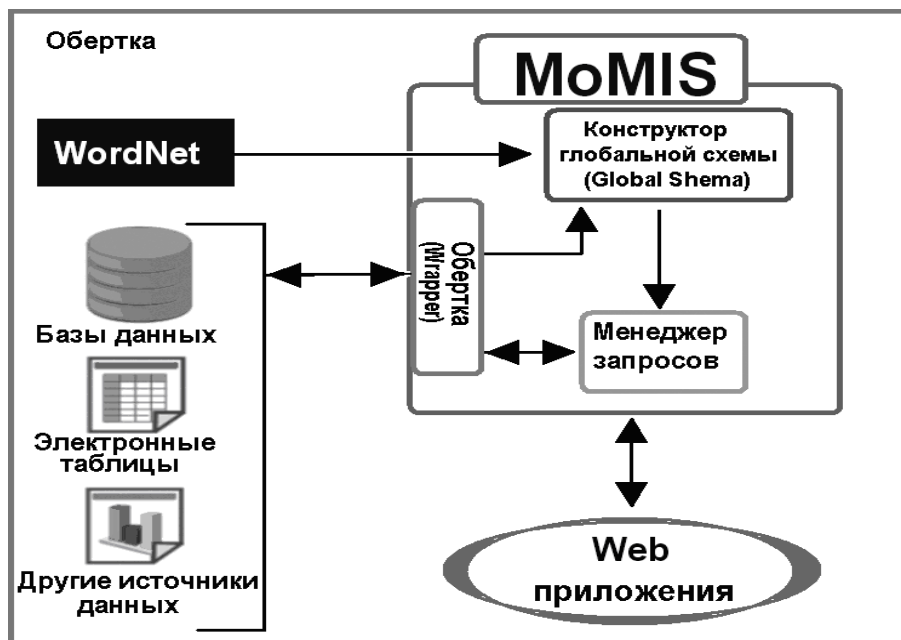


Рис. 3. Схема процесса интеграции данных MOMIS

Инструментарий интеграции данных Karma [21].

Авторы платформы *Karma* называют его инструментом интеграции информации [27, 28]. Система *Karma* позволяет пользователям быстро и легко интегрировать данные из различных источников данных, включая БД, электронные таблицы, текстовые файлы с разделителями, файлы в форматах *XML*, *JSON* и *Web-API (Application Programming Interface)*⁵. Пользователи интегрируют информацию, моделируя ее в соответствии с выбранной онтологией, используя графический пользовательский интерфейс, который автоматизирует большую часть процесса. *Karma* «учится» распознавать отображение данных в классы онтологии, а затем использует онтологию, чтобы предложить модель, которая связывает эти классы. Затем пользователи могут с помощью интерфейса системы провести редактирование автоматически сгенерированной модели. В ходе этого процесса пользователи имеют возможность преобразовывать как сами данные, так и изменять их структуру (вводить новые столбцы, новые классы, связывая их с классами и свойствами из выбранной онтологии и т.п.). После построения модели пользователи могут опубликовать интегрированные данные как *RDF* или сохранить их в собственной БД.

Авторами *Karma* была заявлена возможность в будущем выгрузки данных в файлы *JSON*-формата, что предполагает применение файлового хранилища или БД, пригодных для работы с файлами *JSON*-формата, аналогично популярной в технологии больших данных *MongoDB* [29].

Важной особенностью *Karma*, помимо использования технологий *Semantic Web*, является применение методов *Big Data*, таких как:

- оперирование большими объемами данных с последовательной «подгрузкой» данных по мере обработки данных в пакетном режиме;
- применение алгоритмов распознавания и обучения для подготовки и «очистки» загруженных данных;
- возможность загрузки и оперирования крайне разнородными по структуре и формату данными (например, иерархически структурированными массивами) в собственную реляционную БД с так называемыми «вложенными таблицами» [30].

Karma решает задачу стандартизации формата, сводя все форматы данных к *вложенной* реляционной модели. В этой модели данные представлены в таблицах, где ячейки могут содержать как скалярные значения (текстовые строки, числа и т.д.), так и вложенные таблицы. В качестве хранилища *RDF* используется графовая БД *Sesam RDF Database*.

На рис. 4 представлен фрагмент работы *Karma* с данными из теплофизической БД ТЕРМАЛЬ⁶ – документальной БД, содержащей ссылки на публикации по свойствам чистых веществ [31], и онтологией по теплофизическим свойствам веществ [9].

Апробация работы с наборами данных из БД ТЕРМАЛЬ и онтологией показала хорошие результаты. Система позволяет организовать процедуру связывания концептов выбранной онтологии с атрибутами/ключами тестового набора данных.

Общая концепция предполагает в заключение всей процедуры получение связанного онтологией массива данных: интегрированные данные в виде *RDF*. В дальнейшем в рамках технологии *Semantic Web* можно организовывать *SPARQL* запросы для выборки данных.

⁵ Подробнее о функциях API: Техопедия. – URL: www.techopedia.com/definition/24407

⁶ Трахтенгерц М.С. Свидетельство о государственной регистрации базы данных ТЕРМАЛЬ №2009620063. Зарегистрировано в Реестре баз данных РОСПАТЕНТА 28 января 2009 г.

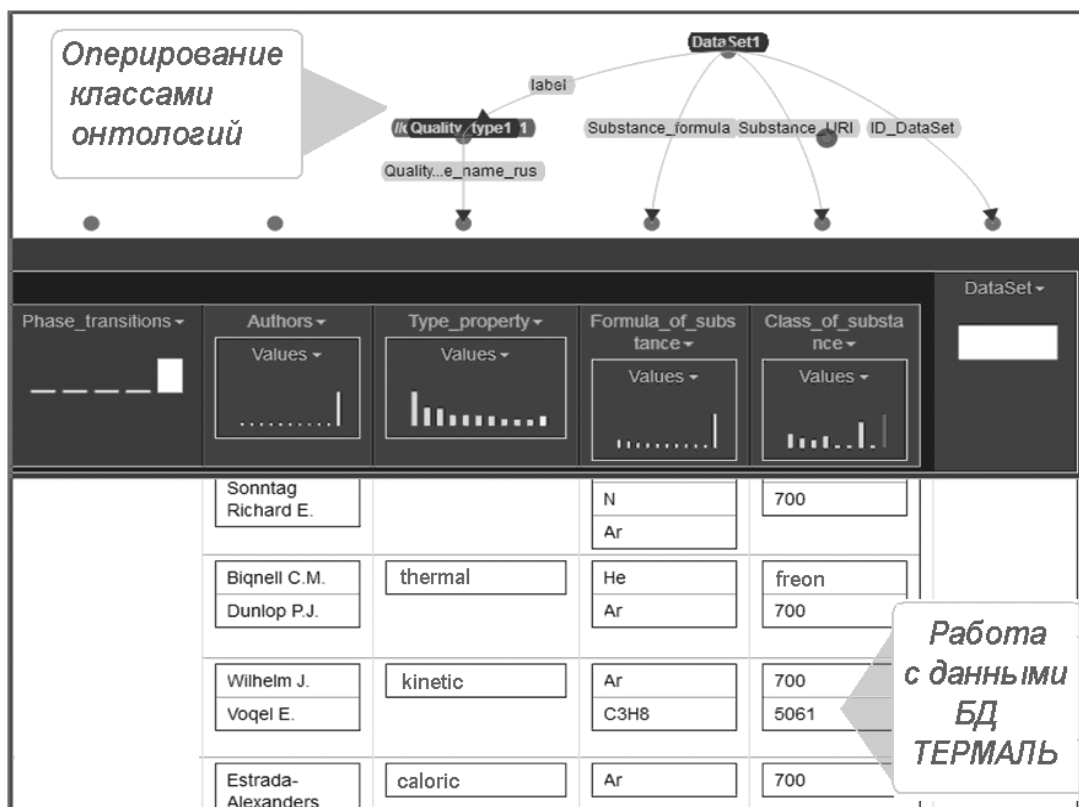


Рис.4. Демонстрация работы *Karma* с данными из БД ТЕРМАЛЬ [31] и онтологий по теплофизическим свойствам веществ

В настоящее время существует скептическое отношение к производительности графовых баз данных (хранилищ *RDF*) на значительном числе записей. Не случайно в работе [28] авторы *Karma* упоминают как одну из нерешенных проблем скорость отклика системы на запросы, предполагая заняться ее решением.

В качестве второй проблемы, которая должна быть решена в будущем, авторы *Karma* справедливо рассматривают необходимость связывания с концептами предметной онтологии значений данных (а не атрибутов/ключей).

ЗАКЛЮЧЕНИЕ

Проведенный анализ технологических платформ использования онтологических моделей и методов работы с большими данными для задач интеграции данных показывает, что существуют различные подходы к решению данной проблемы. Можно отметить, что все подходы не дают полного решения задач интеграции теплофизических данных, но могут служить примером и давать образцы для включения в инструментарий после определенной коррекции. В частности, программная реализация оперирования с онтологиями *OSF* может послужить основой репозитория онтологий, а механизмы и графический интерфейс организации процедуры подготовки и связывания данных и онтологий в *Karma* – рабочим инструментом подготовки данных.

Что касается нашей разработки [14], на сегодняшний день остается неясным, насколько большими объемами исходных данных в *JSON*-формате удастся оперировать в интерактивном режиме, поддерживаемом

в традиционных БД. Преодоление этой проблемы, то есть сохранение производительности систем с нарастанием объема данных, особенно значимо при работе с научными данными, поскольку объектами хранения являются, как правило, большие многомерные массивы.

СПИСОК ЛИТЕРАТУРЫ

1. Allemang D., Dean J. Semantic Web for the working ontologist: effective modeling in RDFS and OWL. 2-ed. – Burlington: Morgan Kaufmann Publishers is an imprint of Elsevier, 2011. – 354 p.
2. Chen Min, Mao Shiwen, Zhang Yin, Leung Victor C.M. Big Data. Related Technologies, Challenges, and Future Prospects. – Cham, Heidelberg, NY, Dordrecht, London : Springer, 2014. – P. 89. DOI:10.1007/978-3-319-06245-7.
3. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Серебряков В.А., Теймуразов К.Б. Технология научных публикаций в среде «Открытых связанных данных» // Научно-техническая информация. Сер. 1. – 2013. – № 12. – С. 1–8; Erkimbaev A.O., Zitserman V.Yu., Kobzev G.A., Serebrjakov V.A., Teymurazov K.B. Publishing Scientific Data as Linked Open Data // Scientific and Technical Information Processing. – 2013. – Vol. 40, № 4. – P. 253–263.
4. Bizer C. Interlinking scientific data on a global scale // Data Science Journal. – 2013. – Vol. 12. – P. GRD16-GRD112.

5. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А. Интенсивное использование цифровых данных в современном естествознании // Научно-техническая информация. Сер. 2. – 2017. – № 9. – С. 9–22; Erkimbaev A.O., Zitserman V.Yu., Kobzev G.A. The Intensive Use of Digital Data in Modern Life Sciences // Automatic Documentation and Mathematical Linguistics. – 2017. – Vol. 51, № 5. – P. 201–213.
6. Кузнецов О.П., Суховеров В.С., Шипилина Л.Б. Онтология как систематизация научных знаний: структура, семантика, задачи // Труды конф. «Технические и программные средства систем управления, контроля и измерения». – М., 2010. – С. 000762-000773.
7. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Сон Э.Е., Сотников А.Н. Интеграция баз данных по свойствам вещества. Подходы и технологии // Научно-техническая информация. Сер. 2. – 2012. – № 8. – С. 1–8; Erkimbaev A.O., Zhizhchenko A.B., Zitserman V.Yu., Kobzev G.A., Son E.E., Sotnikov A.N. Integration of Databases on Substance Properties: Approaches and Technologies//Automatic Documentation and Mathematical Linguistics. – 2012. – Vol. 46, № 4. – P. 170–176.
8. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Серебряков В.А., Шиолашвили Л.Н. Интеграция данных по свойствам веществ и материалов на основе онтологического моделирования предметной области // Журнал «Электронные библиотеки». – 2013. – Т. 16, № 6. – URL: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2013/part6/EZKSS>.
9. Серебряков В.А., Теймуразов К.Б., Хайруллин Р.И., Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Трахтенгерц М.С. Практическая реализация системы интеграции теплофизических данных на основе онтологической модели предметной области // Труды Четвертого Всероссийского симпозиума «Инфраструктура научных информационных ресурсов и систем» (6-8 октября 2014 г.), в 2-х т., т. 1. – СПб, 2014. – С. 87–111.
10. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А., Косинов А.В. Связывание онтологий с базами данных по свойствам веществ и материалов // Научно-техническая информация. Сер. 2. – 2015. – № 12. – С. 1–16.
11. Zhang X., Zhao C., Wang X. A survey on knowledge representation in materials science and engineering: An ontological perspective // Computers in Industry. – 2015. – Vol. 73. – P. 8–22.
12. Киселева Н.Н., Дударев В.А., Земсков В.С. Компьютерные информационные ресурсы неорганической химии и материаловедения // Успехи химии. – 2010. – Т. 79, № 2. – С. 162–188; Kiseleva N.N., Dudarev V.A., Zemskov V.S. Computer information resources in inorganic chemistry and materials science//Russian Chemical Reviews. – 2010. – Vol. 79, № 2. – P. 145–166.
13. Вассерман А.А., Мальчевский В.П. Банки данных и автоматизированные информационные системы по теплофизическим свойствам газов и жидкостей // Технические газы. – 2009. – № 5. – С. 59–66.
14. Erkimbaev A.O., Zitserman V.Yu., Kobzev G.A., Kosinov A.V. Standardization of Storage and Retrieval of Semi-structured Thermophysical Data in JSON-documents Associated with the Ontology // Сб. науч. трудов XIX Международной конференции DAMDID/RCDL'2017 «Аналитика и управление данными в областях с интенсивным использованием данных» (10–13 октября 2017 г.). – М.: МГУ, 2017. – С. 287–292.
15. The JSON Data Interchange Format. Standard ECMA-404. 1st edition/ October 2013. – URL: www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf.
16. Карау Х., Конвински П., Венделл П., Захария М. Изучаем Spark: молниеносный анализ данных / пер. с англ. – М.: ДМК Пресс, 2015. – 304 с.
17. Frenkel M., Chirico R.D., Diky V.V. et al. XML-based IUPAC Standard for experimental, predicted, and critically evaluated thermodynamic property data storage and capture ThermoML: IUPAC recommendations 2006 // Pure and Applied Chemistry. – 2006. – Vol. 78. – P. 541–612.
18. Kaufman J.G., Begley E.F. MatML. A Data Interchange Markup Language // Advanced Materials & Processes. – 2003. – November. – P. 35–36.
19. Open Semantic Framework (OSF). – URL: <http://opensemanticframework.org/>
20. MOMIS DataRIVER. – URL: <http://www.datariver.it/data-integration/momis/>
21. Karma: A Data Integration Tool. – URL: <http://uscisi-i2.github.io/karma/#pub>
22. Bastian E., Matthias J., Werner Q. Ontology-Based Big Data Management // Systems. – 2017. – Vol. 5(3). – P. 45. DOI:10.3390/systems5030045
23. OpenLink Virtuoso Official website. – URL: <https://virtuoso.openlinksw.com/>
24. WordNet. A Lexical Database for English. Princeton University. – URL: <https://wordnet.princeton.edu>
25. Bergamaschi S., Beneventano D., Corni A., Kazazi E., Orsini M., Po L., Sorrentino S. Open Source release of the MOMIS Data Integration System // Proc. of the Nineteenth Italian Symposium on Advanced Database Systems, SEBD (26–29 June, 2011) / eds. G. Mecca, S. Greco. – Maratea, Italy, 2011. – P. 175–186.
26. Bergamaschi S., Beneventano D., Guerra F., Orsini M. Data integration // Handbook of Conceptual Modeling / eds. D.W. Embley, B. Thalheim. – Berlin: Springer, 2011. – P. 443–478.
27. Gupta S., Szekely P., Knoblock C.A., Goel A., Taheriyani M., Muslea M. Karma: A System for Mapping Structured Sources into the Semantic Web // The Semantic Web: ESWC 2012 Satellite Events. ESWC 2012. Lecture Notes in Computer Science, Vol. 7540 / eds. E. Simperl et al. – Berlin, Heidelberg: Springer, 2015. – P. 41.
28. Knoblock C., Szekely P. Exploiting Semantics for Big Data Integration // Ai Magazine. – 2015–Vol. 36(1). – P. 25–38.
29. Hows D., Membrey P., Plugge E., Hawkins T. The Definitive Guide to MongoDB: A complete guide to dealing with Big Data using MongoDB. 3-ed. – NY: Apress, 2015. – 376 p.

30. Wang G., Yang S., Han Y. Mashroom: End-User Mashup Programming Using Nested Tables. In "Proceedings of the 18th International Conference on World Wide Web". – NY: Association for Computing Machinery, 2009. – P. 861–870.
31. Горгороки Е.А., Краевский С.Л., Трахтенгерц М.С., Швальб В.Г., Шпильрайн Э.Э., Якимович К.А. Автоматизированная информационно-поисковая система Теплофизического Центра ИВТАН // Обзоры по теплофизическим свойствам веществ. – 1977. – № 4. – С. 9–130.

Материал поступил в редакцию 29.05.18.

Сведения об авторах

ЕРКИМБАЕВ Адильбек Омирбекович – кандидат технических наук, зав. лабораторией баз данных Объединенного Института высоких температур РАН (ОИВТ РАН), Москва
e-mail: adilbek@ihed.ras.ru

ЗИЦЕРМАН Владимир Юрьевич – кандидат физико-математических наук, ведущий научный сотрудник лаборатории баз данных Объединенного Института высоких температур РАН (ОИВТ РАН), Москва
e-mail: vz1941@mail.ru

КОБЗЕВ Георгий Анатольевич – доктор физико-математических наук, советник Научно-исследовательского центра электрофизики и тепловых процессов Объединенного Института высоких температур РАН (ОИВТ РАН), Москва
e-mail: gkbz@mail.ru

КОСИНОВ Андрей Владимирович – инженер-программист Особого конструкторского бюро Объединенного Института высоких температур РАН (ОИВТ РАН), Москва
e-mail: kosinov@gmail.com

С.Н. Гриняев, Р.А. Злотин, А.И. Милушкин, Д.И. Правиков, И.А. Селионов,
А.Ю. Щербаков, Ю.Н. Щуко

К вопросу о создании универсального защищенного доверенного цифрового актива (токена)

Рассмотрена проблема разработки универсального защищенного доверенного цифрового актива нового поколения на основе универсальной структуры данных и новой парадигмы доверия и защищенности.

Ключевые слова: доверие, токен, блокчейн, код аутентификации, распределенное хранение данных, информационная безопасность, аукцион, шифрование, контроль целостности, смарт-контракт, средства разработки для токенов, цифровой финансовый актив, транзакция, оператор платформы

ВВЕДЕНИЕ

В настоящее время актуален вопрос «цифровой трансформации» российской экономики, формирования доверенной и корректной цифровой среды разработки и применения цифровых национальных технологий. Сегодня в этой области остро ощущается «давление» иностранных криптовалютных технологий, в том числе технологии оборота токенов, цифровых активов, технологии распределенного хранения данных, включая распределенные реестры (блокчейн). Кроме того, регистрируются активные попытки создания на базе технологий распределенного реестра и разного рода криптовалют реальных систем, обеспечивающих решение широкого спектра задач. Вместе с тем, как показал ряд исследований, несмотря на заявленные качества, реализованные системы не обладают свойствами информационной безопасности, что определяет их исследовательский характер.

Неразумное копирование западных технологий может иметь весьма пагубные последствия для технологической безопасности и независимости России, особенно в условиях создания «цифровой экономики», включая влияние на устойчивость российского бизнеса как на национальном уровне, так и на международной арене [1], поэтому необходимо создание технологической платформы для цифровой трансформации, которая удовлетворяла бы следующим требованиям:

- доверенность (универсальная открытая структура и открытый программный код) [2];
- национальная криптографическая локализация (использование национальных стандартов и готовность к получению соответствующих сертификатов у регуляторов);

- универсальность (возможность использования как в технологии цифровых активов, так и в работе с товарами и продукцией, а также для обеспечения государственного заказа и закупок).

Активное изучение технологий распределенного реестра отечественными специалистами привело к появлению собственных разработок в различных научно-технических коллективах. Результатом этого стало объединение усилий представителей академической и вузовской науки, поддержанных ведущими технологическими компаниями. Настоящая работа является результатом не только теоретических разработок, но и их прототипированием, что, в случае поддержки, создает возможность реализации оригинальных отечественных информационных технологий, способных положительно повлиять на развитие цифровой экономики в Российской Федерации.

ЛЕММА БЕЗОПАСНОСТИ СУБЪЕКТНО-ОБЪЕКТНЫХ СИСТЕМ

Установлено, что внедрение технологии блокчейн, несмотря на заверения различных разработчиков, порождает новые виды угроз, которые были достаточно подробно рассмотрены в статье [3] на сайте *Forklog*. В качестве примера можно привести ситуацию со смарт-контрактом, описываемую в различных источниках. В определенный момент выполнения смарт-контракта появляется необходимость получения для дальнейших вычислений значения текущего курса доллара. Технически это реализуется через «ораклиз», который обращается к сайту считающегося доверенным источника. В случае захвата данного сайта злоумышленниками появляется возможность осуществления различного рода атак на систему, на-

чая от недоступности сайта и заканчивая манипуляцией значением курса доллара.

Неформально суть проблемы заключается в том, что пока смарт-контракт оперирует данными и другими элементами внутри блокчейн-платформы (например, мы имеем в виду только операции с криптовалютой), можно говорить об определенной безопасности. Вместе с тем, если попытаться расширить сферу применения распределенных реестров, то неизбежное взаимодействие с внешней средой приводит к возникновению различного рода угроз безопасности.

Формально утверждение о небезопасности взаимодействия с внешними источниками данных можно сформулировать в виде леммы.

Пусть существует система B , состоящая из n субъектов и m объектов:

$$B = (S, O): |S| = n, |O| = m.$$

Для системы задано разграничение доступа в виде Декартова произведения множества субъектов и множества объектов на множество прав доступа:

$$(S \times O) \rightarrow R.$$

Данное отображение полностью описывает возможности множества субъектов по отношению к множеству объектов (в приведенных выше терминах задание такого отображения обеспечивает замыкание системы). Описываемая система считается безопасной, если данное отображение обладает двумя основными свойствами – полнотой и непротиворечивостью. Полнота означает, что право доступа определено для каждой пары (субъект–объект), непротиворечивость означает, что при задании прав доступа не существует такой пары (субъект–объект), для которой конкретное право доступа одновременно и запрещено, и разрешено.

Пусть порождается некий субъект S_i с определенными правами доступа R_{ik} по отношению к существующему объекту O_k . Порождение субъекта S_i в системе B возможно из объекта-источника O_j , при этом субъект порождается некоторым другим субъектом S_e (например, атомом-эксекутором или иной исполнительной средой выполнения смарт-контракта).

Возможны три случая:

1. Субъект S_i совпадает с существующим субъектом ($i \leq n$), при этом права доступа совпадают с существующими правами. Случай тавтологии (тривиальный).

2. Субъект S_i совпадает с существующим субъектом ($i \leq n$), при этом права доступа не совпадают с существующими правами. Таким образом нарушается свойство непротиворечивости.

3. Субъект S_i не совпадает с существующими субъектами ($i > n$), при этом права доступа заданы только для объекта O_k . В этом случае нарушается свойство полноты.

Таким образом, появление для замкнутой системы дополнительной активной сущности приводит к нарушению ее безопасности. Как следствие, большинство блокчейн систем, обеспечивающих оборот и использование криптомонет, в настоящее время не могут считаться безопасными.

ДОВЕРЕННЫЙ ЦИФРОВОЙ АКТИВ *OURCOIN*

С целью демонстрации подходов к созданию технологий распределенного реестра, обладающих свойствами информационной безопасности были сформулированы предложения по созданию защищенной криптовалюты *OURCOIN*.

Криптовалюта *OURCOIN* – цифровой финансовый актив – имущество в электронной форме, созданное платформой *OURCOIN* с использованием криптографических средств, являющееся мерой стоимости, признаваемой таковой неограниченным кругом лиц.

Дадим определения терминов, которые будут на использоваться при описании работы системы.

Токен *OURCOIN* – цифровой финансовый актив – имущество в электронной форме, созданное платформой *OURCOIN* с использованием криптографических средств, удостоверяющее меру прав, предоставляемых офертой проекта *OURCOIN* его владельцу. Токен фиксируется в электронном виде (в виде специализированного файла, либо в виде звена или атома блокчейна), снабжается уникальным номером, имеет номинал, время создания и время действия (срок валидности), сетевое имя владельца и допускает проверку неизменности всех своих параметров.

Участник реестра цифровых транзакций платформы *OURCOIN* (участник; допустимы синонимы: абонент, пользователь, клиент) – физическое или юридическое лицо, прошедшее на платформе *OURCOIN* валидацию (идентификацию личности/принадлежности) и получившее от нее уникальное сетевое имя и осуществляющее цифровые транзакции в соответствии с правилами ведения реестра цифровых транзакций.

Сетевое имя – уникальная последовательность шестнадцатеричных цифр длиной 32 знака (при байтовой записи 16 байт), которая присваивается участнику и связана с его персональным ключом.

Персональный ключ участника – случайное число, вырабатываемое при первичной или повторной регистрации участника, служащее для получения (фиксации) и проверки кодов аутентификации токенов/криптовалюты.

Код аутентификации (КА) токена/криптовалюты – информация, включенная в токен/криптовалюту и позволяющая убедиться в неизменности его параметров, а также восстановить содержание параметров при их повреждении или намеренном изменении. Далее по тексту статьи синонимично с КА используется термин «подпись». Код аутентификации (подпись) может быть личным (сформированным пользователем), транспортным (сформированным для передачи токенов между участниками системы и служащим для проверки неизменности полученного токена) и корневым (сформированным центром учета транзакций, определение которого дано ниже).

Центр эмиссии токенов (ЦЭТ) – юридическое лицо, использующее уникальное программное обеспечение платформы *OURCOIN* с целью выработки первичных обезличенных токенов, характеризующихся только номером, номиналом и временем действия.

Центр учета транзакций (ЦУТ) – юридическое или физическое лицо, являющееся участником реестра цифровых транзакций и осуществляющее дея-

тельность по валидации и учету цифровых записей в реестре цифровых транзакций в соответствии с правилами ведения реестра цифровых транзакций, присваивающее токенам/криптовалюте код аутентификации и направляющее их пользователям.

ЦУТ является прототипом уполномоченного оператора инвестиционной платформы, о котором упоминается в законопроектах по цифровой экономике, принятых Государственной Думой РФ в первом чтении (в частности, проект Федерального закона «Об альтернативных способах привлечения инвестиций (краудфандинге)»). В перспективе ЦУТ может играть роль оператора платформы государственных услуг, нотариальных услуг, оборота отечественной криптовалюты, обеспечения судебной защиты в РФ транзакций, совершаемых в распределенном реестре.

Транзакция – гражданско-правовая сделка по передаче цифровых прав на токен/криптовалюту и/или его части, совершенная участником/участниками. С *технической точки зрения транзакция* представляет собой отраженный в материальном виде в системе процесс передачи части или всего номинала токена от одного абонента к другому. Транзакция характеризуется ее инициатором, получателем, а в части размера – переводом, сдачей и комиссией. При этом сумма номиналов перевода, сдачи и комиссии всегда равна номиналу исходного токена.

Система движения (оборота) токенов/криптовалюты – система, состоящая из абонентов, ЦЭТ и ЦУТ, обеспечивающая жизненный цикл токенов в материальной форме.

Жизненный цикл токена (ЖЦ) – период времени, устанавливаемый платформой OURCOIN с момента его эмиссии и уничтожения.

С технической точки зрения **жизненный цикл токена** включает следующие элементарные компоненты: эмиссию, направление пользователю, участие в транзакциях, завершение существования. В качестве составных элементов ЖЦ выступают объединение и расщепление токенов.

Уничтожение токенов не представляет финансовой опасности для клиентов и не уничтожает меры

стоимости, а служит для обеспечения технических свойств платформы по аналогии с эмиссией и оборотом бумажных денег.

ВИДЫ ТОКЕНОВ В СИСТЕМЕ

Протококен 1 – часть токена, выработанная ЦЭТ и характеризующаяся только номером, номиналом и временем действия. Не может участвовать в транзакциях, поскольку не имеет владельца.

Протококен 2 – часть токена, снабженная именем участника и кодом аутентификации ЦУТ, а также транспортным КА, который дает возможность проверить неизменность токена при его получении абонентом.

Токен имеет свойства протокококов 1 и 2 и снабжен КА и личной подписью владельца. Только токен участвует в транзакциях.

Таким образом, протококен 2 получается при формировании КА ЦУТ и транспортного КА для протококока 1, а токен – при формировании личного КА для протококока 2.

Токен-комиссия – токен, имеющий в качестве КА ЦУТ и кода владельца одно и то же значение, он образуется при транзакциях, за которые ЦУТ взимает комиссию.

Нулевой токен – токен, получающийся в ходе транзакции, когда сдача с транзакции для ее инициатора равна нулю.

Биржевой токен – токен, используемый для биржевых операций с другими цифровыми активами, который характеризуется интегральной стоимостью, переданной абонентом-владельцем абоненту-брокеру или абоненту-бирже, не превышающей совокупной стоимости токенов, принадлежащих данному владельцу; может также использоваться для клиринга биржевых операций в заданный период.

СТРУКТУРА ТОКЕНА

Токен представляет собой защищенные от изменения данные фиксированной длины 128 байт. Структура данных, составляющих токен, представлена в табл.1.

Таблица 1

Структура данных, составляющих токен

Индекс от начала токена в байтах	Длина поля в байтах	Название поля	Примечание
0	4	Заголовок	
4	8	Номер токена	Номер токена, выработанный центром эмиссии токенов (ЦЭТ)
12	16	Контрольная информация E(Mi,Kt)	Представляет собой номер токена, дату его создания и номинал (Mi), зашифрованные на ключе токена Kt, хранящемся в ЦЭТ
28	4	Тип токена	В настоящее время в системе обращается токен с типом 1
32	4	Номинал токена целый (рубли)	
36	4	Номинал токена дробный (копейки)	

Индекс от начала токена в байтах	Длина поля в байтах	Название поля	Примечание
40	4	Параметр эмиссии	В данном проекте равен 256
44	4	Дата создания токена в формате (дд)(мм)(гг-2000) (3 байта, каждая скобка 1 байт)	
48	4	Дата валидности токена в формате (дд)(мм)(гг-2000) (3 байта, каждая скобка 1 байт)	
52	4	Уровень иерархии токена	У прототоконов 1 и 2 равен 0, для токена равен 1, при каждой транзакции увеличивается на единицу
56	8	Резервное поле тела токена	
64	16	Сетевое имя владельца токена	
80	8	Код аутентификации ЦЭТ	
88	8	Транспортный КА	
96	8	КА владельца	
104	24	Резервное поле для КА	

ОПИСАНИЕ ВАРИАНТА ПРОТОКОЛА ДВИЖЕНИЯ ЦИФРОВОГО АКТИВА

1. Центр эмиссии токенов (ЦЭТ) выпускает прототокен 1, который является началом цепочки для транзакций и передаёт его в Центр учета транзакций (ЦУТ).

2. ЦУТ снабжает его своей (корневой) подписью, транспортной подписью пользователя и отправляет пользователю (прототокен 2).

3. Пользователь проверяет транспортную подпись прототокона 2, подписывает своей личной подписью и возвращает токен в ЦУТ.

4. ЦУТ заносит токен в базу данных и одновременно передает в ЦЭТ на архивное хранение.

5. Если формируется платеж, то пользователь создает транзакцию из 4-х звеньев-кандидатов: исходная монета, сдача, перевод и комиссия. Все они подписаны транспортной подписью ЦУТ, а сдача – ещё и личной подписью отправителя платежа.

6. ЦУТ проверяет все четыре звена на правильность (совпадение суммы, корректность транспортной подписи и номера транзакции, а также наличие первого звена в своей базе), после чего передает ЦЭТ запрос на выпуск двух новых прототоконов с суммой сдачи и перевода и номером, совпадающим с номером первого звена.

7. ЦЭТ создает новые прототоконы 1 на перевод и сдачу и направляет их в ЦУТ.

8. ЦУТ подписывает перевод (прототокен 1) корневой и транспортной подписями и передает получателю платежа, а сдачу – также корневой и транспортной подписями и передает отправителю.

9. Отправитель и получатель подписывают сдачу и перевод (прототоконы 2) соответственно своими личными подписями и возвращают токены в ЦУТ.

Транзакция завершена. Далее операции с новыми токенами возможны.

Каждый токен всегда имеет три подписи (корневую, транспортную и личную), прототоконы либо то-

кены с дефектом одного из признаков валидности (подписи, КА) в транзакциях между пользователями не участвуют.

МЕХАНИЗМЫ ТЕХНИЧЕСКОЙ РЕАЛИЗАЦИИ ДОВЕРЕННОГО ТОКЕНА

Как было отмечено, в настоящее время существует целый ряд угроз безопасности блокчейн технологий и реализованных на их основе криптовалют, связанных с их системной архитектурой. Информационная безопасность должна обеспечиваться не только правильно выбранными и реализованными криптографическими схемами, но и другими решениями, обеспечивающими устойчивость и надежность функционирования прикладной системы.

Проблема заключается в том, что обычная монета обладает свойством неделимости и физическая передача означает изъятие у одного лица и появление ее у другого. В отличие от физической монеты токен по своей природе – это двоичная последовательность, которая в ходе обращения многократно копируется. На определенном уровне обобщения считается, что дублированием можно пренебречь. Вместе с тем, как показали исследования, наличие дублей может быть источником уязвимостей для всей системы.

В качестве исходного примера рассмотрим платеж криптовалютой поставщику товара с участием доверенного центра (необходимость существования такого центра обоснована в других работах). Такой платеж означает перечисление криптовалюты с возможной уплатой комиссии и получение сдачи. С точки зрения создания распределенных систем получается система с тремя независимыми точками обработки транзакций. В силу того, что в системе платежей криптовалютами планируется использование стандартной (бытовой) вычислительной техники, без повышенных требований по надежности, постоянное функционирование точек обработки транзакций не гарантируется. Использование сети Интернет для организации взаимодействия точек как наиболее

доступной среды передачи означает использование негарантированных каналов связи.

Как следствие, в процессе осуществления описанного платежа возникает модель угроз, связанных с (не)работоспособностью точки обработки транзакций и каналов связи и, как следствие, со сбоями в прикладной системе, связанными с обработкой в независимых точках сложных транзакций. Так, в выбранном примере владелец криптокошелька должен дожидаться завершения процедуры проверки и реализации токена, чтобы в случае отказа не попытаться использовать его повторно.

Указанные проблемы решаются введением в структуру токена четырех полей и механизмов их обработки: статус токена; признак токена; время изменения статуса токена; лимит времени ожидания в новом статусе. Рассмотрим назначение полей более подробно.

1. Статус токена. Должен принимать следующие значения:

- 1) может быть использован (аналог – монета лежит в кармане);
- 2) в обработке (монета передана для проверки, условное изъятие у владельца);
- 3) не может быть использован (монета доступна владельцу, но проверка показала наличие проблем с монетой);
- 4) монета истрачена.

2. Признак токена. В это поле в точке проверки заносится результат проверки. Например, несовпадение контрольной суммы токена. Более детально значения данного поля будет определено при проектировании модуля проверки.

3. Время изменения статуса токена – точное значение времени, когда токен получил новый статус. Например, начальный статус «может быть использован» присваивается после отправки токена в точку проверки, в криптокошельке значение статуса меняется на «в обработке», устанавливается время изменения статуса, а также лимит ожидания в новом статусе.

4. Лимит ожидания в новом статусе. Данное поле вводится из следующих соображений. Предположим, токен отправлен в точку проверки и на стороне кошелька имеет текущий статус «в обработке»; на стороне точки проверки произошел сбой оборудования, решение по токenu не принято и, возможно, из-за потери данных не будет принято никогда. Для исключения потери монеты вводится лимит ожидания, после окончания которого токен возвращается в исходный статус, но в его признаке ставится отметка, что решение по нему не было принято из-за исчерпания лимита времени. Предполагается, что в данном примере дальнейшие разбирательства должны проводиться с привлечением операторов платформы.

На основании описанной структуры полей возможно предложить алгоритм оплаты криптомонетой со сдачей.

Шаг 0. Исходное состояние. Копия файла с данными токена находится в кошельке. Статус токена – «может быть использован». Признак токена – «отсутствует». Время изменения статуса равно времени, когда последний пользователь получил токен. Лимит времени = -1 (без лимита). Копия файла с данными токена находится в Центре учета транзакций (ЦУТ),

значения полей идентичны значениям в копии в кошельке.

Шаг 1. Запрос на разделение в ЦУТ. Статус токена на стороне кошелька устанавливается «в обработке». Время изменения статуса устанавливается равным времени поступления запроса на разделение. Лимит времени = $t_{\text{пор}}$. Запрос с параметрами разделения отправляется в ЦУТ. Копия файла с данными токена в ЦУТ имеет значения полей, идентичные значениям на шаге 0.

Шаг 2. Прием запроса в ЦУТ. ЦУТ находит копию токена на своей стороне и устанавливает его статус «в обработке». В кошелек отправляется подтверждение статуса «в обработке» в виде квитанции. На основании получения квитанции в копии файла токена на стороне кошелька устанавливается признак «ЦУТ принял заявку».

Шаг 3. Направление запроса в Центр эмиссии токенов (ЦЭТ). На основании параметров заявки ЦУТ запрашивает генерацию трех (в общем случае) монет, сумма номиналов которых равна номиналу исходной монеты. Значения полей такие же, как для шага 2 (признак действует только для стороны кошелька).

Шаг 4. Генерация монет в ЦЭТ. ЦЭТ генерирует три обезличенные монеты с заданными номиналами. Статус всех трех монет – «не может быть использована». Копии файлов всех трех монет пересылаются в ЦУТ.

Шаг 5. Управление монетами в ЦУТ. Получив три монеты из ЦЭТ, ЦУТ проверяет их на соответствие запросу. Если соответствие подтверждается, ЦУТ на своей стороне присваивает исходной монете статус «истрачена», а трем новым монетам назначает новых владельцев с учетом их назначения (получатель, сдача, комиссия) и статус «может быть использована». В ЦЭТ направляется квитанция на подтверждение приема и обработки монет.

Шаг 6. Возврат сдачи. На стороне кошелька отправителя осуществляется прием копии файла криптомонеты-сдачи в кошелек. Кошелек меняет статус исходной монеты с «в обработке» на статус «истрачена».

МОДУЛИ РЕАЛИЗАЦИИ ПЛАТФОРМЫ УПРАВЛЕНИЯ ТОКЕНАМИ

Модуль первичного формирования ключа пользователя, его ключевого контейнера и сетевого имени (*InitUs*)

Формат использования: *InitUser FileUserID UserPIN RealName<RandomString>*,

где: *FileUserID* – имя файла с закрытым при помощи пароля персональным идентификатором пользователя (может быть связано с именем пользователя, но не должно совпадать с именем файла *Realname*);

UserPIN – пароль (пин-код или метод его ввода, например, чтение из *USB*-токена) для закрытия персонального идентификатора пользователя;

RealName – текстовый файл длиной не более 160 байт для размещения данных пользователя;

<RandomString> – необязательный параметр для улучшения работы датчика случайных чисел («разгонная строка»).

Модуль создает файл `random.bin` для дальнейшего использования датчика случайных чисел и файл с закрытым при помощи пароля персональным идентификатором пользователя (фактически – защищенный контейнер для хранения и передачи персонального идентификатора/ключа пользователя). Кроме того, модуль формирует сетевое имя, которое размещается в файле `RealName.net` в читаемом человеком виде (в виде 32 символов шестнадцатеричного кода). Первые 8 цифр определяют также имя файла, в котором сетевое имя находится в бинарном формате.

Модуль возвращает типизированные ошибки, необходимые для интеграции вызовов модулей во внешние приложения или смарт-контракт.

Возвращаемые ошибки:

«-1» – ошибка при тестировании модулей защиты;
«-2» – ошибка формата вызова (неверное количество аргументов);

«-3» – файл идентификатора (`FileUserID`) уже существует или не существует файла реального имени;

«-4» – ошибка формирования случайного числа;

«-5» – ошибка обновления случайного числа;

«-6» – ошибка записи файла пользователя (`FileUserID`);

«-7» – ошибка контрольного чтения файла пользователя (`FileUserID`).

В платформе выделяются два абонента с зафиксированными именами контейнеров – 000 – контейнер Центра эмиссии токенов (ЦЭТ) и 007 – контейнер Центра учета транзакций (ЦУТ).

Пример формирования системы из трех пользователей, ЦЭТ и ЦУТ приведен в следующей последовательности команд (смарт-контракт):

```
md 01
md 02
md 03
md CET
md CRT
copy alisa 01\alisa
initus 01\1 alisa 01\alisa 123
copy *.net 01\*.*
del *.net
copy boris 02\boris
initus 02\2 boris 02\boris 456
copy *.net 02\*.*
del *.net
copy ivan 03\ivan
initus 03\3 ivan 03\ivan 789
copy *.net 03\*.*
del *.net
initus 000 boss
copy 000 CET\000
del 000
copy gentoken.exe CET\gentoken.exe
initus 007 crt
copy 007 CRT\007
del 007
copy sendtok.exe CRT\sendtok.exe
```

Полагаем, что исполняемые модули, копируемые в создаваемые директории, есть в корневой директории. При этом также имеются текстовые файлы `alisa`, `boris` и `ivan`, содержащие реальные имена абонентов в системе.

Модуль эмиссии токенов – *gentoken* (*Generate Token*)

Формат использования: *GenToken UserPIN Номинал токена*, где:

UserPIN – пароль (пин-код или метод его ввода, например, чтение из *USB*-токена) для восстановления персонального идентификатора (ключа) ЦЭТ. Для приведенного макета он будет методом или строкой `boss`;

Номинал токена – десятичное число, не превышающее 999999.

Данный модуль также создает файл `random.bin` для дальнейшего использования датчика случайных чисел и файл `Glob.tcn` с закрытыми при помощи персонального идентификатора пользователя с именем 000 эмитированными токенами и ключами эмиссии, индивидуальными для каждого эмитированного токена. Предполагается, что ключевой контейнер 000 уже создан процедурой (модулем) *UnitUs*. Результатом работы данного модуля будет обезличенный токен заданного номинала, имеющий расширение `*.tc0`.

Модуль передачи токенов пользователям *GetCrtK* (*Get Center Registration Transaction Keys*)

Для передачи токенов пользователям необходимо сформировать транспортные ключи. Этот процесс выполняется либо абонентом самостоятельно, либо в ЦУТ при помощи модуля *GetCrtK* (*Get Center Registration Transaction Keys*).

Формат использования: *GetCrtK Файл сетевого имени UserPIN*, где:

Файл сетевого имени – файл, полученный процедурой *InitUs* длиной 16 байт, содержащий бинарное сетевое имя пользователя (абонента);

UserPIN – пароль (пин-код или метод его ввода, например, чтение из *USB*-токена) для закрытия транспортного ключа Центра эмиссии токенов (ЦЭТ). Этот пароль далее будет использоваться для защиты контейнера с транспортным ключом (имеет расширение `*.007`). Центр учета транзакций (ЦУТ) должен иметь ключи всех пользователей, поэтому пользователи могут выработать транспортные ключи самостоятельно и направить их в ЦУТ, а пароль сообщить ЦУТ отдельно (по смс, письмом или голосом). Либо пользователи высылают в ЦУТ бинарный файл своего сетевого имени, и ЦУТ формирует транспортные ключи пользователей и также отдельно (по другим каналам) сообщает им их пароли. С точки зрения безопасности это равноценная схема, поскольку пользователи не знают пароля друг друга, а ЦУТ является доверенной стороной (доверенным компонентом системы).

Модуль передачи эмитированного токена пользователю

Формат использования: *SendTok[.exe] CrtPIN CrtKeyFile(.007) CrtKeyPIN TokenFile*, где:

CrtPIN – пароль (пин-код или метод его ввода, например, чтение из *USB*-токена) для закрытия контейнера 007 ЦУТ. Этот пароль далее будет использоваться для раскрытия контейнера с ключом ЦУТ для снабжения токена КА ЦУТ;

CrtKeyFile(.007) – имя ключа пользователя, которому передается токен. Он содержит сетевое имя пользователя, которое помещается в соответствующее поле токена;

CrtKeyPIN – пароль, который далее будет использоваться для раскрытия контейнера с транспортным ключом (имеет расширение *.007);

TokenFile – имя обезличенного токена, направляемого пользователю с ключом *CrtKeyFile(.007)*, который получен при помощи описанной выше процедуры *GetCrtK (Get Center Registration Transaction Keys)*.

Результатом работы данного модуля является направленный токен с расширением *.tc7. Этот токен еще не принят пользователем (т.е. не заверен его КА).

Модуль заверения токена

Для заверения токена используется процедура *FixToken[.exe] CrtKeyFile CrtKeyPIN UserFile UserPIN TokenFile*, где:

CrtKeyFile(.007) – имя ключа пользователя, которому передается токен;

CrtKeyPIN – пароль (пин-код или метод его ввода, например, чтение из *USB*-токена) для раскрытия контейнера сетевого ключа пользователя. Указанный в методе *CrtKeyPIN* пароль далее будет использоваться для раскрытия контейнера с транспортным ключом (имеет расширение *.007);

UserFile – имя файла с закрытым при помощи пароля персональным идентификатором пользователя (контейнером) (может быть связано с именем пользователя, но, как было указано, не должно совпадать с именем файла *Realname*);

UserPIN – пароль для раскрытия контейнера пользователя;

TokenFile – имя направленного токена, который заверяется пользователем.

После заверения токен имеет 3 кода аутентификации: ЦУТ (корневой), транспортный и личный пользовательский и может использоваться для проведения транзакций (например, для передачи части его номинала другому пользователю).

РАСШИРЕНИЕ ПОНЯТИЯ ДОВЕРЕННОГО ТОКЕНА В ОБЛАСТИ ДОВЕРЕННОЙ ЦИФРОВОЙ ЕДИНИЦЫ

Развитием описанного подхода является разработка универсальной доверенной цифровой единицы (*Universal Digital Proof Unit – UDPU*), которая является развитием технологии токенов в области учета движения товаров и распределенного хранения, а также обеспечения работы транзакционных систем, включая системы электронных платежей, в том числе и трансграничные.

Универсальная доверенная цифровая единица (УДЦЕ) – это зафиксированная в электронном (цифровом) виде (в виде специализированного файла, либо в виде звена или атома блокчейна) мера учета хранения и перемещения данных, включая персональные цифровые активы, электронные или безналичные деньги, цифровые права, электронные документы и другую цифровую информацию, снабженная уникальным номером, временем создания и време-

нем действия (сроком валидности), сетевым именем владельца и другими параметрами, необходимыми для хранения и обмена информацией и допускающая проверку неизменности всех своих параметров.

Дадим определения терминов, которые будут нами использоваться при описании работы системы.

Владелец УДЦЕ – идентифицированный уникальным сетевым именем участник системы (допустимы синонимы: абонент, пользователь, клиент) системы учета и оборота УДЦЕ, при этом его сетевое имя в обязательном порядке включено в ее состав.

Сетевое имя – уникальная последовательность шестнадцатеричных цифр, однозначно связанная с персональным ключом пользователя и его реальным именем. Сетевое имя вырабатывается при помощи криптографической процедуры.

Персональный ключ пользователя – случайное число, вырабатываемое при первичной или повторной регистрации пользователя, служащее для получения (фиксации) и проверки кодов аутентификации УДЦЕ.

Код аутентификации (КА) УДЦЕ – информация, включенная в состав универсальной доверенной цифровой единицы и позволяющая убедиться в неизменности его параметров, а также восстановить содержание параметров при их повреждении или намеренном изменении.

Центр эмиссии (ЦЭ) – выделенный абонент, вырабатывающий по запросам других абонентов либо самостоятельно первичные обезличенные УДЦЕ.

Центр учета (центр регистрации транзакций, ЦУТ, ЦРТ) – выделенный абонент, снабжающий УДЦЕ кодом аутентификации и направляющий их пользователям. Кроме того, ЦУТ (ЦРТ) регистрирует движение УДЦЕ от одного абонента к другому.

Транзакция – отраженный в материальном виде в системе процесс передачи всей УДЦЕ или ее части от одного абонента к другому. Транзакция характеризуется ее инициатором (отправителем) и получателем.

Система движения (оборота) УДЦЕ – система, состоящая из абонентов, ЦЭ и ЦУТ, обеспечивающая жизненный цикл в материальной форме.

Жизненный цикл – включает следующие элементарные компоненты: эмиссию, регистрацию, направление пользователю, участие в транзакциях, завершение существования. В качестве составных элементов ЖЦ выступают объединение и расщепление УДЦЕ, а также перевод их в машиночитаемую форму в виде двумерного кода.

Применение УДЦЕ позволяет решить такие актуальные задачи в области цифровых активов и финансов, биржевой деятельности и учета движения товаров, как:

- создание токена – цифровой монеты, которая абсолютно устойчива к хищению за счет наличия неизменного имени владельца;
- полное блокирование угроз, связанных с использованием зарубежных криптографических алгоритмов;
- создание цифровых акций и векселей произвольного (при этом неизменяемого и самовосстанавливаемого) номинала;

Преимущества Универсальной доверенной цифровой единицы

№	Свойство	Чем обеспечивается
1	Управляемая анонимность транзакций	Оригинальным протоколом транзакций, в котором содержание транзакции может быть зашифровано или открыто, при полном закрытии никакой информации о транзакции получить невозможно
2	Управляемая анонимность имен	Имена могут образовываться таким образом, что по сетевому имени невозможно определить реальное имя
3	Анонимность УДЦЕ	Содержит только сетевое имя владельца
4	Высокая криптографическая защищенность УДЦЕ и транзакций	Использованием российских криптографических алгоритмов
5	Возможность использования любого блокчейна	Протокольной независимостью от алгоритма работы блокчейна и возможностью помещения данных в звено блокчейна и их извлечения из звена блокчейна любого вида
6	Полный трекинг транзакций с сохранением анонимности владельца	Оригинальной структурой УДЦЕ и их учета в виде впервые используемой блок-матрицы (развитие блокчейна в 2D)
7	Возможность восстановления транзакций и УДЦЕ даже при утрате приватной информации от хранилища (кошелька)	Особенностью протокола транзакций с четырехзвенным подтверждением и использованием ключей парной связи
8	Возможность работы с центром учета транзакций и без него (как централизованная, так и децентрализованная схема)	Особенностью протокола транзакций
9	Возможность управления жизненным циклом УДЦЕ (в частности, предельным сроком жизни по времени)	Оригинальной структурой УДЦЕ и протоколом транзакций
10	Полная защита от повторного использования	Хранением УДЦЕ в блок-матрице
11	Отсутствие необходимости в удостоверяющих центрах	Симметричной ключевой системой
12	Возможность проведения транзакции с подтверждением информации получателем (защита от ошибочных перечислений на кошелек)	Особенностью протокола транзакций с четырехзвенным подтверждением приема
13	Технологическая независимость от лидеров криптовалютного рынка и контролирующих органов стран –инициаторов санкций	Оригинальным протоколом, использованием российских криптоалгоритмов
14	Независимость от действующего рынка криптовалют и токенов	Оригинальным протоколом, использованием российских криптоалгоритмов
15	Широкая область применения	Возможностью трансформации в машиночитаемую форму (двумерный код)

- обеспечение безопасных биржевых торгов при помощи номинирования валют (включая криптовалюты) к УДЦЕ;

- сквозное решение логистических задач как путем снабжения УДЦЕ товаров (в виде маркера двумерного кода), так и хранением УДЦЕ в распределенном реестре;

- защита информации при проведении транзакций, защите персональных и конфиденциальных данных, хранимых в распределенном реестре;

- возможность создания совершенных систем учета произведений искусства и удостоверения их подлинности.

Преимущества Универсальной доверенной цифровой единицы (УДЦЕ) отражены в табл. 2.

ВЫВОДЫ

Предлагаемая технология создания цифрового актива позволяет формировать весь необходимый спектр типов доверенных токенов и операций с ними. Целесообразно криптографические операции реализовывать на основе отечественных криптоалгоритмов для обеспечения свойств доверенности и возможности последующей сертификации и аттестации разработанных решений государственными регуляторами. Свойства доверенности токена или его развития – УДЦЕ – должны быть обеспечены во всех формах их преобразования и обращения, что достигается только благодаря применению соответствующей доверенной цифровой платформы. Таким образом, доверенная технология определяется свойствами доверенной цифровой единицы и доверенной цифровой платформы.

СПИСОК ЛИТЕРАТУРЫ

1. Биктимиров М.Р., Щербаков А.Ю. Проблемы синтеза доверенных систем // Труды ИСА РАН. – 2012. – Т. 53. – С. 264–271.
2. Правиков Д.И., Щербаков А.Ю. К вопросу об изменении парадигмы информационной безопасности // Системы высокой доступности. – 2018. – Т. 14, № 2. – С. 35–39.
3. Васильков Г. «Разбор основных экономико-технических угроз для блокчейн-приложений» // Интернет-журнал «Форклог». – URL: <https://forklog.com/razbor-osnovnyh-ekonomiko-tehnicheskikh-ugroz-dlya-blokchejn-prilozhenij-prodolzhenie>

Материал поступил в редакцию 03.07.18.

Сведения об авторах

ГРИНЯЕВ СЕРГЕЙ НИКОЛАЕВИЧ – доктор технических наук, старший научный сотрудник, декан Факультета комплексной безопасности топливно-энергетического комплекса Российского государственного университета нефти и газа (национальный исследовательский университет) им. И.М. Губкина, Москва
e-mail: gsn@gubkin.pro

ЗЛОТИН РОМАН АЛЬФРЕДОВИЧ – кандидат юридических наук, адвокат Адвокатской палаты Московской области, юридический консультант проекта *OURCOIN*, Москва
e-mail: rzlotin@yandex.ru

МИЛУШКИН АЛЕКСАНДР ИВАНОВИЧ – заместитель директора управляющей компании проекта "Технопарк Пушкино", г. Пушкино Московской области
e-mail: mialiv@techprom.biz

ПРАВИКОВ ДМИТРИЙ ИГОРЕВИЧ – кандидат технических наук, руководитель научно-образовательного центра новых информационно-аналитических технологий Факультета комплексной безопасности топливно-энергетического комплекса РГУ нефти и газа (национальный исследовательский университет) им. И.М. Губкина; научный сотрудник Института комплексной безопасности и специального приборостроения Российского технологического университета (ИКБиСП РТУ), Москва
e-mail: d_pravikov@mail.ru

СЕЛИОНОВ ИГОРЬ АНАТОЛЬЕВИЧ – заместитель директора управляющей компании проекта "Технопарк Пушкино", г. Пушкино Московской области
e-mail: uk@techprom.biz

ЩЕРБАКОВ Андрей Юрьевич – доктор технических наук, профессор, ведущий научный сотрудник ВИНТИ РАН, главный научный сотрудник Центра развития криптовалют и цифровых финансовых активов ВИНТИ РАН; научный консультант проекта "Технопарк Пушкино"
e-mail: x509@ras.ru

ЩУКО ЮЛИЯ НИКОЛАЕВНА – кандидат географических наук, ВРИО директора ВИНТИ РАН, зав. отделом научной информации по комплексным межотраслевым проблемам ВИНТИ РАН, Москва
e-mail: dir@viniti.ru

Рекомендательные системы интернет-сервисов: этапы проектирования и разработки

Приводится обзор наиболее популярных подходов, лежащих в основе создания современных рекомендательных систем интернет-сервисов, а также основные этапы их проектирования и разработки в соответствии с межгосударственным стандартом.

В качестве подходов рассмотрены коллаборативная и контентная фильтрация, их преимущества и недостатки, а также гибридный подход. В этапах проектирования приведены примеры наиболее часто используемых структур данных, алгоритмов формирования и методов оценки качества рекомендаций. Выбор конкретных элементов структуры рекомендательной системы объясняется на примере компании, разрабатывающей рекомендательную систему для интернет-портала.

Ключевые слова: коллаборативная фильтрация, контентная фильтрация, гибридная рекомендательная система, кластеризация, меры близости, отложенная выборка, кросс-валидация, метрики качества

ВВЕДЕНИЕ

В связи с доступностью Интернета, ростом популярности социальных сетей, развитием веб-технологий, облачных программных продуктов и продвинутых систем аналитики современные тенденции ведения бизнеса предполагают переход на электронные площадки с использованием всех возможных цифровых каналов взаимодействия с клиентами. Это способствует как расширению возможностей по продвижению услуг и привлечению клиентов, так и сокращению издержек [1].

Современные технологии позволяют интегрировать все задействованные каналы в единую систему, дающую возможность собирать статистические данные, детально отслеживать каждый этап работы компании и проводить сложную аналитику, что позволяет находить узкие места бизнес-процессов, проводить их оптимизацию и увеличивать прибыль.

В частности, многие компании, предлагающие широкий спектр товаров и услуг, задействуют многофункциональные интернет-порталы. Основная задача такого портала – побудить пользователя сделать покупку, при этом максимально упростив для него процесс выбора из имеющегося ассортимента. Для решения этой задачи необходимо проводить сегментацию пользователей с целью персонализации предложений под их уникальные потребности. При этом остро стоит вопрос об автоматизации как отдельных этапов подобной деятельности, так и процесса принятия решений в целом. В настоящее время основным инструментом персонализации предложений в электронной коммерции является рекомендательная система [2].

ОСНОВНЫЕ ПОНЯТИЯ, ИСПОЛЬЗУЕМЫЕ В РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЕ

Проводить автоматическую обработку данных и прогнозировать интерес пользователя к какому-либо объекту должна рекомендательная система, для работы которой необходимы данные о самих пользователях, имеющихся объектах, и истории их взаимодействия, собранные из всех доступных источников [2].

В настоящей статье будут рассмотрены основные идеи методов коллаборативной и контентной фильтрации. Оба метода оперируют следующими понятиями:

- пользователь – потребитель рекомендательной системы, т.е. тот, для кого составляются рекомендации;
- объект – некоторая сущность, потенциально представляющая интерес для пользователя и являющаяся частью рекомендаций;
- рекомендация – объект, который рекомендательная система выдает пользователю, если этот объект прошел по заданным критериям.

Коллаборативная фильтрация (Collaborative Filtering) – метод предполагает, что похожим пользователям нравятся похожие объекты [3]. Его идея состоит в выделении наиболее близких групп пользователей и объектов с последующим составлением рекомендаций. Существует три типа коллаборативной фильтрации: *model-based*, *memory-based* и гибридный, объединяющий эти типы [4].

I. Model-based подход – предполагает, что просмотры пользователями объектов определяются какими-либо факторами, влияние которых можно описать некоторым законом или моделью. Для формирования рекомендаций составляется модель, хорошо приближающая имеющиеся просмотры пользователей, на основе которой дается прогноз вероятности просмотров других объектов [5].

II. Memory-based подход в отличие от *model-based* подхода не формирует какую-либо модель, а использует исторические данные по просмотрам объектов пользователями и создает матрицы их соотношений. Базовым для коллаборативной фильтрации здесь является подход, основанный на соседстве (*neighbourhood-based*), который подразделяется на два вида [5]:

1) *user-based* подход – в рекомендательную выдачу попадают объекты, понравившиеся похожим пользователям [4];

2) *item-based* подход – в рекомендательную выдачу попадают объекты, близкие к тем, что уже просматривал пользователь [4].

Источники [4, 6] выделяют следующие преимущества и недостатки коллаборативной фильтрации, отраженные в табл. 1.

Таким образом, коллаборативная фильтрация, с одной стороны, менее требовательна к имеющимся данным (достаточно идентификатора пользователя, объекта и факта просмотра) и вследствие этого более простая в эксплуатации, но, с другой, подходит либо для небольших баз данных, обработка которых в режиме реального времени не требует большого объема ресурсов, либо для хорошо масштабируемых систем с быстрым поиском, в которых не нужно

выполнять сжатие данных и терять точность рекомендаций.

Контентная фильтрация (*content-based*) в отличие от коллаборативной фильтрации, опирающейся только на факты взаимодействия пользователей и объектов без учета контекста, использует характеристики объектов.

В этом случае формируются профили пользователей (пол, возраст, интересы и т.д.) и объектов (название, тематика и т.д.) на основе их характеристик [7]. Последующее формирование рекомендаций основывается на сопоставлении пользователей и объектов по схожести их профилей.

В [3] выделяются преимущества и недостатки коллаборативной фильтрации, отраженные в табл. 2.

Таким образом, для контентной фильтрации необходима достаточно развернутая информация о пользователях и объектах, позволяющая соотносить их между собой, но при этом нет необходимости использовать громоздкую статистику просмотров (кроме тех случаев, когда необходимо исключать уже просмотренные пользователем объекты из рекомендаций). Возможность отложенного формирования рекомендаций позволяет применять контентный подход даже в объемных и относительно медленных системах.

Таблица 1

Преимущества и недостатки коллаборативной фильтрации

Преимущества	Недостатки
<ul style="list-style-type: none"> достаточно легка в настройке и использовании не требуется описание объектов и пользователей, явное указание их интересов и предпочтений. Таким образом, нет необходимости проводить сбор и оцифровку дополнительных данных точность рекомендаций увеличивается со временем, так как пользователи и объекты постоянно пополняют статистику просмотров высока вероятность рекомендации принципиально нового для пользователя объекта 	<ul style="list-style-type: none"> расчет рекомендаций возможен в основном в режиме реального времени (в отложенном режиме они утрачивают актуальность) объем и разреженность матриц просмотров пользователей и объектов, что влечет высокую сложность хранения и обработки данных проблема «холодного старта», когда система не знает, что рекомендовать новым пользователям и кому рекомендовать новые объекты, т.к. по ним отсутствует история взаимодействия проблема «белых ворон», т.е. нетипичных пользователей, чьи интересы не совпадают с мнением ни одного выделенного кластера пользователей. Система также не знает, что им предложить

Таблица 2

Преимущества и недостатки контентной фильтрации

Преимущества	Недостатки
<ul style="list-style-type: none"> расчет рекомендаций возможен в отложенном режиме, так как профили пользователей, в основном, постоянны во времени, а новые объекты вводятся относительно нечасто отсутствие проблем холодного старта и «белых ворон» отсутствие необходимости накопления большого объема статистики просмотров 	<ul style="list-style-type: none"> необходимость сбора персональных данных для формирования профилей пользователей и описания объектов сложность оцифровки, обработки и сопоставления текстовой информации профилей пользователей и объектов. тривиальность рекомендаций. Пользователю не будут рекомендоваться принципиально новые для него объекты

Гибридный подход распространен для формирования рекомендаций и включает применение как коллаборативной, так и контентной фильтрации, что позволяет компенсировать их недостатки. Гибридные системы комбинируют применение нескольких методов рекомендаций с помощью разных типов своей внутренней организации [8].

1. **Конвейерная организация.** Задача выработки рекомендации делится на этапы, которые выполняются последовательно. Рекомендательные подсистемы по очереди выдают данные поступающие на вход следующей подсистеме. В итоге последняя подсистема формирует окончательный набор рекомендаций.

2. **Параллельная организация.** Рекомендательные подсистемы работают одновременно, независимо друг от друга. После окончания работ основная подсистема обрабатывает полученные результаты и формирует окончательную окончательный набор рекомендаций.

3. **Монолитная организация.** Представляет собой единый модуль, в котором рекомендательные подсистемы работают совместно. Например, отбирается некоторое множество объектов методом коллаборативной фильтрации, которое корректируется по интересам пользователя методом контентной фильтрации.

ЭТАПЫ ПРОЕКТИРОВАНИЯ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ

Проектирование и разработка любого программного продукта подчиняются определенным этапам жизненного цикла. Классический жизненный цикл включает системный анализ, анализ требований, проектирование, реализацию, тестирование, внедрение и сопровождение [8].

Более детально процессы проектирования и разработки рассмотрены в межгосударственном стандарте ГОСТ 19.102-77 «Единая система программной документации (ЕСПД). Стадии разработки» [9]. Согласно этому документу разработка вычислительных систем и комплексов независимо от их назначения должна в общем случае соответствовать стадиям, отраженным в табл. 3.

В совокупности по всем стадиям (от технического задания до внедрения) можно определить основные этапы разработки, актуальные для рекомендательной системы, которые будут рассмотрены в данной работе:

1. Подготовительный этап, включающий постановку задачи и определение критериев эффективности и качества системы.
2. Определение структуры данных.
3. Выбор подходов формирования рекомендаций.
4. Выбор и разработка алгоритмов.
5. Тестирование и оценка качества.

Таблица 3

Стадии и этапы разработки вычислительных систем

1. Техническое задание	3. Технический проект
Постановка задачи	–
Сбор исходных материалов	–
Выбор и обоснование критериев эффективности и качества разрабатываемой программы	–
Определение структуры входных и выходных данных	Уточнение структуры и формы представления входных и выходных данных
Предварительный выбор методов решения задач	Разработка алгоритма решения задачи
Определение требований к техническим средствам и программе	Окончательное определение конфигурации технических средств и формирование плана мероприятий по разработке и внедрению программ
Согласование и утверждение технического задания	Согласование и утверждение технического проекта
2. Эскизный проект	4. Рабочий проект
Предварительная разработка структуры входных и выходных данных	Программирование и отладка программы
Уточнение методов решения задачи и разработка общего описания алгоритма решения задачи	Разработка, согласование и утверждение программы и методики испытаний
Согласование и утверждение эскизного проекта	Корректировка программы и программной документации по результатам испытаний
5. Внедрение	
Подготовка и передача программы и программной документации для сопровождения и (или) изготовления	

Этап 1. Подготовительный

Рассмотрим основные этапы проектирования и разработки рекомендательной системы на примере компании, предоставляющей консалтинговые и образовательные услуги и организующей мероприятия на интернет-портале.

Часть контента в виде электронных книг и журналов, записей лекций и мастер-классов свободно распространяется на портале для авторизованных пользователей. Платный контент и возможность регистрации на предстоящие курсы также доступны на портале.

Бесплатный контент выполняет две функции:

- консалтинг и ответы на частые вопросы в интересующей пользователя сфере деятельности, а также информирование об актуальных изменениях в законодательстве;
- повышение лояльности клиентов, стимулирование покупки платного контента и записи на курсы.

Для более эффективной работы портала в этих направлениях было решено начать разработку рекомендательной системы, которая должна предлагать пользователям контент, соответствующий их профилю и интересам, определенным на основе их активности на портале. В качестве основного показателя эффективности работы системы выбран показатель конверсии.

Этап 2. Определение структуры данных

Как правило, большая часть данных, используемая в корпоративных CRM-системах, хранится в реляционных базах данных. Однако постепенно набирают популярность *NoSQL* решения, позволяющие распределенно обрабатывать большие объемы данных [10]. Рассматриваемая нами компания использует обе концепции хранения информации.

Всю «контентную» информацию, отображаемую и собираемую на портале, компания хранит в реляционной базе данных. Так как рекомендательные системы оперируют понятиями «пользователь» и «объект», то часть таблиц отведена под эти сущности.

Таблица пользователей содержит базовую информацию, заполняемую ими в профиле при регистра-

ции, при этом каждая отдельная строка описывает отдельного пользователя, идентифицируемого по уникальному *ID*. Аналогичную структуру, имеет таблица с контентом, однако эти данные заполняются уже сотрудниками компании.

Так как предлагаемый клиентам контент довольно специфичен и имеет определенную направленность, то для идентификации и учета этой специфики в рекомендациях профили пользователей и контент помечаются тегами.

Упрощенная схема базы данных, показывающая связи между таблицами пользователей и объектов, приведена на рисунке.

Для реализации быстрого поиска на портале и сбора статистики по просмотрам пользователями контента компания использует поисковую машину, выступающую в качестве *NoSQL* базы данных. Основное отличие от реляционных баз данных состоит в принципиально другой, нетабличной, структуре данных (например, в виде дерева), хорошей масштабируемости и скорости обработки запросов [10].

Этап 3. Выбор подходов формирования рекомендаций

Выше было описано несколько подходов к выработке рекомендаций (коллаборативная, контентная, гибридная), каждый из которых обладает своими преимуществами и недостатками. Кроме того, каждый подход обладает определенными требованиями к имеющимся данным и их структуре [11]. Таким образом, при выборе конкретного подхода необходимо исходить из имеющейся технологической платформы.

В примере с рассматриваемой нами компанией было принято решение о создании гибридной рекомендательной системы с параллельной организацией, так как одна часть данных хранится в «медленной» реляционной базе данных и содержит детальную информацию о пользователях и контенте, пересекающихся по тегам, а другая часть хранится в «быстрой» и масштабируемой *NoSQL* базе и описывает историю взаимодействия пользователей и объектов.



Схема базы данных, описывающая данные о пользователях и объектах

Одна из подсистем будет работать с первой базой и использовать контентную фильтрацию (так как имеются профили пользователей и объектов), а вторая – со второй базой и использовать коллаборативную фильтрацию (так как имеется история взаимодействия).

В итоговую выдачу будет попадать некоторый *TopN* рекомендаций из обеих подсистем. При этом необходимо удалять из финальной выдачи объекты, уже просмотренные пользователем.

Этап 4. Выбор и разработка алгоритмов рекомендательной системы

Выбор конкретного алгоритма обучения рекомендательной системы во многом зависит от самих данных, их структуры и выбранного подхода. Тем не менее, оба метода коллаборативной и контентной фильтрации в рекомендациях опираются на схожесть пользователей и объектов, для определения которой используются методы кластеризации.

В машинном обучении выявление схожих объектов является задачей кластеризации, т.е. задачей выделения в данных содержательной подгруппы или кластера [12]. В свою очередь методы кластеризации используют меры близости для оценки сходства элементов и отнесения их к одной или разным группам.

Наиболее часто используемыми мерами близости являются [4]:

- расстояние Евклида;
- коэффициент Дайса;
- манхэттенское расстояние;
- корреляция Пирсона;
- коэффициент Жаккара;
- косинусная близость;

В рамках кластерного анализа существует множество подходов к выделению содержательных подгрупп в данных, например, метод К-средних, агломеративные и дивизимные методы [13]:

• метод К-средних опирается на предположение о возможном числе различных кластеров. В начале алгоритма задается число кластеров k , на каждой итерации элементы сопоставляются кластерам так, что средние в кластере (для всех элементов) максимально возможно отличаются друг от друга по метрике;

• агломеративные методы характеризуются последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров. На первом шаге алгоритма все объекты являются отдельными кластерами. Затем наиболее похожие объекты объединяются до тех пор, пока все они не будут составлять один кластер;

• дивизимные методы характеризуются последовательным разделением исходного кластера, состоящего из всех объектов, так и соответствующим увеличением числа кластеров. В начале работы алгоритма все объекты принадлежат одному кластеру, который на последующих шагах делится на меньшие кластеры, в результате чего образуется последовательность расщепляющих групп.

Процесс выбора конкретных методов и мер кластерного анализа нетривиален, так как разные меры по-разному оценивают исходные данные, а разные методы по-разному выделяют в них кластеры, что

приводит к различным результатам. Поэтому в каждом конкретном случае исследователь должен делать выбор, исходя из целей кластеризации, физической и статистической природы исходных данных.

Специфика контента, предлагаемого рассматриваемой нами компанией, и интересы пользователей, выделяемые в профиле и идентифицируемые по тега, в общем случае ограничивают возможное число схожих групп пользователей и объектов. Вследствие чего было принято решение использовать кластеризацию по методу К-средних с определенным числом кластеров.

В рамках коллаборативной фильтрации в качестве меры близости была выбрана косинусная близость, так как в [3] она описывается как наиболее подходящая для оценки близости интересов пользователей (чем сильнее они совпадают, тем ближе значение мер к 1). Также в качестве меры для коллаборативной фильтрации распространена корреляция Пирсона [14], однако согласно [15] она не определена для векторов с постоянными значениями, вследствие чего можно потерять рекомендации.

В рамках контентной фильтрации, основывающейся на характеристиках объектов, предпочтение было отдано коэффициенту Дайса. Для его применения необходимо параметризовать профили пользователей и объектов, т.е. разбить их описания на значимые термины. Для решения этой задачи можно использовать алгоритм *TF-IDF* (*term frequency – inverse document frequency*).

Данный алгоритм предполагает следующие операции.

1. Разбиение исходного описания на отдельные слова.
2. Приведение полученных слов к нормальной форме (лемме).
3. Исключение предлогов, артиклей, междометий и дубликатов слов.
4. Оценка частоты вхождения слова по отношению к общему числу слов в описании (TF-блок).
5. Оценка обратной частоты слова, с которой оно встречается в описаниях (IDF-блок). Это позволяет придать больший вес уникальным словам и снизить вес общеупотребительных слов.
6. Расчет меры путем перемножения значений, полученных в TF- и IDF-блоках, и выбор k наиболее значимых слов.

После применения алгоритма, описания профилей пользователей и объектов сравниваются с помощью коэффициента Дайса по степени совпадения выделенных ключевых слов.

Этап 5. Тестирование и оценка качества

Выбор конкретного рекомендательного алгоритма является нетривиальной задачей, так как определить, какой подход оптимально впишется в конкретную предметную область и технологическую платформу разработки довольно трудно. Кроме того, даже в рамках определенного метода существует множество параметров, подлежащих подбору и настройке. Зачастую при разработке рекомендательной системы выбирается несколько подходов, либо один подход с разными параметрами, которые затем сравниваются по качеству даваемых рекомендаций.

Так как основные параметры алгоритма рекомендаций формируются методами машинного обучения, то для оценки его качества используются специальные методы:

1. **Отложенная выборка.** Это один из наиболее распространенных методов оценки качества обученного алгоритма. Его суть состоит в разделении исходных данных на две части – обучающую и тестовую – чаще всего в пропорциях 70/30 или 80/20 процентов. На первой выборке алгоритм обучается, а на второй происходит проверка его качества.

При этом слишком простой, недообученный алгоритм, будет плохо работать на обеих выборках, а переобученный будет хорошо работать только на обучающей выборке (на тестовой качество сильно упадет). Главное достоинство этого метода заключается в относительной простоте реализации.

2. **Кросс-валидация.** Основным недостатком метода отложенной выборки является тот факт, что исходная выборка разбивается на части случайным образом, и при этом все нетипичные значения могут попасть строго в одну из двух частей. При этом алгоритм, обученный только на типичных (или только нетипичных) примерах, не покажет хорошее качество на тестовой части. Для того чтобы избежать такой ситуации используется кросс-валидация. Ее суть состоит в разбиении исходной выборки на k блоков примерно равного размера, при этом каждый блок по очереди будет выступать в качестве тестового множества, а остальные – в качестве обучающего. Показатели качества, полученные на каждом тестовом множестве, усредняются. Метод кросс-валидации позволяет более грамотно оценивать качество, но при этом требует гораздо большего объема ресурсов, так как тестирование каждого алгоритма придется проводить k раз.

Вследствие того, что рассматриваемой нами компанией используется две рекомендательные подсистемы с разными подходами к формированию рекомендаций, применение ресурсоемкой кросс-валидации было признано нецелесообразным, и тестирование качества будет проводиться методом улучшенной¹ отложенной выборки.

3. **Метрики точности и полноты рекомендаций.** Рассмотренные нами методы подходят для тестирования любых систем, в той или иной степени ис-

пользующих алгоритмы машинного обучения. Однако рекомендательные системы имеют свою специфику, так как предлагают пользователю определенный контент, поэтому помимо общей «адекватности» выбранного алгоритма необходимо оценивать качество его содержательной составляющей.

В качестве основных метрик авторы [15, 16] выделяют:

- среднюю абсолютную ошибку;
- среднеквадратическую ошибку;
- долю правильных ответов;
- качество ранжирования;
- точность;
- полноту.

Часть указанных метрик в большей степени подходит для проверки качества рекомендательных систем, если они работают не с просмотрами, а с оценками объектов или их рейтингами.

Так как на портале рассматриваемой нами компании не предусмотрена возможность выставления оценок и формирования рейтинга контента, проверка качества рекомендаций будет производиться на основе доли правильных ответов, точности и полноты. Для их расчета часто используют таблицы сопряженности, сопоставляющие рекомендации системы и фактические просмотры объектов пользователями как, например, в табл. 4.

В этом случае метрики рассчитываются следующим образом [7]:

- Доля правильных ответов (accuracy):

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

- Точность (precision) оценивает долю понравившихся пользователю объектов среди предложенных системой:

$$Precision = \frac{TP}{TP + FP}$$

- Полнота (recall) оценивает долю предложенных системой объектов среди всех понравившихся пользователю:

$$Recall = \frac{TP}{TP + FN}$$

Таблица 4

Таблица сопряженности рекомендаций и просмотров¹

		Пользователь	
		Посмотрел	Не посмотрел
Рекомендательная система	Предложила	True-Positive (TP)	False-Positive (FP)
	Не предложила	False-Negative (FN)	True-Negative (TN)

¹ В улучшенном методе выборка делится на три части: обучающую, валидационную и контрольную. Алгоритмы обучаются на обучающей выборке, их качество измеряется на валидационной выборке, и лучший алгоритм повторно проверяется на контрольной выборке. Это позволяет снизить риск обучения и тестирования на неравномерно разбитых данных и получать более корректную оценку качества.

ЗАКЛЮЧЕНИЕ

В настоящей работе были изложены основные подходы, этапы проектирования и разработки рекомендательных систем. Наиболее используемыми в настоящее время подходами являются коллаборативная и контентная фильтрация, которые опираются на схожесть пользователей, объектов и историю их взаимодействия, а также дополнительную описательную информацию.

Построение рекомендательной системы начинается с определения задач, которые она должна решать, и формирования критериев, по которым будут оцениваться результаты работы. Последующие выбор и разработка подходов, методов и алгоритмов должны опираться на имеющуюся технологическую платформу, структуру и характер пользовательских и контентных данных.

Выбор окончательной структуры рекомендательной системы является нетривиальной задачей. Для ее правильного решения необходимо проводить оценку качества наиболее перспективных алгоритмов с последующим выбором наилучшего.

Рекомендательная система, разработанная рассматриваемой нами консалтинговой компанией, оказалась удачной, так как структура системы, выбранная на этапах проектирования, достаточно легко интегрировалась в имеющиеся системы и технологическую платформу в целом.

Показатели качества по выбранным подходам и алгоритмам показали хороший уровень – около 80–85%. После двух месяцев работы системы конверсия по бесплатному контенту в среднем выросла на 7%, по платным мероприятиям – на 4%.

СПИСОК ЛИТЕРАТУРЫ

1. Сычев А.В. Анализ подходов к построению рекомендательных платформ для мультимедийной персонализации B2C в системах электронной коммерции // Материалы XVI международной научно-методической конференции «Информатика: проблемы, методология, технологии» (11–12 февраля 2016 г.). – Воронеж, 2016. – С. 232–237.
2. Филиппов С.А., Захаров В.Н. Метод повышения пертинентности информации в рекомендательных системах поддержки жизнеобеспечения на основе неявных данных // Системы и средства информатики. – 2016. – № 4(26). – С. 4–18.
3. Гомзин А.Г., Коршунов А.В. Системы рекомендаций обзор современных подходов // Труды Института системного программирования РАН. – 2012. – № 22. – С. 401–418.
4. Пятикоп Е.Е. Исследование метода коллаборативной фильтрации на основе сходства элементов // Наукові праці Донецького Національного Технічного Університету. Серія: «Інформатика, кібернетика та обчислювальна техніка». – 2013. – № 2(18). – С. 109–114.
5. Фёдоровский А.Н., Логачева В.К. Архитектура рекомендательной системы, работающей на основе неявных пользовательских оценок // Труды XIII всероссийской научной конференции

- RCDL'2011 «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (Воронеж, 19–22 октября 2011 г.). – 2011. – № 2 – С. 76–82.
6. Jannach D., Zanker M., Felfernig A., Friedrich G. Recommender Systems: An Introduction. New York: Cambridge University Press, – 2010. – 352 p.
7. Нефедова Ю.С. Архитектура гибридной системы GEFEST (Generation–Expansion–Filtering–Sorting–Truncation) // Системы и средства информатики. – 2012. – № 2(22). – С. 176–196.
8. Коцюба И.Ю., Чунаев А.В., Шиков А.Н. Основы проектирования информационных систем: учеб. пособие. – СПб: Университет ИТМО, 2015. – 206 с.
9. ГОСТ 19.102-77. Единая система программной документации (ЕСПД). Стадии разработки. – Введ. 1980-01-01. – М.: Стандартинформ, 2010. – 4 с.
10. Фаулер М., Салладж Дж. NoSQL: новая методология разработки нереляционных баз данных / пер. с англ. – М.: ООО «И.Д. Вильямс», 2013. – 192 с.
11. Мюллер А., Гвидо С. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными / пер. с англ. – СПб: ООО «Альфа-книга», 2017. – 480 с.
12. Рашка С. Python и машинное обучение / пер. с англ. А.В. Логунова.– М.: ДМК Пресс, 2017. – 418 с.
13. Нейский И.М. Классификация и сравнение методов кластеризации // Интеллектуальные технологии и системы. Сб. учебно-методических работ и статей аспирантов и студентов. Вып. 8. – М.: Изд-во ООО «Элике +», 2008. – С. 111–122.
14. Harrington P. Machine Learning in Action. – NY, Shelter Island: Manning Publications Co, 2012. – С. 384.
15. Игнатов Д.И., Каминская А.Ю., Магизов Р.А. Метод скользящего контроля для оценки качества рекомендательных интернет-сервисов // Труды XII национальной конференции по искусственному интеллекту с международным участием КИИ-2010 (20–24 сентября 2010 г., г. Тверь, Россия). Том 1. – М.: Физматлит, 2010. – С. 175–182.
16. Федоренко В.И., Киреев В.С. Анализ подходов к построению гибридных рекомендательных систем в задаче рекомендации фильмов // Международный электронный научно-технический журнал «Теория. Практика. Инновации». – 2017. – № 6(18). – С. 44–50.

Материал поступил в редакцию 14.05.18.

Сведения об авторах

НЕФЕДОВ Юрий Викторович – кандидат экономических наук, доцент, доцент кафедры Управления информационными системами и программирования РЭУ им. Г. В. Плеханова, Москва
e-mail: Nefedov.YV@rea.ru

ГАЛЬПЕРИНА Ирина Александровна – магистр информатики и вычислительной техники, разработчик BI-систем, ООО "ЭЛКОД", Москва
e-mail: irinagal13@gmail.com

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

УДК 001.4:316.4(=11/8)

И.Ю. Кухно, Ю.В. Сложеникина, А.В. Растягаев

Авторский термин и авторская терминология пассионарной теории этногенеза

Доказывается гипотеза, согласно которой создание словаря понятий и терминов теории этногенеза Л.Н. Гумилева как «сферы фиксации» нового оригинального знания обусловлено невозможностью влить «молодое вино в старые мехи».

Показано, что любая авторская теория устанавливает причинно-следственные, иерархические, линейные и прочие отношения, описываемые системой понятий и терминов, которые находятся друг с другом в отношениях выводимости, сопоставленности/противопоставленности, включения и др. Единичный авторский термин может рассматриваться только как компонент авторской терминологии. Изолированного авторского термина не существует.

Ключевые слова: термин, терминология, авторская интенция, терминообразование, Л.Н. Гумилев, теория этногенеза

Концепция пассионарности «до настоящего времени порождает самые противоречивые оценки исследователей ... И если Д. Лихачев, например, утверждал, что открытие пассионарности – одно из крупнейших достижений отечественной науки, ставящее Гумилева в один ряд с такими выдающимися учеными, как В. Вернадский, К. Циолковский, А. Чижевский» [1, с. 202], то другие ученые считают, что далеко не все созданное Л.Н. Гумилевым будет принято научной общественностью.

Один из спорных вопросов пассионарной теории – её метаязык. Анализ словариков исторических словарей показал, что авторская терминология Л.Н. Гумилева находится на периферии принятой в историческом дискурсе терминологии. Так, в «Словаре исторических терминов» В.Г. Згурского зафиксированы всего 4 термина в значениях, принадлежащих терминологии этногенеза: *диахронизм, истина, историзм, пассионарии* [2]; в словаре В.С. Симакова – 10: *пассеизм, популяция, симбиоз, справедливость, творчество, футуризм, химера, этногенез, этнология, явление* [3].

Современная парадигма терминоведения актуализировала коммуникативно-прагматический аспект изучения феномена языка для специальных целей. Л.Ю. Буянова считает, что в основе этого подхода лежит общий для языка вообще и языка для специальных целей, в частности, принцип антропоцентризма. Создатель термина рассматривается как языковая личность, а его речевые акты – как имеющие

установку, намерения (интенциональность), явные или скрытые цели, тактику и правила научной коммуникации, пресуппозицию [4, с. 168].

На основе традиционного учения о речевой деятельности, разработанного отечественными учеными А.Н. и А.А. Леонтьевыми, Л.С. Выготским, Ю.В. Сложеникина предложила коммуникативно-прагматическую модель терминологической деятельности. Терминовтворчество рассматривается наряду с другими видами деятельности и характеризуется с позиций мотива, цели, эвристической направленности и последовательности фаз (ориентировка, планирование, реализация плана, контроль). Процесс терминообразования можно представить следующим образом: 1) мотив, потребность; 2) целеполагание; 3) анализ пропозиции; 4) формулирование авторской интенции; 5) создание фрейма; 6) моделирование понятийной структуры термина; 7) предикация; 8) номинализация; 9) выбор модуса; 10) оценка, авторская рефлексия; 11) узуализация, дефинирование [5, с. 90-95].

Терминовтворчество имеет преднамеренный характер: «особенность терминов заключается в том, что они не даны в языке сами по себе, а творятся по мере осознания их необходимости» [6, с. 25]. Философско-гносеологический подход к термину подчеркивает две его функциональные особенности: термины, наряду с функцией фиксации, выполняют функцию открытия нового знания. Движение науки создает условия для развития терминов и терминологий. Возможность специального слова изменяться

позволяет ему развиваться и подниматься до высоты современного научного осмысления того или иного концепта.

Изучение языковой личности как субъекта продуцирования и развития научного знания в ментально-языковом аспекте связано в так называемым парадоксальным мышлением. Под парадоксом понимают гипотезу, предположение, расходящееся с общепринятым и устоявшимся мнением по какому-либо вопросу. Ученый – автор новой теории – стремится к устранению противоречий, аргументации оригинальной точки зрения. Н.И. Береснева, С.Л. Мишланова, Е.О. Голева видят созидательный смысл «разрушения» ортодоксальной теории: «Существующие противоречия как бы изнутри подрывают научную теорию. Но этот недостаток оборачивается в положительный момент для самой теории ...и, как следствие, приводит к развитию науки» [7, с. 21].

Выдающийся российский историк, географ и философ Л.Н. Гумилев предложил новый ракурс в исторической науке. В 1989 г. вышла в свет монография ученого «Этногенез и биосфера Земли». Эта работа считается его основополагающим трудом. Советская наука категорически отвергла парадоксальную точку зрения исследователя. Рукопись была депонирована в шкаф научных хранений Всесоюзного института технической информации АН СССР в г. Люберцы. На момент создания книги Л.Н. Гумилев еще не мог предложить точные дефиниции терминов этногенеза. Сам автор считал эту книгу своим главным произведением, именно в ней он сформулировал и подробно развил основные положения теории этногенеза и учение о пассионарности. В качестве приложения в книге впервые появляется толковый словарь понятий и терминов. Автор этого словаря В.Ю. Ермолаев – аспирант и ученик Л.Н. Гумилева и первый Председатель его Фонда [8].

В 1993 г. появился еще один «Словарь понятий и терминов теории этногенеза Л.Н. Гумилева» [9]. Его автор В.А. Мичурин – выпускник географического факультета ЛГУ, ученик Л.Н. Гумилева. Словарь напечатан в книге «Этносфера: история людей и история природы». Нужно отметить, что оба словаря готовились при непосредственном участии Л.Н. Гумилева и были опубликованы под его редакцией. В книге «Гумилев сын Гумилева» С. С. Беляков пишет о том, что «настоящим автором словаря был Гумилев. Просто внимательный и настойчивый ученик просил разъяснить смысл каждого термина, умел правильно задать вопрос, а Гумилев разъяснял» [10, с. 179-180].

Словарь В.Ю. Ермолаева – это классический толковый терминологический сборник, в котором левая часть словарной статьи – сам термин, а правая – его дефиниция. Словарь В.А. Мичурина больше тяготеет к энциклопедическому, в нём правая сторона словарной статьи, помимо определения, содержит разнообразные примеры из текстов самого Л.Н. Гумилева.

Предпосылками создания словарей понятий и терминов Л.Н. Гумилева послужили следующие события. В 1976 г. «Вестник ЛГУ» прервал публикацию серии статей Л.Н. Гумилева «Ландшафт и этнос». Главный редактор журнала профессор Л.Е. Смирнов устно ответил автору, что отклоняет статью, потому что она

дискуссионна [11, с. 634]. В адрес Л.Н. Гумилева последовал ряд обвинений от историков, этнографов и социологов в географическом детерминизме, биологизме, отсутствии четкого определения понятия «этнос», вольности в обращении с терминами. Вопрос, связанный с терминопотреблением, стоял довольно остро, и некоторые оппоненты позволяли себе некорректные высказывания в адрес ученого. По мнению А. Янова, «Гумилев и ему подобные до такой степени привыкли к «эзопову языку», что он постепенно стал для них родным» [12, с. 114]. Обилие специальных слов в тексте вызывало недоумение, раздражение и отторжение.

В настоящее время в словарной литературе отсутствует определение авторского термина. Рассуждения на эту тему встречаются либо в монографиях, либо в научных статьях. Н.М. Азарова причисляет к авторским (конкретно – философским) терминам слова, «появившиеся первоначально в текстах одного философа и актуализирующие в настоящее время связь с текстом этого философа» [13, с. 62]. Ученый для квалификации авторских терминов прибегает к понятию так называемого «живого слова». В понятии «живого слова» актуализируется оппозиция любому специальному термину. Авторские термины в целом рассматриваются Н.М. Азаровой как необщепринятые.

Основываясь на теоретических разработках Н.М. Азаровой, Н.В. Козловская предложила классификацию авторских терминов русского философского текста и подразделила их на четыре основных типа: 1) авторские термины, образованные морфологическим (*бытиебоязнь, богооставленность, безмысленность* и др.) и синтаксическим (*цветущая сложность, конкретная метафизика, конец истории* и др.) способами; 2) образованные лексико-семантическим способом (терминологизация общеязыковой лексемы): *отцы, сыны, беспочвенность* и др.; 3) переосмысленные термины других наук: *нация, литургия, троичность* и др.; 4) лексемы и специальные философские термины многоаспектного характера, оригинально интерпретируемые конкретным философом (термины в дискурсивной функции): *свобода, бытие, истина* и др. [14, с. 187].

Попытку подойти к определению авторского термина предприняла В.Д. Табанакова в монографии «Авторский термин: знаю, интерпретирую, перевожу». Потребности максимально точного научного перевода обратили внимание исследователя на семантическую природу термина и потребовали обращения к анализу дискурса. Принципиально новая мысль В.Д. Табанаковой заключается в следующем: «Авторский термин – это такое специальное понятие (выраженное любой структурой предикативного характера), за которым стоит собственно авторское понимание и собственно авторская интерпретация» [15, с. 168].

С нашей точки зрения, представления об авторском вкладе должны выходить на уровень гипотезы, теории. Ученый рассматривается прежде всего как основоположник оригинальной научной концепции. Эта концепция, в терминологии А.А. Реформатского, ограничивается системой координат, «координатной сеткой», четко отделяющей креативную концепцию

от общепринятой. Авторская теория устанавливает причинно-следственные, иерархические, линейные и проч. отношения, описываемые системой понятий и терминов. Эти термины находятся друг с другом в отношениях выводимости, сопоставленности/противопоставленности, включения и др. Единичный авторский термин, на наш взгляд, может рассматриваться только как компонент авторской терминологии. Изолированного авторского термина не существует [16, с. 57].

Можно утверждать, что оригинальная авторская гипотеза строится вокруг одного или нескольких (немногих) базовых понятий. Центральным для пассионарной теории Л.Н. Гумилева является понятие специфически представленного *этногенеза*. В рамках изучения исторической науки теория этногенеза Л.Н. Гумилева демонстрирует спектр разнообразных проблем: закономерности изменений в этногенезе, присущие любому этносу, единообразие всех процессов этногенеза (ученый выделил семь фаз этногенеза: *фаза подъема, акматическая фаза, фаза надлома, инерционная фаза, фаза обскурации, фаза регенерации и мемориальная фаза*) и др. Понятие этногенеза является терминологическим ключом (понятие А.А. Реформатского), с помощью которого можно открыть авторский замысел историка.

Сформулируем дефиницию: *авторский термин* – созданный в рамках креативной научной концепции специальный знак, элемент системы понятий, взаимосвязанный в другими терминами и понятиями оригинальной гипотезы, соотносимый носителями языка для специальных целей с конкретной авторской идеей, необщепринятой научной картиной мира.

Рассмотрим способы образования терминов пассионарной теории этногенеза Л.Н. Гумилева.

1. Терминологизация общеупотребительных слов.

А. Переосмысление слова. Первую группу терминов этногенеза составляют слова русского литературного языка (как иноязычные, так и исконно русские по происхождению), употребленные Л.Н. Гумилевым в функции терминов: *адаптация, бездна, биосфера, воля, деяния, идеал, красота, ксения, мироощущение, мозаичность, ностальгия, рассудок, смерть, событие, творчество, химера, этногенез, этнология, этнос*, а также некоторые другие. Приведем пример (табл. 1).

Красота – термин, образовавшийся на базе общеупотребительного слова.

Общеупотребительное слово может определять как форму, так и содержание. У Л.Н. Гумилева термин имеет более конкретное значение: «комплекс форм», то есть он связан с внешними характеристиками. Если на уровне дифференциальных сем слово связано с обозначением очень сильной положительной эмоции, высшей степени удовольствия («наслаждение»), то термин *красота* отсылает к мыслительной деятельности: то, что нравится без предубеждения («предвзятость»).

Смерть – значение термина коренным образом разнится с общеупотребительным. По Л.Н. Гумилеву смерть оказывается одним из способов существования.

Б. Переосмысление словосочетания: *историческая судьба, история культуры, логика событий, полоса свободы, пусковой момент, традиции культуры*. Обратимся к примерам.

Пусковой момент – в терминологическом словаре Л.Н. Гумилева «момент пассионарного толчка в абсолютной временной шкале, являющейся точкой отсчета «0» для диахронической шкалы». И в общелитературном, и в специальном употреблении *пусковым моментом* называется начальный этап движения, действия. Разница в значении связана с процессом конкретизации терминологического значения. Она проявляется в наличии дополнительных сем в дефиниции термина: субъектной и временной.

Традиции культуры – Л.Н. Гумилев предлагает дефиницию: «сумма знаний и представлений, передаваемая по ходу времени от этноса к этносу». Разница между литературным и специальным употреблением заключается в толковании понятия «культура». В узусе – это совокупность различного рода достижений народа, цивилизации; у Л.Н. Гумилева – это знания и представления.

2. Вторичная терминологизация (транстерминологизация). Под транстерминологизацией понимают переход специальной единицы из одной терминологии в другую. В данном процессе важно наделение термина новой, отличной от прежней дефиницией, то есть в результате транстерминологизации возникают специальные омонимы. Мы зафиксировали два структурных типа терминов, заимствованных Л.Н. Гумилевым из других наук для нужд метаязыка теории этногенеза: термины-слова и термины-словосочетания.

Таблица 1

Общеупотребительное значение	Термин Л.Н. Гумилева
<i>Красота</i> – все красивое, прекрасное, все то, что доставляет эстетическое и нравственное наслаждение	<i>Красота</i> – комплекс форм, нравящихся без предвзятости
Общеупотребительное значение	Термин Л.Н. Гумилева
<i>Смерть</i> – прекращение жизнедеятельности организма)	<i>Смерть</i> – способ существования биосферных феноменов, при котором происходит отделение пространства от времени

А. Переосмысление термина-слова. В данную группу мы включаем единицы, которые уже существовали в тех или иных специальных сферах как термины. Л.Н. Гумилев закрепил за устоявшимися формами новые дефиниции. Среди них: *актуализм, аннигиляция, дивергенция, консорция, констелляция, пассаизм, саморегуляция, симбиоз, старение, сукцессия, флуктуация, футуризм*, некоторые другие. Рассмотрим пример (табл. 2).

Симбиоз – термин появился во второй половине XIX века. Авторство принадлежит российским естествоиспытателям П.А. Кропоткину и К.Ф. Кесслеру.

Общая сема определяет бытийность, но в одном случае речь о взаимовыгодном сожительстве, а в другом – об изолированности. Можно сказать, что гумилевский термин антонимичен биологическому по семе вместе/врозь.

Старение – термин, характеризующий процесс жизнедеятельности организма.

У Л.Н. Гумилева процесс, названный *старением*, касается не только отдельного индивида, но распространяется на совокупность людей, этническую систему. Для термина этнологии важна сема причины, обозначенная в дефиниции: «потеря инерции пассионарного толчка».

Б. Переосмысление терминологического словосочетания. Словосочетания, ставшие субстратом в тер-

миносистеме этнологии: *время историческое, инкубационный период, сигнальная наследственность, энтропийный процесс*. Рассмотрим пример (табл. 3).

Сигнальная наследственность – биологический термин.

В целом, Л.Н. Гумилев взял за основу дефиницию биологического термина, сузив ее результативной семой «приспособление».

Специальное словосочетание *энтропийный процесс* входит в систему физических и химических терминов.

Оба термина объединены идеей процессуальности. Различие на семантическом уровне описывается следующим образом: в общепринятом термине речь идет о норме и аномалии; в термине Л.Н. Гумилева эта аномалия конкретизируется 2 признаками: необратимость, лишение.

3. Авторские неологизмы. Ряд терминов не фиксируется ни общеупотребительными, ни терминологическими словарями русского языка. Они созданы Л.Н. Гумилевым только для потребностей нового научного направления. Терминологические неологизмы, предложенные Л.Н. Гумилевым для трансляции положений этнологии, относятся к двум видам: 1) образованные морфологическим способом; 2) образованные синтаксическим способом.

Таблица 2

Значение научного термина	Термин Л.Н. Гумилева
<i>Симбиоз</i> – биол. Различные формы совместного существования разноименных организмов, составляющих симбионтную систему	<i>Симбиоз</i> – сосуществование двух и более этносов в одном регионе, когда каждый занимает свою экологическую нишу
Значение научного термина	Термин Л.Н. Гумилева
<i>Старение</i> – биол. Закономерный разрушительный процесс возрастных изменений организма, ведущий к снижению его адаптационных возможностей	<i>Старение</i> – потеря инерции пассионарного толчка в этнической системе на персональном и этническом уровне

Таблица 3

Значение научного термина	Термин Л.Н. Гумилева
<i>Сигнальная наследственность</i> – биол. Передача поведенческих навыков потомству через условный рефлекс подражания	<i>Сигнальная наследственность</i> – передача навыков потомству через условный рефлекс подражания, формирующий стереотип поведения как высшую форму адаптации
Значение научного термина	Термин Л.Н. Гумилева
<i>Энтропийный процесс</i> – физ. отклонение системы от эталонного состояния	<i>Энтропийный процесс</i> – необратимый процесс потери энергии

А. Термины, образованные морфологическим способом:

- Префиксация: *антисистема*, *сверхнапряжение*, *субэтнос*.

По Л.Н. Гумилеву, *антисистема* – системная целостность людей с негативным мироощущением. Ученый в дефиниции оставляет основное значение слова *система* – «целостность людей», а отрицательную семантику связывает с человеческим фактором, ограничивая понятие семей «чувства, состояния, ощущения». Таким образом, в толкование термина вводится качественный конкретизатор с негативным, отрицательным компонентом за счет словообразовательного элемента *-анти*.

Субэтнос. Словообразовательный элемент субэтимологически аккумулирует значение под-, ниже, внизу, снизу, т.е. имеет значение неполноты, вторичности. У Л.Н. Гумилева *субэтнос* – это этническая система, являющаяся элементом структуры этноса.

- Суффиксация: *биполярность*, *гармоничники*, *пассионарность*, *пассионарий*.

Гармоничники (*гармоничные особи*). В терминологии Л.Н. Гумилева – это особи, пассионарный импульс которых равен по величине импульсу инстинкта самосохранения. Термин указывает на соразмерность поведения индивида. Если в общеобиходном значении термин имеет категориальное значение «слаженность, соразмерность, стройность чего-либо», то в терминологии этногенеза семантически слово сужается, речь идет только о людях, конкретно – об уравновешенности потребности в изменениях, стремлении к самосохранению. Отмечаем суффиксы: *-н* и *-ик*, характерные для обозначения лица по свойству или признаку, определяющему его отношение к предмету, занятию. Л.Н. Гумилев именуется явление, метафорически сближая образ жизни небольшой группы людей и наименование лица по значению имени.

Пассионарии – особи, пассионарный импульс поведения которых превышает величину импульса инстинкта самосохранения. Общей корневой морфемой у литературного слова *пассионарный* и термина *пассионарий* является морф *пассионар*. Термин образован с помощью суффикса *-ий*. В теории словообразования такие производные слова характеризуются как отадеквативные существительные со значением лица, названного по признаку, представленному в производящей основе.

- Бесфлексивный способ: *геохор*.

Ранее флексивный способ в отечественном словообразовании не выделялся. В современной теории

деривации – это морфологический способ образования производных слов при помощи окончания, выражающего в этом случае грамматическое и словообразовательное значения. Поскольку в нашей ситуации производное слово образуется усечением флексии, мы сочли возможным назвать этот способ словопроизводства безфлексивным.

Геохор – термин, созвучный с географическим «геохора». В данном случае в метаязыке Л.Н. Гумилева изменилась морфологическая принадлежность термина: нулевое окончание перевело лексему в имя существительное мужского рода второго склонения. Термин претерпел и семантические изменения. Родственность в значении присутствует, но у Л.Н. Гумилева термин демонстрирует сужение значения, называя только часть от целого, в то время как географический термин определяет геологическую систему. В толковании общепринятого термина *геохора* (геосистема) указывается на принципиальную разнородность составляющих системы, у Л.Н. Гумилева, наоборот, речь идет об однородном пространстве (участок земной поверхности). Можно сказать, что термин Гумилева антонимичен принятому в геологии.

Б. Термины, образованные синтаксическим способом: *аберрация близости*, *аберрация дальности*, *аберрация состояния*, *генетический «дрейф»*, *диахроническая шкала*, *императив поведения*, *импульс инстинкта*, *импульс поведения*, *импульс сознания*, *пассионарная индукция*, *пассионарное напряжение*, *пассионарное поле*, *пассионарный импульс*, *пассионарный признак*, *пассионарный толчок*, *разнообразие этносферы*, *родина этноса*, *этническая диагностика* и некоторые др.

Л.Н. Гумилев терминологизировал понятие *аберрация*. Как термин лексема используется в физике, астрономии, биологии. У слова есть вещественное значение, когда речь идет о неком теле или изображении. На основе этого конкретного значения развилось переносное значение «заблуждение, отклонение от истины». Л.Н. Гумилев заимствовал лексему, назвав *аберрацией* преувеличение, расплывчатость. На базе родового термина он создал 3 видовых: *аберрация близости*, *аберрация дальности*, *аберрация состояния* (табл. 4). Дифференциальная сема связана с временем фиксации события, она сужает значение за счет временных маркеров: «недавние события», «далекие явления», «долгоидущий процесс», «медленность восприятия».

Таблица 4

Значение научного термина

Аберрация – биол. В морфологии и физиологии всякое отклонение в строении и функциях от типичного образа

Термин Л.Н. Гумилева

Аберрация близости – преувеличение грандиозности недавних событий.

Аберрация дальности – расплывчатость далеких явлений.

Аберрация состояния – восприятие наблюдателем динамики долгоидущего процесса

Большое гнездо терминов связано с авторским определением *пассионарный*. Именно это слово отождествляется с теорией этногенеза. Данная группа терминов представлена в совокупности следующим образом: *пассионарная индукция, пассионарное напряжение, пассионарное поле, пассионарный импульс, пассионарный признак, пассионарный толчок*. Этимология слова *пассионарность* восходит к лат. *passionem* – «страдание, страсть». Специальное прилагательное выражает, по Л.Н. Гумилеву, специфическое для каждого этноса и внутренне присущее этносу и отдельному индивиду стремление к обновлению и развитию, осуществляемое независимо от внешних обстоятельств.

Ряд терминов начинаются с лексемы *этническая* *-ий), (ие): диагностика, дивергенция, доминанта, иерархия, история, регенерация, традиция, контакты, гомеостаз, субстрат, поле*. Такое обширное гнездо терминов обуславливается ключевым понятием пассионарной теории – «этнос» Л.Н. Гумилев неоднократно предлагал новые, дополненные, переработанные дефиниции этого термина. Для ученого этнос определялся через понятие энергетической системы (или структуры), устанавливающей разнообразные отношения с другими человеческими сообществами.

Термины пассионарной теории этногенеза Л.Н. Гумилева образованы на базе существующих корней, слов и словосочетаний общелитературной и специальной лексики. Среди гумилевских терминов встречаются все структурные типы, которые характерны для лексики русского национального языка, поскольку деривация в языке науки основана на системе словообразования литературного языка. В конце своего жизненного и научного пути, публикуя «Биографию научной теории, или Автонекролог», Л.Н. Гумилев сетовал, что пассионарная теория этногенеза «не вызвала никакого интереса у историков-источниковедов, филологов и востоковедов. А жаль. Она и там нашла бы применение» [17, с. 215]. Представленная работа явилась первой попыткой филологического осмысления оригинальной исторической концепции и метаязыка ее описания.

СПИСОК ЛИТЕРАТУРЫ

1. Пушкин С.Н. Концепция пассионарности в творчестве Л.Н. Гумилева // Вестник Нижегородского университета им. Н.И. Лобачевского. – 2012. – № 1-3. – С. 202–207.
2. Згурский Г.В. Словарь исторических терминов. – М.: Эксмо, 2008. – 457 с.
3. Словарь исторических терминов / сост. В.С. Симмаков; под ред. А.П. Крюковских. – СПб.: Лита, 1998. – 464 с.
4. Буянова Л.Ю. Термин как единица логоса. – 3-е изд., стер. – М.: Флинта: Наука, 2013. – 218 с.
5. Сложеникина Ю.В. Основы терминологии: Лингвистические аспекты теории термина. – М.: Либроком, 2013. – 120 с.
6. Володина М.Н. Теория терминологической номинации. – М.: Изд-во МГУ, 1997. – 197 с.

7. Береснева Н.И., Мишланова С.Л., Голева Е.О. Парадоксальные конструкции языка науки // Исторические, философские, политические и юридические науки, культурология и искусствоведение. Вопросы теории и практики. – 2015. – № 1–1 (51). – С. 21–23.
8. Ермолаев В.Ю. Толковый словарь понятий и терминов // Гумилев Л.Н. Этногенез и биосфера Земли / под ред. Л.Н. Гумилева. – Л.: Изд-во ЛГУ, 1989. – С. 391–399.
9. Мичурин В.А. Словарь понятий и терминов теории этногенеза Л.Н. // Гумилев Л.Н. Этносфера: история людей и история природы Гумилева / под ред. Л.Н. Гумилева. – М.: Экопрос, 1993. – С. 493–542.
10. Беляков С.С. Гумилев сын Гумилева. – М.: Астрель, 2013. – 800 с.
11. Гумилев Л.Н. Этногенез и биосфера Земли. – М.: Ин-т ДИ-ДИК, 1997. – 637 с.
12. Янов А. Учение Льва Гумилева // Свободная мысль. – 1992. – № 17. – С. 104–116.
13. Азарова Н.М. Типологический очерк языка русских философских текстов XX в. – М.: Логос:Гнозис, 2010. – 250 с.
14. Козловская Н.В. Типы авторских терминов в русской философии (на материале произведений Н.Ф. Федорова и Н.А. Бердяева) // Сибирский филологический журнал. – 2016. – № 1. – С. 185–192.
15. Табанакова В.Д. Авторский термин: знаю, интерпретирую, перевожу. – Тюмень: Изд-во ТюмГУ, 2013. – 208 с.
16. Сложеникина Ю.В., Растягаев А.В., Кухно И.Ю. Авторский термин: к определению понятия // Онтология проектирования. – 2018. – Т. 8, 1. – С. 49–57.
17. Гумилев Л.Н. Биография научной теории, или Автонекролог // Знамя. – 1988. – № 4. – С. 202–216.

Материал поступил в редакцию 16.05.18.

Сведения об авторах

КУХНО ИРИНА ЮРЬЕВНА – аспирант кафедры филологии и массовых коммуникаций филологического факультета Московского городского педагогического университета (Самарский филиал), г. Самара e-mail: irina.kuhno@yandex.ru

СЛОЖЕНИКИНА Юлия Владимировна – доктор филологических наук, доцент, декан филологического факультета Московского городского педагогического университета (Самарский филиал), профессор кафедры филологии и массовых коммуникаций, г. Самара e-mail: goldword@mail.ru

РАСТЯГАЕВ АНДРЕЙ ВИКТОРОВИЧ – доктор филологических наук, доцент, заведующий кафедрой филологии и массовых коммуникаций филологического факультета Московского городского педагогического университета (Самарский филиал), г. Самара e-mail: avr67@yandex.ru

ВНИМАНИЮ ПОДПИСЧИКОВ!

С 2018 года возобновляется издание информационного бюллетеня «Иностранная печать об экономическом, научно-техническом и военном потенциале государств-участников СНГ и технических средствах его выявления» серии «Экономический и научно-технический потенциал» (56741) взамен информационного бюллетеня «Экономика и управление»

Периодичность выхода – 12 номеров в год. Объем 48 уч.-изд. л. в год.

В бюллетене освещаются материалы иностранной печати по широкому спектру вопросов, касающихся сфер экономического и научно-технического развития России и стран СНГ: общие вопросы, финансы, промышленность, рынки, сельское хозяйство, космос, транспорт и связь, природные ресурсы, трудовые ресурсы, внешние торгово-экономические и научные связи

Оформить подписку на информационный бюллетень, начиная с любого номера, можно в ВИНТИ РАН по адресу: 125190, Россия, Москва, ул. Усиевича, 20,

Телефоны: (499) 151-78-61; (499) 155-42-85

Факс: (499) 943-00-60;

E-mail: contact@viniti.ru; sales@viniti.ru

База данных (БД) ВИНИТИ РАН

Федеральная база отечественных и зарубежных публикаций по естественным, точным и техническим наукам, генерируется с 1981 г., обновляется ежемесячно, пополнение составляет около 1 млн документов в год. Тематическое наполнение соответствует реферативному журналу ВИНИТИ. Для поиска одновременно по всем или нескольким тематическим фрагментам генерируется единая Политематическая БД.

БД ВИНИТИ РАН в сети INTERNET

Сервер ВИНИТИ - <http://www.viniti.ru> – обеспечивает on-line доступ к Базе данных ВИНИТИ РАН круглосуточно без выходных.

На основе БД ВИНИТИ РАН предоставляются следующие услуги:

- Диалоговый поиск научно-технической информации **в режиме on-line**;
- **Демо-версия**, позволяющая ознакомиться с основными функциями поисковой системы, составом данных, формами представления документов и получить навыки работы с системой;
- **Поисковые эксперты ВИНИТИ** выполняют тематический поиск по разовым или постоянным запросам, а также окажут **консультационные услуги**.

БД ВИНИТИ РАН на CD-ROM

Любые наборы тематических фрагментов БД ВИНИТИ или их разделов за любой период с 1981 г., а также **проблемно-ориентированные выборки** из БД ВИНИТИ по актуальным направлениям научных исследований могут быть предоставлены на договорной основе **в поисковой системе (ИПС) "Сокол"**, работающей под управлением Microsoft Windows и обеспечивающей следующие возможности:

- **Чтение** документов в режиме последовательного просмотра или выборочно по оглавлению за весь период заказанной ретроспективы
- **Поиск** документов по автору, заглавию, источнику, ключевым словам или словосочетаниям, реферату, рубрикам, году издания, стране, языку и т.д. (всего более 20 признаков)
- **Словарь** системы поможет правильно подобрать термины для поиска и выбрать глубину их усечения.
- Для **уточнения поиска** можно дополнительно использовать год издания документа, язык текста документа, рубрики, шифры тематических разделов БД.
- Выполненные **запросы можно сохранять** для их последующего использования и/или редактирования.

125190, г. Москва, ул. Усиевича, 20, БД ВИНИТИ РАН.

Отдел взаимодействия с потребителями – (499) 155-45-25, (499) 152-58-81

E-mail: csbd@viniti.ru, sales@viniti.ru

WWW: <http://www.viniti.ru>

УВАЖАЕМЫЕ КОЛЛЕГИ!

ВИНИТИ РАН предлагает Вашему вниманию Реферативный Журнал в электронной форме

РЖ в электронной форме (ЭлРЖ) выпускается по всем разделам естественных, технических и точных наук.

Каждый номер ЭлРЖ является полным аналогом печатного номера РЖ по составу описаний документов, их оформлению и расположению. Он сопровождается оглавлением, указателями.

ЭлРЖ представляет собой информационную систему, снабженную поисковым аппаратом и позволяющую пользователю на персональном компьютере:

- читать номер РЖ, последовательно листая рефераты;
- просматривать рефераты отдельных разделов по оглавлению;
- обращаться к рефератам по указателям авторов, источников, ключевых слов;
- проводить поиск документов по словам и словосочетаниям;
- выводить текст описаний документов во внешний файл.

ЭлРЖ в версии Windows Вы можете получить за текущий год с любого номера, а также за предыдущие годы.

Подробную информацию Вы можете получить:

Адрес: 125190, Россия, Москва, ул. Усиевича, 20, ВИНТИ РАН

Коммерческое управление

Телефон/Факс: 8 (499) 155-45-25, 8 (499) 152-58-81

E-mail: contact@viniti.ru, sales@viniti.ru