

# Институциональный репозиторий публикаций как активный компонент в деятельности современного научного сообщества

В.Н. Добрынин, В.В. Кореньков, И. А. Филозова, Е. Н. Черемисина

**Аннотация** — Статья посвящена исследованию институциональных репозиториев как активных компонентов в деятельности научной организации. Обосновывается необходимость адаптации репозиториев к вызовам цифровой трансформации, таким как увеличение объема данных, разнообразия их форматов и источников, необходимость междисциплинарного подхода. Представлен анализ целевой аудитории репозиториев научных публикаций, изучены информационные потребности ее представителей. Описан портрет современного исследователя с учетом возраста, интересов в научных работах и мотивации. Обсуждаются сложности извлечения знаний из полных текстов публикаций и взаимосвязей между ними. Изложена методология извлечения знаний из текстов научных работ. Описана технология анализа научных работ на основе концептуального шаблона, которая позволяет выявлять и анализировать ключевые аспекты работы, структурируя их с помощью терминов и ассоциаций, а также количественных и качественных показателей. Применение данной технологии проиллюстрировано на примере концептуального шаблона *Проблемы, сформулированные автором*. Представлена общая архитектура информационно-аналитической системы на основе описанной технологии. Рассмотрены задачи, которые она позволяет решать, и услуги, предоставляемые такой системой целевой аудитории. Обсуждаются возможные

перспективы развития системы. Подчеркивается необходимость активного использования репозиториев для развития эффективного взаимодействия научных коллективов.

**Ключевые слова** — анализ научных текстов, репозитории научных публикаций, портрет исследователя, концептуальный шаблон, информационно-аналитический сервис.

## I. ВВЕДЕНИЕ

Современное научное сообщество сталкивается с рядом вызовов, оказывающих существенное влияние на структуру и содержание их информационных потребностей. Первоначально основная задача репозиториев научных публикаций заключалась в обеспечении долговременного хранения и доступности результатов исследований. Однако цифровая трансформация стимулировала возникновение и развитие новых требований.

## II. ВЫЗОВЫ ЦИФРОВОЙ ТРАНСФОРМАЦИИ

Рассмотрим основные факторы, определяющие информационные потребности членов современного научного сообщества.

Рост объема данных и многообразие их источников требуют от исследователя значительных временных ресурсов и усилий по поиску, отбору и анализу информации. Современные исследования, как правило, происходят на стыке нескольких областей науки, что усложняет данные процессы. Появление новых методологических подходов стимулирует ученых быть погруженными в несколько дисциплин одновременно. Новые инструменты анализа данных требуют приобретения и постоянного совершенствования дополнительных знаний и навыков для их эффективного применения. Социальные сети и платформы активно размещают значимые для научных исследований материалы, но поиск пертинентной информации в огромном потоке контента требует усилий [1]. Реализация принципов Открытой науки расширяет возможности исследователей. Ученые активно используют открытые данные, открытое программное обеспечение, что требует знаний и навыков работы с различными форматами и стандартами, а также критического осмысления материала [2]. Оценка надежности и валидности

Статья получена 10 марта 2025.

И. А. Филозова — начальник Группы развития и сопровождения информационных систем общепитетутского назначения, Лаборатория информационных технологий, Объединенный институт ядерных исследований (141980 г. Дубна, Московская обл., ул. Жолио-Кюри, 6); старший преподаватель, Государственный университет «Дубна», Институт системного анализа и управления (141980, Россия, Московская обл., г. Дубна, ул. Университетская, 19), ORCID: <https://orcid.org/0000-0003-3441-7093>, [fia@jinr.ru](mailto:fia@jinr.ru).

В. В. Кореньков — научный руководитель Лаборатория информационных технологий, Объединенный институт ядерных исследований (141980 г. Дубна, Московская обл., ул. Жолио-Кюри, 6), доктор технических наук; заведующий кафедрой распределенных информационно-вычислительных систем, Государственный университет «Дубна», Институт системного анализа и управления (141980, Россия, Московская обл., г. Дубна, ул. Университетская, 19), ORCID: <https://orcid.org/0000-0002-2342-7862>, [korenkov@jinr.ru](mailto:korenkov@jinr.ru).

В. Н. Добрынин — профессор, Государственный университет «Дубна», Институт системного анализа и управления (141980, Россия, Московская обл., г. Дубна, ул. Университетская, 19), кандидат технических наук, старший научный сотрудник, [i@vdobrynin.ru](mailto:i@vdobrynin.ru).

Е. Н. Черемисина — заведующая отделением ФГБУ «Всероссийский научно-исследовательский геологический нефтяной институт» (117105 Москва, Варшавское ш., д. 8), доктор технических наук, академик РАН, профессор, ORCID: <https://orcid.org/0000-0002-6041-8359>, [head@geosys.ru](mailto:head@geosys.ru).

открытых источников информации становится важной частью исследования. Исследователи должны быть в курсе юридических аспектов использования данных и цитирования источников, соблюдать этические нормы, связанные с использованием личных данных и ходом ведения исследований [3-4].

Сложность информационных потребностей современных ученых и исследователей требует от них не только глубоких знаний в своей области, но и способности адаптироваться к быстро меняющимся условиям и требованиям. Использование технологий, понимание междисциплинарных аспектов и осознание этических норм становятся ключевыми факторами успешного ведения исследований.

Исходя из вышесказанного, роль современных репозиториев научных публикаций не должна ограничиваться их базовыми функциями. Традиционные репозитории часто не отражают динамику научных процессов и коллективной деятельности, в рамках которых появляются научные результаты. Репозитории научных публикаций должны быть важными инструментами в современном научном процессе, интегрированными в научные исследования и взаимодействие между учеными [5][6]. Такой подход способствует не только улучшению результатов, но и развитию научной культуры в целом.

### III. ЦЕЛЕВАЯ АУДИТОРИЯ РЕПОЗИТОРИЕВ НАУЧНЫХ ПУБЛИКАЦИЙ

Современный научный исследователь — это человек, занимающийся изучением, анализом и генерацией новых знаний в своей области науки. В зависимости от опыта, сферы деятельности и формата работы, можно выделить несколько типов исследователей. Исходя из общих аспектов научной деятельности составим портрет современного научного исследователя с учетом возраста, интересов и мотивации (см. таблицу 1) [7-12].

Представители всех категорий исследователей используют в качестве информационных источников научные статьи как основной источник информации об актуальных исследованиях и их результатах; монографии, представляющие собой фундаментальные труды, систематизированные знания [13]; патенты, удостоверяющие инновационные технические решения, которые могут быть применимы в исследованиях [14]; конференции как информационные площадки для обсуждения новых идей, личных контактов и обмена мнениями; научные семинары — форма внутренней научной коммуникации для регулярных обсуждений проблем, формирования экспертного мнения [15].

Чаще всего исследователи обращаются к научным публикациям различных жанров в силу следующих обстоятельств: 1) при подготовке публикаций с их соавторством к изданию для поиска источников, подтверждающих сформулированные положения и выдвинутые гипотезы; 2) при разработке новых методик для изучения существующих подходов; 3) в междисциплинарных исследованиях для изучения возможностей применения знаний из других научных дисциплин; 4) при анализе новых тенденций для отслеживания актуальных направлений науки; 5) при

экспертной оценке, например, при рецензировании работ других ученых и специалистов.

Исследователи изучают содержание научных работ с определенными целями: получение новых знаний на основе анализа современных открытий; проверка гипотез путем сопоставления новых данных с теоретическими ожиданиями; разработка новых методов на основе изучения методологических инноваций; поиск научных пробелов через выявление нерешенных вопросов в некоторой области науки; сравнение результатов собственных исследований с уже опубликованными данными; подготовка обучающих материалов — лекций, курсов, семинаров, обзоров — интеграция существующих знаний в образовательный процесс.

С точки зрения читателя в научных работах исследователи стремятся найти ответы на вопросы, которые могут расширить границы знаний, предложить новые теории или решить практические проблемы: Какие новые результаты представлены в работе? Насколько надежны методы исследования? Как работа соотносится с текущими тенденциями в науке? Существуют ли потенциальные противоречия с другими исследованиями? Какие практические выводы можно сделать из результатов? Как исследование можно применить в смежных областях? Какие перспективы дальнейших исследований открывает данная работа? и др.

Разные представители целевой аудитории имеют разные интересы в научных работах. Так, теоретики в научных исследованиях, как правило, интересуются фундаментальными концепциями (новые теории, модели, принципы), обоснованием гипотез (логические и математические доказательства), корреляциями и закономерностями (статистический анализ, аналитические прогнозы), историческим развитием теории (изменения научных парадигм), критическим разбором существующих теорий (ограничения, противоречия, парадоксы) .

Для ученого-экспериментатора важен анализ научных работ для извлечения информации, которая может быть применена в его собственных исследованиях. Основные сведения, представляющие интерес для экспериментатора, касаются методологии исследования; результатов эксперимента и их сравнительный анализ; исходные теории и модели, на которых базируется исследование; информация о предшествующих актуальных исследованиях; проблемы и ограничения, с которыми столкнулись авторы; оценка значимости исследования в контексте современных научных тенденций.

Состояние целевой аудитории не является статичным во времени. Ее динамика проявляется в изменении интересов (смещение фокуса внимания на новые научные проблемы и направления), росте компетенций (повышении уровня знаний и навыков в определенных областях), технологических изменениях (влиянии новых технологий на способы потребления и анализа научной информации), междисциплинарности (повышении интереса к междисциплинарным исследованиям и

подходам).

Таблица 1. Портрет исследователя.

Тип	Возраст (лет)	Характеристика	Интересы	Мотивация	Источники информации
Молодой ученый	20-35	Начинающий исследователь, выпускник магистратуры, аспирант (молодой постдокторант), работающий над своей первой значимой научной работой.	Освоение методологии, поиск актуальных тем, анализ фундаментальных работ.	Получение опыта, завоевание признания в научном сообществе, защита диссертации, публикация в высокорейтинговых журналах.	Учебные пособия, фундаментальные статьи, научные семинары, публикации ведущих исследователей в своей области.
Опытный специалист	35-50	Ученый с профессиональным опытом, исследователь, имеющий значительное количество публикаций, защитивший кандидатскую/докторскую диссертацию, обладающий опытом работы в проектах.	Анализ новых результатов, поиск грантов, участие в международных конференциях.	Развитие собственной научной карьеры, сотрудничество с коллегами, руководитель молодых ученых.	Рецензируемые статьи, монографии, международные конференции, патенты.
Ученый-новатор	40-60	Исследователь, сформировавший собственную научную школу, разрабатывающий уникальные методологии.	Ведение междисциплинарных исследований, создание новых теорий, работа с грантами и финансированием науки.	Развитие своей школы, внедрение инноваций, руководство крупными проектами.	Публикации и конференции по междисциплинарным исследованиям, крупные исследовательские центры, социальные сети и профессиональные сообщества
Исследователь в команде	30-60	Специалист, работающий в коллективе ученых в лаборатории, институте или международном проекте.	Разработка совместных исследований, участие в кооперативных проектах, использование коллективных ресурсов.	Достижение командных целей, публикации совместных работ, применение методов коллег.	Внутренние отчеты о ходе исследований, результаты предварительных экспериментов, совместные рабочие документы, публикации в научных журналах, командные совещания и обсуждения, профессиональные сети
Ученый-эксперт	50+	Ведущий авторитетный специалист в своей области, влияющий на развитие науки, работающий в экспертных советах и международных комиссиях.	Оценка тенденций в науке, участие в формировании научной политики, участие в редколлегиях ведущих научных журналов.	Научное лидерство, консолидация знаний, наставничество нового поколения ученых.	Аналитические отчеты, документация крупных научных проектов, международных научных организаций.

Современный научный исследователь ориентируется на широкий спектр источников, мотивов и условий при изучении научных работ. В зависимости от уровня опыта, профессиональных аспектов своей деятельности он может быть теоретиком, практиком, новатором или экспертом, но его основная цель – поиск новых знаний и их интеграция в науку. Для достижения данной цели требуется глубокий анализ содержания научных работ поскольку сведения, интересующие исследователя, не описываются стандартными библиографическими

метаданными и могут быть выявлены только из полного текста публикаций.

#### IV. ПРОБЛЕМЫ АНАЛИЗА НАУЧНЫХ ТЕКСТОВ ДЛЯ ВЫЯВЛЕНИЯ СКРЫТЫХ ЗНАНИЙ С ТОЧКИ ЗРЕНИЯ ДЕЯТЕЛЬНОСТИ НАУЧНОГО СООБЩЕСТВА

Результаты современных научных исследований публикуются в огромных объемах и в разных источниках (репозитории, журналы, конференции), что

усложняет выявление скрытых взаимосвязей между работами. Из-за отсутствия в репозиториях автоматизированных инструментов для анализа ссылок, терминологии и исследовательских направлений важные взаимосвязи между научными статьями могут оставаться скрытыми. Современная наука активно развивается на стыке дисциплин, но репозитории часто не имеют инструментов для выявления скрытых пересечений между областями знаний. Различные хранилища научной информации используют разные форматы представления данных, что мешает эффективному автоматизированному анализу. Несмотря на то, что во многих журналах оформление публикаций подчинено единым правилам на структуру научной статьи (стандарт IMRAD - Introduction, Methods, Results and Discussion) [16], научные тексты представляют знания в свободной форме, что затрудняет их обработку и интерпретацию методами машинного анализа. Имеются серьезные трудности с извлечением информации из графиков, таблиц, иллюстраций и других визуальных представлений данных, которые часто содержат скрытые закономерности. Репозитории как правило не имеют встроенных развитых средств для выявления скрытых трендов в научных публикациях, научных пробелов и прогнозирования будущих направлений исследований. Открытая наука способствует росту числа публикаций, но в то же время создает вызов в проверке достоверности представленных результатов, что затрудняет выявление ложных и слабых научных работ. Традиционные репозитории недостаточно связаны с экспертными системами и платформами научного общения, что мешает коллективному анализу скрытых знаний. Доступ к некоторым исследованиям ограничен из-за авторских прав, политики держателей цифровых ресурсов, этических норм, что препятствует комплексному анализу научных данных. Вышеперечисленные признаки позволяют сформулировать следующие проблемные ситуации:

1. Сложность выявления скрытых взаимосвязей между работами из-за большого объема и фрагментированности информации, отсутствия в репозиториях автоматизированных инструментов.
2. Трудности в анализе междисциплинарности.
3. Отсутствие единых стандартов структурирования знаний для представления их в научных публикациях.
4. Ограничения в анализе неструктурированных данных, представленных в виде графиков, таблиц и других визуализаций.
5. Отсутствие в репозиториях инструментов для идентификации тенденций и прогнозирования исследований, обнаружения научных пробелов.
6. Трудности проведения оценки надежности информации.
7. Отсутствие интеграции с экспертными системами.
8. Наличие этических и правовых барьеров.

Обозначенные проблемные ситуации создают основы для возникновения определенных проблем. Так, из первой проблемной ситуации может вытекать следующая проблема. Несмотря на то, что в настоящее время существует широкий спектр частных технологий для извлечения скрытых знаний из фрагментированной информации, тем не менее отсутствуют обобщенные

подходы к извлечению знаний из совокупности фрагментированных текстов.

## V. МЕТОДОЛОГИЯ ИЗВЛЕЧЕНИЯ ЗНАНИЙ ИЗ НАУЧНЫХ ПУБЛИКАЦИЙ

Извлечение знаний из полных текстов научных публикаций представляет собой сложный, многогранный процесс, включающий ряд последовательных взаимосвязанных этапов, инструментов и подходов.

### 1. Подготовка данных.

#### 1.1 Сбор данных и выбор текстов для анализа.

Источниками научных публикаций могут служить как институциональные, так и тематические репозитории, а также научные тексты, отобранные экспертами. Поддерживаемые текстовые форматы (с возможностью включения новых по мере необходимости): pdf, docx, txt, odt.

#### 1.2 Предобработка.

На данном шаге выполняется очистка текста для удаления шума (лишних символов, тегов разметки, пробелов и т.д.) и нормализация для приведения текста к унифицированной форме (например, лемматизация и стемминг). Полученные материалы сохраняются для дальнейшего анализа.

### 2. Извлечение явных знаний.

#### 2.1 Текстовый анализ.

Поиск ключевых слов с использованием таких методов как TF-IDF (Term Frequency-Inverse Document Frequency), для определения значимых терминов.

Тематическое моделирование с помощью алгоритмов, таких как LDA (Latent Dirichlet Allocation), для выявления основных тем в текстах.

#### 2.2 Структурирование информации.

Выделение и тегирование ключевых понятий, таких как авторы, даты, методики и т.д.

Организация хранилища для выделенных артефактов.

### 3. Извлечение скрытых знаний.

#### 3.1 Моделирование данных.

Анализ текстов: Применение методов машинного обучения и обработки естественного языка (NLP) для выявления взаимосвязей и паттернов в текстах [17-23].

Классификация и кластеризация: Использование алгоритмов для автоматической классификации публикаций и группировки их по схожести.

#### 3.2 Выявление паттернов.

На данном шаге идентифицируются повторяющиеся структуры, закономерности и взаимосвязи в данных. Паттерны представляют собой закономерности, которые могут быть обнаружены в данных. Примеры паттернов:

- Темы: основные идеи, обсуждаемые в текстах.
- Структуры: форматы и стили, используемые в публикациях.
- Связи: взаимосвязи между различными публикациями, авторами и исследовательскими направлениями.

Сетевой анализ: анализ цитирований, авторских сетей и междисциплинарных связей.

Анализ взаимосвязей: применение графовой аналитики для нахождения скрытых зависимостей и влияний между исследованиями.

### 4. Визуализация и интерпретация.

#### 4.1 Визуализация данных.

Создание разнообразных графических представлений: графиков, диаграмм, карт и других визуализаций, чтобы продемонстрировать выявленные паттерны и связи.

Интерактивные инструменты: платформы для интерактивного анализа данных (Tableau, Gephi, Apache Superset и др.).

#### 4.2 Интерпретация результатов.

Анализ выводов: обсуждение значимости извлеченных знаний и их потенциального влияния на научное сообщество.

Формулирование рекомендаций: выработка предложений по применению выявленных знаний в новых исследованиях и/или практических приложениях.

### 5. Управление.

#### 5.1 Документирование процесса.

Отчетность: составление отчетов о методике, используемых алгоритмах и достигнутых результатах.

Обратная связь: проведение обсуждений в научном сообществе для дальнейшего улучшения и уточнения методологии.

#### 5.2 Управление изменениями.

Корректировка методологии с учетом полученной обратной связи и доработка методов извлечения знаний на основе новых данных и технологий.

Современные технологии и методы позволяют значительно улучшить качество извлекаемой информации из текстов научных работ, что способствует повышению ее практической значимости.

## VI. ТЕХНОЛОГИЯ АНАЛИЗА НАУЧНЫХ РАБОТ НА ОСНОВЕ КОНЦЕПТУАЛЬНОГО ШАБЛОНА

Данная технология анализа научных работ направлена на систематизацию и количественную оценку научных публикаций с использованием концептуального (концептуально-метрического) шаблона. Она позволяет выделить и проанализировать ключевые аспекты работы, структурируя их с помощью терминов и ассоциаций, а также количественных и качественных показателей. При разработке технологии были использованы методы и приемы формирования адаптивных семантических сетей научных текстов (вопрос-ответ-реакция) [24-28].

Рассмотрим основные компоненты и последовательность реализации этой технологии.

#### 1. Структура концептуального шаблона.

Концептуальный шаблон обозначается как

$$KШ = \{T, C, R, G, S, M\},$$

где

T — тема/название шаблона.

C — множество понятий (терминов), относящихся к теме шаблона.

R — множество отношений (связей) между понятиями.

G — цель (описание предполагаемого результата анализа).

S — шкала измерений характеристик.

M — методы количественной оценки характеристик.

##### 1.1 Множество понятий.

Множество понятий шаблона обозначается как:

$$C = \{c_1, c_2, \dots, c_n\},$$

где  $c_i$  — понятие (термин) области применения,  $i \in \{1, 2, \dots, n\}$

##### 1.2 Характеристики понятий.

Каждое понятие с описывается набором характеристик  $A: A = \{a_1, a_2, \dots, a_n\}$

где  $a_i$  — измеримая характеристика, которая определяется через функцию:

$a_i = f(p)$ , где  $p$  — сегмент научного текста или текст целиком.

#### 2. Процесс анализа.

##### 2.1 Определение темы и целей.

Шаг 1: Определение темы концептуального шаблона (T) и описание целей (G) анализа. Например, целью может быть анализ тенденций в определенной области науки.

##### 2.2 Выделение понятий.

Шаг 2: Выделение множества понятий (C) из текста, анализ которых приведет к достижению цели шаблона.

##### 2.3 Определение отношений.

Шаг 3: Установление отношений (R) между выделенными понятиями. Это может включать типы отношений, такие как "является частью", "зависит от" или "связано с".

##### 2.4 Оценка характеристик.

Шаг 4: Определение набора характеристик A для каждого понятия  $c_i$ . Каждая характеристика  $a_i$  должна быть связана с показателями, которые можно измерить.

##### 2.5 Выбор методов количественной оценки.

Шаг 5: Определение методов количественной или качественной оценки M для каждой характеристики.

#### 3. Измерение и анализ.

##### 3.1 Определение шкалы измерений.

Шаг 6: Установка шкалы измерений (S) для каждой характеристики, например, периодические шкалы (раз в месяц) или шкалы оценки (от 1 до 5).

##### 3.2 Вычисление характеристик

Шаг 7: Применение функции для вычисления характеристик:

$$a_i = f(p),$$

где  $p$  — сегменты текста, касающиеся каждого понятия.

##### 3.3 Интерпретация результатов.

Шаг 8: Анализ полученных результатов. На основе измеренных характеристик можно формулировать выводы о значении и влиянии понятий на исследование.

#### 4. Модификация шаблона.

##### 4.1 Гибкость и адаптация.

Концептуальный шаблон является модифицируемым, что позволяет добавлять или удалять характеристики и методы. Это требует пересмотра существующих данных и адаптации анализа в зависимости от конкретных целей или новых научных направлений.

Технология анализа научных работ на основе концептуального шаблона предоставляет системный подход к извлечению знаний из научных текстов. Она способствует как получению количественных оценок, так и качественному пониманию артефактов и значимости научных работ, что может значительно повысить качество исследований и их применение в практике.

#### A. Пример концептуального шаблона Проблемы

Рассмотрим применение технологии анализа научных текстов на примере концептуального шаблона *Проблемы, сформулированные автором*. Проблемы, затронутые в публикации, могут быть классифицированы по различным основаниям: по

степени очевидности – проблемы явные и скрытые, по уровню сложности – простые и сложные, по сфере распространения – общие и частные, по степени изученности – хорошо изученные и малоизученные, по междисциплинарности – локальные и комплексные и др. Составим шаблон для выявления из текста простых и сложных проблем. Решение простых проблем возможно с помощью известных методов. Это пропущенные знания. Для решения сложных проблем

на текущий момент отсутствуют подходящие методы или технологии. Для поиска их решения требуются глубокий анализ, междисциплинарный подход и длительные исследования.

Цель шаблона – выявление простых и сложных проблем из научного текста.

Свойства, их характеристики, методы их измерений, оценки и шкалы шаблона представлены в таблице 2.

Таблица 2. Свойства и характеристики шаблона Простые и сложные проблемы.

Свойство	Характеристика	Шкала измерений	Метод измерений	Метод оценки	Формула расчета
Наличие пробелов в знаниях	Частота фраз «не исследовано», «остается открытым», «требует дальнейшего изучения»	Количественная (количество на 1000 слов)	NLP-анализ (поиск ключевых слов)	Пороговый анализ (если частота > X, проблема значима)	$F = \frac{N_2}{N_1} \times 1000$ где N <sub>2</sub> – число фраз, N <sub>1</sub> – общее число слов в тексте
Наличие междисциплинарных связей	Процент терминов из разных областей	Относительная (процентное соотношение)	Тематическое моделирование (LDA, BERT)	Выявление тематических кластеров	$M = \frac{N_d}{N_t} \times 100$ где N <sub>d</sub> – число терминов из разных дисциплин, N <sub>t</sub> – общее число терминов.
Наличие методологических ограничений	Частота фраз «не удалось», «ограничение метода», «не применимо»	Количественная (количество на 1000 слов)	NLP-анализ Контекста	Анализ смысловой нагрузки фраз	$L = \frac{N_l}{N_t} \times 1000$ где N <sub>l</sub> – число фраз, N <sub>t</sub> – общее число слов в тексте.
Соответствие текста структуре IMRAD	Отклонение от IMRAD	Категориальная (доля несоответствий)	Классификация структуры текста	Сравнение с эталонными статьями	$S = \frac{N_d}{N_t} \times 1000$ где N <sub>d</sub> – число отклонений, N <sub>t</sub> – общее число секций IMRAD.
Достоверность информации	Количество цитат, доля ссылок на источники	Относительная (процентное соотношение) Качественная (надежность источников)	Анализ библиографии	Сравнение с НМБД (Scopus, WoS, elibrary)	$C = \frac{N_c}{N_t} \times 100$ где N <sub>c</sub> – число цитат, N <sub>t</sub> – общее число ссылок.

Также на основе значений характеристик таблицы 2 необходимо рассчитать степень уверенности в отнесении проблемы к простой или сложной. Метод оценки степени достоверности:

#### 1. Выявление пробелов в знаниях (F)

$F = 0.0$  → высокая уверенность в том, что проблема существует (100%).

$F > 0.0$  → средняя или низкая уверенность.

#### 2. Междисциплинарные связи (M)

$M < 1\%$  → низкий уровень междисциплинарности, высокая уверенность в наличии проблемы (90–100%).

$M > 1\%$  → средняя уверенность (50–70%).

#### 3. Методологические ограничения (L)

$L = 0.0$  → высокая уверенность, что проблема требует новых подходов (80–100%).

$L > 0.0$  → проблема скорее простая (50–70%).

#### 4. Соответствие IMRAD (S)

$S = 100\%$  → отсутствие проблем со структурой, высокая уверенность в необходимости унификации

знаний (90–100%).

$S < 100\%$  → средняя уверенность (60–80%).

#### 5. Достоверность информации (C)

$C < 0.5\%$  → низкий уровень цитируемости, высокая уверенность в проблеме (80–100%).

$C > 0.5\%$  → средняя уверенность (50–70%).

Проблемы с уверенностью 85–95% потенциально могут быть сложными, если выявятся новые факторы.

#### *B. Применение концептуального шаблона Проблемы для анализа текста*

Для описанного в разделе VI А концептуального шаблона была выполнена программная реализация на языке python и применена для анализа текста данной статьи. Результаты анализа представлены в таблицах 3 и 4.

Таблица 3. Выявленные проблемы

Тип проблемы	Выявленная проблема	Свойство (Характеристика)	Значение	Обоснование
Простые проблемы (пропущенные знания)	Отсутствие автоматизированных инструментов анализа научных текстов в репозиториях	Наличие пробелов в знаниях (F)	0.0	В тексте не содержатся фразы о наличии автоматизированных инструментов для анализа научных текстов
	Недостаток инструментов для выявления междисциплинарных взаимосвязей	Наличие междисциплинарных связей (M)	0.25%	Низкий процент междисциплинарных терминов указывает на слабую интеграцию знаний из разных областей
	Разные форматы представления данных в научных хранилищах	Наличие методологических ограничений (L)	0.0	В тексте отсутствуют упоминания о стандартизации форматов и инструментах их унификации
	Отсутствие стандартов структурирования знаний	Соответствие текста структуре IMRAD (S)	100%	Полное соответствие IMRAD говорит о следовании структуре, но отсутствие упоминаний о единых стандартах структурирования указывает на проблему
	Проблемы обработки неструктурированных данных (графики, таблицы)	Наличие методологических ограничений (L)	0.0	Нет упоминаний о методах автоматизированной обработки графиков, таблиц и других визуальных данных
	Ограниченность инструментов для прогнозирования научных трендов	Достоверность информации (C)	0.29%	Низкий уровень цитирования указывает на недостаточную аналитическую обработку и прогнозирование тенденций
Сложные проблемы (требуют новых знаний и технологий)	Выявление скрытых взаимосвязей между научными публикациями	Наличие междисциплинарных связей (M)	0.25%	Междисциплинарные термины встречаются редко, что указывает на отсутствие инструментов для выявления скрытых взаимосвязей
	Разработка методов оценки надежности научных данных	Достоверность информации (C)	0.29%	Низкий уровень цитирования может указывать на недостаточные механизмы верификации научных данных
	Интеграция репозитория с экспертными системами	Выявление пробелов в знаниях (F)	0.0	В тексте нет упоминаний об использовании экспертных систем для анализа научных данных
	Разработка универсальных подходов к анализу фрагментированных текстов	Выявление пробелов в знаниях (F)	0.0	В тексте не выявлены методики, позволяющие обобщать знания из разрозненных источников
	Совершенствование методов анализа естественного языка (NLP) для научных текстов	Выявление методологических ограничений (L)	0.0	Отсутствие упоминаний о развитии NLP-инструментов для обработки научных публикаций
	Разработка интеллектуальных систем, предсказывающих перспективные научные направления	Достоверность информации (C)	0.29%	Низкое количество ссылок может говорить о недостатке механизмов прогнозирования и автоматического выявления научных трендов

Таблица 4. Выявленные проблемы с оценкой достоверности

<i>Выявленная проблема</i>	<i>Рассчитанное значение</i>	<i>Степень достоверности (0–100%)</i>	<i>Категория (простая/сложная)</i>
Отсутствие автоматизированных инструментов анализа	F = 0.0	100%	Простая
Недостаток инструментов выявления междисциплинарных связей	M = 0.25%	95%	Простая
Разные форматы представления данных	L = 0.0	100%	Простая
Отсутствие стандартов структурирования знаний	S = 100%	90%	Простая
Проблемы обработки графиков и таблиц	L = 0.0	100%	Простая
Ограниченность инструментов прогнозирования трендов	C = 0.29%	85%	Простая
Выявление скрытых взаимосвязей между публикациями	M = 0.25%	95%	Сложная
Оценка надежности научных данных	C = 0.29%	85%	Сложная
Интеграция репозиторий с экспертными системами	F = 0.0	100%	Сложная
Разработка методов анализа фрагментированных текстов	F = 0.0	100%	Сложная
Совершенствование NLP для научных текстов	L = 0.0	100%	Сложная
Интеллектуальные системы предсказания трендов	C = 0.29%	85%	Сложная

Результаты такого анализа могут быть полезны разным группам научного сообщества, включая исследователей, разработчиков технологий обработки данных, специалистов по научной аналитике и организаторов научной деятельности. Молодым ученым и аспирантам возможность проведения такого анализа облегчит выбор актуальных тем исследований, обоснование новизны работы. Для опытных исследователей представляет интерес выявление трендов, междисциплинарных связей. Анализ позволит находить новые перспективные направления и эффективнее планировать исследования. Администраторы репозиторий и баз данных заинтересованы в постоянном улучшении структуры данных, автоматизации поиска. Возможности подобного аналитического инструмента позволят им более эффективно разрабатывать интеллектуальные сервисы анализа научных публикаций. Специалисты в области научной политики нуждаются в инструментах оценки достоверности публикаций, влиянии трендов на науку.

Результаты анализа могут быть использованы при формировании стратегии научного развития и принятии решения о финансировании перспективных направлений.

## VII. АРХИТЕКТУРА ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОЙ СИСТЕМЫ ИНТЕЛЛЕКТУАЛЬНОГО ОБЩЕНИЯ С НАУЧНЫМ СООБЩЕСТВОМ

Архитектура информационно-аналитической системы (ИАС) для интеллектуального общения с научным сообществом, построенная на технологии анализа научных работ с использованием концептуального шаблона, включает в себя несколько ключевых компонентов. Эти компоненты обеспечивают сбор, обработку, анализ и представление научной информации, что позволяет исследователям эффективно взаимодействовать друг с другом и развивать научное знание.

### 1. Общая структура системы.

Архитектура ИАС представлена в виде следующих основных блоков:

#### 1.1 Блоки архитектуры.

Пользователи: Научные работники, исследователи, студенты, эксперты и другие участники научного сообщества. Делятся на 2 основные категории: 1) пользователь - лицо, осуществляющее запуск аналитических запросов по хранящимся в системе концептуальным шаблонам; 2) эксперт - лицо, осуществляющее разработку концептуальных шаблонов и пополнение библиотеки шаблонов.

Интерфейс пользователя (UI): Интуитивно понятный интерфейс, обеспечивающий доступ к функциям системы, таким как поиск, анализ и представление результатов. Взаимодействие пользователей с ИАС осуществляется посредством автоматизированных рабочих мест (АРМ) эксперта и пользователя.

АРМ эксперта обеспечивает реализацию двух основных функций: 1) модификация доступных в системе концептуальных шаблонов; 2) описание нового концептуального шаблона. В режиме редактирования шаблона эксперт может добавлять, удалять или изменять его свойства, их характеристики и методы оценки. При отсутствии в системе подходящего для решения некоторой задачи анализа концептуального шаблона, эксперт формирует его через описание свойств, их характеристик, методов оценки и шкал измерения. Формализованное описание шаблона сохраняется в системе в библиотеку заданий для реализации. Разработчик посредством API извлекает описание нового шаблона и реализует его. После чего готовый шаблон интегрируется в систему и становится доступным в библиотеке шаблонов.

АРМ пользователя реализует функции поиска (выбора из списка доступных) и конструирования аналитических запросов и сохранение их в библиотеке запросов. При сборке запроса пользователь формирует его через конструктор аналитических запросов, выбирая 1) концептуальный шаблон, его свойства и характеристики из библиотеки доступных шаблонов, 2) файлы с текстами для анализа. После завершения сборки запускается модуль анализа, аналитический запрос запускается в библиотеку запросов и доступен для повторного использования. Результаты анализа отображаются на графическом интерфейсе как в текстовой форме, так и в форме визуализаций. Пользователь имеет возможность сохранить результаты анализа в виде отчета в указанный файл.

Система ввода данных: Модуль для загрузки научных публикаций и других материалов в систему, поддерживающий различные форматы. Данный блок обеспечивает импорт библиографических метаданных и полных текстов научных статей из внешних источников (репозитории и баз данных). Требуемые данные могут поступать в систему через общую шину данных или хранилище цифровой экосистемы.

Хранилище данных: Центральное хранилище, где аккумулируются данные о научных работах, концептуальных шаблонах и результатах анализа. Для организации хранения текстов научных публикаций, их метаданных и результатов анализа может быть использована реляционная и/или нереляционная базы данных.

Модуль анализа: Основной компонент системы, который реализует технологии анализа научных работ на основе концептуального шаблона, обеспечивает извлечение понятий, их характеристик и метрик. Данный блок включает использование методов

машинного обучения для анализа текстов, выявления паттернов и метрик.

Результаты анализа: Отчетный вывод о результатах, включая анализ исследовательских трендов, взаимосвязей и рекомендаций.

База знаний: Ресурс, который хранит накопленные знания и полезные факты, извлеченные из научных публикаций в виде структурированных данных: концептуальные шаблоны, аналитические запросы, метрики. Доступ к знаниям обеспечивается инструментами для поиска и извлечения полезной информации из базы знаний.

Модуль обратной связи: Система для взаимодействия с пользователями, позволяющая получать отзывы и улучшать систему на основе пользовательского опыта. Сбор обратной связи обеспечивается механизмами для сбора отзывов от пользователей о работе системы и полученных результатах. На основе собранной обратной связи происходит непрерывный процесс обновления и доработки системы, поддерживая ее адаптивность к изменяющимся требованиям пользователей.

## 2. Применение системы.

### 2.1 Интерактивные отчеты.

Модуль анализа предоставляет пользователям отчеты о результатах анализа как в текстовой, так и в графической форме.

### 2.2 Рекомендации по исследованиям.

На основе результатов анализа модуль формирует рекомендации по дальнейшим направлениям исследований.

### 2.3 Сетевое сотрудничество.

Система может поддерживать создание научных сетей, основанных на общей тематике и интересах пользователей, способствуя более активному научному взаимодействию.

Диаграмма на рис.1 иллюстрирует структуру информационно-аналитической системы интеллектуального общения с научным сообществом, включая ее ключевые блоки, связи и внешние источники данных. Информационно-аналитическая система, реализованная на изложенной выше архитектуре, позволяет решать следующие задачи:

## 1. Задачи по обработке и структурированию научных текстов

1.1. Автоматизированный сбор и предобработка данных путем импорта метаданных научных публикаций и их полных текстов из различных репозиториях и хранилищ.

### 1.2. Извлечение явных знаний:

- Выявление ключевых понятий, авторов, дат, методик, ключевых слов, проблем, вопросов и других артефактов, присутствующих в тексте.

- Классификация публикаций по тематическим направлениям.

### 1.3. Выявление скрытых знаний и взаимосвязей:

- Автоматизированный анализ текстов для обнаружения закономерностей.

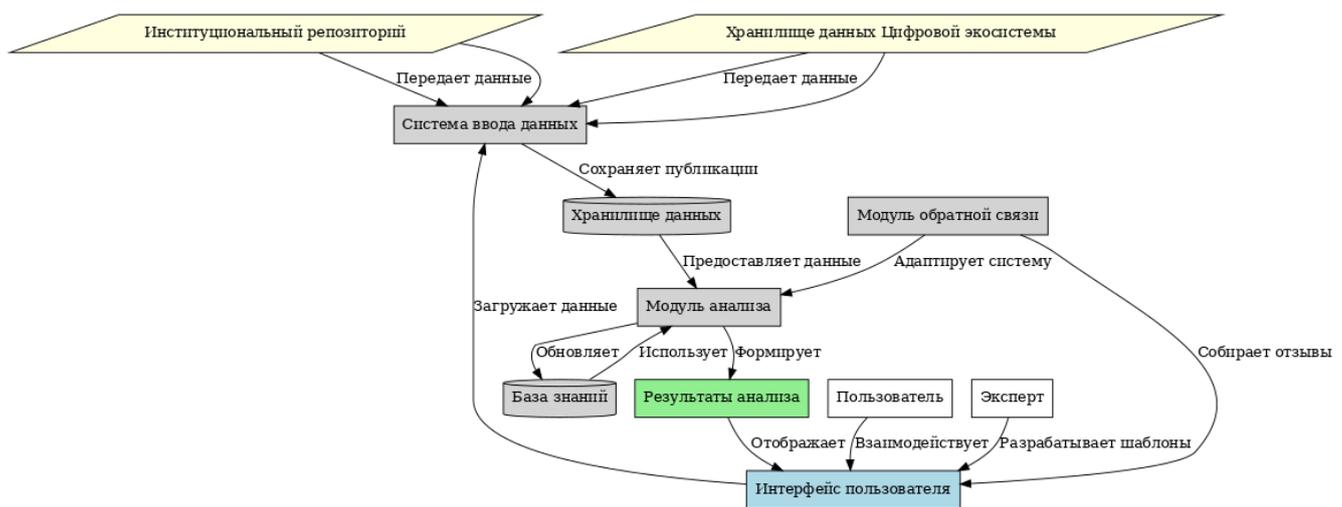


Рис.1. Диаграмма компонентов ИАС

- Определение скрытых связей между статьями, авторами и исследовательскими направлениями.
  - Выявление научных пробелов и перспективных тем для исследований.
2. Задачи по анализу и представлению научных данных.
- 2.1. Тематический анализ и моделирование данных:
- Тематическое моделирование для выделения основных тем.
  - Кластеризация публикаций и выявление трендов научных исследований.
- 2.2. Прогнозирование и рекомендации:
- Формирование персонализированных рекомендаций по научным работам.
  - Прогнозирование новых трендов и направлений исследований.
3. Задачи по взаимодействию с пользователями.
- 3.1. Создание и редактирование концептуальных шаблонов.
- Экспертам предоставляется возможность разрабатывать, редактировать и настраивать шаблоны анализа.
- 3.2. Формирование аналитических запросов.
- Пользователи могут создавать запросы для выявления информации по концептуальным шаблонам.
- 3.3. Визуализация и интерпретация данных:
- Графическое представление результатов анализа (графы, диаграммы, карты знаний).
  - Интерактивные отчеты с возможностью фильтрации и уточнения данных.
- 3.4. Сбор обратной связи и адаптация системы:
- Улучшение качества анализа на основе пользовательских оценок и замечаний.
  - Постоянное обновление базы знаний и алгоритмов анализа.
4. Задачи по обеспечению работы базы знаний:
- 4.1. Хранение и управление извлеченной информацией:
- Организация базы знаний с возможностью быстрого доступа к извлеченным данным.
  - Актуализация информации по мере появления новых публикаций.
- 4.2. Интеграция с научными экосистемами:
- Объединение данных из различных репозиторий и цифровых платформ.

- Обеспечение доступа к аналитическим данным для исследователей, научных организаций и грантовых фондов.

Информационно-аналитическая система, основанная на концептуальных шаблонах и методах анализа научных текстов, решает широкий спектр задач — от сбора и обработки данных до выявления скрытых знаний и поддержки научного взаимодействия.

Успешное выполнение перечисленных задач является основой для предоставления качественных услуг целевой аудитории ИАС. Перечень услуг, предоставляемых системой научному сообществу, включает:

1. Услуги по управлению и анализу научных данных.
  - 1.1. Автоматизированный поиск и сбор научных публикаций. Обеспечивается доступ к научным статьям из институциональных репозиторий и цифровых хранилищ.
  - 1.2. Обнаружение скрытых взаимосвязей между публикациями. Предоставляются аналитические инструменты для выявления междисциплинарных связей, научных пробелов и новых направлений исследований.
  - 1.3. Создание концептуальных шаблонов для анализа научных текстов. Предоставляются возможности разработки и редактирования экспертами шаблонов, настройки их свойств, характеристик и метрик для систематизированного анализа публикаций.
  - 1.4. Формирование аналитических отчетов. Генерация отчетов, визуализаций данных различных видов (диаграммы, графики, карты знаний) по результатам анализа.
2. Услуги по поддержке исследовательской деятельности.
  - 2.1. Рекомендательная система научных публикаций. Предоставляются персонализированные рекомендации на основе интересов пользователя, наиболее релевантные работ в заданной тематике.
  - 2.2. Автоматическая проверка качества научных работ. Предоставляются результаты анализа качества публикаций по соответствию жанру научной публикации, структуре научного текста (IMRAD),

репутации журналов и др. признакам. Определение несоответствий, дублирующих или сомнительных публикаций.

2.3. Интерактивный анализ и прогнозирование научных трендов.

Выявление новых направлений исследований с помощью машинного обучения. Прогнозирование динамики развития научных дисциплин.

2.4. Поддержка подготовки научных публикаций.

Автоматизированная проверка соответствия работы научным стандартам и требованиям издательства.

3. Услуги по взаимодействию и научному сотрудничеству.

3.1. Создание и управление научными профилями пользователей.

Персональные кабинеты для исследователей с историей аналитических запросов.

Инструменты для отслеживания научной активности коллег и соавторов.

3.2. Формирование научных сетей и коллабораций.

Поиск потенциальных соавторов по схожести исследований.

4. Услуги по повышению прозрачности и доступности научных данных.

4.1. Открытый доступ к анализируемым данным:

Размещение данных в репозитории открытого доступа при соблюдении правовых ограничений. Публикация отчетов и аналитических материалов.

4.2. Мониторинг научной активности и метрик.

Инструменты для отслеживания уровня цитируемости и индекса влияния публикаций, анализа динамики публикационной активности по различным научным областям.

4.3. Обратная связь и адаптация системы:

Возможность пользователям оставлять отзывы и предложения по улучшению функционала. Постоянное совершенствование алгоритмов анализа и механизмов поиска.

Система предоставляет комплексные услуги для научного сообщества, охватывая поиск, анализ, прогнозирование, научное сотрудничество и мониторинг исследовательской деятельности. Это делает ее важным инструментом для поддержки современных научных исследований и управления знаниями.

## VIII. ПЕРСПЕКТИВЫ РАЗВИТИЯ ИАС НА ОСНОВЕ ТЕХНОЛОГИИ КОНЦЕПТУАЛЬНЫХ ШАБЛОНОВ

Развитие системы может идти в нескольких ключевых направлениях, включая технологическое совершенствование, расширение функциональности и интеграцию с глобальными научными экосистемами.

1. *Технологическое развитие и улучшение аналитических возможностей.*

1.1. Использование методов глубокого обучения и искусственного интеллекта (ИИ).

Развитие алгоритмов машинного обучения для предсказания новых научных направлений. Применение нейросетей для автоматической аннотации и категоризации публикаций. Разработка интеллектуальных ассистентов для поддержки исследователей.

1.2. Углубленный анализ естественного языка (NLP) в научных публикациях.

Улучшение распознавания терминов и структур научных статей. Определение контекста исследований для более точной классификации публикаций. Расширенные возможности поиска с учетом семантического анализа.

1.3. Внедрение технологий семантического веба и графовых баз данных.

Использование онтологий и знаний о предметных областях для выявления сложных взаимосвязей. Развитие механизма связи между различными научными доменами.

2. *Интеграция и взаимодействие с международными научными экосистемами*

2.1. Интеграция с внешними источниками - реферативными базами данных, цифровыми библиотеками, репозиториями открытого доступа.

Автоматическое обновление профилей пользователей. Улучшенная система индексации публикаций и цитируемости.

2.2. Развитие репозитория открытых данных и открытого кода.

Создание специализированных платформ для обмена данными между исследователями. Совершенствование механизмов лицензирования и правового регулирования открытой науки.

2.3. Поддержка мультидисциплинарных исследований и коллабораций.

Разработка инструментов для автоматического поиска междисциплинарных пересечений. Формирование научных сетей и команд на основе анализа профилей ученых.

3. *Развитие пользовательского опыта и инструментов для исследователей.*

3.1. Персонализированные рекомендации для ученых.

Умная подборка статей, конференций и возможных коллабораций.

Индивидуальные предложения по грантам и конкурсам.

3.2. Развитие интерактивных инструментов визуализации данных.

Создание динамических карт знаний и графов взаимосвязей между исследованиями.

Улучшение механизмов интерактивного анализа тенденций в научных областях.

3.3. Поддержка автоматизированного рецензирования статей.

Инструменты для предварительного анализа качества и новизны научных работ, обнаружения потенциальных этических и методологических нарушений.

4. *Автоматизация и интеллектуальная поддержка научного процесса.*

4.1. Генерация автоматизированных научных обзоров.

Автоматический анализ и обобщение публикаций по заданной тематике.

Составление отчетов и аналитических справок по ключевым темам.

4.2. Развитие цифровых научных ассистентов.

Виртуальные помощники, которые смогут помогать в проведении исследований, анализе данных и подготовке научных работ.

Автоматизация управления библиографией и подготовка публикаций к подаче в журналы.

## 5. Улучшение прозрачности, безопасности и открытости научных данных.

5.1. Развитие блокчейн-технологий для проверки подлинности научных данных.

Разработка децентрализованных механизмов защиты от плагиата и фальсификации данных. Автоматический аудит происхождения информации.

5.2. Совершенствование механизмов борьбы с дезинформацией в научных исследованиях. Выявление фальшивых и слабо научных публикаций.

Автоматическая проверка источников и их надежности.

Перспективы развития системы направлены на ее интеллектуализацию, интеграцию с научными информационными системами, развитие персонализированных рекомендаций и автоматизацию исследовательского процесса. Внедрение новых технологий (таких как ИИ, NLP, блокчейн) позволит повысить качество научного взаимодействия, а также ускорить процессы поиска, анализа и прогнозирования научных тенденций.

## IX. ЗАКЛЮЧЕНИЕ

В научном сообществе происходит переосмысление роли институциональных репозиториях в современном научном процессе. Вместо пассивного архива репозитории должны становиться активными инструментами анализа и распространения знаний, способствующими взаимодействию исследователей. Это требует внедрения современных технологий обработки данных, машинного обучения и интеллектуального поиска.

Ключевыми аспектами активного использования институциональных репозиториях являются:

*Разнообразие целевой аудитории.* Исследователи разных уровней (от молодых ученых до ведущих экспертов) используют репозитории с различными целями, такими как поиск актуальных исследований, подготовка собственных публикаций, выявление научных трендов.

*Интеграция с научными процессами.* Современные репозитории выходят за рамки традиционных архивов и становятся инструментами интеллектуального анализа, позволяя исследователям выявлять закономерности и скрытые знания.

*Эффективность анализа на основе концептуального шаблона.* Использование данного подхода в репозиториях позволяет не только структурировать научные знания, но и выявлять новые направления исследований, повышая их практическую значимость.

*Развитие интеллектуальных информационных сервисов.* Информационно-аналитические системы, интегрированные с репозиторием, обеспечивают автоматизированную обработку и визуализацию научных данных, облегчая исследователям доступ к релевантным публикациям.

Дальнейшие исследования могут быть направлены на разработку новых моделей организации репозиториях, их интеграцию с глобальными научными экосистемами и анализ их влияния на эффективность научных коммуникаций.

## БИБЛИОГРАФИЯ

- [1] Тягунов А. М. Информационные потребности как отношения в информационном поле //Образовательные ресурсы и технологии. – 2021. – №. 2 (35). – С. 50-56.
- [2] Suber P. Open access. – The MIT Press, 2012. – С. 256.
- [3] Скларова А. М. Этические нормы и принципы научной деятельности //Библиосфера. – 2006. – №. 3. – С. 26-29.
- [4] Леонтович О. А. Этика научных исследований //Известия Волгоградского государственного педагогического университета. – 2011. – Т. 62. – №. 8. – С. 99-102.
- [5] Филозова И. А., Заикина Т. Н. О роли и месте институционального репозитория в цифровой экосистеме научной организации //International Journal of Open Information Technologies. – 2024. – Т. 12. – №. 10. – С. 86-101.
- [6] Балуткина Н. А., Стукалова А. А. Институциональные репозитории в России и за рубежом: обзор публикаций //Библиотекосведение. – 2022. – Т. 71. – №. 2. – С. 193-206.
- [7] Лебедев С. А. Научная деятельность. Основные понятия. – М.:Прспект, 2021. - 136 с. - ISBN: 978-5-392-33672-2.
- [8] Иванченко О. С. Молодые ученые в России и проблемы их профессионализации в научно-исследовательском дискурсе // Гуманитарий Юга России. 2020. Том 9. № 6. С. 99-110. DOI: <https://doi.org/10.18522/2227-8656.2020.6.7>.
- [9] Бесова М. И., Курилов С. Н., Родин А. Б. Роль трансдисциплинарности в контексте подготовки специалистов-исследователей // Медицина. Социология. Философия. Прикладные исследования. – 2022. – №. 4. – С. 73-77.
- [10] Научные коммуникации: ученый в современном обществе / И. И. Замощанский, А. М. Конашкова, И. В. Красавин [и др.] // Известия Уральского федерального университета. Сер. 3, Общественные науки. — 2016. — № 1 (149). — С. 30-41
- [11] Хеххаузен Х. Мотивация и деятельность. СПб.: Питер; М.: Смысл, 2003. 860 с..
- [12] Сидорова И. М. и др. Метафизика научных коммуникаций: ценности и идеалы //Альманах современной науки и образования. – 2017. – №. 4-5.
- [13] Куулар М. Ч. Функции монографии как документа в системе научных коммуникаций // Библиосфера. 2022. № 4. С. 37–45. <https://doi.org/10.20913/1815-3186-2022-4-37-45>.
- [14] Калинин А. Использование патентной информации в инновационной деятельности //Наука и инновации. – 2020. – №. 4 (206). – С. 24-27.
- [15] Фирстов В. Г. Роль и функции научных семинаров в развитии инновационной деятельности университетов //Инновации и инвестиции. – 2019. – №. 4. – С. 26-29 ..
- [16] Nair P. K. R. et al. Organization of a research paper: The IMRAD format //Scientific writing and communication in agriculture and natural resources. – 2014. – С. 13-25.
- [17] Тучкова Н. П., Атаева О. М. Подходы к извлечению знаний в научных предметных областях //Информационные и математические технологии в науке и управлении. – 2020. – №. 2 (18). – С. 5-18.
- [18] Бобамурадов О. Д., Рахимов Н. О. Этапы извлечения знаний из электронных информационных ресурсов //Евразийский Союз Ученых. – 2015. – №. 10-2 (19). – С. 130-133
- [19] Starukhin, Y., & Diukarev, V. (2024). AUTOMATION OF TEXT DATA PROCESSING USING NLP. *The American Journal of Engineering and Technology*. <https://doi.org/10.37547/tajet/volume06issue07-04>.
- [20] Muralidharan, B., Beadles, H., Marzban, R., & Mupparaju, K. (2024). Knowledge AI: Fine-tuning NLP Models for Facilitating Scientific Knowledge Extraction and Understanding. *ArXiv*, abs/2408.04651. <https://doi.org/10.48550/arXiv.2408.04651>.
- [21] Vigenesh, M., R., Patil, V., Kumar, S., Geyi, N., & Chattopadhyay, S. (2023). Assessing the Ability of AI-Driven Natural Language Processing to Accurately Analyze Unstructured Text Data. *2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, 1-5. <https://doi.org/10.1109/SMARTGENCON60755.2023.10442594>.
- [22] Zala, K., Acharya, B., Mashru, M., Palaniappan, D., Gerogiannis, V., Kanavos, A., & Karamitsos, I. (2024). Transformative Automation: AI in Scientific Literature Reviews. *International Journal of Advanced Computer Science and Applications*. <https://doi.org/10.14569/ijacsa.2024.01501122>.
- [23] Liu, Y., Cao, M., Blodgett, S., Cheung, J., Olteanu, A., & Trischler, A. (2023). Responsible AI Considerations in Text Summarization Research: A Review of Current Practices. *ArXiv*, abs/2311.11103. <https://doi.org/10.48550/arXiv.2311.11103>.

- [24] Добрынин В. Н., Филозова И. А. Поиск на основе логико-семантической сети «вопрос-ответ-реакция» в научных информационных фондах //Системный анализ в науке и образовании. – 2011. – №. 1. – С. 18-30. Режим доступа: URL: <https://sanse.uni-dubna.ru/index.php/sanse/article/view/408>
- [25] Добрынин В. Н., Филозова И. А. Семантический поиск в научных электронных библиотеках //Информатизация образования и науки. – 2014. – №. 2. – С. 111-127. Режим доступа: URL: [https://www.elibrary.ru/download/elibrary\\_21404653\\_18672006.pdf](https://www.elibrary.ru/download/elibrary_21404653_18672006.pdf)
- [26] Добрынин В. Н., Филозова И. А. Структурно-функциональное обеспечение электронного семантического сборника образовательных ресурсов (Structural and Fuctional Support of the Semantic Digital Book) //RCDL. – 2012. – С. 252-257. Режим доступа: URL: <http://ceur-ws.org/Vol-934/paper39.pdf>
- [27] Добрынин В. Н., Филозова И. А. Технология формирования каталога информационного фонда //Компьютерные исследования и моделирование. – 2015. – Т. 7. – №. 3. – С. 661-673. DOI: 10.20537/2076-7633-2015-7-3-661-673. Режим доступа: URL: <https://www.mathnet.ru/links/b74d04c2b464b5cdc7a089dc41893720/crm233.pdf>
- [28] Добрынин В. Н., Филозова И. А. Создание, поддержка и развитие модели интерпретации смыслов //CEUR Workshop Proceedings. – 2016. – С. 189-196. Режим доступа: URL: <http://ceur-ws.org/Vol-1787/189-196-paper-32.pdf>

# Institutional repository of publications as an active component in the modern scientific research

Vladimir Dobrynin, Vladimir Korenkov, Irina Filozova, Evgenia Cheremisina

**Abstract** — The article is devoted to the study of institutional repositories as active components in the life of a scientific organization. The need to adapt repositories to the challenges of digital transformation (an increase of data volumes, the diversity of their formats and sources, the interdisciplinary approach). A portrait of a modern researcher is described, considering age, interests and motivation. The difficulties of extracting knowledges from of the publications are discussed. A methodology for extracting knowledges from scientific texts is presented. The technology for analyzing scientific papers based on a conceptual template is described. This approach allows to identify keynote aspects of the work with their quantitative and/or qualitative indicators. The conceptual template *Problems formulated by the author* is illustrated. The architecture of an information and analytical system based on the described technology is presented. The problems that it allows to solve and the services for the target audience are considered. Prospects for the development of the system are discussed. The need for active use of repositories to develop effective interaction between researchers is emphasized.

**Keywords** — analysis of scientific texts, repositories of scientific publications, researcher portrait, conceptual template, analytical service.

Filozova Irina A., Head of Group, Laboratory of Information Technologies, Joint Institute for Nuclear Research (6 Joliot-Curie St Dubna Moscow Region Russia 141980); Senior Lecturer, Dubna State University, Institute of System Analysis and Control (19 Universitetskaya St Dubna Moscow region Russia 141982), ORCID: <https://orcid.org/0000-0003-3441-7093>, [fia@jinr.ru](mailto:fia@jinr.ru)

Korenkov Vladimir V., Scientific Director, Laboratory of Information Technologies, Joint Institute for Nuclear Research (141980, Dubna, Moscow region, Joliot-Curie str., 6), Doctor of Technical Sciences; Head of the Department of Distributed Information and Computing Systems, Dubna State University, Institute of Systems Analysis and Control (141980, Russia, Moscow region, Dubna, Universitetskaya str., 19), ORCID: <https://orcid.org/0000-0002-2342-7862>, [korenkov@jinr.ru](mailto:korenkov@jinr.ru).

Dobrynin Vladimir N., Professor, Dubna State University, Institute of Systems Analysis and Control (141980, Russia, Moscow region, Dubna, Universitetskaya str., 19),

Candidate of Technical Sciences, Senior Researcher, [i@vdbrynin.ru](mailto:i@vdbrynin.ru).

Cheremisina Evgenia N., Head of the department of the All-Russian Scientific Research Geological Oil Institute (117105 Moscow, Varshavskoe shosse, 8), Doctor of Technical Sciences, Academician of the Russian Academy of Natural Sciences, Professor, ORCID: <https://orcid.org/0000-0002-6041-8359>, [head@geosys.ru](mailto:head@geosys.ru).

## REFERENCES

- [1] Tyagunov A. M. Information needs as relationships in the information field // Educational resources and technologies. - 2021. - No. 2 (35).
- [2] Suber P. Open access. – The MIT Press, 2012. – C. 256.
- [3] Sklyarova A. M. Ethical norms and principles of scientific activity // Bibliosphere. - 2006. - No. 3. - P. 26-29.
- [4] Leontovich O. A. Ethics of scientific research // News of the Volgograd State Pedagogical University. - 2011. - V. 62. - No. 8. - P. 99-102.
- [5] Irina Filozova, Tatiana Zaikina On the role and place of the institutional repository in the digital ecosystem of a scientific organization //International Journal of Open Information Technologies. – 2024. – Vol. 12. – No. 10. – P. 86-101.
- [6] Balutkina N.A., Stukalova A.A. Institutional Repositories in Russia and Abroad: Review of Publications. Bibliotekovedenie [Russian Journal of Library Science]. 2022;71(2):193-206. (In Russ.) <https://doi.org/10.25281/0869-608X-2022-71-2-193-206>
- [7] Lebedev S. A. Scientific activity. Basic concepts. - M.: Prospect, 2021. - 136 p. - ISBN: 978-5-392-33672-2.
- [8] Ivanchenko O.S. Young Scientists in Russia and Problems of Their Professionalization in the Research Discourse. Humanities of the South of Russia. 2020. Vol. 9. No. 6. P. 99-110. DOI: <https://doi.org/10.18522/2227-8656.2020.6.7> (in Russ.).
- [9] Besova M. I., Kurilov S. N., Rodin A. B. The role of transdisciplinarity in the context of training research specialists // Medicine. Sociology. Philosophy. Applied research. - 2022. - No. 4. - P. 73-77.
- [10] [Scientific communications: scientist in modern society / I. I. Zamoshchansky, A. M. Konashkova, I. V. Krasavin [et al.] // Bulletin of the Ural Federal University. Ser. 3, Social Sciences. - 2016. - No. 1 (149). - P. 30-41
- [11] Heckhausen H. Motivation and activity. SPb.: Piter; M.: Smysl, 2003. 860 p.
- [12] Sidorova I. M. et al. Metaphysics of scientific communications: values and ideals // Almanac of modern science and education. - 2017. - No. 4-5.
- [13] Kuular M.C. Functions of a Monograph as a Document in Scholarly Communications. Bibliosphere. 2022;(4):37-45. (In Russ.) <https://doi.org/10.20913/1815-3186-2022-4-37-45>
- [14] Kalinin A. Use of patent information in innovation activities // Science and Innovation. - 2020. - No. 4 (206). - P. 24-27.
- [15] Firstov V. G. The role and functions of scientific seminars in the development of innovative activities of universities // Innovations and Investments. - 2019. - No. 4. - P. 26-29.
- [16] Nair P. K. R. et al. Organization of a research paper: The IMRAD format //Scientific writing and communication in agriculture and natural resources. – 2014. – C. 13-25.

- [17] Tuchkova N. P., Ataeva O. M. Approaches to knowledge extraction in scientific subject areas // *Information and Mathematical Technologies in Science and Management*. - 2020. - No. 2 (18). - P. 5-18.
- [18] Bobamuradov O. D., Rakhimov N. O. Stages of knowledge extraction from electronic information resources // *Eurasian Union of Scientists*. - 2015. - No. 10-2 (19). - P. 130-133.
- [19] Starukhin, Y., & Diukarev, V. (2024). AUTOMATION OF TEXT DATA PROCESSING USING NLP. *The American Journal of Engineering and Technology*. <https://doi.org/10.37547/tajet/volume06issue07-04>.
- [20] Muralidharan, B., Beadles, H., Marzban, R., & Mupparaju, K. (2024). Knowledge AI: Fine-tuning NLP Models for Facilitating Scientific Knowledge Extraction and Understanding. *ArXiv*, abs/2408.04651. <https://doi.org/10.48550/arXiv.2408.04651>.
- [21] Vigenesh, M., R., Patil, V., Kumar, S., Geyi, N., & Chattopadhyay, S. (2023). Assessing the Ability of AI-Driven Natural Language Processing to Accurately Analyze Unstructured Text Data. *2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, 1-5. <https://doi.org/10.1109/SMARTGENCON60755.2023.10442594>.
- [22] Zala, K., Acharya, B., Mashru, M., Palaniappan, D., Gerogiannis, V., Kanavos, A., & Karamitsos, I. (2024). Transformative Automation: AI in Scientific Literature Reviews. *International Journal of Advanced Computer Science and Applications*. <https://doi.org/10.14569/ijacsa.2024.01501122>.
- [23] Liu, Y., Cao, M., Blodgett, S., Cheung, J., Olteanu, A., & Trischler, A. (2023). Responsible AI Considerations in Text Summarization Research: A Review of Current Practices. *ArXiv*, abs/2311.11103. <https://doi.org/10.48550/arXiv.2311.11103>.
- [24] Dobrynin V. N., Filozova I. A. Search based on the logical-semantic network "question-answer-reaction" in scientific information collections // *Systems analysis in science and education*. - 2011. - No. 1. - P. 18-30. Access mode: URL: <https://sanse.unidubna.ru/index.php/sanse/article/view/408>
- [25] Dobrynin V. N., Filozova I. A. Semantic search in scientific electronic libraries // *Informatization of education and science*. - 2014. - No. 2. - P. 111-127. Access mode: URL: [https://www.elibrary.ru/download/elibrary\\_21404653\\_18672006.pdf](https://www.elibrary.ru/download/elibrary_21404653_18672006.pdf)
- [26] Dobrynin V. N., Filozova I. A. Structural and Fuctional Support of the Semantic Digital Book // *RCDL*. - 2012. - P. 252-257. Access mode: URL: <http://ceur-ws.org/Vol-934/paper39.pdf>
- [27] Dobrynin V. N., Filozova I. A. Technology of formation of the catalog of the information fund // *Computer research and modeling*. - 2015. - Vol. 7. - No. 3. - P. 661-673. DOI: 10.20537/2076-7633-2015-7-3-661-673. Access mode: URL: <https://www.mathnet.ru/links/b74d04c2b464b5cdc7a089dc41893720/crm233.pdf>
- [28] Dobrynin V. N., Filozova I. A. Creation, support and development of a model of interpretation of meanings // *CEUR Workshop Proceedings*. - 2016. - P. 189-196. Access mode: URL: <http://ceur-ws.org/Vol-1787/189-196-paper-32.pdf>