

Федеральное государственное автономное образовательное учреждение высшего  
образования  
«Национальный исследовательский ядерный университет «МИФИ»  
(НИЯУ МИФИ)

**АРТАМОНОВ Алексей Анатольевич**

**Модели, методы и технологии интеллектуального анализа  
информационных объектов в научно-технических и социально значимых задачах**

Специальность 2.3.1 – Системный анализ,  
управление и обработка информации, статистика

**Автореферат**  
диссертации на соискание ученой степени  
доктора технических наук

Научный консультант  
д.т.н., с.н.с. Кореньков Владимир Васильевич

Москва – 2025

## ВВЕДЕНИЕ

### Актуальность работы

В современных условиях стремительного роста объёмов информации и развития технологий искусственного интеллекта необходима разработка новых подходов к интеграции и анализу данных для поддержки принятия решений в научно-технической и социальной сферах. В различных предметных областях накапливаются огромные массивы разнородных данных: научные публикации, патенты, отчёты, а также цифровые следы человеческой деятельности в социальных сетях и других онлайн-сервисах. Интеллектуальный анализ данных (ИАД) стал одним из ключевых инструментов для извлечения знаний из таких массивов. Он сочетает методы машинного обучения, статистического анализа, обработки неструктурированных данных и визуальной аналитики, что открывает возможности выявления скрытых закономерностей и принятия более обоснованных решений на основе имеющихся данных. Классические подходы зачастую не справляются с актуальными требованиями: данные поступают непрерывно и в разнообразных форматах, содержащаяся в них информация слабо структурирована, а существующие системы нередко ориентированы на заданные типы данных. Это приводит к необходимости разработки новых методологических принципов и комплексных инструментов для интеллектуального анализа информации.

В научно-технической сфере задача анализа данных важна для отслеживания развития научных направлений, выявления технических трендов и управления знаниями. В социальной сфере интеллектуальный анализ цифровых профилей позволяет решать социально значимые задачи, например, раннее обнаружение групп риска (по поведенческим индикаторам в соцсетях) или противодействие информационным угрозам. Однако междисциплинарный характер подобных задач выявляет ряд проблем.

Во-первых, необходимо обеспечить единый подход к описанию и хранению разнородных данных, чтобы можно было интегрировать информацию из разных источников (научные статьи, посты в соцсетях, базы данных и т.д.) в единую систему.

Во-вторых, требуется поддерживать высокое качество и достоверность данных: без очистки, нормализации и верификации исходной информации выводы анализа могут оказаться неверными.

В-третьих, сложность алгоритмов ИАД порождает проблему интерпретируемости и доверия – пользователям и экспертам важно понимать, на каких фактах основаны те или иные прогнозы или рекомендации.

В-четвертых, при работе с социальными данными возникают дополнительные ограничения, связанные с конфиденциальностью и этикой использования персональной информации.

Актуальность работы обусловлена потребностью в новых моделях, методах и технологиях, обеспечивающих эффективный интеллектуальный анализ информационных объектов различной природы, учитывая перечисленные вызовы. Разработка интегративной системы интеллектуального анализа

данных позволит систематизировать и объединять неструктурированные данные из научно-технической и социальной областей, извлекать из них новые знания с высокой степенью достоверности, а также автоматизировать значительную часть трудоёмких аналитических процессов, что будет способствовать повышению качества управления научными исследованиями и социальными проектами, раннему выявлению важных тенденций и решению практических задач в интересах науки, экономики и общества. Факторологически подтверждается, что объём и разнообразие данных будут только расти, а потому создание интеллектуальных инструментов для их анализа является актуальной и стратегически значимой научной задачей.

#### **Степень научной разработанности темы исследования.**

Фундаментальные основы интеллектуального анализа данных заложены в трудах G. Piatetsky-Shapiro, R. Agrawal, J. Han, T.M. Mitchell, W.F. Frawley, B. A. Дюка, В. В. Коренькова, Ю. С. Сахарова, И. Г. Благовещенского и А. В. Замятиной. Эти исследователи сформировали концептуальный аппарат и методологические принципы Data Mining, разработали базовые алгоритмы поиска ассоциативных правил, классификации и кластеризации, создав теоретический фундамент для извлечения знаний из данных.

В области интеграции и управления разнородными данными значительный вклад внесли А. И. Аветисян, В. И. Будзко, Ю. А. Зеленков, А. Е. Янковская, M. Stonebreaker, J. Gray, R. Kimball, I.F. Ilyas, D. Abaid. Разработанные этим научным сообществом подходы, включая концепций озер данных и хранилищ больших объемов данных, онтологического инжиниринга и управления метаданными, обеспечивают технические основы для консолидации разноформатных данных, однако остаются ограниченными в решении задач семантической интеграции разнородных информационных объектов.

Качество и предобработка данных исследовались в работах C. Fan, E. E. Simoudis, H.G. Miller, J.M. Hellerstein, D. Barbara, Д. Барбары, В. Б. Яковлева, С. В. Ключарева, А. Н. Тихонова, М. Г. Шеразадишвили и П. С. Бондаренко. Разработанные этим сообществом методы очистки, восстановления и верификации информации создают важный задел для обеспечения достоверности анализа, однако носят фрагментарный характер и не образуют целостной технологии сквозной предобработки.

Проблемы интерпретируемости и этики в интеллектуальном анализе данных активно исследуются К. В. Воронцовым, А. Ю. Дьяконовым, С. В. Мирошниковым, K.T. Lundberg, M.T. Ribeiro, S.M. Fatemi, S.M. Hosseini, C.D. Waitz. Их работы в области объяснимого искусственного интеллекта и дифференциальной приватности создают теоретическую базу для построения доверяемых аналитических систем, однако не решают в полной мере проблему «семантического разрыва» между сложными моделями ИАД и потребностями экспертов в интерпретируемых результатах.

В диссертационной работе предложено комплексное решение, включающее формальную модель представления цифровых информационных объектов, методы автоматизированного сбора, обработки, насыщения и хранения данных, а также специализированные программные средства для их анализа и

визуализации. Интеграция разнородных источников информации на основе единой модели и применение совокупности методов ИАД позволит получать новые знания и решения в междисциплинарных задачах быстрее и надёжнее, чем при использовании разрозненных подходов.

**Объектом исследования** являются цифровые информационные объекты, формируемые в процессе научно-технической деятельности и социальной коммуникации, а именно данные научных публикаций, патентных описаний, отчетов, записей социальных сетей и иных источников, значимые для решения научно-технических и социальных задач.

**Предметом исследования** являются методы, модели, алгоритмы и программные средства интеллектуального анализа данных цифровых информационных объектов, обеспечивающие интеграцию разнородных данных в единую модель, насыщение характеристик цифровых объектов и выявление неявных знаний для решения научно-технических и социально значимых задач.

**Цель работы** состоит в разработке и обосновании системы интеллектуального анализа информационных объектов, объединяющей модели представления данных, методы автоматизированного извлечения значений характеристик из разнородных ресурсов, методы насыщения данных, механизмы визуальной аналитики для выявления явных и неявных закономерностей и поддержки принятия решений в научно-технических и социально значимых задачах.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Обобщение модели комплексного цифрового информационного объекта, объединяющей разные виды характеристик (статические, динамические, вычисляемые) и отношения между объектами.
2. Разработка методов и алгоритмов автоматизированного извлечения информации из разнородных источников (научные статьи, веб-документы) и насыщения данных (выделение ключевых сущностей, распознавание и нормализация физических величин, обработка изображений и таблиц, определение международных альянсов) для наполнения модели цифрового информационного объекта, обеспечивающих учёт специфики научно-технической информации.
3. Разработка методов идентификации целевых социальных объектов, в том числе алгоритмов анализа текстовых полей профилей цифровых объектов для обнаружения лингвистических индикаторов; методов анализа динамических характеристик активности цифровых объектов для выявления аномальных паттернов; метода расчета обобщённой вычисляемой характеристики, интегрирующей значения характеристик модели цифрового информационного объекта, и определения ее порогового значения для отнесения цифровых объектов к целевой группе.
4. Выбор и обоснование подходов к визуальному анализу научно-технической информации, позволяющих выявлять тенденции, ведущие организации, неявные связи, кластеры тематических направлений,

международное сотрудничество на основе построения интерактивных аналитических панелей, обеспечивающих интеграцию различных цифровых объектов; построению графовых моделей, отражающих связи между объектами.

5. Проектирование и реализация программных средств, интегрирующих предложенные модели и методы в единый программно-аналитический комплекс, обеспечивающий функциональность поиска, выявление как явных, так и неявных связей между объектами, построение аналитических срезов научно-технологической и социальной сфер.

**Научная новизна** состоит в разработке и обосновании нового комплексного подхода к анализу разнородной информации, включающего оригинальные модели и методы. Основные научные результаты, полученные автором, заключаются в следующем:

1. **Предложена и обоснована обобщённая аналитическая модель комплексного цифрового информационного объекта**, объединяющая статические, динамические и вычисляемые характеристики, а также систему связей между объектами. Разработанная модель обеспечивает единое представление разнородной научно-технической и социальной информации, что упрощает подготовку данных к анализу и повышает точность и воспроизводимость аналитических результатов за счёт стандартизации структуры данных.

2. **Разработан новый методический аппарат насыщения данных из неструктурированных разнородных источников.** Предложены оригинальные методы автоматизированного извлечения и насыщения данных, отличающиеся учётом специфики научно-технической информации, в том числе: метод распознавания физических величин и единиц измерения с их приведением к единому стандарту (СИ); метод обработки нетекстовых элементов научных документов – извлечение и структурирование содержимого таблиц и подписей к рисункам и др. Новизна указанных методов заключается в адаптации технологий обработки данных (NLP, OCR, геокодирование и др.) к задачам научно-технического контента, что позволяет существенно обогатить исходные данные и извлечь новые знания, недоступные при стандартной обработке информационных материалов.

3. **Предложена методика идентификации целевых объектов в социальной среде**, использующая методы семантического анализа текстовой информации профиля цифрового объекта (посты, комментарии, описания), обработки естественного языка и анализа тональности, количественной оценки динамических характеристик активности пользователя (частота и время публикаций, смена аудитории), ранжирования и нормирования множества характеристик профиля с учётом их значимости.

4. **Предложен метод построения интерактивных аналитических панелей**, позволяющий работать со слабоструктуризованными массивами научной информации для проведения сравнительного анализа динамики развития различных областей науки, комплексного обзора научно-технологического ландшафта: выявление лидеров и динамики

публикационной активности, основных исследовательских трендов, картографирования международного сотрудничества. Впервые реализован междисциплинарный анализ публикационной активности, показавший эффективность при выявлении скрытых тематических акцентов и точек роста.

**5. Сформирована система интеллектуального анализа данных для решения широкого спектра актуальных научно-технических и социально значимых задач**, базирующаяся на обобщённой модели описания цифровых информационных объектов, методах преобразования информации из различных источников и визуальной аналитики больших объемов разнородных данных.

**Теоретическая значимость** работы заключается в развитии методологических основ интеллектуального анализа данных применительно к разнородным информационным объектам в научно-технической и социальной сферах. Предложенные автором модели формализуют новый подход к представлению знаний – через комплексный цифровой информационный объект с множеством разнотиповых (статических, динамических и вычисляемых) характеристик и связей. Разработанные модели и методы расширяют аппарат интеллектуального анализа данных: введены новые вычисляемые характеристики и метрики для анализа социальных и научных данных, предложены новые методы извлечения знаний из неструктурированных данных. Полученные в работе обобщения (например, категоризация характеристик социальных профилей, формулы нормировки признаков, концепция эволюции модели цифрового объекта) могут служить основой для дальнейших исследований по интеграции данных и развитию теории представления знаний. Полученные результаты закладывают методологический фундамент для развития гибких аналитических систем, объединяющих машинное обучение, базы знаний и механизмы визуализации. Теоретическая ценность работы состоит в междисциплинарном подходе, который объединяет концепции информатики (структуры данных, алгоритмы), искусственного интеллекта (машинное обучение, NLP), социологии (психографический анализ) и научометрического анализа в рамках единой научной парадигмы. Таким образом, диссертация вносит значимый вклад в развитие теории интеллектуального анализа данных, предлагая новые модели и методы, расширяющие границы применимости интеллектуального анализа данных.

**Практическая значимость результатов исследования** подтверждается их внедрением и опытным применением в ряде проектов. Разработанные модели и методы легли в основу программных средств, использованных для решения прикладных задач в интересах ведущих научных организаций. С 2018 года автор являлся руководителем или ответственным исполнителем 10 хоздоговорных работ в интересах Министерства образования и науки Российской Федерации, организаций контура Госкорпорации Росатом (НИИ «Графит», ВНИИА им. Н.Л. Духова, ФГУП «РФЯЦ-ВНИИТФ им. академ. Е.И. Забабахина»), Фонда перспективных исследований, Российского энергетического агентства. В рамках государственного задания Министерства

образования и науки РФ №2.12915.2018.12.1 «Разработка и апробация информационной системы комплексной антисуициdalной интернет-профилактики» был реализован и прошёл тестирование метод идентификации целевых социальных профилей, показав работоспособность в реальных условиях мониторинга социальных сетей. По государственному заданию Министерства науки и высшего образования Российской Федерации №3466-22 «Создание учебно-методических материалов по финансовой безопасности для школьников и студентов, в том числе для передачи указанных учебно-методических материалов в зарубежные страны-партнеры Международного сетевого института в сфере противодействия отмыванию доходов, полученных преступным путем, и финансированию терроризма» апробированы методы построения интерактивных аналитических панелей и проведено исследование состояния научно-технических разработок по направлению «Финансовая безопасность». По договору №349ГС1ЦТС10-D5/80243 от 12.12.2022 «Разработка и тестирование прототипа мультиагентной системы обработки и представления неструктурированных массивов данных» с Фондом содействия инновациям создан программный комплекс «СИА.Атташе», предназначенный для интеллектуального анализа научно-технической информации, прошедший апробацию по договорам № 2024-sia-dgk-1 от 15.04.2024 и № 2024-sia-dgk-2 от 15.05.2024. Практический эффект от использования комплекса выражается в существенном сокращении времени на сбор и обработку данных, например, по договору № 1707 от 29 августа 2022 г. по выполнению НИР «Разработка программы выборки данных по свойствам и структурам облученных реакторных материалов из мировых источников информации» с ВНИИА им. Н.Л. Духова за три месяца были выполнены работы по сбору и обработке более 40 тысяч научных публикаций по материалам реакторных исследований. В результате заказчику было передано ~8700 точек, описывающих требуемые свойства реакторных материалов. Оценочно, решение такой задачи в ручном режиме заняло бы более года. Таким образом, практическая ценность работы состоит в том, что её результаты и разработки доведены до прикладных решений, используемых для ускорения и улучшения аналитических процессов в науке и промышленности. Разработанные методы и технологии могут быть оперативно адаптированы под новые задачи – от мониторинга технологических направлений до систем поддержки принятия решений в социальной сфере – что подтверждает их практическую значимость и универсальность.

**Методы исследования.** В диссертации использован комплекс методов, соответствующий междисциплинарному характеру поставленных задач. Теоретической основой послужили методы системного анализа (для формализации модели информационного объекта и постановки требований к ней), теория графов и сетевой анализ (при разработке схем представления связей между объектами), методы машинного обучения и обработки естественного языка (NLP) (для извлечения сущностей из текстов, классификации и кластеризации данных), методы визуализации данных и

человеко-машинного взаимодействия (для реализации интерактивных аналитических панелей). Для реализации алгоритмов применялись современные технологии программирования и работы с данными: инструменты веб-скрапинга и парсинга HTML-документов для сбора информации из интернет-источников, библиотеки компьютерного зрения для извлечения содержимого из изображений и PDF-документов, платформы хранения и поиска по неструктурированным данным (NoSQL базы, Elasticsearch) и фреймворки визуализации данных (например, Kibana, D3.js).

**На защиту выносятся** следующие положения:

**1. Обобщённая модель комплексного цифрового информационного объекта**, обеспечивающая унифицированное представление различных по природе данных (научные статьи, социальные профили) за счёт выделения статических, динамических и вычисляемых характеристик объекта и использования графовой схемы для хранения взаимосвязей. Применение данной модели повышает точность и воспроизводимость анализа данных за счёт уменьшения количества ошибочных или неоднозначных трактовок и облегчения слияния данных из разных источников.

**2. Методы интеллектуальной обработки данных:** метод автоматического выделения информативных признаков (ключевых слов, физических параметров) из текстовых массивов научно-технической информации; метод распознавания структурированных элементов (таблиц, рисунков) в научных публикациях и их конвертации в цифровую форму; метод унификации геопространственной привязки организаций по месту работы авторов и выявления на этой основе научных сообществ и международного сотрудничества. Предложенные методы обеспечивают наполнение информационной модели достоверными и актуальными данными в автоматизированном режиме.

**3. Методика аналитического описания и идентификации социальных объектов**, включающая разбиение характеристик профиля на статические и динамические, введение системы весовых коэффициентов значимости признаков и вычисление интегрального показателя соответствия профиля целевому образу. Пороговое правило по этому показателю позволяет выделять целевые группы пользователей. Предложенная методика продемонстрировала эффективность при решении задачи обнаружения групп риска в социальных сетях.

**4. Система интеллектуального анализа информационных объектов**, включающая концепцию комплексного цифрового информационного объекта с многоуровневой структурой характеристик и связей, совокупность методов извлечения и насыщения данных из разнородных источников и механизмы построения специализированных аналитических инструментов для анализа публикационной активности и обеспечивающие многомерное представление и исследование состояния науки в приоритетных направлениях.

**5. Специализированные программные средства интеллектуального анализа**, в частности: интерактивный инструмент построения графовых представлений, позволяющий выделять взаимосвязанные кластеры

публикаций, авторов и организаций и обнаруживать в них неявные связи; инструмент автоматизированного построения научно-технологических ландшафтов, формирующий на основе массива публикаций карту исследовательского пространства с выделением ключевых тематик и динамики их развития.

**6. Программный комплекс интеллектуального анализа данных**, созданный в рамках работы, интегрирует процессы сбора, обработки, хранения и анализа информации из разнородных источников, обеспечивает высокую гибкость и масштабируемость системы при добавлении новых типов данных или методов анализа.

**Апробация результатов работы.** Основные результаты работы представлены на международных и всероссийских научных конференциях, семинарах и школах, в частности, на: Международной конференции GRID-2025 и GRID-2023 (Россия), Международной конференции по компьютерной графике и зрению GraphiCon-2023 и GraphiCon-2022 (Россия), Научная сессия НИЯУ МИФИ-2024 (Россия), Международном симпозиуме по ядерной электронике и вычислительной технике NEC'2019 (Будва, Черногория), ежегодной конференции SPBPU IDE-2020 (Санкт-Петербург, 2020). Результаты работы в части анализа больших объемов научно-технической информации докладывались на научно-технических советах ГК «Ростех», ФГУП «РосРАО», в части решения социально значимых задач – на Межведомственной рабочей группе Министерства науки и высшего образования.

**Тематика исследований** соответствует паспорту специальности 2.3.1 – Системный анализ, управление и обработка информации, статистика, по разделам: П.2 – формализация и постановка задач системного анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта; П.5 – разработка специального математического и алгоритмического обеспечения систем анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта; П.12 – визуализация, трансформация и анализ информации на основе компьютерных методов обработки информации.

**Личный вклад.** Все основные результаты, изложенные в диссертации, включая постановки задач и их алгоритмические решения и созданное программное обеспечение, получены автором лично или выполнены под его научным руководством и при непосредственном участии.

**Публикации.** Основные положения диссертационной работы опубликованы в 47 печатных работах, из них 9 статей в изданиях, индексируемых в библиографических и реферативных базах данных Web of Science и/или Scopus, 4 статьи в изданиях, рекомендованных ВАК при Министерстве науки и высшего образования Российской Федерации для опубликования основных научных результатов диссертаций на соискание ученой степени доктора наук, 16 статей в материалах международных конференций, 2 учебно-методических пособия. По научно-техническим разработкам в составе коллектива авторов получено 6 свидетельств о

регистрации баз данных и **10** свидетельств о регистрации программ для ЭВМ в Федеральной службе по интеллектуальной собственности Российской Федерации.

**Структура и объем диссертации.** Диссертация содержит введение, шесть глав, заключение, перечень используемых сокращений, список используемой литературы, 3 приложений. Работа состоит из **287** страниц, из них основной текст **224** страницы, включая **76** рисунков, **12** таблиц и список литературы, содержащий **178** наименований.

## **ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ**

**Во введении** дано обоснование актуальности выбранной автором темы диссертации, сформулирована цель и задачи работы, аргументирована ее научная новизна и практическая ценность.

**В первом разделе** представлен обзор актуальных методов и технологий, применяемых для интеллектуального анализа больших объёмов данных и проанализированы особенности научно-технической и социальной информации, как объекта такого анализа. Отмечено, что современная парадигма больших данных характеризуется непрерывностью поступления данных, их многообразием по форматам и источникам, а также высокими требованиями к скорости обработки, что существенно влияет на выбор методов анализа: например, огромный объём данных требует масштабируемых распределённых алгоритмов, разнообразие структур – универсальных моделей данных, высокая скорость генерации и передачи данных – возможности обработки в режиме реального времени, а достоверность – встроенных механизмов фильтрации шума и ошибок. Отдельно рассматривается растущая роль визуализации данных: интерактивные графики, карты знаний и панели мониторинга облегчают интерпретацию результатов и помогают выявлять неявные связи между объектами. Сделан вывод о необходимости разработки методологического подхода, объединяющего универсальную модель данных, методы и алгоритмы работы с данными, и интерактивные аналитические инструменты, чтобы преодолеть выявленные ограничения существующих методов.

**Во втором разделе** изложено разработанное автором описание информационного объекта. В работе рассматриваются информационные объекты, представленные в виртуальной среде Интернет – цифровые объекты – и отдельно стоит задача соотнесения информационного объекта (образа физического объекта) с объектом реального мира, которая не рассматривается в рамках работы.

Предложена классификация информационных объектов: социальные объекты и научно-технические объекты. Отмечено, что эти группы пересекаются – например, учёные (социальные объекты «персона») фигурируют, как авторы научно-технических документов. Поэтому единая модель данных должна охватывать обе области, чтобы анализ мог проводиться в том числе на стыке социальной и научно-технической информации.

Сформулированы требования к модели цифрового объекта. Во-первых, она должна унифицировать представление данных разнородной природы – то есть позволять хранить и обрабатывать в едином формате сведения о текстовых документах, изображениях, профилях и т.д. Во-вторых, модель должна учитывать изменение характеристик во времени. В-третьих, требуется выделять вычисляемые характеристики – показатели, которые не хранятся явно, а вычисляются из базовых данных. В-четвёртых, модель должна быть расширяемой и модульной: при появлении новых типов данных или метрик не должно требоваться кардинального изменения всей структуры; достаточно добавить новую характеристику или тип связи. Пятое требование – универсальность и переносимость: модель должна применяться в различных предметных областях, обеспечивая тем самым преемственность подходов и возможность интеграции данных из разных источников (научные репозитории, социальные сети, государственные реестры и др.).

В соответствии с этими принципами, автором предложена базовая информационная модель цифрового объекта:

$$Obj = \langle ID, S, D \rangle, \quad (1)$$

где  $ID$  – уникальный идентификатор объекта,

$S$  – множество статических характеристик (свойства, значения которых либо постоянны, либо изменяются крайне редко):

$$S = \{(s_1, \tau_1), (s_2, \tau_2) \dots, (s_k, \tau_k)\}, \quad (2)$$

где  $s_i$  – наименование  $i$ -ой статической характеристики, (например, автор, название организации),  $\tau_i$  – тип данной характеристики (строка, число и др.),  $i = 1, 2, \dots, k$ .

$D$  – множество динамических характеристик (свойства, значения которых могут меняться во времени или под воздействием внешних факторов):

$$D = \{(d_1(t_1), t_1, \tau_1), (d_2(t_2), t_2, \tau_2), \dots, (d_l(t_l), t_l, \tau_l)\}, \quad (3)$$

где  $d_j(t_j)$  – значение  $j$ -ой динамической характеристики в момент времени  $t_j$ ,  $t_j$  – значение времени,  $\tau_j$  – тип данной характеристики (строка, число),  $j = 1, 2, \dots, l$ .

Базовая модель задаёт первичное описание объекта и достаточна для отслеживания изменений его состояния. Решение аналитических задач зачастую невозможно на основе только статических и динамических характеристик, что приводит к необходимости насыщения модели дополнительными характеристиками.

Автором вводится аналитическая модель цифрового объекта:

$$AOBJ = \langle ID, S, D, F, Rel \rangle, \quad (4)$$

где  $F$  – совокупность вычисляемых характеристик (значения, которые не собираются непосредственно из информационных ресурсов, а рассчитываются исходя из значений статических и динамических характеристик):

$$F = \left\{ (f_1, \varphi_{f_1}), (f_2, \varphi_{f_2}), \dots, (f_p, \varphi_{f_p}) \right\}, \quad (5)$$

где  $f_j$  – наименование вычисляемой характеристики,  $\varphi_{f_j}$  – формула или алгоритм, определяющий, как вычислять  $f_j$  на основе статических ( $S$ ) и/или динамических полей ( $D$ ),  $j = 1, 2, \dots, p$ .

$Rel$  – множество связей (отношений) данного объекта с другими объектами (значения, которые позволяют описывать связи данного объекта с другими объектами в информационном пространстве):

$$Rel = \{(r_1, Obj_{j_1}), (r_2, Obj_{j_2}), \dots, (r_m, Obj_{j_m})\}, \quad (6)$$

где  $r_i$  — тип связи (например, «Автор», «Владелец», «Состоит в группе», «Цитирует»),  $Obj_{j_i}$  — конкретный объект, с которым установлена данная связь,  $i = 1, 2, \dots, m$ .

Детализировано аналитическая модель цифрового объекта выглядит следующим образом:

$$AOBJ = \langle ID, (s_1, \tau_1), \dots, (s_k, \tau_k), (d_1(t_1), t_1, \tau_1), \dots, (d_l(t_l), t_l, \tau_l), \\ (f_1, \varphi_{f_1}), \dots, (f_p, \varphi_{f_p}), \{(r_1, Obj_{j_1}), \dots, (r_m, Obj_{j_m})\} \rangle. \quad (7)$$

При проведении аналитических исследований встречаются ситуации, когда анализируемые объекты обладают вложенными объектами. Классический пример — статья, содержащая в себе список авторов (каждый из которых сам по себе объект), ссылки на литературу (каждая ссылка может быть объектом) и приложения (отдельные документы).

В работе введено понятие комплексного цифрового информационного объекта, обладающего собственными характеристиками и набором вложенных объектов:

$$CAObj = \langle ID, S, D, F, Rel_{complex} \rangle, \quad (8)$$

где  $Rel_{complex} = \{(r_1, \{AOBJ_{j_1}, \dots, AOBJ_{j_n}\}), \dots, (r_m, \{AOBJ_{j_1}, \dots, AOBJ_{j_k}\})\}$ ,  $r_i$  — тип связи,  $i = 1, \dots, m$ .

Особенностью комплексной модели цифрового информационного объекта является то, что вычисляемые характеристики ( $F$ ) рассчитываются не только на статических ( $S$ ), динамических ( $D$ ) характеристиках комплексного объекта  $CAObj$ , но и на аналогичных характеристиках вложенных объектов. Отметим, что аналитическая цифровая модель является частным случаем комплексной информационной модели цифрового объекта.

Выделены особенности и преимущества модели  $CAObj$ : ясность структуры (строгое разделение типов характеристик упрощает управление данными), гибкость (можно добавлять новые поля или связи без пересмотра общей схемы), аналитическая полезность (автоматический пересчёт  $F$  при изменениях  $D$  позволяет оперативно получать актуальные метрики), масштабируемость (сеть объектов через  $Rel$  формирует граф, пригодный для расширенного анализа, вплоть до построения онтологий).

Таким образом схема преобразований цифрового объекта представлена на рисунке (Рисунок 1), где:

$\varphi$  – функции/алгоритмы расчета характеристик,

$r$  – типы связей между объектами,

$A(x)$  – агрегирующие и комбинированные функции.

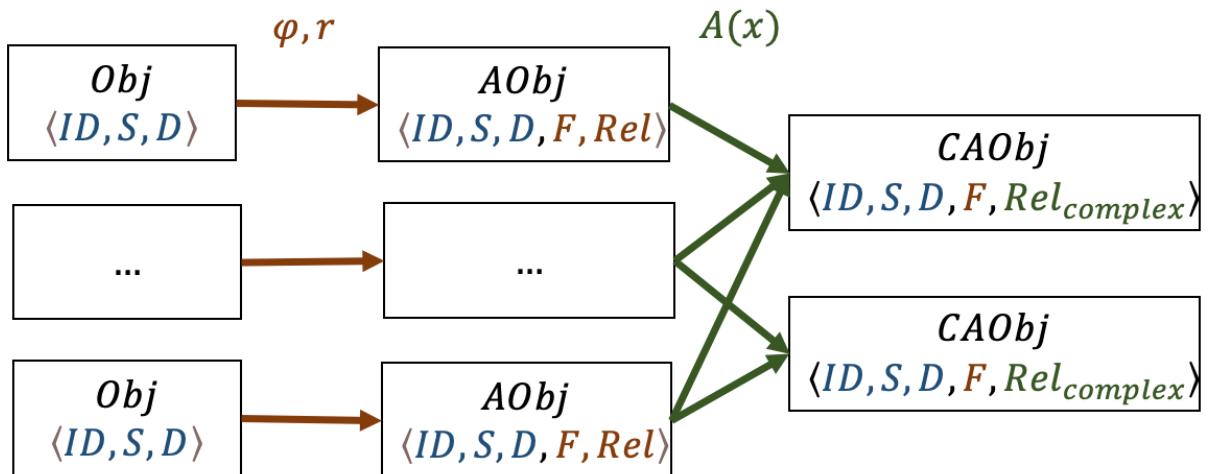


Рисунок 1 Схема преобразований цифрового объекта

Модель комплексного цифрового информационного объекта является методологически обоснованным подходом к стандартизации описания данных. Она задает единый формат представления данных, повышает точность и достоверность результатов анализа (за счёт устранения неоднозначностей и дублирования), а также служит фундаментом для разработки единых алгоритмов работы с информацией из разнородных информационных источников.

**В третьем разделе** приведены разработанные автором методы преобразования данных из разнородных информационных ресурсов, которые формируют методику наполнения моделей цифрового объекта, состоящую из трёх основных этапов:

1. Извлечение данных – получение статических и динамических характеристик, например, название публикации, авторы, аффилиации авторов, аннотация, ключевые слова (заданные авторами), основной текст статьи, информация о финансировании, список литературы, название журнала, год публикации, ссылки (DOI, URL).

2. Насыщение данных – вычисление и добавление новых сведений на основе извлечённой информации (ключевые слова текста, химические упоминания, геоданные и т.п.). В работе рассматриваются несколько направлений насыщения: выделение из полного текста ключевых слов и фраз; распознавание в тексте физических элементов и упоминаемых единиц измерения с их нормализацией; обработка аффилиаций авторов для определения геолокации (координат) организаций и приведение названий стран к единому виду; на основе стран аффилиаций – определение, в какие международные научные альянсы входят соответствующие страны.

3. Организация хранения результирующих данных о цифровом объекте. Собранные сведения организуются в единую структуру и сохраняются в целевое хранилище, что предполагает приведение данных к заранее определённой схеме, включающей поля для всех типов информации (метаданные, текст, ключевые слова, таблицы, изображения, химические сущности, единицы и пр.).

Автором разработаны методы извлечения данных из информационных ресурсов в зависимости от типа входного формата. Предложен алгоритм конвертации PDF в текст, очистки текста. Оговорены трудности, возникающие при распознавании PDF (ошибки OCR, разбиение по колонкам) и способы их преодоления. Предложен метод обработки веб-документов (HTML-страниц) – для случаев, когда данные доступны только через веб-интерфейс. Разработаны парсеры HTML с использованием шаблонов под конкретные сайты.

Важной задачей при анализе научных публикаций является сохранение связей между текстом и иллюстрациями. Реализован модуль, выделяющий из PDF-документов встроенные изображения, определяющий их тип (рисунок, диаграмма, фотография) по характеру подписи и контекста. Эти изображения помещаются в структуру данных как вложенные объекты, а их метаданные (например, подпись к рисунку) используются для дальнейшего анализа. Аналогично извлекаются таблицы из документов: либо в виде CSV-данных (при наличии машиночитаемого формата в PDF), либо путём распознавания (OCR) при сканированных таблицах. Полученные табличные данные также прикрепляются к объекту.

Автором предложено несколько методов насыщения данных. Разработан алгоритм автоматического выделения ключевых словосочетаний из полного текста научной публикации. Реализованный программный инструмент, принимает на вход текст статьи и параметры (максимальная длина фразы и требуемое число ключевых фраз), и выдает упорядоченный список ключевых словосочетаний. В основе инструмента лежит модифицированный метод YAKE (Yet Another Keyword Extractor). Оригинальный метод YAKE вычисляет для кандидатов фраз веса по совокупности статистических показателей (позиция в тексте, частота отдельных слов, их распространённость и др.). В работе алгоритм дополнен эвристиками, учитывающими особенности научных текстов: исключение из рассмотрения ключевых фраз, полностью входящих в состав более длинных фраз; игнорирование определенных слов, заранее заданных как нерелевантные для данной предметной области. Этот механизм позволяет добиться разнообразия ключевых словосочетаний и исключить избыточные дубли.

Разработан алгоритм, автоматически находящий упоминания числовых значений с единицами измерения в тексте статьи, и приводящий их к единой системе. Алгоритм с помощью регулярных выражений и словаря единиц выделяет такие конструкции, разделяет число и единицу, затем на основе базы знаний о единицах переводит значение в единицы СИ. Для этого создан справочник единиц и их коэффициентов пересчета (включая температурные шкалы, где требуется аддитивная поправка, как  $^{\circ}\text{C} \rightarrow \text{K}$ ). Все найденные и нормализованные величины собираются в единую структуру данных. В результате обработки всего текста статьи формируется список физических величин с их значениями в стандартизированном виде, что облегчает дальнейший количественный анализ.

В разделе описываются методы обработки метаданных авторов и организаций. Сформулирована задача унификации аффилиаций, поскольку

авторы научных статей могут указывать название учреждения по-разному (сокращения, разный язык), необходим алгоритм приведения собранных данных к стандартизованному виду. Предложено решение, где для каждой аффилиации из статьи алгоритм ищет ближайшее совпадение с известными названиями (метрики сравнения строк с учётом сокращений). Результатом работы является идентификация официального названия организации и страны, которые затем сохраняются как атрибуты автора. Также разработан алгоритм геокодирования – определение географических координат организации по её названию через публичные гео-сервисы. Полученные широта и долгота используются далее, например, для построения карт научного сотрудничества.

Описан метод выявления международных научных объединений по аффилиациям. Он заключается в анализе совместных указаний авторов из разных стран в одной публикации. Если в статье присутствуют авторы из, условно, 3–4 разных стран, алгоритм фиксирует факт международной кооперации. Накапливая такие данные по множеству публикаций, можно выделить устойчивые альянсы – группы организаций или стран, часто вместе публикующихся. В работе построена база данных таких международных связей: для каждой групп стран рассчитано число совместных публикаций. Эти сведения, будучи помещёнными в модель (как особые отношения между объектами типа «страна» и «публикация»), позволяют ответить на вопросы о степени интеграции научных сообществ разных стран.

Рассмотрены задачи хранения данных о цифровом информационном объекте. Обосновано, что традиционные реляционные СУБД не столь удобны для хранения комплексной модели цифрового информационного объекта. Выбор сделан в пользу документно-ориентированных баз данных (в работе в качестве основы используется Elasticsearch).

На основе предложенных моделей цифрового объекта разработана стандартизованная структура данных, в которой объединяются все извлечённые и обогащённые сведения на примере научной публикации. Эта структура реализована в формате JSON и приведена в обобщённом виде в таблице (Таблица 1). Выбор формата JSON обусловлен тем, что он поддерживает вложенность структур, легко читается человеком и обрабатывается программно.

Таблица 1 Структура данных научных публикаций

Ключ	Содержание
title	название публикации
published	тип публикации, журнал, дата публикации, язык, DOI, библиографическое описание
authors	список авторов (каждый автор: имя, уникальный идентификатор при наличии, электронная почта при наличии)
affiliations	список аффилиаций авторов: названия организаций, подразделений, страна, координаты организации (широта, долгота)

countries	список стран, к которым относятся организации авторов (основные страны аффилиаций)
abstract	аннотация
keywords	ключевые слова публикации (заданные авторами)
full_text	данное поле является составным и включает в себя другие: plain – полный текст публикации keywords – ключевые слова текста публикации (по наибольшей встречаемости), tables – выделенные таблицы, images – выделенные изображения, alliances – альянсы стран (в какие межгосударственные союзы входят страны организаций авторов), chemical_units – химические элементы, measurement_units – единицы измерения и их величины
acknowledgements	благодарности (если указаны в публикации)
funding	сведения о финансировании исследования (гранты, контракты, если указаны)
references	библиографический список источников, цитируемых в публикации
updated_at	дата и время последнего обновления информации о публикации
created_at	дата и время первоначального создания записи о публикации
labels	метки или категории, присвоенные публикации (например, принадлежность к определённому тематическому кластеру)

Использование предложенных автором методов обеспечивает повышение полноты данных для дальнейшего интеллектуального анализа: в частности, автоматически дополняются такие поля, которые вручную зачастую не обрабатывались (географическая локализация аффилиаций, извлечение числовых показателей из текста и т.д.).

**В четвертом разделе** рассматривается применение разработанных автором моделей и методов для решения социально значимых задач. Объектом исследования выступают цифровые профили пользователей в социальных сетях, а целью – разработка метода автоматизированной идентификации определённых категорий или состояний пользователей (например, выявление пользователей, склонных к деструктивному поведению).

Аналитическая модель цифрового объекта в социальной сфере рассматривается как объект, обладающий статическими (*S*), динамическими (*D*) характеристиками, а также множеством связей (*Rel*). Выявлено, что цифровой профиль можно устойчиво описать 41 статической характеристикой, которые можно разделить на 5 категорий (основное, контакты, деятельность, интересы, жизненная позиция).

К характеристикам, содержащим основную информацию о профиле персоны, может быть добавлена такая информация как: аудио и видео записи, фотография профиля (аватар) и текстовый статус. Все эти данные могут изменяться с течением времени существования цифрового профиля, поэтому выделим их в качестве динамических характеристик ( $D$ ).

Существенное значение на анализ цифрового профиля оказывают связи с другими объектами ( $Rel$ ) – связь с такими объектами как друзья, подписчики, вхождение в группы. Аналитическое описание цифрового профиля персоны в социальной среде, согласно введенной формуле (2), обладает следующей размерностью:

$$|S| = 41, |D| = 4, |Rel| = 3. \quad (9)$$

В разделе представлена авторская методика формирования и заполнения аналитической модели цифрового профиля социального объекта для решения задач идентификации в информационной среде.

На первом этапе происходит ранжирование списка фактических характеристик по степени их важности с последующим преобразованием отранжированного ряда посредством формальной процедуры в ряд весовых коэффициентов, нормированных в интервале  $[0,1]$ ; для каждой характеристики аналитической цифровой модели разрабатывается собственная методика и алгоритм измерений для сопоставления разнотиповых значений.

На втором этапе каждой характеристике присваивается детерминированная оценка – численная, балльная или основанная на тезаурусе с указанием диапазона возможных значений. Однако полученные детерминированные оценки, являясь неаддитивными величинами различной размерности, не подлежат суммированию в интегральном критерии, что обуславливает введение для каждого критерия относительных значений, преобразуемых в безразмерные. В качестве интегральной оценки объекта используется сумма произведений относительной значимости критерия на относительное значение его характеристики.

На третьем этапе устанавливается нормативное значение критерия и формулируется правило маркировки объектов, согласно которому все объекты со значением критерия, превышающим нормативное, маркируются как целевые.

Аналитическая модель цифрового объекта, предложенная автором, предоставляет формальную основу для решения аналитических задач, однако для практического применения её необходимо «привязать» к конкретным задачам идентификации. Решение аналитических задач базируется на синтезе методов обработки исходных данных и разработке методов расчета вычисляемых характеристик ( $F$ ) для каждого цифрового информационного объекта. Рассмотрим некоторые примеры расчета вычисляемых характеристик в реальных аналитических задачах.

Одной из базовых задач, решаемых в социальной среде, является соотнесение цифрового профиля с различными социальными группами, например, школьник, студент.

Идентификация школьника по данным в социальной среде возможна по значениям двух характеристик – наличие школы ( $\varphi_{boolen}$ ), целевой возраст ( $\varphi_{in\_range}$ ).

Примем, что перечисленные характеристики являются равнозначными, так как значение возраста информационного цифрового профиля должно попасть в заданный диапазон (от 6 до 18 лет) и в профиле должна быть указана школа; тогда их веса в методике соотнесения цифрового профиля к социальной группе школьника одинаковы  $v_{birth} = v_{nsch} = 0,5$ .

$$f_{sch} = \varphi_{sch}(s_{birth}, s_{nsch}) = \\ v_{birth} \cdot \varphi_{in\_range}(s_{birth}) + v_{nsch} \cdot \varphi_{boolen}(s_{nsch}), f_{sch} \in [0,1], \quad (10)$$

где  $s_{birth}$  – дата рождения,  $s_{nsch}$  – поле номера школы.

В разделе описаны примеры решения сложных идентификационных задач в социальной среде с помощью разработанной методики.

В работе предложены методы обработки статических и динамических характеристик аналитической модели цифрового объекта для решения задач идентификации девиантного поведения. Кроме методов анализа наличия и отсутствия информации в полях разработан метод анализа текстовых полей на основе заданных словарей ключевых слов.

Например, текстовый статус, динамично изменяется в цифровом профиле, является уникальной лакмусовой полоской (краткий эмоциональный выпад часто заметно опережает по динамике все остальные изменения), анализировался на основе составленного корпуса ключевых слов, включающего в себя 583 значения, разделенных на 11 типов. К каждому значению составлены соответствующие поисковые образы для обеспечения точности поиска информации. Данный тезаурус является динамичным и возможно его изменение (пополнение/удаление) в соответствии с изменениями анализируемого информационного поля (графическое представление тезауруса – Рисунок 2).

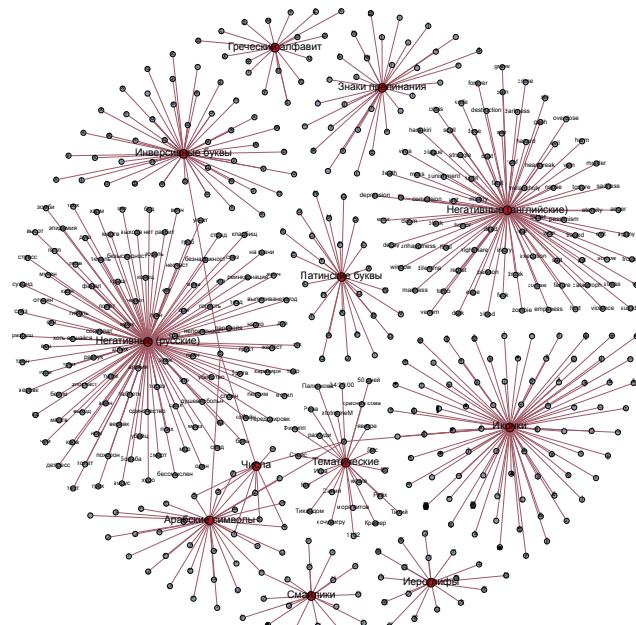


Рисунок 2 Графическое представление базового тезауруса

Для определения значения вычисляемой характеристики отнесения к «группе риска»  $f_{destr}$  разработана 28-критериальная балльная модель:

$$f_{destr} = \alpha f_S + \beta f_D, \alpha + \beta = 1, \quad (11)$$

где  $f_S = \sum_{i=1}^{21} w_i x_i$  — взвешенная сумма статических характеристик,  $f_D = \sum_{j=1}^7 w_j x_j$  — взвешенная сумма динамических характеристик,  $\beta = 0,45$  (по умолчанию) — экспериментально выявленный баланс, при котором точность и полнота характеристики достигают оптимального компромисса.

Для определения порогового значения  $f_{destr}$  проведены расчеты значений характеристики по каждому цифровому профилю эмпирической базы ( $f_{d\_trainset}$ ), полученной на основе анализа 100 подтвержденных цифровых профилей девиантного поведения. Распределение  $f_{d\_trainset}$  по обучающей выборке оказалось с минимальным зафиксированным значением  $f_{d\_trainset}_{min} = 0,176$  и пиком плотности в области 0,32.

Финальное интегральное значение вычисляемой характеристики отнесения школьника к «группе риска» вычисляется по формуле:

$$f_{comb} = \lambda_1 f_{destr} + \lambda_2 (1 - f_{bot}) + \lambda_3 f_{sch}, \sum_{i=1}^3 \lambda_i = 1, \quad (12)$$

где  $f_{destr}$  — вычисляемая характеристика девиантного поведения,  $f_{bot}$  — вычисляемая характеристика «робота» в цифровой среде,  $f_{sch}$  — вычисляемая характеристика «школьника» в цифровой среде.

По умолчанию  $\lambda_1 = 0,6, \lambda_2 = 0,3, \lambda_3 = 0,1$ . Таким образом, приоритет отдаётся рисковой составляющей, затем — проверке на техническую достоверность страницы, и затем — возрастному фактору.

Предложенная методика была реализована в качестве цифровой платформы по сбору, анализу и идентификации целевых профилей, склонных к девиантному поведению. Цифровая платформа реализована на языке Python 3.5. В качестве веб-фреймворка используется Django 2.0.2, для хранения и управления данными задействована база данных PostgreSQL 10. Система написана как кроссплатформенная, разработка ядра системы проходила на операционной системе Windows 10, а испытания проводились на операционной системе Ubuntu Server 16.04.

Проведены тестовые испытания методики идентификации целевых объектов на выборке из 800 цифровых профилей. В результате поисковых испытаний определено целевое значение вычисляемой характеристики  $f_{destr} = 0,3$ , при которой однозначно идентифицируется цифровой объект, принадлежащий к девиантному поведению, а также подтверждены методики расчета вычисляемых характеристик «школьник» по параметрам — возраст, наличие школы и ее номера в аккаунте.

Предложенный автором подход к решению задач идентификации целевых объектов использовался в государственном задании Минобрнауки России №2.12915.2018/12.1 «Разработка и апробация информационной системы комплексной антисуициdalной интернет-профилактики», реализованном под руководством автора.

Предложенная автором аналитическая модель и методика идентификации целевых объектов в социальной среде может применяться для решения прикладных социальных задач, требующих автоматизированного анализа поведения и характеристик пользователей. Интеграция разнородных данных профиля и использование пороговой метрики позволяют достичь высоких показателей обнаружения целевых объектов. Апробация на реальных данных подтвердила работоспособность предложенного подхода.

**В пятом разделе** рассматривается применение предложенных автором моделей и методов в организации научно-технической информации для решения аналитических задач. Автором предложен подход к построению интерактивных аналитических панелей по определённому тематическому направлению на основе формирования озера данных для нескольких выбранных научных областей. В качестве примера рассмотрены: технологии больших данных, медицина и финансовая безопасность. Каждое направление имеет свои особенности в плане доступности и структуры данных, однако применена единая схема анализа.

При рассмотрении направления «Финансовая безопасность» сделан вывод о том, что направление не имеет прямого соответствия в традиционных рубрикаторах и требует первоначального тематического обзора. В качестве стартового источника выбран профильный журнал *Journal of Money Laundering Control*, из которого за период 1997–2022 гг. извлечено 1008 статей. На основе ключевых слов и ссылок из этого корпуса статей определён расширенный пул источников – специализированные издания, конференции, отчёты – по которому далее выполнен сбор данных. Все собранные публикации (более 11 тысяч статей) загружены в хранилище ElasticSearch и подвергнуты предложенным в третьем разделе процедурам насыщения (выделены ключевые слова, выявлены аффилиации, построены связи по цитированию и соавторству).

В работе детально описываются механизмы взаимодействия с озером данных через панель визуализации: для выбранной тематики в Kibana создано 14 интерактивных визуализаций (Рисунок 3).

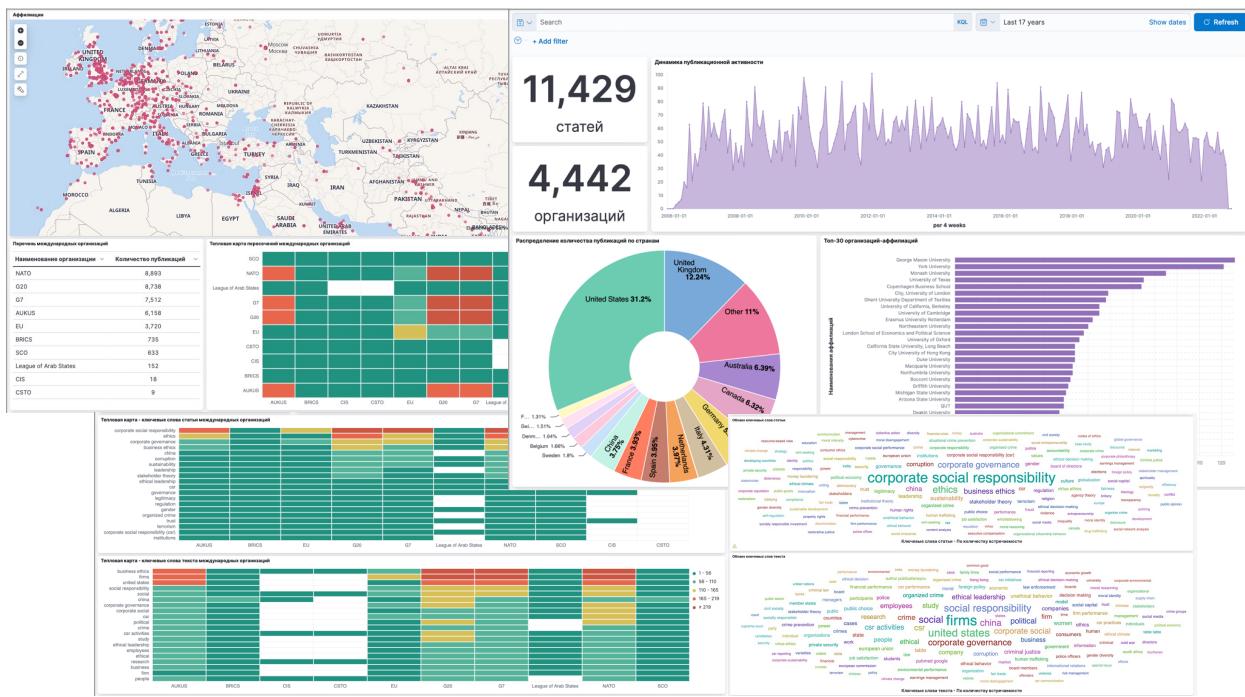


Рисунок 3 Визуализации данных по направлению  
«Финансовая безопасность»

Среди них – метрика общего числа публикаций по теме, число организаций-авторов, временная шкала динамики публикаций по годам, географическая карта распределения публикаций по странам, облако наиболее частотных ключевых слов, а также таблица крупных международных организаций с указанием количества публикаций участников. Аналитику доступна функциональность интерактивной фильтрации: например, выделение определённого временного периода на графике автоматически перестраивает все остальные визуализации под выбранный интервал; выбор конкретной страны на карте позволяет отфильтровать публикации этой страны и сразу увидеть связанные с ней организации и авторов.

Разработанная интерактивная аналитическая панель по «Финансовой безопасности» эффективна, как инструмент визуальной аналитики: эксперт, не погружённый в технические детали, может быстро получить ответы на ключевые вопросы (кто лидеры по публикационной активности, каков тренд интереса к теме, какие институты и страны наиболее активно сотрудничают и пр.).

В совокупности построенное озеро данных и интерактивная аналитическая панель наглядно продемонстрировали значимость комплексного подхода для новой области знаний. Несмотря на междисциплинарный характер темы (пересечение экономики, права, информационных технологий), разработанная система успешно справилась с интеграцией данных и выявлением скрытых закономерностей. Апробация результатов работ происходила в рамках Государственного задания Министерства науки и высшего образования Российской Федерации №3466-22 «Создание учебно-методических материалов по финансовой безопасности для школьников и студентов, в том числе для передачи указанных учебно-методических материалов в зарубежные страны-партнеры Международного сетевого института в сфере

противодействия отмыванию доходов, полученных преступным путем, и финансированию терроризма».

Предложенный подход апробирован по медико-биологической тематике, в части интеграции научных публикаций и данных доказательной медицины для поддержки клинических решений. Данная область кардинально отличается от информационной сферы по характеру источников информации: помимо научных статей, большой интерес представляют клинические рекомендации, данные о лекарственных средствах, клинические исследования, фармакологические базы знаний и т. п. Целью проводимой работы являлась разработка методов искусственного интеллекта для систем поддержки принятия врачебных решений (СППВР) на основе большого массива доказательных данных о применении лекарств при коморбидных (сочетанных) заболеваниях. Исследование проводилось в рамках гранта Российского научного фонда № 23-75-30012 «Снижение рисков полифармакотерапии с использованием искусственного интеллекта и анализа Больших данных о лекарственных препаратах и их взаимодействиях», в котором автор являлся ответственным исполнителем. На основе предложенной автором аналитической модели цифрового объекта реализована интеграция разнородных медицинских ресурсов.

Полученный в результате интеграции набор медицинских данных является многокомпонентным. Во-первых, удалось собрать полный массив национальных клинических рекомендаций – 415 документов в структурированном виде (общим объемом ~126,4 МБ). Каждая рекомендация была автоматически разобрана на разделы, таблицы и списки рекомендаций, что облегчает последующую идентификацию упоминаний конкретных препаратов и методов терапии. Во-вторых, из Государственного реестра лекарственных средств извлечены сведения обо всех зарегистрированных лекарственных средствах (на апрель 2023 г.) – несколько тысяч наименований. По каждому лекарственному препарату хранится его регистрационный номер, состав, производители, а также тексты инструкций по применению (если они доступны). В-третьих, из международных ресурсов по взаимодействию лекарств получены ценные данные о нежелательных сочетаниях препаратов. Так, с сайта Drugs.com было выгружено 5 433 JSON файлов (объемом ~4,11 ГБ) – по числу уникальных действующих веществ, для которых указаны взаимодействия. Формат одного такого файла представляет собой словарь с ключами: name (название вещества), source (ссылка на страницу источника) и interactions (список всех найденных взаимодействий с указанием типа взаимодействия, степени риска, описания и пр.). Аналогично, по базе Medscape собрано 2 583 JSON файлов (152,4 МБ) со структурированной информацией о взаимодействиях каждого препарата.

Важным результатом является демонстрация практической ценности собранных данных для разработки СППВР. На основе базы знаний, наполненной описанными данными, был создан прототип рекомендательной системы, предлагающей врачу информацию о возможных взаимодействиях назначаемых лекарств и о наличии клинических рекомендаций по

интересующему случаю. Этот прототип подтвердил, что межотраслевой синтез (сочетание научной литературы и официальных руководств) дает более полную поддержку решений: например, система не только предупреждает о риске одновременного назначения двух лекарств, но и сразу ссылается на соответствующую клиническую рекомендацию Минздрава и на свежие исследования из PubMed по данному сочетанию.

В разделе рассматривается пример реализованной базы данных свойств облученных реакторных материалов и решение задачи надежности материалов научно-технической информации при формировании профиля объекта типа сталь. Апробация методики проходила в рамках договора с ВНИИА им. Н.Л. Духова № 1707 от 29 августа 2022 г. по выполнению НИР «Разработка программы выборки данных по свойствам и структурам облученных реакторных материалов из мировых источников информации».

С использованием предложенных в разделе 3 методов сформирована специальная коллекция публикаций по радиационному материаловедению (более 40 тыс. научных публикаций), из каждой статьи извлечены числовые данные о свойствах материалов после облучения (например, предел прочности, твёрдость, изменение размера и т.д., вместе с условиями эксперимента). Реализована база данных с удобным интерфейсом поиска по материалам и параметрам, позволяющая специалистам быстро находить все доступные результаты по конкретному материалу или режиму облучения (Рисунок 4).

База данных стала прямым практическим результатом использования системы интеллектуального анализа данных, где применены все основные компоненты: унификация данных (каждая запись стандартизована по набору свойств), насыщение (выделены физические величины – дозы, температуры, времена из текстов статей), анализ (сделаны сводные таблицы и графики для разных материалов). База данных зарегистрирована как объект интеллектуальной собственности, а ее содержимое используется специалистами госкорпорации Росатома при планировании экспериментов и оценке эксплуатационных границ материалов.

The image shows two screenshots of the Article Items Review software interface. The left screenshot displays a search results page for 'Journal of Nuclear Materials'. It includes filters for 'Selected journal(s)', 'Select year(s)', 'Processing status', and a 'FILTER' button. The results table shows columns for 'TITLE', 'JOURNAL', 'YEAR', 'TABLES', 'IMAGES', and a green checkmark column. Several articles are listed, such as 'Structural assessment of a whole toroidal sector of the HELIAS 5-B breeding blanket' from Fusion Engineering and Design in 2021, and 'Thermodynamic modelling of the C-U and B-U binary systems' from Journal of Nuclear Materials in 2001. The right screenshot shows a detailed view of an article titled 'MECHANICAL ALLOYING AND MILLING' from the 'ScienceDirect, Scopus' database. It includes sections for 'Authors', 'Affiliations', 'Abstract', 'Keywords', and 'Acknowledgements'. Below this is a table titled 'Table 1. Departure from equilibrium achieved in different non-equilibrium processing techniques' with columns for 'Technique', 'Effective quench rate (K/H)', 'Maximum departure from equilibrium (K/Jmol)', and 'Ref. [15,16]'. The table lists various techniques like solid state quench, rapid solidification, mechanical alloying, and condensation from vapor, along with their respective values and references.

Рисунок 4 Примеры интерфейса БД по облученным реакторным материалам

Предложенные автором методы интеллектуального анализа доказала свою эффективность для структурирования и анализа больших коллекций научно-технических данных. Построенные на их основе инструменты визуальной аналитики обеспечивают исследователям возможность самостоятельного изучения огромных массивов публикаций, выявления трендов и закономерностей. Предложенный подход ориентирован на слабоструктурированные данные (не требуя заранее строго нормализованных баз) и, благодаря гибкости NoSQL-схемы, может быть оперативно перенастроен под другую предметную область. Подчёркивается практическая значимость предложенных моделей и методов в решении различных аналитических задач: от мониторинга научных направлений (в интересах научных фондов и министерств) до создания тематических баз знаний (в интересах отраслевых научных организаций).

**В шестом разделе** рассмотрены разработанные в ходе исследования программные инструменты, направленные на решение прикладных задач анализа информации. Предложен программный инструмент выявления явных и неявных связей между цифровыми объектами посредством графового представления совокупности объектов и их характеристик. Графы естественным образом моделируют сети связей: вершины могут обозначать объекты или значения характеристик, а рёбра – наличие определённого отношения между ними. Визуальное представление данных в виде графа предоставляет аналитику наглядную «карту» взаимосвязей, позволяющую интерактивно исследовать структуру данных.

Пусть  $Q$  – множество объектов  $AObj$ , то есть

$$Q = \{AObj = (a_1, a_2, \dots, a_M) | a_i \text{ – характеристика объекта } AObj, i = 1, 2, \dots, M\}, \quad (13)$$

где  $a_i \in S \cup D \cup F \cup Rel$ .

Каждому объекту  $AObj = (a_1, a_2, \dots, a_M) \in Q$  поставим в соответствии граф  $G = < V, E >$ , где множество вершин

$$V = \{v_1, v_2, \dots, v_T\}, T \geq M, \quad (14)$$

составляют значения характеристик  $a_1, a_2, \dots, a_M$  объекта  $AObj$ , и множество ребер

$$E = \{(v_i, v_j) | \exists AObj \in Q : AObj(a_i) = v_i \text{ и } AObj(a_j) = v_j\}, \quad (15)$$

то есть две вершины  $v_i, v_j \in V$  соединены ребром, если найдется объект  $AObj = (a_1, a_2, \dots, a_M) \in Q$ , такой что значение характеристики  $a_i$  объекта  $AObj$  равно  $v_i$  и значение характеристики  $a_j$  объекта  $AObj$  равно  $v_j$ .

Правило построения графа определяется аналитиком, то есть заранее указывается, какие именно характеристики и каким образом следует связывать между собой для выявления интересующих связей.

Каждой вершине из множества  $V$  поставим в соответствие ее тип – характеристику из множества  $\{a_1, a_2, \dots, a_M\}$ . Тогда множество вершин

$$V = K_1 \cup K_2 \cup \dots \cup K_M, \quad (16)$$

где  $K_i = \{AObj(a_i) | AObj \in Q\}$  – множество вершин типа  $a_i$ ,  $i = 1, 2, \dots, M$ .

Таким образом, граф  $G = \langle V, E \rangle$  можно представить в виде гетерогенного графа

$$G = \langle K_1 \cup \dots \cup K_L, E \rangle, \quad (17)$$

где  $L \in \{1, 2, \dots, M\}$ , при этом если  $v_i, v_j \in K_W$ ,  $W \in \{1, \dots, L\}$ , то  $(v_i, v_j) \notin E$  (характеристики и их количество  $L$  задает аналитик).

Разработанный программный инструмент использован для решения ряда актуальных научно-технических задач, продемонстрировавших его практическую применимость. Были собраны данные 48 835 научных публикаций Объединённого института ядерных исследований (ОИЯИ) – крупного международного межправительственного научного центра, связанных с сотрудниками Института за 1957-2024 года. На основе этих данных построен график по двум полям за 2023-2024 гг.:

- 1 тип – affiliations.name (название аффилиации, организации),
- 2 тип – keyword (ключевые слова публикации).

Вершины первого типа (организации) отображены красным цветом, второго типа (ключевые слова) – розовым. Всего в графике 7088 узлов и 89575 ребер (Рисунок 5). Полученный график отражает взаимосвязи организаций по тематическим направлениям исследований: узлы-организации соединены с узлами-терминами, что означает участие данной организации в работах по соответствующей тематике. Граф разбивается на несколько кластеров, каждый из которых соответствует определённой тематике исследований. Анализ этих кластеров позволяет выявить, с какими организациями ОИЯИ взаимодействует по каждой тематике.

Каждый кластер на рисунке (Рисунок 5) представляет совокупность организаций, связанных через общие ключевые слова, то есть совместную научную тематику. Например, один из крупнейших кластеров соответствует направлению физики частиц. Для его детального изучения был использована функциональность фильтрации по запросу: «quark gluon plasma» OR «jets» (Рисунок 6).

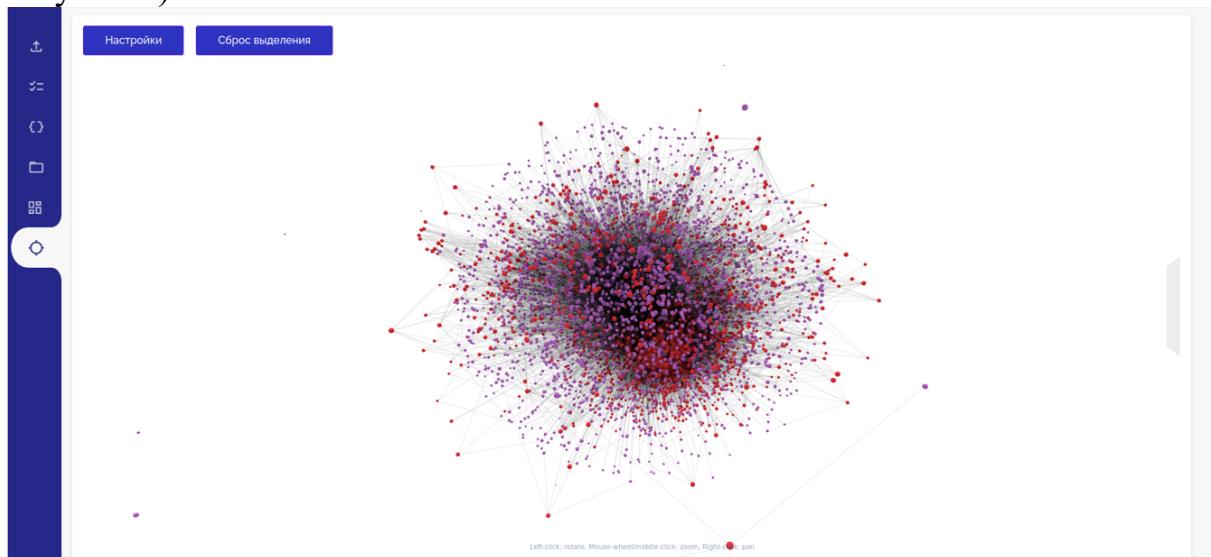


Рисунок 5 Граф публикационной активности ОИЯИ

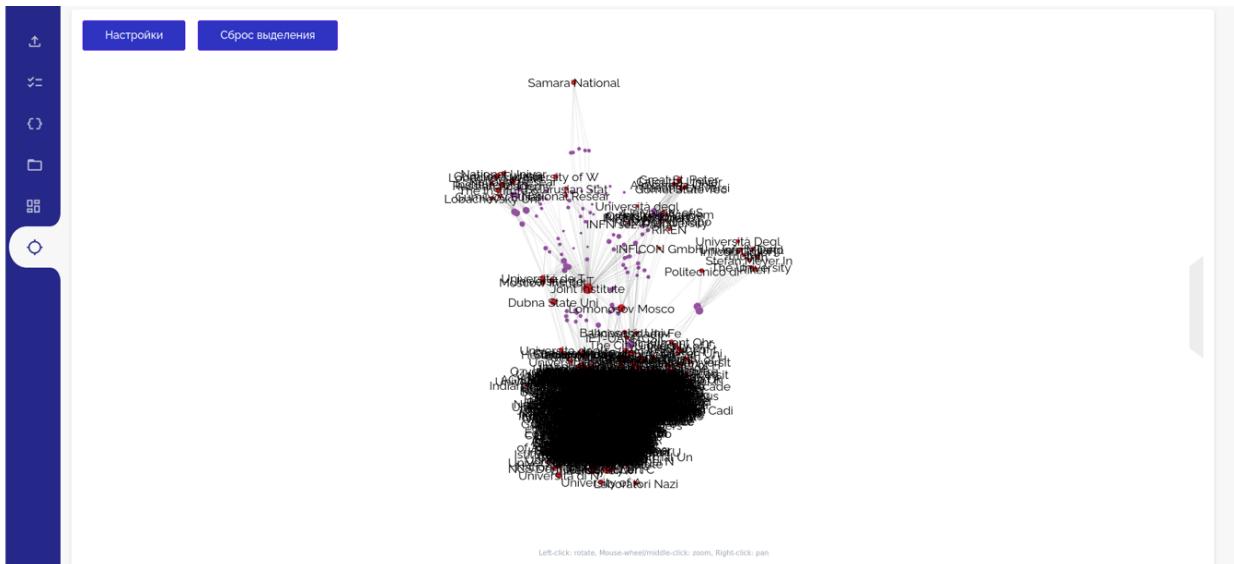


Рисунок 6 Фильтрация по запросу: «quark gluon plasma» OR «jets»

Использование разработанного инструмента для анализа деятельности организаций позволяет определить ключевых партнёров, основные совместные направления исследований, а также выявить скрытые связи между партнёрами (например, через трети страны или общие проекты). Эти сведения имеют практическую ценность для формирования стратегии международного сотрудничества: зная, какие организации и страны наиболее активно вовлечены в совместные публикации, можно планировать укрепление связей с ними, а также искать новые точки роста (например, тематические ниши, где есть потенциал для расширения коллаборации).

Автором разработана методика построения научно-технологического ландшафта для поддержки принятия обоснованных решений и стратегическом планировании.

Под научно-технологическим ландшафтом (НТЛ) понимается текущее состояние и развитие научно-технических областей в конкретном регионе или стране, то есть коллективная экосистема научных знаний, технологических достижений, организаций и специалистов, определяющих прогресс в этих областях.

Для построения НТЛ модель исходного цифрового объекта соотносится с определённым временным интервалом и тематикой. В результате совокупность записей представляет собранные статистические данные вида – тема, дата, количество публикаций (Рисунок 7).

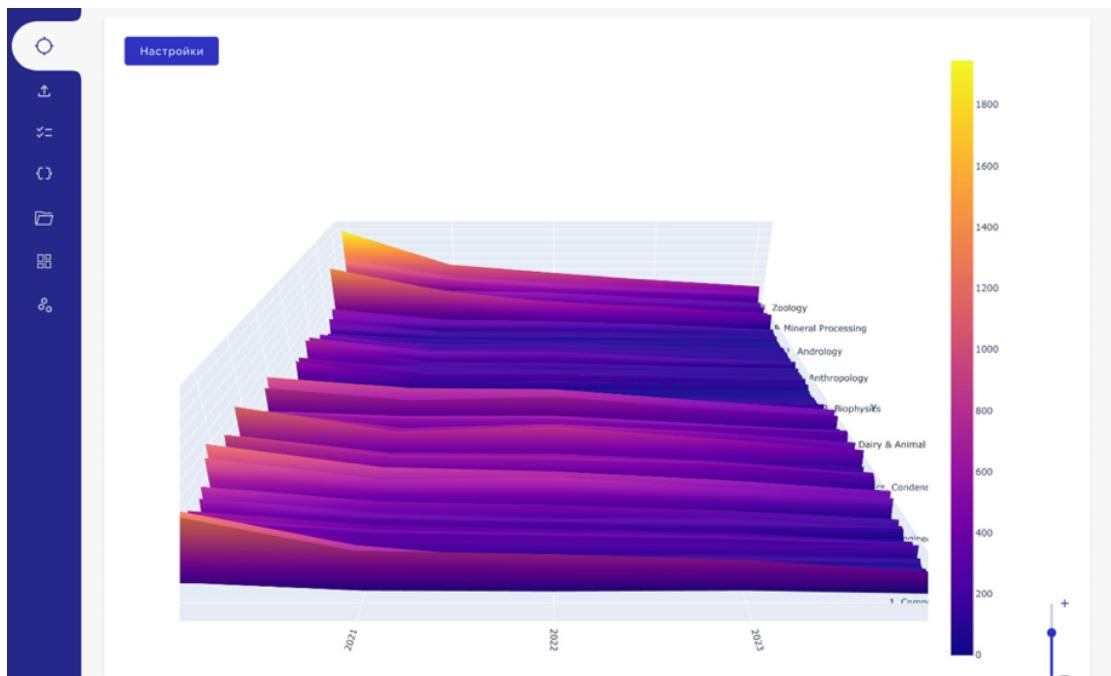


Рисунок 7 Визуализация НТЛ по нескольким странам

Пользователь может выбрать сразу несколько стран и формировать их совмещённый НТЛ за заданный период. Инструмент объединяет данные по выбранным странам и отображает единую поверхность, показывающую суммарный ландшафт. Для непосредственного сравнения двух наборов данных предусмотрен режим с двумя графиками (Рисунок 8). Например, построены параллельно ландшафты Японии и Республики Корея за 2010–2021 гг. на отдельных графиках. Анализируя их, эксперт может увидеть различия: в какие годы в каждой стране был пик публикаций по определённой теме, какая тематика имеет больший относительный вес в одной стране по сравнению с другой и т.д. Оказалось, что подобное визуальное сравнение очень эффективно для выявления национальной специфики научных приоритетов.

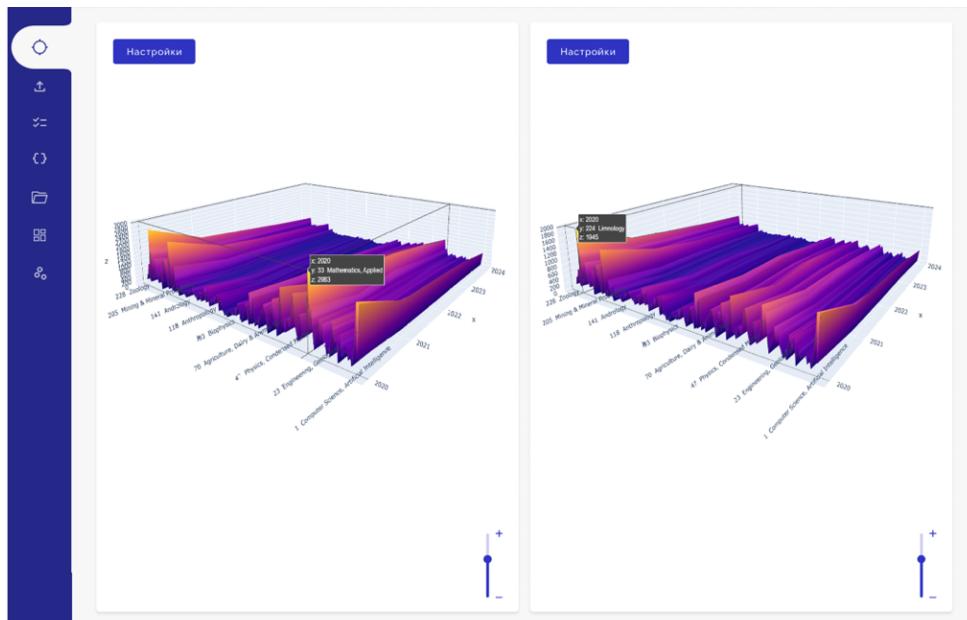


Рисунок 8 Визуализация сравнения НТЛ двух стран  
(Япония и Южная Корея)

Рассмотренные практические примеры подтверждают, что предложенный автором инструмент позволяет оценить динамику научной активности, выявить ключевые темы и проследить изменение количества публикаций по тематикам исследований во времени. Благодаря многоуровневой фильтрации, можно переключаться от макро-анализа (все области науки по стране) к микро-анализу (отдельная тема по нескольким странам). Программный инструмент полезен для стратегического планирования: карты знаний и технологий помогают выявить, насколько полно охвачена та или иная сфера и перспективные направления дальнейших исследований.

В разделе сформулирована целостный подход к интеллектуальному анализу информационных объектов в научно-технических и социально значимых задачах, объединяющий все разработанные автором элементы – модели цифровых объектов ( $Obj$ ,  $AObj$ ,  $CAObj$ ), методы преобразования данных из разнородных информационных ресурсов; функции, алгоритмы расчета характеристик, методы визуализации – в единую систему (Рисунок 9).

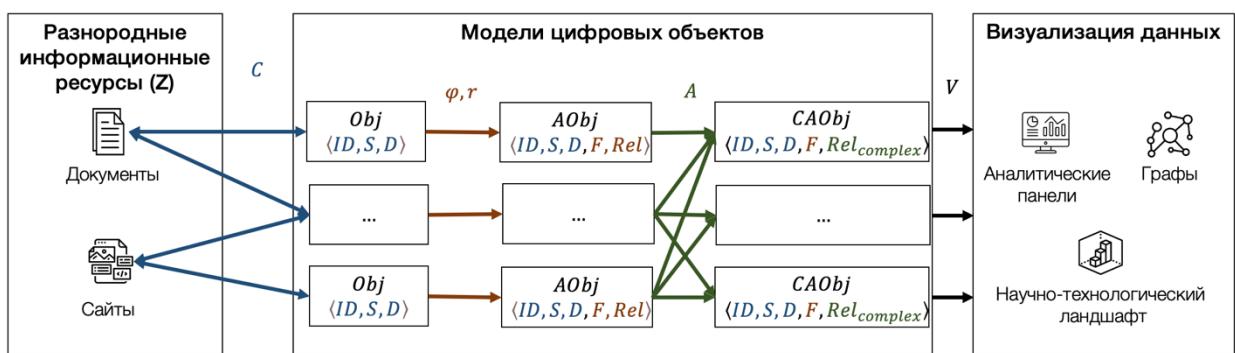


Рисунок 9 Графическое представление системы интеллектуального анализа информационных объектов

Предложенную автором систему интеллектуального анализа информационных объектов можно представить в виде следующий математической модели:

$$\left\{
 \begin{array}{l}
 \begin{aligned}
 C: Z \rightarrow Obj, \text{ где} \\
 & Z - \text{множество разнородных} \\
 & \text{информационных ресурсов,} \\
 & Obj - \text{базовая модель цифрового} \\
 & \text{объекта,} \\
 & C - \text{методы преобразования данных;} \\
 \\ 
 M: Obj \rightarrow CAObj, \text{ где} \\
 & CAObj - \text{модель комплексного} \\
 & \text{цифрового объекта,} \\
 & M = A \circ r \circ \varphi - \text{функции, алгоритмы} \\
 & \text{расчета характеристик;} \\
 & (CAObj = AObj \text{ при } A = \emptyset)
 \end{aligned}
 \end{array}
 \right. \quad (18)$$

$$\begin{aligned}
 & K - \text{решенные аналитические задачи,} \\
 & V - \text{методы визуализации.}
 \end{aligned}$$

Применение системы интеллектуального анализа данных на практике позволило повысить производительность труда аналитиков в разы, сократить время подготовки аналитических материалов с месяцев до недель и обеспечить принятие решений на основе более полного фактического материала. Полученные в диссертации результаты открывают перспективы для дальнейших исследований: например, углублённая интеграция методов искусственного интеллекта (в частности, использование глубоких нейронных сетей для анализа текстов в связке с предлагаемой моделью), расширение применения в новых областях (гуманитарные исследования, экономический мониторинг) и создание обучающих программ для специалистов, которые будут пользоваться такими инструментами.

Шестой раздел аккумулирует весь опыт работы и демонстрирует, что создан не просто набор разрозненных алгоритмов, а целостная система интеллектуального анализа современной информации. Это соответствует поставленной цели работы и свидетельствует о решении всех намеченных задач исследования. Практические примеры показали, что программный комплекс позволяет решать прикладные задачи анализа информации – от сбора данных до наглядного представления результатов. Внедрение комплекса и его частей в различных организациях подтверждает эффективность предложенных решений и их востребованность.

**В заключении** отражены основные результаты, полученные в диссертационной работе.

**Приложения** содержат акты о внедрении результатов диссертационной работы, пример структуры полученных данных научной публикации по результатам извлечения и насыщения данных.

## ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

В диссертации разработаны методы, модели и технологии интеллектуального анализа цифровых информационных объектов в научно-технических и социально значимых задачах, агрегирующая модели представления цифрового объекта, методы извлечения и насыщения данных, программные инструменты визуализации разнородной информации. Основные результаты диссертационной работы заключаются в следующем:

1. Разработана система интеллектуального анализа данных, объединяющая формальную модель представления цифровых объектов, методы автоматизированной обработки данных и специализированные аналитические инструменты. Предложенная система обеспечивает единый подход к работе с разнородной научно-технической и социальной информацией, что подтверждено её успешным применением в нескольких предметных областях.

2. Предложена модель комплексного цифрового информационного объекта, включающая уникальный идентификатор, статические, динамические и вычисляемые характеристики, а также связи (отношения) с другими объектами. Показано, что использование данной модели повышает полноту представления информации о каждом объекте, облегчает слияние

данных из разных источников и снижает неоднозначность интерпретации сведений.

3. Разработан набор методов извлечения и насыщения данных из слабоструктурированных информационных источников (PDF-документов, веб-страниц, социальных сетей). В их числе: алгоритмы парсинга текстов научных статей с восстановлением структуры, методы выделения значимых сущностей (ключевых слов, физических величин) из текста, процедура унификации аффилиаций авторов и геокодирования организаций, а также алгоритмы выделения изображений и таблиц из документов. Комплексное применение этих методов позволяет автоматизированно преобразовывать массивы документов в структурированную базу знаний, пригодную для дальнейшего анализа.

4. Разработана методика аналитического описания социальных объектов и выявления целевых групп пользователей. В рамках методики предложено представлять цифровой профиль пользователя как совокупность статических и динамических характеристик, ранжировать эти характеристики по значимости и вычислять интегральный критерий принадлежности профиля к искомой категории. Методика апробирована при решении практических задач и показала высокую эффективность: автоматическая идентификация целевых профилей достигла точности, сопоставимой с экспертной, значительно сократив при этом объём ручной работы специалистов.

5. Выполнен комплексный анализ научно-технической информации с использованием разработанных инструментов. Реализованы интерактивные панели визуализации, отражающие динамику и географию развития таких областей, как технологии больших данных, биомедицинские исследования, финансовая безопасность и др. Получены новые научно-практические выводы: определены лидирующие страны и организации в указанных областях, выявлены основные тематические кластеры и тренды эволюции тематик, показана степень международного сотрудничества. Полученные результаты подтверждают, что интеграция разнородных данных (научные публикации, проекты, нормативные документы) в едином информационном пространстве значительно расширяет возможности экспертного анализа и принятия решений.

6. Применение программного комплекса в проектах по анализу научно-технической информации (например, создание базы данных свойств материалов для ГК «Росатом») позволило сократить время обработки данных в несколько раз и извлечь из текстов десятки тысяч структурированных фактов. В социально значимых задачах внедрение алгоритмов привело к повышению оперативности реагирования и более точному выявлению целевой аудитории.

7. Реализованы специализированные программные средства интеллектуального анализа данных, в части построения графовых представлений, позволяющих выявлять неявные связи между объектами и автоматизированного построения научно-технологического ландшафта, позволяющего определить и сравнить динамику изменения научно-технологической области стран.

8. Разработан программный комплекс интеллектуального анализа данных, обеспечивающий автоматизированный сбор данных из сети Интернет, их хранение в документно-ориентированном хранилище, обогащение метаданными и аналитическими признаками, а также интерактивную визуализацию результатов через веб-интерфейсы. Структура программного комплекса спроектирована с учётом горизонтальной масштабируемости и модульности, что подтверждено испытаниями: система успешно обработала свыше 50 тысяч документов и легко расширяется для новых типов объектов.

9. Результаты диссертационной работы апробированы в рамках выполнения научно-исследовательских и опытно-конструкторских работ по Государственным заданиям Министерства науки и высшего образования Российской Федерации, Фонда перспективных исследований, организаций контура Госкорпорации Росатом (НИИ «Графит», ФГУП «РФЯЦ-ВНИИТФ им. академ. Е.И. Забабахина», ФГУП ВНИИА им. Н.Л. Духова), Российского энергетического агентства. Научные и технические положения, изложенные в диссертации, использованы в проектах других систем, имеющих специальное назначение.

## ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

### I. Публикации, представленные в международных базах цитирования Scopus и Web of Science

1. Ulizko M.S., **Artamonov A.A.**, Tukumbetova R.R., Antonov E.V., Vasilev M.I. Critical Paths of Information Dissemination in Networks // Scientific Visualization. 2022. Vol. 14. No. 2. pp. 98-107.
2. **Artamonov A.A.**, Ionkina K.V., Kirichenko A.V., Lopatina O.L., Tretyakov E.S., Cherkasskiy A.I. Agent-based search in social networks International // Journal of Civil Engineering and Technology. 2018. Vol. 9. No. 13. pp. 28-35.
3. Onykiy B.N., **Artamonov A.A.**, Tretyakov E.S., Ionkina K.V. Visualization of large samples of unstructured information on the basis of specialized thesauruses // Scientific Visualization. 2017. Vol. 9. No. 5. pp. 54-58.
4. Ulizko M.S., Antonov E.V., **Artamonov A.A.**, Tukumbetova R.R. Visualization of graph-based representations for analyzing related multidimensional objects // Scientific Visualization. 2020. Vol. 12. No. 4. pp. 133-142.
5. Ulizko M. S., Antonov E. V., Grigorieva M. A., Tretyakov E.S., Tukumbetova R.R., **Artamonov A.A.**. Visual analytics of twitter and social media dataflows: a casestudy of Covid-19 rumors // Scientific Visualization. 2021. Vol. 13. No. 4. pp. 144-163.
6. Onykiy B.N., Antonov E.V., **Artamonov A.A.**, Tretyakov E.S. Information Analysis Support for Decision-Making in Scientific and Technological Development // International Journal of Technology. 2020. Vol. 11. No. 6. pp. 1125-1135.
7. Antonov E.V., **Artamonov A.A.**, Rudik A.V., Malugin M.I. Trend Visualization of Academic Field: Proposed Method and Big Data Review // Scientific Visualization. 2022. Vol. 14. No. 2. pp. 62-76.

8. Sokolov I., Antonov E., **Artamonov A.** Evaluation of Named Entity Recognition Software Packages for Data Mining // Physics of Particles and Nuclei. Vol. 55. No. 3. pp. 557-559.

9. Malugin M., Antonov E., **Artamonov A.** Designing a System for Monitoring the Publication Activity of the Scientific Organization // Physics of Particles and Nuclei. 2024. Vol. 55. No. 3. pp. 554-556.

## **II. Публикации в журналах, включенных в перечень периодических изданий ВАК Российской Федерации**

10. Оныкий Б.Н., **Артамонов А.А.**, Третьяков Е.С., Черкасский А.И., Ионкина К.В. Индуктивные модели обучения поисковых агентов, работающих в социальных сетях // Системы высокой доступности. 2020. Т. 16. № 1. С. 5-13.

11. Антонов Е. В., **Артамонов А. А.**, Орлов А. В., Николаев В. С., Захаров В. П., Хохлова М. В., Концевая Ю. М., Бонарцев А. П., Воинова В. В. Обработка научно-технической информации в междисциплинарных исследованиях методами математико-лингвистического направленного поиска на примере области изучения биоматериалов для тканевой инженерии // International Journal of Open Information Technologies, 2022, №10 (11). С. 134-140.

12. Сафиков Д.И., **Артамонов А.А.**, Фомина Ю.Е., Черкасский А.И. Статистическая модель поиска целевых объектов в социальной сети // International Journal of Open Information Technologies. 2018. Т. 12. № 10. С. 71-77.

13. Тукумбетова Р.Р., Улизко М.С., Коренькова Т.В., **Артамонов А.А.** Сравнение методов классификации данных в машинном обучении на примере научных публикаций по ядерному топливному циклу // Системы высокой доступности. 2025. Т. 21. № 1. С. 25-38.

## **III. Публикации материалов конференций**

14. Koshlan D.I., Tretyakov E.S., Korenkov V.V., Onykij B.N., **Artamonov A.A.** Proceedings of the VIII International Conference "Distributed Computing and Grid-technologies in Science and Education" (GRID 2018) // Agent technology situational express analysis in assessment of technological development level of the BRICS countries. Dubna, Moscow region, Russia. 2018. Vol. 2267. pp. 436-440.

15. Fomina J., Safikanov D., **Artamonov A.**, Tretyakov E. Postproceedings of the 10th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2019 (Tenth Annual Meeting of the BICA Society), held August 15-19, 2019 in Seattle, Washington, USA // Parametric and semantic analytical search indexes in hieroglyphic languages. Washington. 2020. Vol. 169. pp. 507-512.

16. Cherkasskiy A., **Artamonov A.**, Cherkasskaya M., Leonova N. 2020 Annual International Conference on Brain-Inspired Cognitive Architectures for Artificial Intelligence: Eleventh Annual Meeting of the BICA Society // Methods for identifying an information object in social networks. 2021. Vol. 190. pp. 137-141.

17. Cherkasskiy A.I., Cherkasskaya M.V., **Artamonov A.A.**, Galin I.Y. User Group Classification Methods Based on Statistical Models // Studies in Computational Intelligence. 2022. Vol. 1032. pp. 69-74.
18. Ulizko M., Pronicheva L., **Artamonov A.**, Tretyakov E. Brain-Inspired Cognitive Architectures for Artificial Intelligence: BICA\*AI 2020 // Complex Objects Identification and Analysis Mechanisms. 2020. pp. 517-526.
19. Tretyakov E.S., Tukumbetova R.R., **Artamonov A.A.** Methodology of analysis of similar objects with the use of modern visualization tools // Mechanisms and Machine Science. 2020. Vol. 80. pp. 113-119.
20. Inkina V.A., Antonov E.V., **Artamonov A.A.**, Ionkina K.V., Tretyakov E.S., Cherkasskiy A.I. Proceedings of the 27th International Symposium Nuclear Electronics and Computing (NEC'2019) // Multiagent information technologies in system analysis. Budva, Becici, Montenegro. 2019. pp. 195-199.
21. Ulizko M., Antonov E., **Artamonov A.**, Tukumbetova R. Graph Visualization of the Characteristics of Complex Objects on the Example of the Analysis of Politicians // CEUR Workshop Proceedings – 2020. 2020. Vol. 2744. pp. short8-1 — short8-9.
22. Ulizko M., Tretyakov E., Tukumbetova R., **Artamonov A.**, Esaulov M. Visualization of Dataflows: a Casestudy of COVID-19 Rumors // CEUR Workshop Proceedings. 2021. Vol. 3027. pp. 259-267.
23. Рудик А.В., Антонов Е.В., **Артамонов А.А.** 33-я Международная конференция по компьютерной графике и машинному зрению // Инструменты оценки научно-технологического ландшафта страны. Москва. 2023. С. 256-265.
24. Ulizko M., **Artamonov A.**, Fomina J., Antonov E., Tukumbetova R. Proceedings of the International Conference on Computer Graphics and Vision “Graphiccon” // Clustering Thematic Information in Social Media. Ryazan. 2022. pp. 403-413.
25. **Artamonov A.**, Vasilev M., Tukumbetova R., Ulizko M. 2022 Annual International Conference on Brain-Inspired Cognitive Architectures for Artificial Intelligence: The 13th Annual Meeting of the BICA Society // Multiagent System for Monitoring, Analysis and Classification of Data from Procurement Services. 2022. Vol. 213. pp. 96-100.
26. **Artamonov A.A.**, Kshnyakov D.O., Danilova V.V., Cherkasskiy A.I., Galin I.Y. 8th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2017 (Eighth Annual Meeting of the BICA Society), held August 1-6, 2017 in Moscow, Russia // Methodology for the Development of Dictionaries for Automated Classification System. Moscow. 2018. Vol. 123. pp. 57-62.
27. **Artamonov A.A.**, Ionkina K.V., Tretyakov E.S., Timofeev A.I. Postproceedings of the 9th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2018 (Ninth Annual Meeting of the BICA Society), held August 22-24, 2018 in Prague, Czech Republic // Electronic document processing operating map development for the implementation of the data

management system in a scientific organization. Prague, Czech Republic. 2018. Vol. 145. pp. 248-253.

28. Kontsewaya Y., Antonov E., **Artamonov A.** 2020 Annual International Conference on Brain-Inspired Cognitive Architectures for Artificial Intelligence: Eleventh Annual Meeting of the BICA Society // Evaluating the Effectiveness of Machine Learning Methods for Spam Detection. 2021. Vol. 190. pp. 479-486.
29. Ulizko M.S., Tukumbetova R.R., Artamonov A.A., Antonov E.V., Ionkina K.V. Biologically Inspired Cognitive Architectures 2023 // Data Preparation for Advanced Data Analysis on Elastic Stack. 2024. pp. 884-893.

#### **IV. Свидетельства о государственной регистрации баз данных**

30. Свидетельство о государственной регистрации базы данных № 2018621999 Российская Федерация. База данных маркированных аудио-объектов/ Оныкий Б.Н., Черкасский А.И., Проничева Л.В., Суслина А.С., **Артамонов А.А.**, Ионкина К.В., Третьяков Е.С., Инкина В.А., Лопатина Е.О.; правообладатель ФГАОУ ВО «Национальный исследовательский ядерный университет «МИФИ» (НИЯУ МИФИ). – № 2018621999; опубл. 11.12.2018 г.
31. Свидетельство о государственной регистрации базы данных № 2017620244 Российской Федерации. Интегральная база данных агентного поиска информации в сети интернет по тематическому направлению аддитивные технологии / Петровский В.Н., Ионкина К.В., **Артамонов А.А.**, Галин И.Ю., Третьяков Е.С., Оныкий Б.Н., Проничева Л.В., Данилова В.В., Кшняков Д.О., Суслина А.С., Черкасский А.И., Быковский Д.П. правообладатель ФГАОУ ВО «Национальный исследовательский ядерный университет «МИФИ» (НИЯУ МИФИ). - №2017620244, 27.02.2017.
32. Свидетельство о государственной регистрации базы данных № 2022620590 Российской Федерации. База данных изображений и метаданных научных публикаций по облученным ядерным материалам за 2014-2018 года / Антонов Е.В., Тукумбетова Р.Р., Чернов И.И., Михальчик В.В., Улизко М.С., Стальцов М.С., **Артамонов А.А.**, Рудик А.В., Малугин М.И.; правообладатель ФГАОУ ВО «Национальный исследовательский ядерный университет «МИФИ» (НИЯУ МИФИ). – № 2022620590; опубл. 18.03.2022 г.
33. Свидетельство о государственной регистрации базы данных № 2023621854 Российской Федерации. База данных сгенерированных объектов информационно-учебного полигона для подготовки специалистов в сфере международных отношений / Антонов Е.В., Ионкина К.В., Кондратько В.О., Смирнова Е.А., **Артамонов А.А.**; правообладатель ФГАОУ ВО «Национальный исследовательский ядерный университет «МИФИ» (НИЯУ МИФИ). – № 2023621854; опубл. 07.06.2023 г.
34. Свидетельство о государственной регистрации базы данных № 2023621163 Российской Федерации. Рецензированные учебно-методические материалы по финансовой безопасности по укрупненным группам специальностей / Норкина А.Н., **Артамонов А.А.**, Морозов Н.В., Антонов Е.В., Улизко М.С., Ионкина К.В., Соколов И.Д., Мальцев М.В.;

правообладатель ФГАОУ ВО «Национальный исследовательский ядерный университет «МИФИ» (НИЯУ МИФИ). – № 2023621163; опубл. 11.04.2023 г.

35. Свидетельство о государственной регистрации базы данных № 2024623588 Российской Федерации. База данных компаний незападных стран – крупнейших торговых партнеров России / Силаев Н.Ю., **Артамонов А.А.**, Бондарев И.М., Аршба Б.Б., Гладышева А.И., Громяк И.В., Жабина Д.П., Лебедев Д.В., Таран В.Е., Тукумбетова Р.Р., Улизко М.С., Фомин М.Ю. правообладатель ФГАОУ ВО «Московский государственный институт международных отношений Министерства иностранных дел Российской Федерации» (МГИМО МИД РФ). – № 2024623588; опубл. 16.08.2024 г.

#### **V. Свидетельства о государственной регистрации программ для ЭВМ**

36. Свидетельство о государственной регистрации программы для ЭВМ № 2017613798 Российской Федерации. Программа автоматизированного квазиреферирования научно-технической информации / Третьяков Е.С., Ионкина К.В., Данилова В.В., Кшняков Д.О., **Артамонов А.А.**, Оныкий Б.Н., Проничева Л.В., Суслина А.С., Суслина И.В., Толстая П.М.; правообладатель ФГАОУ ВО «Национальный исследовательский ядерный университет «МИФИ» (НИЯУ МИФИ). – № 2017613798; опубл. 31.03.2017 г.

37. Свидетельство о государственной регистрации программы для ЭВМ № 2018665872 Российской Федерации. Конвейерное определение тематической принадлежности научно-технической документации / Оныкий Б.Н., **Артамонов А.А.**, Ионкина К.В., Третьяков Е.С., Свистунов А.С., Лопатина Е.О., Зиновьева М.Г., Черкасский А.И., Проничева Л.В., Суслина А.С., Иванченко А.М.; правообладатель ФГАОУ ВО «Национальный исследовательский ядерный университет «МИФИ» (НИЯУ МИФИ). – № 2017613798; опубл. 11.12.2018 г.

38. Свидетельство о государственной регистрации программы для ЭВМ № 2022669767 Российской Федерации. Программа выборки данных по свойствам и структурам облученных реакторных материалов / **Артамонов А.А.**, Стальцов М.С., Антонов Е.В., Чернов И.И., Улизко М.С., Тукумбетова Р.Р., Ионкина К.В.; правообладатели Российская Федерация, от имени которой выступает Государственная корпорация по атомной энергии «Росатом» и ФГУП «Всероссийский научно-исследовательский институт автоматики им. Н.Л. Духова» (ФГУП «ВНИИА»). – № 2022669767; опубл. 25.10.2022 г.

39. Свидетельство о государственной регистрации программы для ЭВМ № 2022664709 Российской Федерации. Программа автоматизированного сбора тендеров по заданной тематике / Тукумбетова Р.Р., **Артамонов А.А.**, Улизко М.С., Антонов Е.В., Ионкина К.В., Васильев М.И.; правообладатель ФГАОУ ВО «Национальный исследовательский ядерный университет «МИФИ» (НИЯУ МИФИ). – № 2022664709; опубл. 03.08.2022 г.

40. Свидетельство о государственной регистрации программы для ЭВМ № 2023617975 Российской Федерации. Программный комплекс формирования и экспертизы учебно-методических материалов / Норкина А.Н., **Артамонов**

**А.А.**, Морозов Н.В., Антонов Е.В., Ионкина К.В., Тукумбетова Р.Р., Улизко М.С., Соколов И.Д.; правообладатель ФГАОУ ВО «Национальный исследовательский ядерный университет «МИФИ» (НИЯУ МИФИ). – № 2023617975; опубл. 18.04.2023 г.

41. Свидетельство о государственной регистрации программы для ЭВМ № 2024616345 Российской Федерации. Программа автоматизированного сбора и обработки публикаций в области биомедицины / Тремасов Г. М., Антонов Е.В., **Артамонов А.А.**, Тукумбетова Р.Р., Хвостова М.О., Соколов И.Д., Патрушев К.А., Чупрыгин С.С.; правообладатель ФГАОУ ВО «Национальный исследовательский ядерный университет «МИФИ» (НИЯУ МИФИ). – № 2024616345; опубл. 19.03.2024 г.

42. Свидетельство о государственной регистрации программы для ЭВМ № 2024616882 Российской Федерации. Программа автоматизированного выделения значений и единиц измерения физических величин из полнотекстовых материалов / Хвостова М.О., Антонов Е.В., Тукумбетова Р.Р., Соколов И.Д., Тремасов Г.М., **Артамонов А.А.**, Матвеева А.Р., Андреев М.Н.; правообладатель ФГАОУ ВО «Национальный исследовательский ядерный университет «МИФИ» (НИЯУ МИФИ). – № 2024616882; опубл. 26.03.2024 г.

43. Свидетельство о государственной регистрации программы для ЭВМ № 2024668072 Российской Федерации. Программа бинарной классификации текстовой информации на основе современных нейросетевых технологий / Антонов Е.В., **Артамонов А.А.**, Ионкина К.В., Кучинов В.П., Соколов И.Д., Тукумбетова Р.Р., Улизко М.С., Черкасский А.И.; правообладатель Акционерное общество «Прорыв». – № 2024668072; опубл. 01.08.2024 г.

44. Свидетельство о государственной регистрации программы для ЭВМ № 2024690021 Российской Федерации. Мультиагентная система сбора, обработки и анализа слабоструктурированных данных / Николаев В.С., **Артамонов А.А.**, Улизко М.С., Антонов Е.В., Кателевский Д.Н.; правообладатель – ООО «Системы информационной аналитики». – № 2024690021; опубл. 11.12.2024 г.

45. Свидетельство о государственной регистрации программы для ЭВМ № 2024690979 Российской Федерации. Аналитический фреймворк обработки и представления научно-технической информации / **Артамонов А.А.**, Тукумбетова Р.Р., Соколов И.Д., Зрелова Д.П., Коренькова Т.В., Хвостова М.О., Вуйкович А.Д., Антонов Е.В., Улизко М.С., Черкасский А.И.; правообладатель ФГАОУ ВО «Национальный исследовательский ядерный университет «МИФИ» (НИЯУ МИФИ). – № 2024690979; опубл. 04.12.2024 г.

## **VI. Учебно-методические пособия**

46. **Артамонов А.А.**, Тукумбетова Р.Р., Антонов Е.В., Сафikanов Д.И. Технологии и средства создания и ведения онтологий в информационно-аналитических системах (учебно-методическое пособие). Москва: НИЯУ МИФИ, 2023. 42 с.

47. Артамонов А.А.; Тукумбетова Р.Р.; Ионкина К.В.; Улизко М.С. Современные технологии и средства построения графа знаний (учебно-методическое пособие). Москва: НИЯУ МИФИ, 2023. 44 с.