

Министерство науки и высшего образования Российской Федерации  
 федеральное государственное автономное  
 образовательное учреждение высшего образования  
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Инженерная школа информационных технологий и робототехники  
 Направление подготовки 09.04.02 «Информационные системы и технологии»  
 Отделение информационных технологий

### МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Алгоритмическое и программное обеспечение анализа открытых данных баз цитирований научных публикаций

УДК 004.415.2:004.65:001.891

Студенты:

Группа	ФИО	Подпись	Дата
8ИМ92	Непомнящих Дарья Сергеевна		
8ИМ92	Сулягина Анна Андреевна		

Руководитель ВКР:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Савельев Алексей Олегович	к.т.н.		

### КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Верховская Марина Витальевна	к.э.н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ООД ШБИП	Сечин Андрей Александрович	к.т.н.		

### ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
09.04.02 «Информационные системы и технологии», профиль «Геоинформационные системы»	Шерстнев Владислав Станиславович	к. т. н., доцент		

Министерство науки и высшего образования Российской Федерации  
 федеральное государственное автономное  
 образовательное учреждение высшего образования  
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Инженерная школа информационных технологий и робототехники  
 Направление подготовки 09.04.02 «Информационные системы и технологии»  
 Отделение информационных технологий

УТВЕРЖДАЮ:  
 Руководитель ООП

\_\_\_\_\_

(Подпись) (Дата) (Ф.И.О.)

### ЗАДАНИЕ на выполнение выпускной квалификационной работы

В форме:

магистерской диссертации (бакалаврской работы, дипломного проекта/работы, магистерской диссертации)
--

Студентам:

Группа	ФИО
8ИИМ92	Непомнящих Дарья Сергеевна
8ИИМ92	Сутягина Анна Андреевна

Тема работы:

Алгоритмическое и программное обеспечение анализа открытых данных баз цитирований научных публикаций	
Утверждена приказом директора (дата, номер)	15.02.2021, №46-5/с

Срок сдачи студентом выполненной работы:

--	--

### ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<b>Исходные данные к работе</b> <i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i>	1. Интеграция с программным интерфейсом приложений библиографической базы данных Scopus.
--	--

<p><b>Перечень подлежащих исследованию, проектированию и разработке вопросов</b></p> <p><i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p>	<ol style="list-style-type: none"> <li>1. Анализ научной и технической документации, описывающей предметную область, а также современных систем оценки научной деятельности.</li> <li>2. Проектирование системы анализа открытых данных баз цитирований научных публикаций.</li> <li>3. Разработка программного обеспечения системы анализа открытых данных баз цитирований научных публикаций.</li> <li>4. Апробация функций системы анализа открытых данных баз цитирований научных публикаций.</li> <li>5. Оценка конкурентоспособности разработки, расчет затрат на проведение исследования.</li> <li>6. Анализ условий труда исполнителей проекта.</li> </ol>
<p><b>Перечень графического материала</b></p> <p><i>(с точным указанием обязательных чертежей)</i></p>	<p>Концептуальные схемы разрабатываемого программного обеспечения.</p>

**Консультанты по разделам выпускной квалификационной работы**

*(с указанием разделов)*

Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Доцент ОСГН ШБИП, к.э.н., Верховская Марина Витальевна
Социальная ответственность	Доцент ООД ШБИП, к.т.н., Сечин Андрей Александрович
Раздел на иностранном языке	Доцент ОИЯ ШБИП, к.ф.н., Коротченко Татьяна Валериевна

**Названия разделов, которые должны быть написаны на русском и иностранном языках:**

Аналитический обзор научной и технической документации предметной области

<b>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</b>	15.01.2021
---	------------

**Задание выдал руководитель:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Савельев Алексей Олегович	к.т.н.		

**Задание приняли к исполнению студенты:**

Группа	ФИО	Подпись	Дата
8ИМ92	Непомнящих Дарья Сергеевна		
8ИМ92	Сутягина Анна Андреевна		

Министерство науки и высшего образования Российской Федерации  
 федеральное государственное автономное  
 образовательное учреждение высшего образования  
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Инженерная школа информационных технологий и робототехники  
Направление подготовки 09.04.02 «Информационные системы и технологии»

Уровень образования: магистр

Отделение информационных технологий

Период выполнения: весенний семестр 2020 /2021 учебного года

Форма представления работы:

магистерская диссертация
--------------------------

(бакалаврская работа, дипломный проект/работа, магистерская диссертация)

### КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	
--	--

Дата контроля	Название раздела (модуля) /вид работы (исследования)	Максимальный балл
12.04.2021	Раздел 1. Аналитический обзор научной и технической документации предметной области	20
26.04.2021	Раздел 2. Проектирование системы анализа открытых данных баз цитирований научных публикаций	20
10.05.2021	Раздел 3. Разработка программного обеспечения анализа открытых данных баз цитирований научных публикаций	20
24.05.2021	Раздел 4. Апробация системы анализа открытых данных баз цитирований научных публикаций	20
31.05.2021	Раздел 5. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	10
01.06.2021	Раздел 6. Социальная ответственность	10

**СОСТАВИЛ:**

**Руководитель ВКР**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Савельев Алексей Олегович	к.т.н.		

**СОГЛАСОВАНО:**

**Руководитель ООП**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Шерстнев Владислав Станиславович	к.т.н.		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА  
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ,  
РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»**

Студентам:

<b>Группа</b>	<b>ФИО</b>
8ИМ92	Непомнящих Дарья Сергеевна
8ИМ92	Сутягина Анна Андреевна

<b>Школа</b>	Инженерная школа информационных технологий и робототехники	<b>Отделение (НОЦ)</b>	Отделение информационных технологий
<b>Уровень образования</b>	Магистратура	<b>Направление/специальность</b>	09.04.02 «Информационные системы и технологии»

**Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:**

1. Стоимость ресурсов научного исследования (НТИ): материально-технических, энергетических, финансовых, информационных и человеческих	Затраты на материалы – 1362 руб.; Затраты на амортизацию – 44000 руб.; Затраты на ЗП – 419202,55 руб.; Затраты на отчисления во внебюджетные фонды – 113603,92 руб.; Накладные расходы – 92506,95 руб.
2. Нормы и нормативы расходования ресурсов	Районный коэффициент – 30%; Коэффициент дополнительной заработной платы – 12%; Накладные расходы – 16%.
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	Коэффициент отчислений на уплату во внебюджетные фонды – 27,1%.

**Перечень вопросов, подлежащих исследованию, проектированию и разработке:**

1. Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения	Анализ конкурентных технических решений; SWOT-анализ. FAST-анализ Диаграмма Исикава
2. Планирование и формирование бюджета научных исследований	Формирование плана и графика разработки: – определение структуры и трудоемкости работ; – создание диаграммы Ганнта. Формирование бюджета затрат на разработку: – материальные затраты; – затраты на амортизацию оборудования; – заработная плата (основная и дополнительная); – отчисления во внебюджетные фонды; – накладные расходы.
3. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности разработки	Определение потенциального эффекта разработки.

**Перечень графического материала:**

- Оценочная карта конкурентных технических решений;
- Матрица SWOT-анализа и интерактивная матрица проекта;
- Функционально-стоимостная диаграмма;

- Диаграмма Исикава;
- График проведения НТИ;
- Диаграмма Гантта;
- Бюджет НТИ;
- Оценка ресурсной, финансовой и экономической эффективности НТИ.

Дата выдачи задания для раздела по линейному графику:	01.03.2021
---	------------

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Верховская Марина Витальевна	к.э.н.		

Задание приняли к исполнению студенты:

Группа	ФИО	Подпись	Дата
8ИМ92	Непомнящих Дарья Сергеевна		
8ИМ92	Сулягина Анна Андреевна		

## ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студентам:

Группа	ФИО
8ИИМ92	Непомнящих Дарья Сергеевна
8ИИМ92	Сутягина Анна Андреевна

Школа	Инженерная школа информационных технологий и робототехники	Отделение (НОЦ)	Отделение информационных технологий
Уровень образования	Магистратура	Направление/специальность	09.04.02 «Информационные системы и технологии»

Тема ВКР:

Алгоритмическое и программное обеспечение анализа открытых данных баз цитирований научных публикаций	
<b>Исходные данные к разделу «Социальная ответственность»:</b>	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Объект исследования: информационная веб-система. Область применения: органы управления научной деятельностью организаций.
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
<b>1. Правовые и организационные вопросы обеспечения безопасности:</b> <ul style="list-style-type: none"> <li>– специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства;</li> <li>– организационные мероприятия при компоновке рабочей зоны.</li> </ul>	<ul style="list-style-type: none"> <li>– Трудовой кодекс Российской Федерации от 30.12.2001 N 197ФЗ</li> <li>– Федеральный закон "Об обязательном социальном страховании от несчастных случаев на производстве и профессиональных заболеваний" от 24.07.1998 N 125-ФЗ</li> <li>– Федеральный закон "О техническом регулировании" от 27.12.2002 N 184-ФЗ</li> <li>– ГОСТ 12.2.032–78 «ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования»</li> <li>– ГОСТ 21889–76 «Кресло человека-оператора. Общие эргономические требования»</li> <li>– СанПиН 2.2.2/2.4.1340-03. Гигиенические требования к персональным электронно-вычислительным машинам и организация работы</li> </ul>
<b>2. Производственная безопасность:</b> 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	Вредные факторы: <ul style="list-style-type: none"> <li>– Отклонение показателей микроклимата</li> <li>– Недостаточная освещенность рабочей</li> </ul>

	зоны – Повышенный уровень электромагнитных излучений – Повышенный уровень шума на рабочем месте
<b>3. Экологическая безопасность:</b>	Литосфера, атмосфера, гидросфера: – Отсутствие или неправильная утилизация люминесцентных ламп – Отсутствие или неправильная утилизация бытового мусора (канцелярские, пищевые отходы)
<b>4. Безопасность в чрезвычайных ситуациях:</b>	Возможные ЧС: Внезапное обрушение здания, аварии на коммунальных системах жизнеобеспечения населения, пожар, угроза пандемии. Наиболее типичная ЧС: Возникновение пожара

<b>Дата выдачи задания для раздела по линейному графику</b>	
---	--

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ООД ШБИП	Сечин Андрей Александрович	к.т.н.		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ИМ92	Непомнящих Дарья Сергеевна		
8ИМ92	Сутягина Анна Андреевна		

## Планируемые результаты обучения

Код	Результаты обучения
<b>09.04.02 «Информационные системы и технологии» (Геоинформационные технологии)</b>	
P1	Воспринимать и самостоятельно приобретать, развивать и применять математические, естественно-научные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте.
P2	Владеть и применять методы и средства получения, хранения, переработки и представления информации посредством современных компьютерных технологий, в том числе в глобальных компьютерных сетях.
P3	Демонстрировать способность выстраивать логику рассуждений и высказываний, основанных на интерпретации данных, интегрированных из разных областей науки и техники, выносить суждения на основании неполных данных, анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями.
P4	Анализировать и оценивать уровни своих компетенций в сочетании со способностью и готовностью к саморегулированию дальнейшего образования и профессиональной мобильности. Демонстрировать способность к самостоятельному обучению новым методам исследования, способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения.
P5	Владеть современными коммуникативными технологиями, в том числе на иностранном языке для академического и профессионального взаимодействия.
P6	Использовать на практике умения и навыки в организации исследовательских, проектных работ и профессиональной эксплуатации современных программных и информационных систем, в управлении коллективом. Способность организовывать и эффективно руководить работой команды проекта при разработке программных и информационных систем.
P7	Разрабатывать стратегии проектирования, критерии эффективности и ограничения применимости новых методов и средств проектирования и разработки программных систем.
P8	Планировать и проводить теоретические и экспериментальные (численные) исследования в области создания программных систем. Оценивать и выбирать вариант архитектуры информационной системы.
P9	Владеть методами и средствами обработки и анализа пространственных данных, методами и средствами разработки хранилищ данных и программного обеспечения геоинформационных систем и технологий.
P10	Владеть современными инструментальными CASE-средствами программирования и технологиями управления и обработки данными. Использовать их при разработке требований, проектировании и создании программного обеспечения геоинформационных систем и технологий.
P11	Осуществлять проектирование и разработку геоинформационных систем и технологий в среде корпоративных и глобальных информационно-телекоммуникационных систем.
P12	Осуществлять управление процессами внедрения/сопровождения (модернизации, интеграции) информационных, геоинформационных систем и технологий.

## Реферат

Выпускная квалификационная работа 139 с., 56 рис., 22 табл., 29 источников, 3 приложения.

Ключевые слова: анализ, публикационная активность, наукометрические показатели, база цитирований, система поддержки принятия решений.

Объектом исследования данной работы является анализ публикационной активности научной организации.

Цель работы – разработка алгоритмического и программного обеспечения анализа открытых данных баз цитирований научных публикаций. Данная цель обусловлена необходимостью иметь структурированную информационную базу для поддержки процессов принятия управленческих решений в сфере развития научной деятельности.

В процессе исследования был проведен анализ возможностей использования открытых данных баз цитирований для поддержки процессов принятия решений в сфере управления научной деятельностью. Были разработаны модуль интеграции с API Scopus и модуль формирования и визуализации статистических данных по научной организации, научному сообществу, отдельному автору.

В результате исследования были формализованы требования к алгоритмическому и программному обеспечению.

Область применения: управление научной деятельностью организаций.

Экономическая эффективность/значимость работы: разработка имеет экономическую значимость, так как направлена на повышение эффективности научной деятельности организаций наукоемких отраслей, что позволит повысить конкурентоспособность организации.

В дальнейшем планируется расширение системы дополнительными модулями для интеграции с другими популярными зарубежными и

отечественными базами цитирований, такими как Web Of Science, Google Scholar и РИНЦ.

## Оглавление

Реферат .....	10
Обозначения и сокращения .....	16
Введение .....	17
1. Аналитический обзор научной и технической документации предметной области .....	18
1.1 Анализ предметной области.....	18
1.2 Основные наукометрические показатели публикационной активности	19
1.3 Современные системы оценки научной деятельности .....	23
1.4 Спецификации требований к разрабатываемому решению ....	29
1.5 Выводы по аналитическому обзору научной и технической документации предметной области.....	32
2. Проектирование системы анализа открытых данных баз цитирований научных публикаций .....	33
2.1 Концептуальное проектирование системы оценки научной деятельности.....	33
2.2 Алгоритмическое обеспечение анализа данных базы данных Scopus	35
2.3 Функциональное проектирование системы оценки научной деятельности.....	36
3. Разработка программного обеспечения анализа открытых данных баз цитирований научных публикаций.....	37
3.1 Краткое описание используемых технологий.....	37
3.1.1 Фреймворк Laravel.....	37
3.1.2 Язык программирования Python 3 .....	37
3.1.3 Фреймворк Dash .....	39
3.1.4 СУБД PostgreSQL .....	40

3.2	Импорт исходных данных .....	40
3.3	Описание формирования статистики .....	43
3.4	Описание реализованной визуализации публикационной активности	45
3.5	Описание реализованных возможностей системы оценки научной деятельности .....	46
3.6	Реализация функции интеграции с инфраструктурой научного сообщества	57
3.7	Реализация функции построения графа популярных научных областей	59
3.8	Реализация функции построения карты журналов с публикационной активностью .....	61
4.	Апробация системы анализа открытых данных баз цитирований научных публикаций .....	63
4.1.	Апробация функций анализа научной активности организации	63
4.2.	Апробация функций анализа научной активности журнала ...	69
4.3.	Апробация функций анализа научной активности научного сообщества	70
4.4.	Апробация функций анализа научной активности автора .....	72
4.5.	Заключение по разделу .....	77
5.	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение .....	79
5.1.	Предпроектный анализ .....	79
5.1.1.	Потенциальные потребители результатов исследования ...	79
5.1.2.	Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения .....	80
5.1.3.	FAST-анализ .....	82
5.1.4.	Диаграмма Исикава .....	85

5.1.5. SWOT-анализ .....	86
5.2. Планирование научно-исследовательских работ .....	88
5.2.1. Разработка графика проведения научного исследования... ..	88
5.3. Бюджет научно-технической разработки .....	93
5.3.1. Расчет затрат на материалы .....	93
5.3.2. Расчет затрат на специальное оборудование .....	94
5.3.3. Расчет заработной платы .....	94
5.3.4. Дополнительная заработная плата исполнителей темы .....	96
5.3.5. Отчисления во внебюджетные фонды .....	96
5.3.6. Накладные расходы .....	97
5.3.7. Бюджет затрат .....	98
5.4. Определение ресурсной, финансовой, бюджетной, социальной и экономической эффективности исследования .....	98
5.4.1. Определение финансовой эффективности исследования ..	99
5.4.2. Определение показателя ресурсоэффективности .....	99
5.4.3. Интегральный показатель эффективности .....	100
5.5. Заключение по разделу .....	101
6. Социальная ответственность .....	103
6.1. Правовые и организационные вопросы обеспечения безопасности .....	103
6.2. Производственная безопасность .....	105
6.2.1. Анализ вредных и опасных факторов, которые могут возникнуть при эксплуатации объекта исследования .....	105
6.2.2. Отклонение показателей микроклимата рабочего помещения .....	106
6.2.3. Недостаточная освещенность рабочего помещения .....	107
6.2.4. Повышенный уровень электромагнитных излучений .....	110

6.2.5. Повышенный уровень шума .....	111
6.3. Экологическая безопасность.....	113
6.4. Безопасность в чрезвычайных ситуациях .....	114
6.5. Выводы по разделу .....	116
Список используемых источников.....	118
Приложение А .....	122
Приложение Б.....	135
Приложение В .....	139

## **Обозначения и сокращения**

В работе использованы следующие обозначения и сокращения:

- API – Application Programming Interface
- БД – База данных
- СУБД – Система управления базами данных
- SQL – Structured Query Language
- JSON – JavaScript Object Notation
- РИНЦ – Российский Индекс Научного Цитирования
- SJR – SCImago Journal Ranking
- SNIP – Source Normalized Impact Per Paper
- JCR – Journal Citation Reports
- СПбГУ – Санкт-Петербургский государственный университет

## Введение

Различные научные исследования научной организации или совокупности научных организаций имеют разный отклик в научной среде, какие-то исследования могут цитироваться на протяжении еще долгого времени исследователями, а какие-то вызывают сравнительно меньший интерес. Для того чтобы научной организации осуществлять свою научную деятельность наиболее эффективно, необходимо иметь возможность проанализировать свой путь в науке, выявить наиболее удачные партнерства и наиболее интересные среди научных организаций темы исследований.

На сегодняшний день вопросам повышения эффективности научной деятельности уделяется большое значение, разрабатываются критерии оценки и показатели, которые демонстрируют, как работают отдельные ученые, группы ученых, институты и ВУЗы. Используя библиографические базы данных, можно получить данные о различных журналах, публикациях и авторах. Это могут быть как количественные показатели, например, количество цитирований публикации или индекс Хирша автора, так и качественные, например, ключевые слова публикации или области интереса автора. Основываясь на данных библиографических баз, можно построить такие аналитические инструменты, которые помогут принимать управленческие решения в области развития научной деятельности наиболее быстро и результативно. Подобные инструменты должны систематизировать данные со всех библиографических баз так, чтобы конечный пользователь для вынесения экспертной оценки мог полагаться на результаты работы интеллектуальных аналитических инструментов и карт науки.

Это дает возможность применения метода принятия управленческих решений на основе анализа больших данных для поиска наиболее эффективного пути развития научной деятельности. Такой подход обеспечивает обоснованность и объективность принятых решений и, как правило, качественно улучшает управленческую деятельность.

Целью работы является разработка алгоритмического и программного обеспечения анализа открытых баз цитирований научных публикаций.

Поставленные задачи:

1. Выполнить аналитический обзор научной и технической документации.
2. Составить спецификацию требований к разрабатываемому решению.
3. Спроектировать систему анализа открытых баз цитирований научных публикаций.
4. Разработать систему анализа открытых баз цитирований научных публикаций.
5. Произвести анализ научной деятельности организации на примере Томского политехнического университета.

## **1. Аналитический обзор научной и технической документации предметной области**

### **1.1 Анализ предметной области**

Управление научной деятельностью включает в себя организацию, планирование, управление и координацию человеческих и материальных ресурсов для достижения определенных результатов в исследованиях с точки зрения состава и объема работ, затрат, времени и качества. Вопрос повышения эффективности управления научной деятельностью обоснован необходимостью определения приоритетных направлений научной деятельности, создания и освоения новых технологий, формирования и развития научных организаций и научных коллективов.

Современные публикации часто пишутся авторами из различных стран, объединенными общими научными интересами. Важным аспектом управления научной деятельностью является поиск партнеров для совместного решения актуальных научно-технических задач, поскольку

такие задачи обычно являются междисциплинарными и требуют участия различных специалистов для решения.

Кроме задач планирования и организации научной деятельности в процессы управления также входит формирование отчетности по итогам принятия каких-либо управленческих решений. Данный этап позволит оценить эффективность принятого решения и планировать дальнейшее развитие.

Для принятия управленческих решений в области развития научной деятельности необходимо анализировать большие объемы данных о уже проведенных исследованиях и прогнозировать ситуацию в научной среде на будущее. Необходимость новых подходов к управлению научными организациями обсуждается также в Стратегии научно-технического развития Российской Федерации [1]. В данной работе представлен способ информационного обеспечения для процессов управления научной деятельностью на основе наукометрических показателей.

## **1.2 Основные наукометрические показатели публикационной активности**

На сегодняшний день высокая оценка публикационной активности и научной результативности ученого и организации, к которой он принадлежит, – пропуск в мир авторитетной науки и объемной господдержки.

Публикационная активность – это результат научно-исследовательской деятельности автора или научного коллектива или иного коллективного актора исследовательского процесса (организация, регион, страна), воплощенный в виде научной публикации, например, журнальной статьи, статьи в коллективном сборнике, доклада в трудах научной конференции, авторской или коллективной монографии, опубликованного отчёта по НИР [2].

Наукометрические показатели – это количественные индексы, используемые для оценки состояния и перспективности научно-исследовательской деятельности авторов и организаций, их сравнения и ранжирования в различных рейтингах.

Основные наукометрические показатели публикационной активности авторов:

- Количество публикаций – показатель научной производительности ученого, сообщества, организации, отражающий исследования, которые превратились в публикации в рецензируемом журнале/журналах.
- Цитируемость – это количество ссылок, которое получила публикация. Это показатель научной влиятельности или авторитетности, отражающий, насколько публикация автора повлияла на дальнейшие исследования.
- Суммарная цитируемость выражается, как следует из названия, в сумме всего количества ссылок на работы автора в публикациях других авторов.
- Средняя цитируемость на одну статью – это показатель для оценки средней влиятельности статей отдельного ученого или научного сообщества, или организации или целой страны. Рассчитывается путем деления количества цитирований на количество публикаций [2].
- Индекс Хирша (h-индекс) – это количественная характеристика продуктивности ученого, основанная на количестве его публикаций и количестве цитирований этих публикаций: ученый имеет индекс  $h$ , если он опубликовал  $h$  статей, на каждую из которых сослались как минимум  $h$  раз. Индекс Хирша был разработан, чтобы получить более адекватную оценку научной продуктивности исследователя, чем могут дать такие простые характеристики, как общее число

публикаций или общее число цитирований. Индекс Хирша рассчитывается автоматически с помощью специальных приложений в реферативных базах данных Scopus, Web of Science, РИНЦ.

Основные наукометрические показатели публикационной активности журналов, в которых печатаются результаты научных исследований:

- **Импакт-фактор журнала** – это отношение числа ссылок, которые получил журнал в текущем году на статьи, опубликованные в этом журнале в предыдущие два года, к числу статей, опубликованных в этом журнале в эти же два предшествующих года. Импакт-фактор высчитывается в базах данных Web of Science (аналитический модуль Journal Citation Reports), РИНЦ. Для расчета классического импакт-фактора используется массив публикаций за 3 года, включая 2-летнее окно цитирования. Импакт-фактор является мерой, определяющей частоту, с которой цитируется средне цитируемая статья журнала и может служить индикатором «востребованности» и «признанности» журнала в научной среде. Публикация в журнале, который имеет высокие показатели, повышает рейтинг автора статьи.
- **SJR** – рейтинг журналов, используемый Scopus, учитывающий не только общее количество цитирований, но и взвешенные показатели цитирований по годам и качественные показатели, такие как авторитетность ссылок. В целом SJR не очень сильно отличается от привычного импакт-фактора, весьма привлекая более широким спектром журналов и полностью открытым характером.
- **SNIP** – еще более продвинутый показатель, используемый Scopus. Этот показатель учитывает еще и уровень цитирований в каждой научной области, поэтому может быть использован для сравнения публикаций в разных научных направлениях. Основные

особенности расчета этого показателя заключаются в следующем. В показателе учитываются ссылки, сделанные в текущем году, на статьи, вышедшие в течение трех предыдущих лет. Публикационное окно равно 3 года, окно цитирования равно 1 год, типы документов одинаковы для всех этапов подсчета показателя. Вводится специальное определение «индивидуальной области науки» для журнала, или «окружения журнала»: все статьи, опубликованные в текущем году (в любом издании), которые хотя бы однажды цитировали выпуски журнала, вышедшие за последние десять лет. Для определения потенциала цитирования (это среднее число позиций, средняя «длина» списков цитируемой литературы в статьях «окружения») подсчитывается среднее число ссылок в статьях, составляющих «окружение журнала». Но учитываются только те ссылки, которые:

- а) ведут на статьи, вышедшие в течение трех предыдущих лет;
- б) ведут на статьи, имеющиеся в базе данных, по которой идет расчет.

- Индекс оперативности (*immediacy index*) – это показатель количества ссылок на публикации журнала, полученные в год публикации. Отражает то, насколько быстро становятся известны в научном мире статьи, опубликованные в журнале. Вычисляется как отношение числа полученных журналом в некотором году ссылок на статьи, вышедшие в том же самом году, к суммарному числу статей, вышедших за этот год в журнале [3].

Существующие в настоящее время наукометрические показатели не идеальны, но они позволяют количественно оценить публикационную активность автора, сообщества или организации.

### **1.3 Современные системы оценки научной деятельности**

Для хранения результатов научно-исследовательской деятельности существуют библиографические базы данных, а для отслеживания публикационной активности – сервисы наукометрии.

В данной работе были рассмотрены наиболее популярные системы оценки научной деятельности, использующие библиографические базы данных Scopus, Web of Science и РИНЦ.

#### **1. Интегрированная модульная платформа SciVal**

SciVal – это интегрированная модульная платформа, которая обеспечивает представление и оценку результатов исследований по всем научным областям. Инструмент охватывает более 4600 организаций из 220 стран и позволяет организациям оптимизировать стратегическое вложение средств, а также эффективно определять дальнейшие направления исследовательской работы и принимать рациональные решения при выборе персонала и партнеров.

Инструментарий SciVal базируется на данных библиографической базы данных Scopus и позволяют визуализировать результаты научной деятельности, сопоставлять и отслеживать результаты деятельности любых организаций, стран и predetermined групп, а также выявлять существующие и потенциальные возможности совместной деятельности на основе данных публикаций и цитирования.

Базовый инструментарий SciVal позволяет осуществлять:

- визуализацию результатов научно-исследовательской деятельности, определять уникальные компетенции отдельных организаций и стран,
- выявлять сильные стороны и многопрофильные области исследования, создавать наглядные отчеты, проводить детальные анализы за счет выбора необходимых комбинаций областей исследований и метрик для достижения поставленных целей,

- выявлять и анализировать существующие и потенциальные возможности совместной деятельности на основе публикаций и цитирований.

Помимо этого, SciVal позволяет создавать собственные области исследований для изучения и мониторинга, которые бы отражали фокусную тематику, а также осуществлять анализ исследовательских групп для создания уникальных центров превосходства.

Новый модуль Trends, реализованный в текущей версии SciVal, позволяет ответить на вопросы о том, какие направления в той или иной предметной области являются наиболее актуальными, а также о том, какие страны, организации и научные коллективы добились наибольших достижений в выбранном направлении [4].

## 2. Инструмент для оценки научной деятельности InCites

InCites – это легко настраиваемый наукометрический инструмент для оценки результатов научных исследований, который дает возможность отслеживать эффективность исследований и сопоставлять научную производительность с другими организациями и странами.

InCites использует в качестве источника данные индекса цитирования Web of Science и позволяет производить широкий анализ места организации в процессе исследования, а также делать выборочные срезы данных, которые демонстрируют определенные аспекты научной производительности.

Наукометрический инструмент InCites имеет возможность выявлять самые продуктивные совместные проекты. Наукометрические данные по странам помогут определить, где быстрее всего развивается научная деятельность, и выявить возможности нового сотрудничества.

InCites позволяет анализировать продуктивность институтов, сопоставлять деятельность организации и ее работников с деятельностью коллег в России и по всему миру [5].

Санкт-Петербургский государственный университет сотрудничает с InCites, поэтому все работники и обучающиеся университета имеют возможность пользоваться инструментом со всех компьютеров СПбГУ и по логинам и паролям с домашних компьютеров.

В InCites предусмотрена возможность создания как стандартных, так и пользовательских отчетов с заданным набором параметров. В основном пункте меню «Research performance profiles» доступны следующие возможности:

1. overview and summary metrics (общий обзор и сводные показатели);
2. productivity and researcher output (продуктивность отдельных исследователей);
3. collaboration and research networks (сотрудничество и исследовательские сети);
4. specialization and field strengths (специализация и сильные стороны);
5. trends and time series analysis (цитирование по годам, тренды, динамика);
6. impact and citation ranking (импакт и рейтинг цитирований).

Все данные для составления Research performance profiles обновляются ежеквартально.

### 3. Аналитическая система «Research in view»

В конце февраля 2011 г. «Thomson Reuters» выпустила новый аналитический продукт – систему «Research in view», также как и «InCites» ориентированную на организацию. Технологически «Research in view» отличается от системы «InCites»: «InCites» – это веб-приложение, «Research in view» – это корпоративная система, которая устанавливается на сервере организации, но через клиент-сервер связана с серверами «Web of knowledge».

«Research in view» дает возможность не только анализировать научную производительность самой организации или отдельных ее сотрудников (авторов публикаций), но и вводить в анализ организационную структуру

учреждения. Тем самым в ней можно анализировать публикационную активность лаборатории, кафедры, факультета, отдела и других структурных подразделений организации. «Research in view» реализует практически те же возможности, что и «InCites», но при этом организация и ее отдельные подразделения могут вводить новые (внешние) данные по публикациям помимо тех, которые система заимствует из библиографических источников, расположенных на платформе «Web of knowledge».

Таким образом, организация может анализировать результативность своей научной деятельности, деятельности своих научных подразделений и своих сотрудников, опираясь на совокупный массив данных, часть которых берется с «Web of knowledge», а другая часть вводится в систему самой организацией и отражает иные виды публикаций – монографии, научные отчеты, патенты, диссертации, учебники и др.

#### 4. Информационно-аналитическая система Наука-МГУ («Истина»)

Система «Истина» предназначена для перманентного сбора и систематизации, хранения и анализа наукометрической информации в вузах и научных организациях с целью подготовки и принятия управленческих решений.

Для автоматизации учета выходящих публикаций в мае 2012 года была введена в эксплуатацию (в пилотном режиме) система научной информации "Наука-МГУ" (предыдущее название "Истина"). Эта система позволяет сотрудникам самим заносить свои вышедшие научные публикации, а также другие достижения в научно-педагогической области [6].

Цели системы:

- сбор, систематизация и хранение информации, характеризующей результаты деятельности научных организаций;
- статистический и семантический анализ данных, выдача информации по запросам различных категорий пользователей: отдельных ученых, ученых секретарей, руководителей подразделений.

Основные принципы:

1. Распределенное наполнение: вся информация вводится децентрализованно, конкретными зарегистрировавшимися пользователями.
2. Публичность: вся введенная информация открыта и доступна всем пользователям интернета.
3. Верифицируемость: введенные публикации в высоко рейтинговых журналах проходят ручную проверку (при определении стимулирующих надбавок), для учета цитирований публикации проходят автоматическую проверку в Web of Science (при расчете наукометрических показателей при переизбрании) [7].

Введенные сотрудниками Московского государственного университета имени М.В. Ломоносова публикации систематизируются и анализируются в системе «Наука-МГУ» специальным инструментарием с обобщенной информацией по публикациям среди подразделений и областей наук.

Система позволяет добавлять в полуавтоматическом режиме библиографические записи из внешних баз данных, таких как:

- Web of Knowledge / Web of Science;
- ResearcherID.com;
- Scopus;
- eLibrary.ru;
- PubMed;
- MathSciNet и другие базы данных математических статей;
- Сайты издательств (Elsevier, Springer и др.) [8].

##### 5. Карта российской науки

Карта российской науки – это информационная система, которая предназначена для регулярного автоматического обновления информации об научных авторах и организациях (включая показатели их деятельности), осуществления статистического анализа научно-исследовательской

активности и обеспечения основы для создания аналитических материалов о состоянии российского сектора научных исследований и разработок. Информационная система «Карта российской науки» использует только общедоступные данные, официально полученные из следующих источников:

- ООО «Научная электронная библиотека» (научные публикации, входящие в Российский индекс научного цитирования);
- Thomson Scientific Inc. (научные публикации, индексируемые в базе данных Web of Science, и массив данных по зарубежным патентам, выданным российским организациям);
- Федеральное государственное бюджетное учреждение «Федеральный институт промышленной собственности» (информация по патентам на изобретения, полезные модели и промышленные образцы);
- Федеральное государственное бюджетное учреждение науки «Российская книжная палата» (информация по монографиям, учебникам для вузов, учебным пособиям для вузов и сборникам научных трудов).

В «Карте российской науки» представлены сведения за период 2007—2016 гг., с последующей актуализацией.

Данный проект с самого начала является неоднозначным и критикуемым. Цель системы – оптимизировать контроль над научной деятельностью в России со стороны государства средствами самих учёных.

Согласно заявлению Совета по науке при Министерстве образования и науки РФ от 31.01.2017 о конкурсе научных проектов, выполняемых в рамках госзадания в подведомственных МОН вузах, в результате проведения оценки деятельности по работе с информационной системой было установлено неудовлетворительное качество инструмента за четыре года своего существования. Совет призвал Министерство образования и науки РФ впредь не использовать «Карту российской науки» для каких-либо целей, а

использовать перечень общепринятых баз данных по различным областям [9].

В целом из всего вышесказанного можно сделать следующие выводы:

1. В общем случае все продукты представляют собой онлайн-сервисы по выводу наукометрических индикаторов для подбираемых пользователями наборов сущностей авторов, организаций, стран и т.д. Они снабжены встроенными средствами визуализации данных, которые облегчают и ускоряют анализ, а также позволяют экспортировать таблицы с показателями.

2. Ценность SciVal и InCites обусловлена коммерческим характером деятельности Thomson Reuters и Elsevier, по сути, это аналитические надстройки над Web of Science (WoS) и Scopus, которые вынесены в отдельные дорогие продукты. При этом в самих WoS и Scopus возможности для аналитики и экспорта специально существенно ограничены.

3. В SciVal с InCites есть множество ограничений и лимитов на выгрузку данных, поэтому для крупномасштабных исследований все равно необходимо закупать выгрузки, подключать API или вручную выкачивать сырые данные кусочками разрешенного размера из WoS и Scopus.

4. Еще одним недостатком современных инструментов SciVal и InCites является то, что они основаны на зарубежных публикациях и в них практически не учтены российские журналы.

5. Российский инструмент система «Наука-МГУ» направлена на формирование отчетов и представление данных в различных форматах, а также анализ статистических показателей по подразделениям и по тематикам для сотрудников Московского государственного университета имени М.В. Ломоносова.

#### **1.4 Спецификации требований к разрабатываемому решению**

Основные требования к разрабатываемому программному обеспечению:

- Должна быть функция автоматической выгрузки данных с базы Scopus.
- Должна быть функция приведения всех данных к единому виду и сохранения в базе данных системы в том числе:
  - a. функция формирования метаданных (описание полей и типов для хранения данных всех сущностей базы данных);
  - b. функция приведения полученных данных с API библиографических баз данных к соответствию сформированным метаданным;
  - c. функция сохранения данных в базе данных системы.
- Должна быть актуализация данных при обновлении используемых библиографических баз данных, посредством формирования ежедневных запросов к API библиографических баз данных для проверки появления новых данных.
- Должна быть функция создания научного сообщества в организации.
- Должна быть функция выгрузки обработанных данных публикационной активности организации, сообщества, журнала и автора на веб-страницу.
- Должна быть функция визуализации статистических данных в виде интерактивных диаграмм и графиков.
- Должна быть функция реализации интерактивной карты журналов с публикационной активностью научной организации.
- Должна быть функция реализации интерактивных графов партнеров автора, популярных научных областей.

Возможности системы анализа открытых баз цитирований научных публикаций представлены на диаграмме вариантов использования для пользователя (рисунок 1).



Рисунок 1 – Диаграмма вариантов использования

Комментарии к диаграмме:

1. Параметрический поиск предполагает поиск автора, издания или научного сообщества по имени автора, заглавию издания и названию научного сообщества.
2. Просмотр публикационной активности автора предполагает анализ показателей: число публикаций и цитирований, среднее число цитирований, индекс Хирша, количество изданий, в которых публикуется автор, а для издания – SJR (индекс рейтинга издания) и индекс Хирша. Также публикационная активность автора предполагает анализ графиков изменения количества публикаций/ цитирований по годам, диаграммы сравнения индекса Хирша автора с индексом Хирша других авторов, диаграммы публикационной активности автора в изданиях, таблицы с публикациями автора, а для издания – графики изменения рейтинга издания по годам.

3. Создание научного сообщества подразумевает объединение авторов, чтобы просматривать статистику не только по всей научной организации или отдельному автору, но и по созданным объединениям авторов.
4. Просмотр публикационной активности научного сообщества предполагает анализ диаграммы сравнения числа публикаций сотрудников сообщества, графика публикационной активности сообщества и диаграммы публикационной активности сообщества в журналах.
5. Просмотр научной деятельности ТПУ предполагает анализ научной деятельности университета по разным показателям, например, просмотр наиболее популярных тематик исследований и областей интересов, наиболее цитируемых авторов.
6. Карта журналов с публикационной активностью ТПУ предполагает просмотр географического местоположения журналов, в которых публикуются авторы ТПУ, с отображением количества публикации в журнале.
7. Поиск партнеров по авторам предполагает просмотр соавторов выбранного автора в виде графа, а по тематике – просмотр наиболее популярных авторов и их соавторов в интересующей тематике исследований, сравнение авторов по индексу Хирша в научной области.

### **1.5 Выводы по аналитическому обзору научной и технической документации предметной области**

Проведенный аналитический обзор предметной области, в том числе обзор существующих систем наукометрии позволил выявить основные недостатки популярных систем наукометрии и составить требования к разрабатываемой системе. Рассмотренные аналоги предоставляют широкий спектр статистических данных, однако, главный недостаток данных систем для использования их в российских научных организациях это

ограниченность источников данных, а именно отсутствие среди них библиографической базы данных РИНЦ.

Опираясь на приведенный аналитический обзор, было выполнено проектирование системы с учетом всех выявленных требований.

## **2. Проектирование системы анализа открытых данных баз цитирований научных публикаций**

### **2.1 Концептуальное проектирование системы оценки научной деятельности**

Разрабатываемая система предназначена для автоматизации процесса оценки эффективности публикационной деятельности в рамках научных организаций.

Данная система включает в себя получение данных с библиографических баз данных, вычисление статистик на основе собранных данных, обработку результатов статистических исследований, визуализацию данных в виде графиков, диаграмм и карт науки, формирование отчетности по полученным данным. Также отдельным модулем системы является интеграция со структурой научного сообщества. В рамках данного модуля предоставлена возможность воспроизвести в системе организационную структуру научного сообщества, включающую в себя различные объединения авторов, и в результате чего произвести оценку научной деятельности по разным элементам организационной структуры научного сообщества.

На рисунке 2 представлен прототип интерфейса настройки автоматической выгрузки данных. Данный интерфейс позволяет определить научные сообщества, по которым будут собираться данные, установить временной период в годах, по которому будут получены публикации, а также указать периодичность запуска автоматической синхронизации.

Добавить научное сообщество +

Укажите временной период для синхронизации:

2019 г. ✓ 2021 г. ✓

Периодичность синхронизации:

неделя  месяц  год

Томский политехнический университет

Казанский федеральный университет

Сохранить

Рисунок 2 – Схема работы команды синхронизации публикаций

На рисунке 3 представлена концептуальная схема, отражающая текущие сущности, на которых строится система. Публикации, авторы и источники выгружаются с базы Scopus, все виды статистик формируются приложением на основе загруженных данных. Сущность объединений авторов необходима для интеграции приложения с организационной структурой научного сообщества, например, для создания научных коллективов или лабораторий.

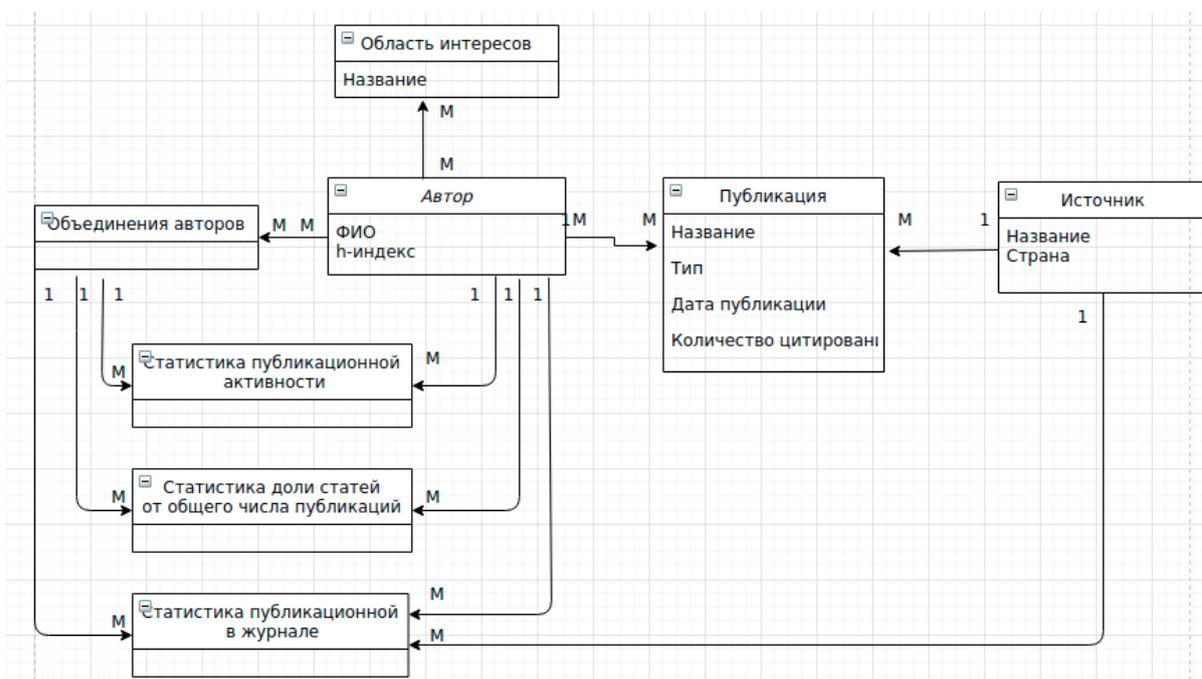


Рисунок 3 – Концептуальная схема системы

## 2.2 Алгоритмическое обеспечение анализа данных базы данных Scopus

Обобщенный алгоритм получения и обработки данных, полученных из библиографической базы данных Scopus, представлен на рисунке 4.

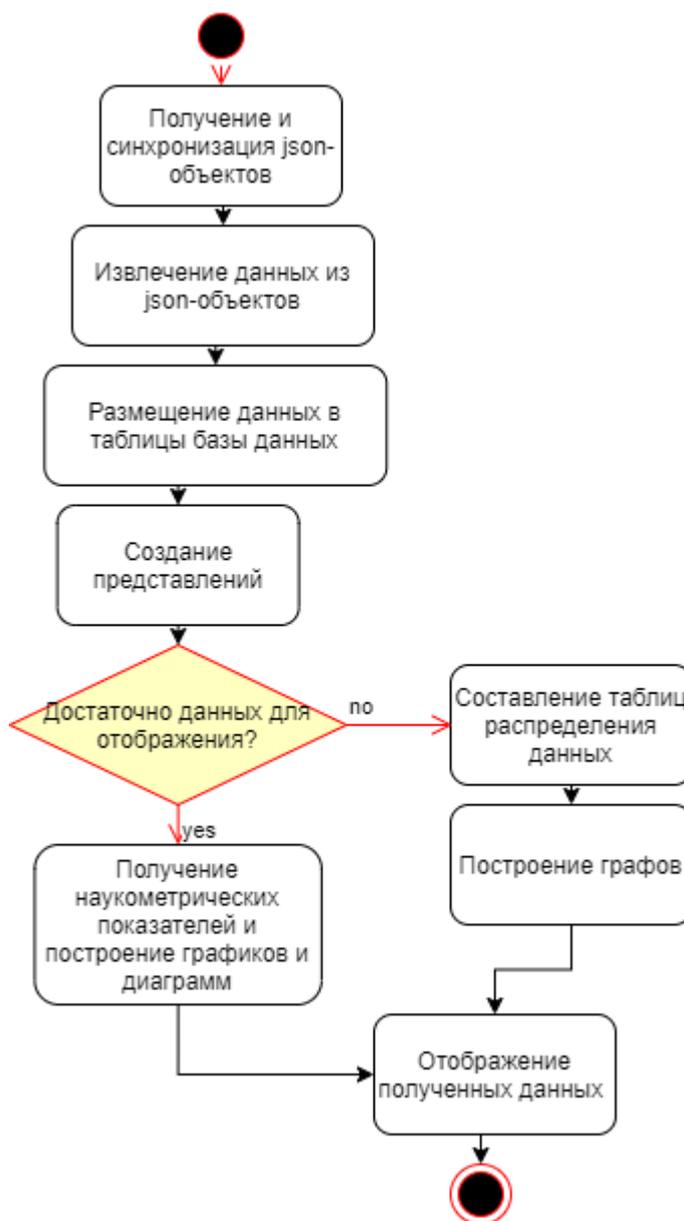


Рисунок 4 – Обобщенный алгоритм получения и обработки данных

После получения данных с помощью API Scopus в виде json-объектов, используя язык программирования PHP, из полученных объектов извлекаются необходимые данные и показатели об авторах, публикациях и изданиях и помещаются в базу данных, затем происходит обновление представлений базы данных, формирующих необходимую статистику, и

полученные данные учитываются в статистике. Далее созданные представления используются для отображения наукометрических показателей, построения графиков и диаграмм. В некоторых случаях, например, для построения графов, необходимо составить таблицы распределения данных. Для составления общей статистики и отслеживания динамики изменения различных показателей системы необходимы числовые характеристики этих показателей: объем, размах измерения, мода измерения и среднее или среднее арифметическое.

### 2.3 Функциональное проектирование системы оценки научной деятельности

Целью разработки системы анализа открытых баз цитирований научных публикаций является предоставление набора инструментов пользователю для отслеживания публикационной активности организации в целом и ее составляющих (авторы, подразделения, сообщества).

Функциональная схема системы оценки научной деятельности представлена на рисунке 5.



Рисунок 5 – Функциональная схема системы оценки научной деятельности

### **3. Разработка программного обеспечения анализа открытых данных баз цитирований научных публикаций**

#### **3.1 Краткое описание используемых технологий**

##### **3.1.1 Фреймворк Laravel**

Приложение на фреймворке Laravel версии 8 представляет серверную часть системы и реализует следующий функционал:

- автоматический сбор данных с базы Scopus;
- формирование статистик;
- формирование отчетности в виде документов формата Excel.

Рассмотрим подробнее основные используемые возможности фреймворка Laravel.

Автоматический сбор данных происходит посредством выполнения консольной команды, для чего Laravel предоставляет интерфейс командной строки artisan. Консольная команда выполняется запрос к API Scopus и сохраняет необходимые полученные данные. Запуск команды происходит с помощью планировщика задач Laravel, позволяющего задавать расписание запуска задач.

Формирование отчетности в формате Excel реализовано с использованием библиотеки Laravel Excel. Библиотека позволяет экспортировать SQL-запросы в Excel, при этом предоставляя возможности стилизации заголовков, размеров и прочих параметров ячеек.

##### **3.1.2 Язык программирования Python 3**

Python – простой и удобный язык программирования, включающий большое количество специализированных библиотек с множеством уже готовых программных конструкций.

Выбор языка программирования в первую очередь осуществлялся из соображений экономии времени разработки основных программных функций. К тому же библиотеки для анализа данных и поиска оптимальных

решений на их основе включают различные алгоритмы машинного обучения, статистические инструменты, методы визуализации и интеллектуального анализа данных.

На языке программирования Python выполнен следующий функционал:

1. визуализация статистических данных в виде интерактивных диаграмм и графиков;
2. реализация интерактивных графов;
3. реализация интерактивной карты журналов с публикационной активностью научной организации;
4. реализация веб-приложения с отображением полученных данных и сформированных статистик.

При разработке программного обеспечения использовались следующие библиотеки Python:

- plotly и networkx – для построения интерактивных графиков, диаграмм и графов,
- pyscorp2 – для получения данных из базы данных,
- pandas – для работы с данными из базы данных,
- folium – для построения карты журналов с публикационной активностью научной организации.

Plotly – это библиотека для создания интерактивных визуализаций и управления ими. В этой написанной на JavaScript библиотеке есть множество встроенных приложений для машинного обучения и анализа данных, что упрощает реализацию и визуализацию разных алгоритмов.

Преимущества использования Plotly:

- интерактивные диаграммы и анимации на Python;
- построение карт, научных или финансовых графиков и 3D-диаграмм;
- работает offline;
- отдельные объекты осей не создаются при построении нескольких графиков вместе.

Plotly – это не только библиотека для Python, но ещё и JS, это значит, что любые графики, можно показывать и на сайте (если он на Python).

NetworkX – это программный пакет на языке Python для создания, управления и изучения структуры, динамики и функций сложных сетей. Он используется для изучения больших сложных сетей, представленных в виде графов с узлами и ребрами [10].

Psycopg2 – самая популярная библиотека для работы с PostgreSQL, выполнения SQL-запросов и других операций с базой данных.

Pandas – это библиотека Python, которая предоставляет широкие возможности для анализа данных. Данные чаще всего хранятся в форме табличек – например, в форматах .csv, .tsv или .xlsx. С помощью библиотеки Pandas такие табличные данные очень удобно загружать, обрабатывать и анализировать с помощью SQL-подобных запросов. В экосистеме Python, pandas является наиболее продвинутой и быстроразвивающейся библиотекой для обработки и анализа данных.

Folium – это библиотека для визуализации географических данных и информации, которая содержит координаты и местоположения. В данной работе библиотека использовалась для построения карты журналов с публикационной активностью организации по координатам (широта, долгота) стран, к которым принадлежат журналы.

### **3.1.3 Фреймворк Dash**

Dash – это фреймворк для построения веб-приложений на Python. Также используется для создания современных элементов интерфейса, таких как диаграммы, таблицы и интерактивные функции ввода и вывода. Он написан с использованием Flask, React.js, Plotly.js и является инструментом для интерактивной визуализации, которую можно интегрировать в веб-сайт.

Фреймворк Dash содержит модули `dash_html_components`, в котором находятся все компоненты HTML, и `dash_core_components`, который отвечает за работу с диаграммами Plotly. Также был использован модуль `dash-`

bootstrap-components, в котором находятся компоненты Bootstrap для Plotly Dash, упрощающие создание последовательно стилизованных приложений со сложными адаптивными макетами [11].

### **3.1.4 СУБД PostgreSQL**

В качестве СУБД используется PostgreSQL – свободная объектно-реляционная система управления базами данных. Данная СУБД поддерживает работу с данными типа json, что повсеместно используется при выгрузке данных, поскольку ответы на запросы к API Scopus извлекаются в формате json-объектов. Помимо этого, PostgreSQL удобен для составления представлений, что используется для формирования статистических данных внутри системы.

### **3.2 Импорт исходных данных**

Для сбора данных система включает в себя следующие команды:

- Команда синхронизации интересующих научных сообществ по наименованию сообществ.
- Команда синхронизации публикаций выгруженных сообществ с возможностью указать интересующие годы публикационной активности сообщества.
- Команда синхронизации авторов выгруженных публикаций (в том числе и соавторов из других научных сообществ).
- Команда синхронизации источников выгруженных публикаций.

На рисунках 6 и 7 представлены схемы работы команд синхронизации научных сообществ и публикаций, остальные команды работают по аналогичным схемам.

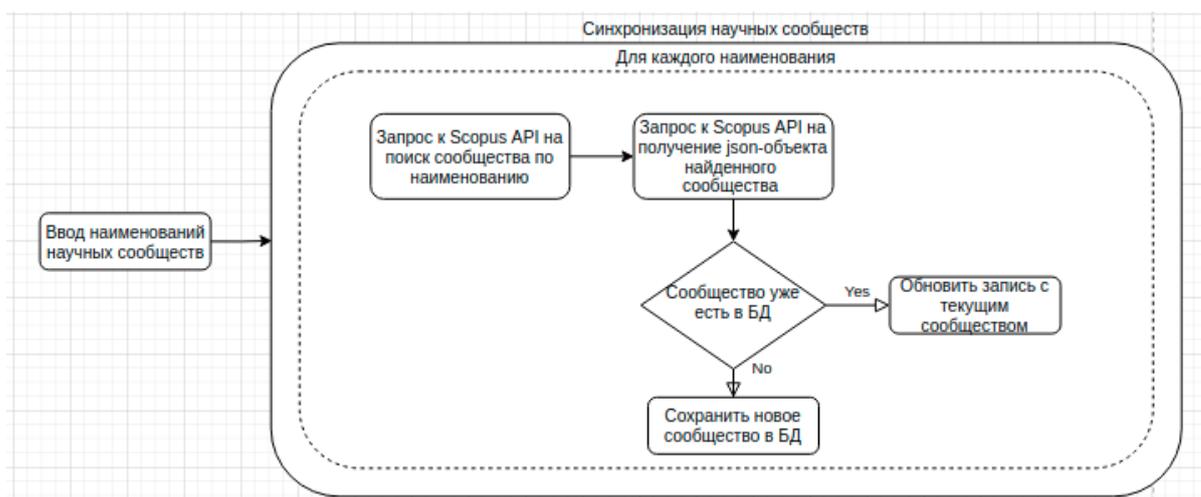


Рисунок 6 – Концептуальная схема работы команды синхронизации научных сообществ

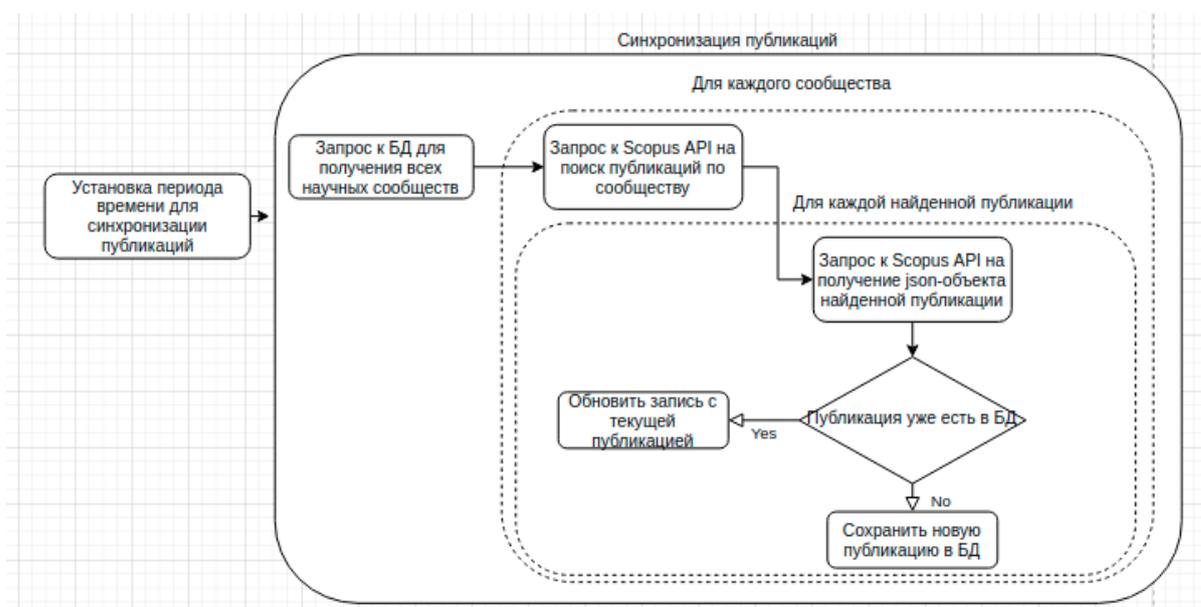


Рисунок 7 – Концептуальная схема работы команды синхронизации публикаций

В результате работы вышеперечисленных команд заполняются соответствующие таблицы в базе данных, данные для заполнения получают посредством отправки необходимых запросов к API Scopus.

Для того чтобы избежать дублирования записей, каждая таблица хранит идентификаторы Scopus. Если в ответе на запрос приходит запись с идентификатором, уже имеющимся в базе данных системы, новая запись не создается, вместо этого происходит обновление полей уже имеющейся записи.

В ответе на запросы к API Scopus извлекаются json-объекты, описывающие научные сообщества, публикации, авторов и источники. При выгрузке в БД отдельно сохраняются необходимые на текущий момент поля json-объекта и сам json-объект. При возникновении необходимости получить дополнительные сведения о выгруженных объектах можно не запрашивать повторно данные, а извлечь их из хранимого json-объекта. Это поможет ускорить процесс расширения хранимой информации, поскольку длительность получения данных через запросы может измеряться в днях из-за большого количества данных и имеющихся ограничений со стороны Scopus на количество запросов.

По каждому научному сообществу сохраняется json-объект с информацией о сообществе из Scopus и наименование.

По каждой публикации сохраняется json-объект с информацией о публикации из Scopus, название публикации, год публикации, количество цитирований и тип публикации.

По каждому автору сохраняется json-объект с информацией об авторе из Scopus и ФИО. По выгруженным публикациям рассчитывается количество публикаций, количество цитирований автора и количество документов, в которых цитируется автор, на текущий день. Данную информацию можно получать непосредственно из json-объекта автора, однако, в таком случае она будет актуальна только в рамках одной библиографической базы данных.

По каждому источнику сохраняются json-объект с информацией об источнике из Scopus, название источника и тип источника.

Чтобы избежать лишних запросов во время выполнения автоматической синхронизации, сущность публикации имеет 2 флага – `is_author_sync` и `is_source_sync`. По умолчанию эти флаги для всех полученных публикаций имеют значение `false`. Как только по публикации закончилась выгрузка авторов, флаг `is_author_sync` принимает значение `true`,

и в следующий раз по данной публикации авторы повторно не будут запрашиваться. Аналогично работает флаг `is_source_sync`.

На рисунке 8 представлена полученная схема базы данных, включающая только таблицы. Полная схема базы данных со всеми представлениями приведена в приложении.

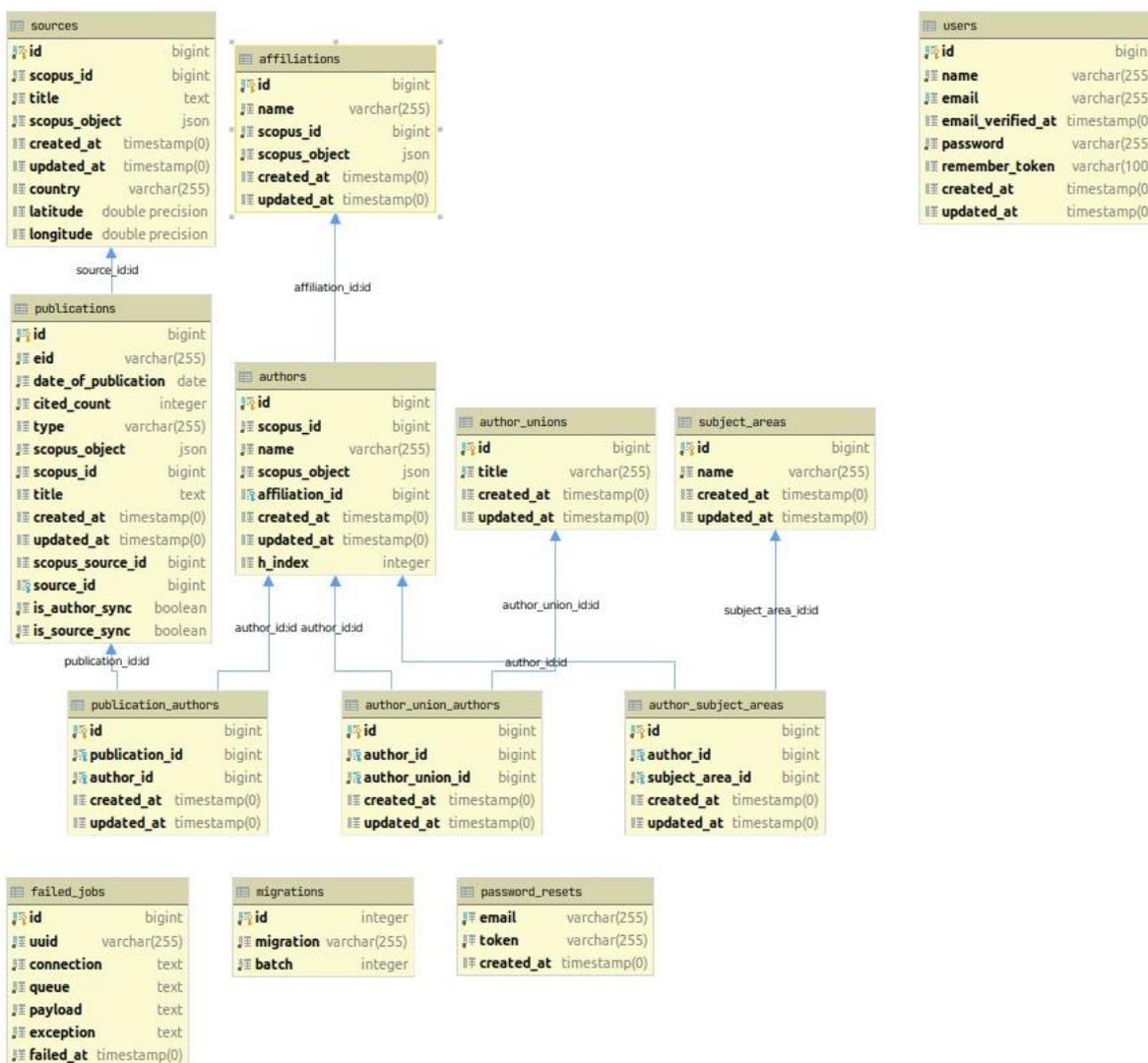


Рисунок 8 – Схема базы данных (без представлений)

### 3.3 Описание формирования статистики

На примере использования данных по Томскому политехническому университету было сформировано несколько возможных вариантов статистических таблиц.

Для формирования статистики по университету и отдельным авторам на основе имеющихся данных средствами SQL созданы материализованные представления. Примеры скриптов генерации данных статистик представлены в приложении. Реализованы следующие статистики:

1. Статистика публикационной активности по университету. Данная статистика хранит в себе общее число публикаций, общее число цитирований и среднее значение цитируемости на весь университет за каждый год публикационной активности университета.
2. Статистика публикационной активности по отдельным авторам. Данная статистика хранит в себе всех авторов, их число публикаций, число цитирований и среднее значение цитируемости за каждый год публикационной активности автора.
3. Статистика по числу публикаций типа «статья» по университету. Данная статистика хранит в себе процент числа публикаций, имеющих тип «статья», от общего числа публикаций на весь университет за каждый год публикационной активности университета.
4. Статистика по числу публикаций типа «статья» по отдельным авторам. Данная статистика хранит в себе всех авторов, процент числа публикаций, имеющих тип «статья», от общего числа публикаций автора за каждый год публикационной активности автора.
5. Статистика по публикациям в различных источниках по отдельным авторам. Данная статистика хранит всех авторов, журналы, где они публикуются, и число статей, опубликованных автором в данном журнале за все время публикационной активности автора.

Для каждой из статистик реализован функционал выгрузки в отчет формата Excel с помощью библиотеки Laravel Excel. При выгрузке можно указать интересующий год публикационной активности и получить отчет с данными за этот год. На рисунке 9 представлен фрагмент отчета формата Excel статистики по публикациям автора в различных журналах.

	А	В	С
1	Автор	Кол-во публикаций	Журнал
2	Artamonov A.		9 2018 9th International Conference on Information, Intelligence, Systems and Applications, ISA 2018
3	Ghete V.M.		4 2018 9th International Conference on Information, Intelligence, Systems and Applications, ISA 2018
4	Audurier B.		2 2018 9th International Conference on Information, Intelligence, Systems and Applications, ISA 2018
5	Zaichenko T.N.		1 2018 9th International Conference on Information, Intelligence, Systems and Applications, ISA 2018
6	Xu M.		1 2018 9th International Conference on Information, Intelligence, Systems and Applications, ISA 2018
7	Maksimov I.V.		1 2018 9th International Conference on Information, Intelligence, Systems and Applications, ISA 2018
8	Liu C.		1 2018 9th International Conference on Information, Intelligence, Systems and Applications, ISA 2018
9	Maglitteri A.		1 2018 9th International Conference on Information, Intelligence, Systems and Applications, ISA 2018
10	Basels I.M.		9 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
11	Alduganov D.		1 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
12	Igumnov L.O.		1 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
13	Li M.		1 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
14	Kerbizi A.		1 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
15	Shen J.		1 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
16	Nicker K.		1 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
17	Vodova Y.M.		1 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
18	Goldouzian B.		3 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
19	Lutsenko A.S.		1 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
20	Allyarova I.S.		1 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
21	Schofbeck R.		4 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
22	Cornelis T.		3 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
23	Moiseeva Y.		1 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
24	Ivanov Y.F.		2 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
25	Szuerits S.		1 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
26	Amato S.		9 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
27	Verbeke W.		1 2019 International Siberian Conference on Control and Communications, SIBCON 2019 - Proceedings
28	Bogomazova E.		1 ACS applied materials & interfaces
29	Kurjata R.P.		1 ACS applied materials & interfaces
30	Ryzhakova N.		2 ACS applied materials & interfaces
31	Bencivenni G.		9 ACS applied materials & interfaces
32	Pavlov S.		2 ACS applied materials & interfaces
33	Troyan P.E.		1 ACS applied materials & interfaces
34	Zakharova A.A.		1 ACS applied materials & interfaces
35	Ustinov A.M.		1 ACS applied materials & interfaces
36	Gräslund T.		1 ACS applied materials & interfaces
37	Orlov A.		1 ACS applied materials & interfaces
38	Kral Z.		1 ACS applied materials & interfaces
39	Campi D.		1 ACS applied materials & interfaces
40	Henschel C.L.		1 ACS applied materials & interfaces
41	Kryuchkov D.		1 Advanced Energy Materials
42	Brossa Gonzalo A.		2 Advanced Energy Materials

Рисунок 9 – Фрагмент отчета формата Excel

### 3.4 Описание реализованной визуализации публикационной активности

Обязательной частью анализа является визуализация. Визуализация данных помогает представить большие и сложные наборы данных в простом и наглядном виде, а именно в виде графиков, диаграмм и графов.

В виде интерактивных графиков и диаграмм можно представить изменение и сравнение наукометрических показателей объектов исследования. Такие графики и диаграммы обладают приятным видом и удобны для длительного исследования данных.

В данной работе для оценки научной деятельности ТПУ будут использоваться следующие графические представления:

1. Линейный график публикационной активности помогает визуализировать данные в компактном и точном формате, что позволяет легко воспринимать информацию, отмечать тренды или соотношение показателей.

2. Столбчатая диаграмма сравнения наукометрических показателей значительно упрощает сравнение сопоставимых данных одной категории в рамках поставленных ограничений.
3. Круговая диаграмма публикационной активности упрощает восприятие, потому что отношения частей от целого показаны на ней наглядным образом.
4. Граф позволяет отобразить не только показатели, но и связь между ними.
5. Карта с публикационной активностью позволяет показать не только количественные показатели, но и географическое расположение объектов исследования.

### **3.5 Описание реализованных возможностей системы оценки научной деятельности**

Разработка веб-приложения велась на языке программирования Python 3 с использованием фреймворка Dash для представления пользовательского интерфейса и с использованием фреймворка Laravel для получения данных.

Веб-приложение содержит следующие страницы:

- публикационная активность ТПУ,
- авторы,
- издания,
- научные сообщества,
- партнеры,
- приоритетные направления,
- карта журналов.

Пример стартовой страницы, на которой представлена публикационная активность ТПУ, представлен на рисунке 10. Публикационная активность представлена круговыми и столбчатыми диаграммами, построенными по статистическим данным авторов университета.

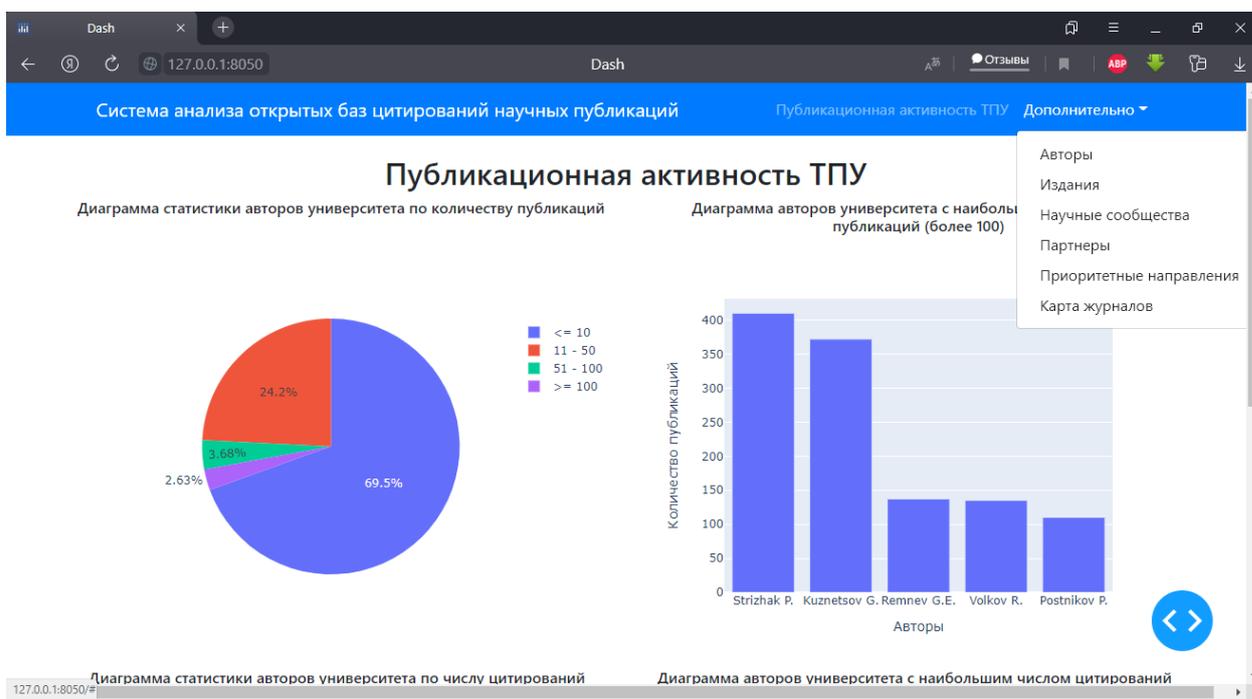


Рисунок 10 – Стартовая страница

На странице «Авторы» размещены поисковая строка для нахождения, интересующего автора и список авторов ТПУ в виде ссылок (рисунок 11).

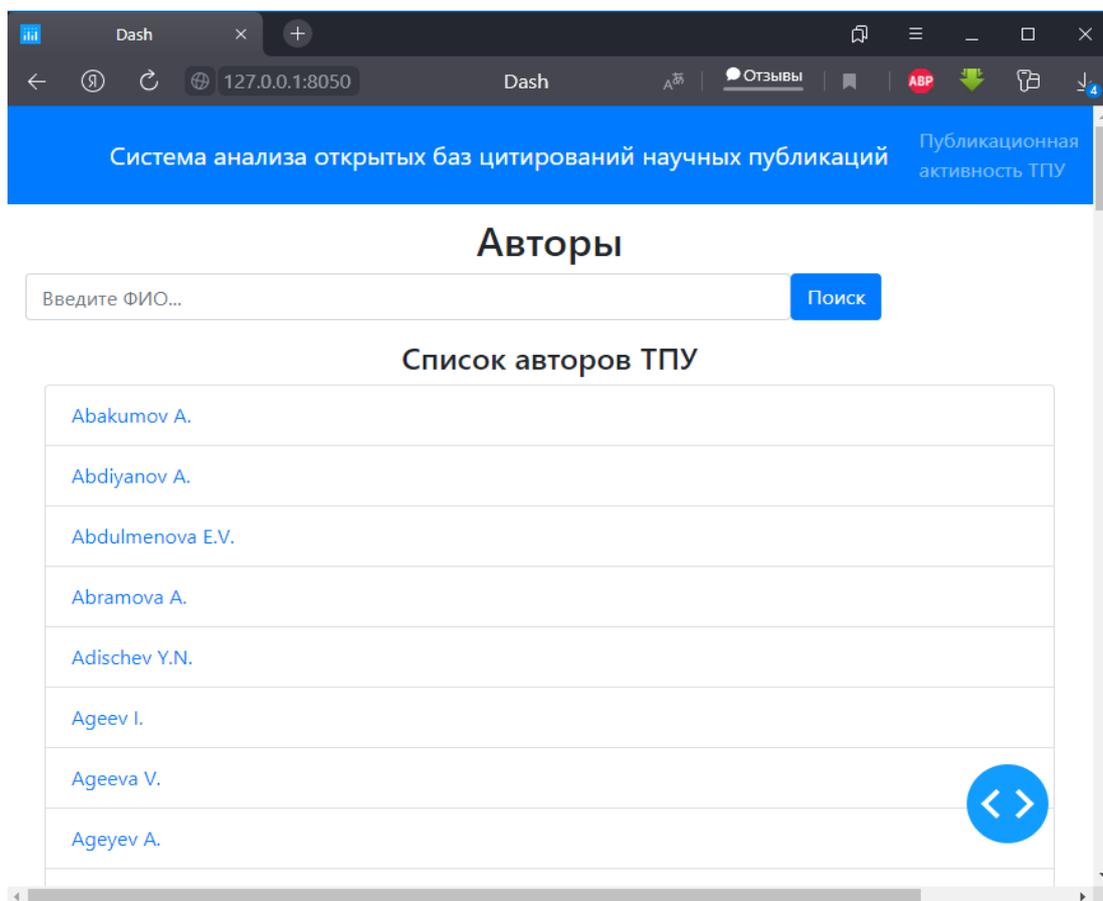


Рисунок 11 – Страница «Авторы»

После выбора автора на странице отображается публикационная активность автора, которая содержит наукометрические показатели автора, области интересов, графики и диаграммы публикационной активности, а также публикации автора. Примеры страницы представлены на рисунках 12-14.

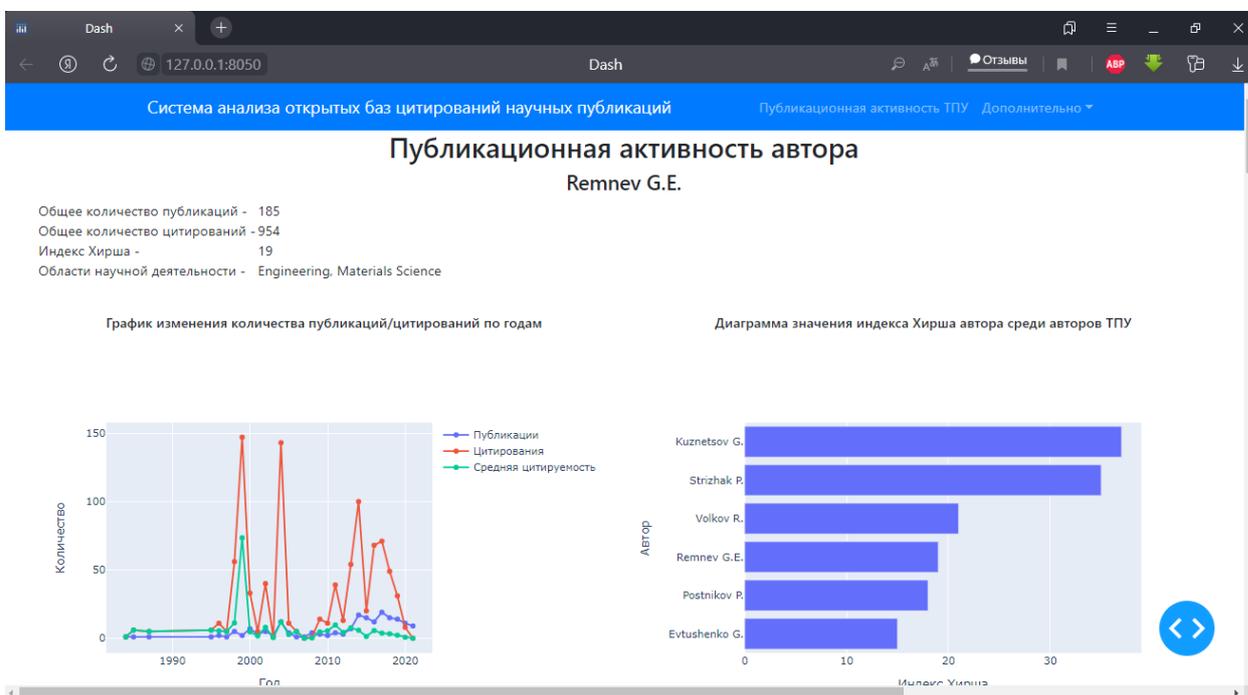


Рисунок 12 – Страница «Авторы»

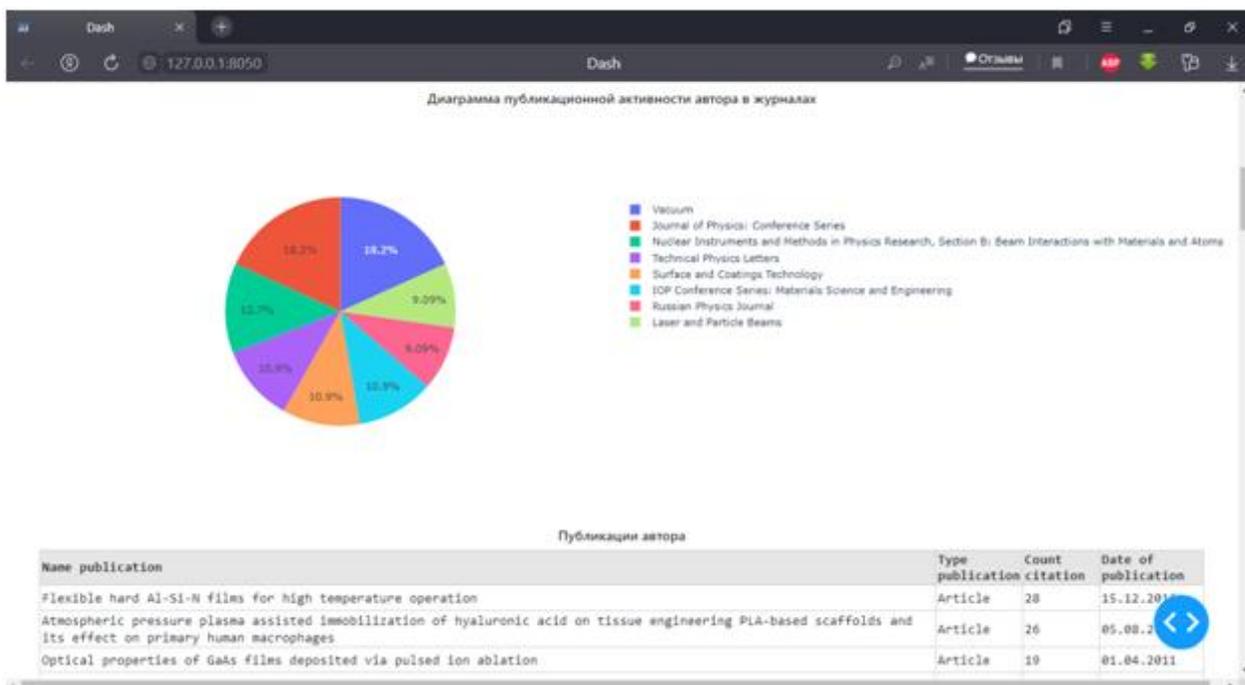


Рисунок 13 – Страница «Авторы»

Name publication	Type publication	Count citation	Date of publication
Flexible hard Al-Si-N films for high temperature operation	Article	28	15.12.2016
Atmospheric pressure plasma assisted immobilization of hyaluronic acid on tissue engineering PLA-based scaffolds and its effect on primary human macrophages	Article	26	05.08.2017
Optical properties of GaAs films deposited via pulsed ion ablation	Article	19	01.04.2011
Effect of energy on the formation of flexible hard Al-Si-N films prepared by magnetron sputtering	Article	16	01.01.2016
Pulse plasma-chemical synthesis of ultradispersed powders of titanium and silicon oxide	Article	15	24.10.2013
Progress in plasma focus research at the Kurchatov Institute	Conference Paper	15	01.01.2014
Formation of nanoscale carbon structures in the surface layer of metals under the impact of high intensity ion beam	Article	15	15.08.2014
A high-repetition rate pulsed electron accelerator	Article	13	20.08.2013
Characteristics of pulsed plasma-chemical synthesis of silicon dioxide nanoparticles	Article	13	01.01.2014
Plasma-chemical synthesis of composite nanodispersed oxides	Article	12	15.12.2011
A high-repetition rate pulsed electron accelerator	Conference Paper	12	01.12.2012
The Astra repetitive-pulse electron accelerator	Article	12	01.01.2013
Physicochemical properties of the surface of a zirconium alloy modified by a pulsed ion beam	Article	12	01.01.2014
Effect of temperature on radiation resistance of $TiO_2$ powders during heating and modification by $SiO_2$ nanoparticles	Article	12	01.10.2014
Effects on structure and properties of $Zr_{55}Al_{10}Cu_{30}Ni_5$ metallic glass irradiated by high intensity pulsed ion beam	Article	11	15.09.2014
Size effect in AlN/SiN multilayered films irradiated with helium and argon ions	Article	11	15.11.2014
A high-voltage pulse generator for electric-discharge technologies	Article	10	01.01.2013
Effect of emission current delay on the efficiency of electron beam production	Article	10	01.09.2013
Formation of silicon carbide and diamond nanoparticles in the surface layer of a silicon target during short-pulse carbon	Article	10	01.09.2013

Рисунок 14 – Страница «Авторы»

Аналогично странице «Авторы» построена страница «Издания», после выбора издания на странице отображается публикационная активность издания, которая содержит наукометрический показатель (индекс Хирша), страна размещения журнала и график изменения рейтинга журнала (индекс SJR). Пример страницы представлен на рисунке 15.

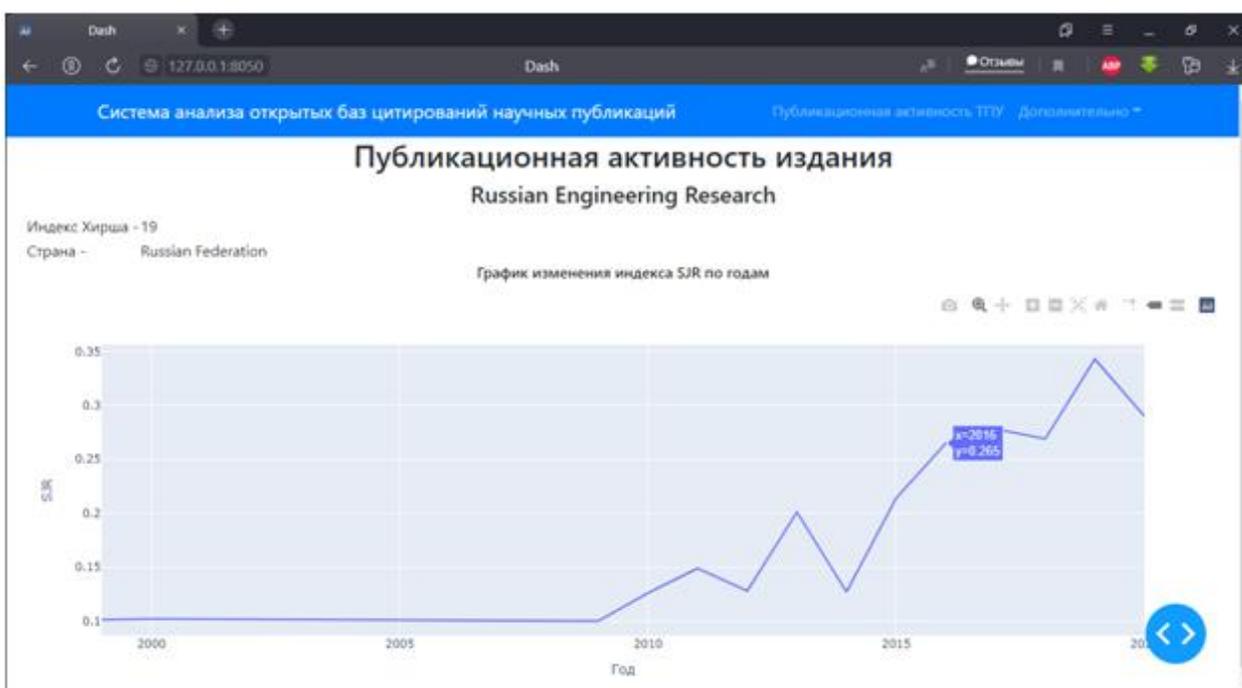


Рисунок 15 – Страница «Издания»

На рисунке 16 представлена страница «Научные сообщества», на которой размещены форма создания научного сообщества и список уже существующих научных сообществ ТПУ.

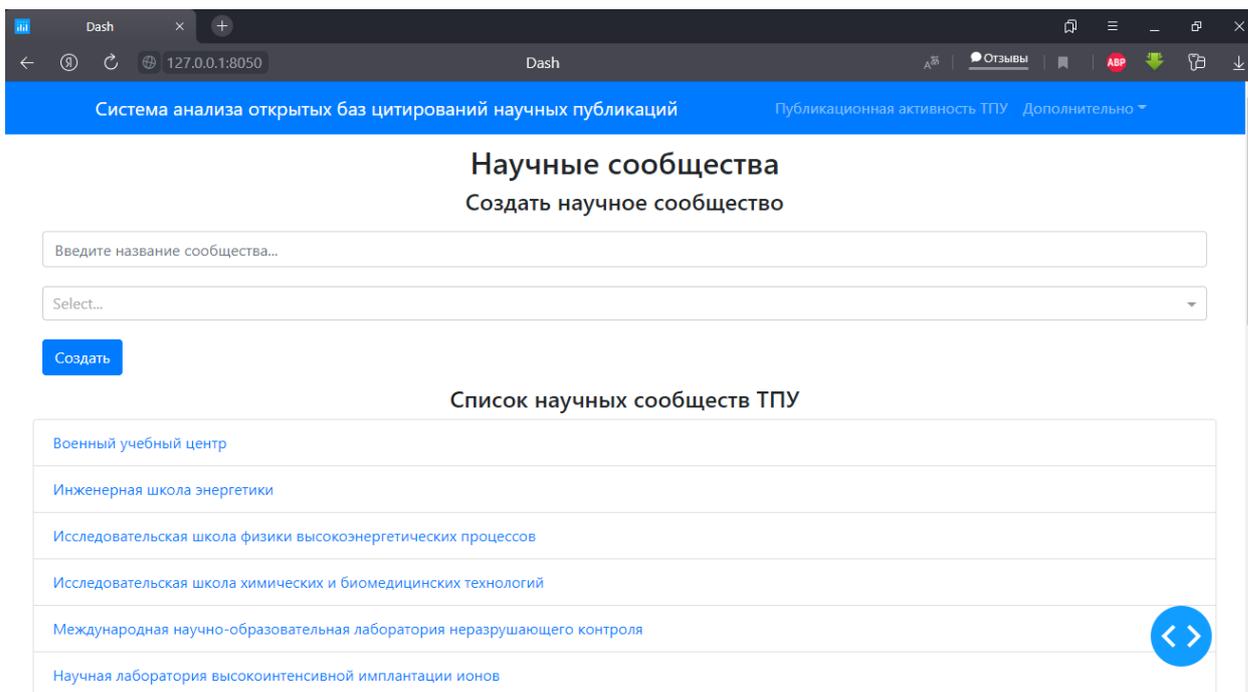


Рисунок 16 – Страница «Научные сообщества»

После выбора научного сообщества из списка, пользователь переходит на страницу публикационной активности выбранного сообщества (рисунки 17-18). На данной странице размещены список сотрудников сообщества с их значением индекса Хирша, диаграмма сравнения количества публикаций сотрудников, график и диаграмма публикационной активности сообщества.

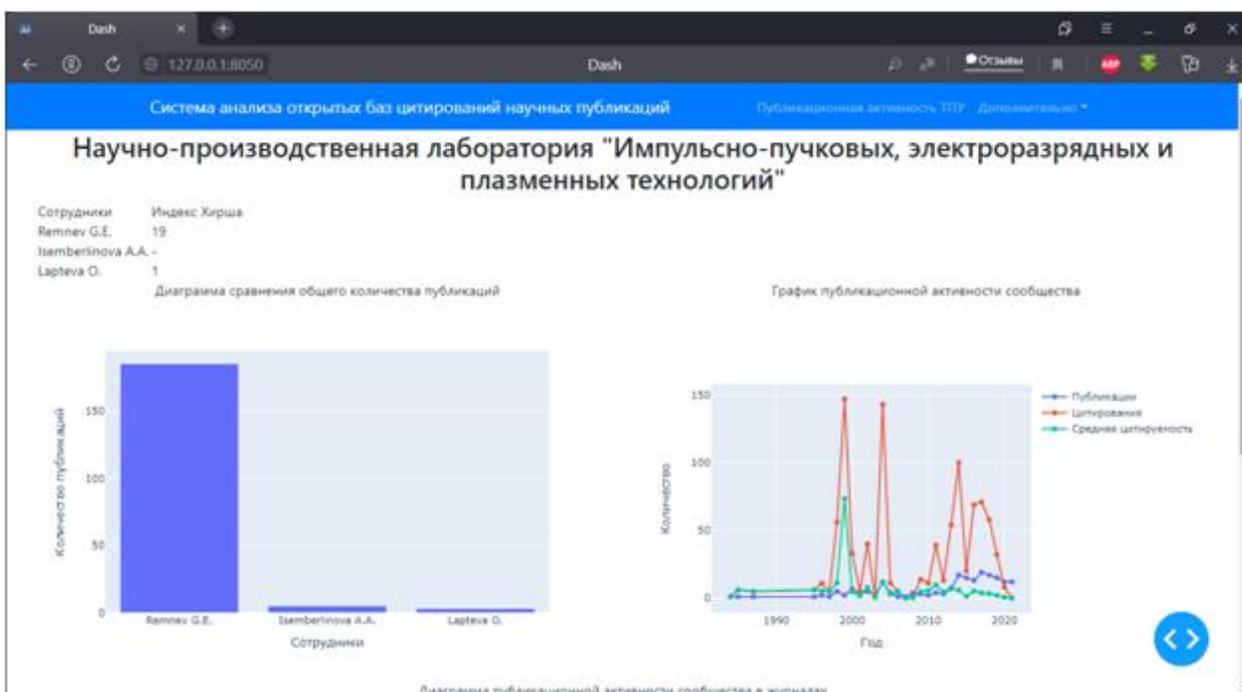


Рисунок 17 – Страница «Научные сообщества»

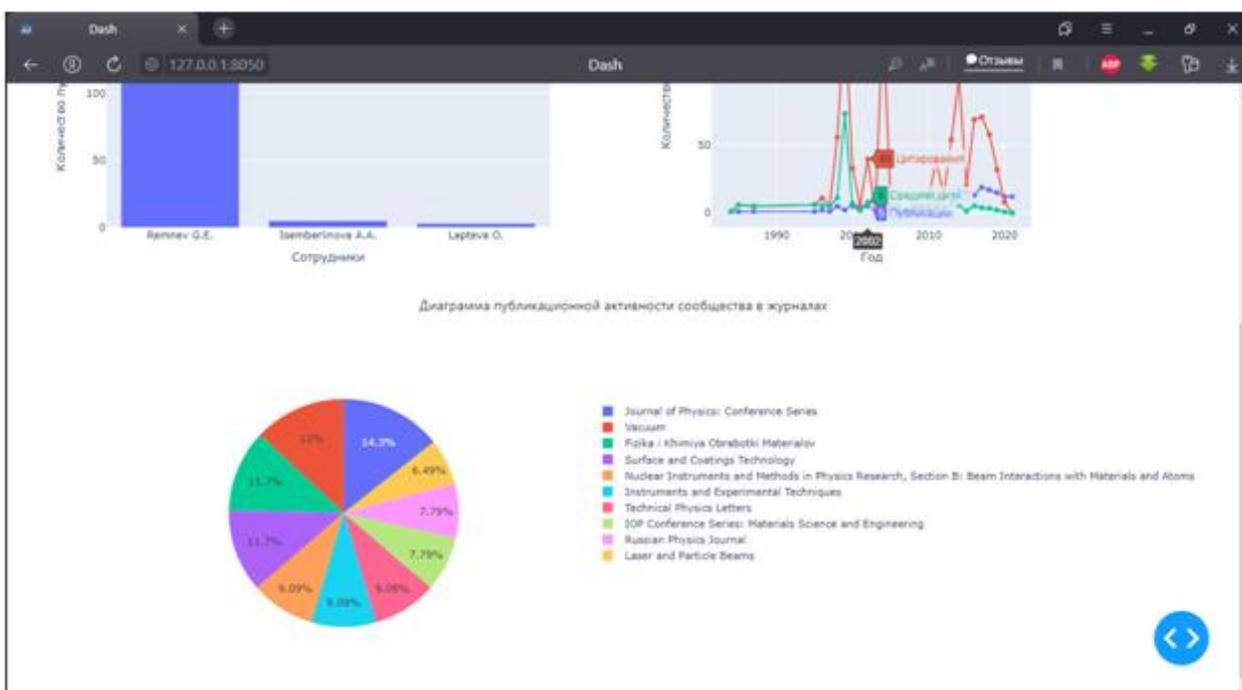


Рисунок 18 – Страница «Научные сообщества»

На странице «Поиск партнеров» пользователю предложено выполнить поиск по научной области или по соавторам какого-либо автора (рисунок 19).

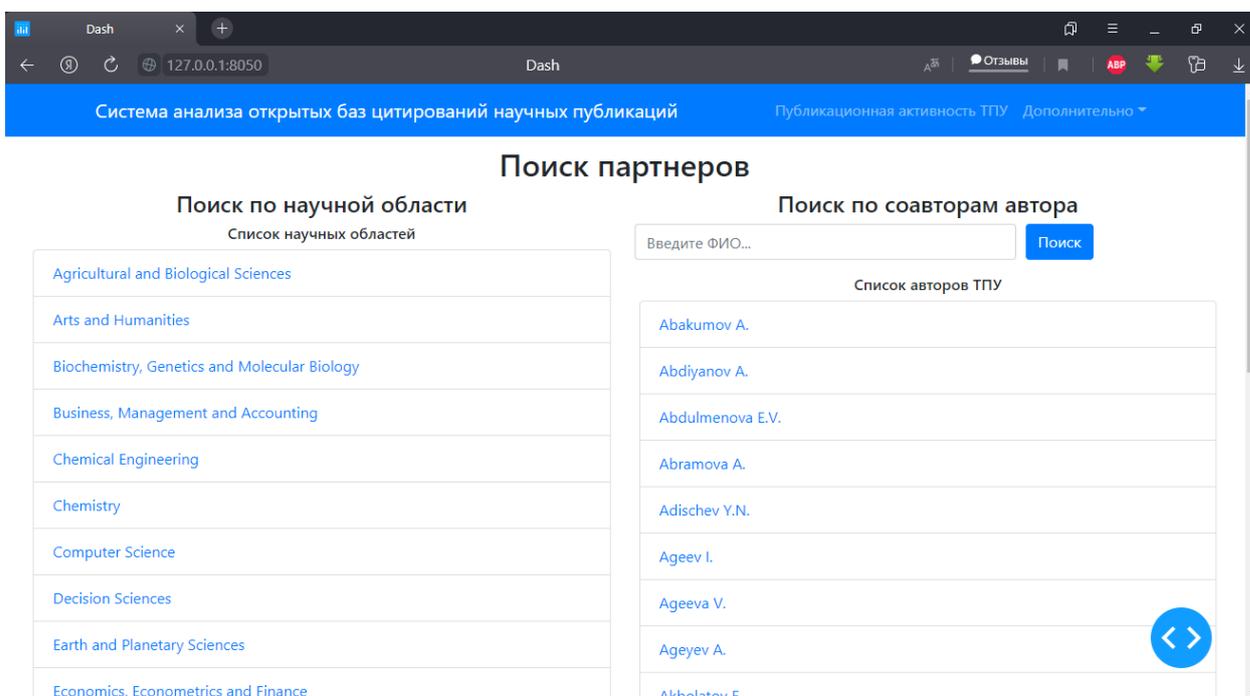


Рисунок 19 – Страница «Поиск партнеров»

На странице «Поиск по научной области» размещены диаграмма сравнения индекса Хирша авторов, а также ползунок изменения диапазона значения индекса Хирша, что позволяет пользователю рассматривать авторов, как с наименьшими значениями h-индекса, так и с наибольшими (рисунок 20).

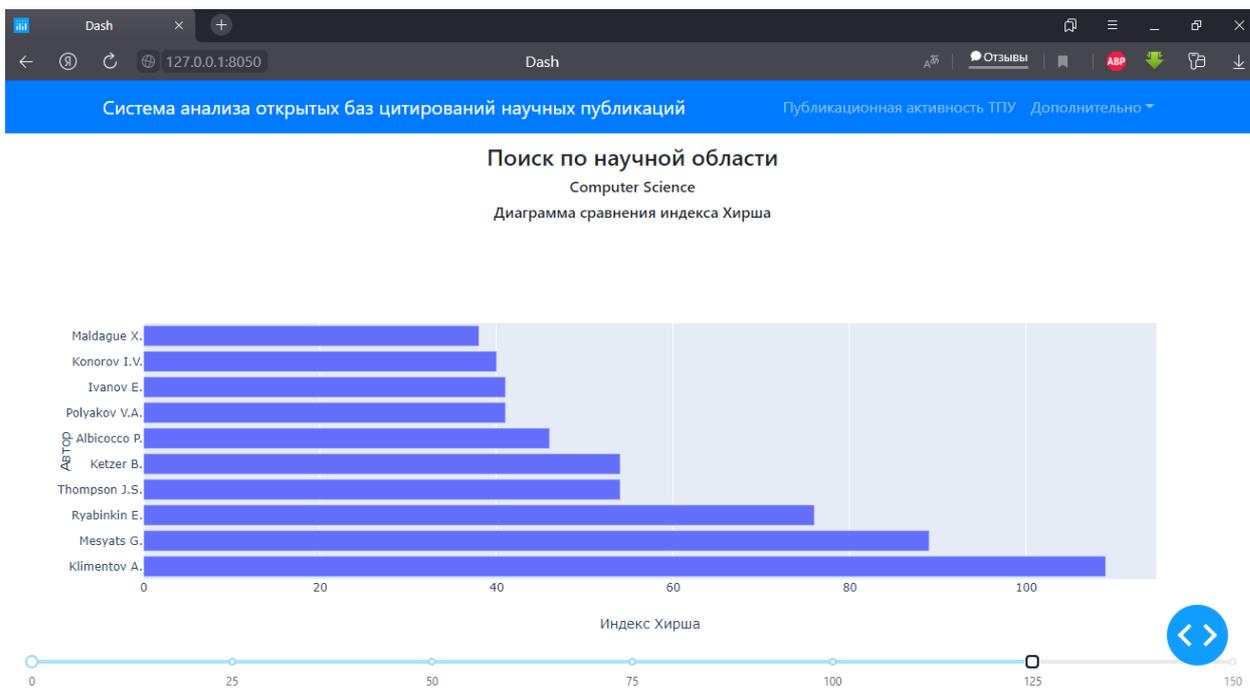


Рисунок 20 – Страница «Поиск по научной области»

На странице «Поиск по соавторам автора» размещен граф соавторов выбранного автора (рисунок 21).

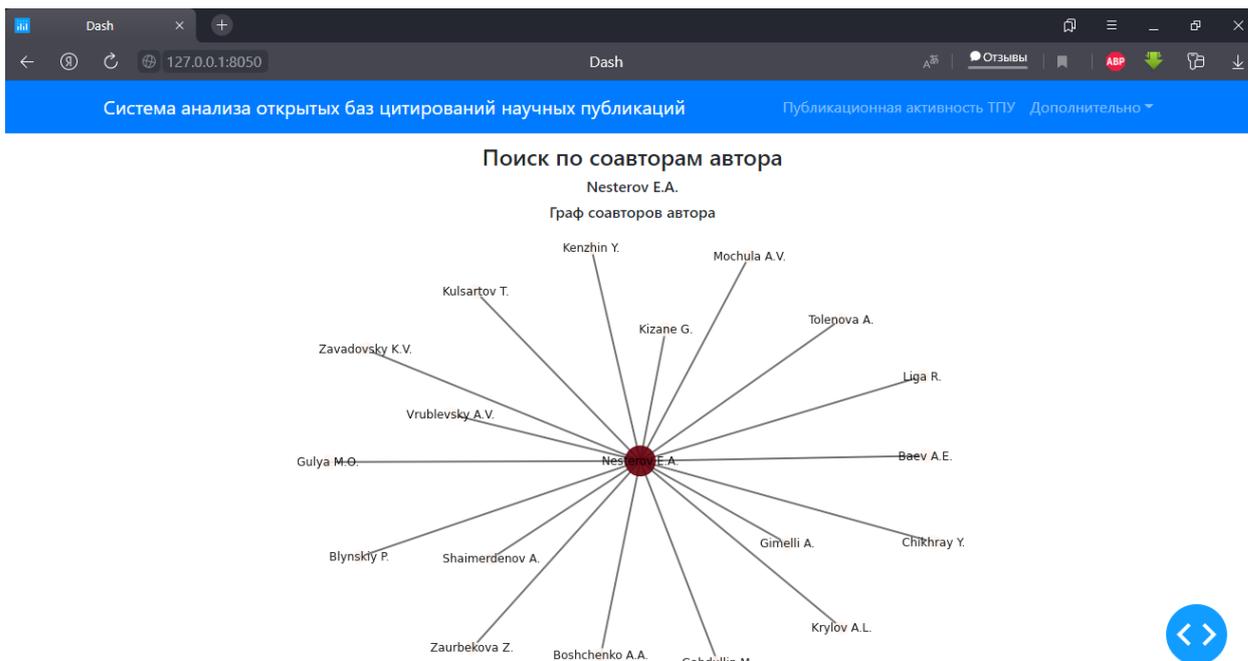


Рисунок 21 – Страница «Поиск по соавторам автора»

На рисунке 22 представлена страница «Приоритетные направления», на которой размещен граф приоритетных научных направлений, позволяющий визуально выделить наиболее популярные области исследования.

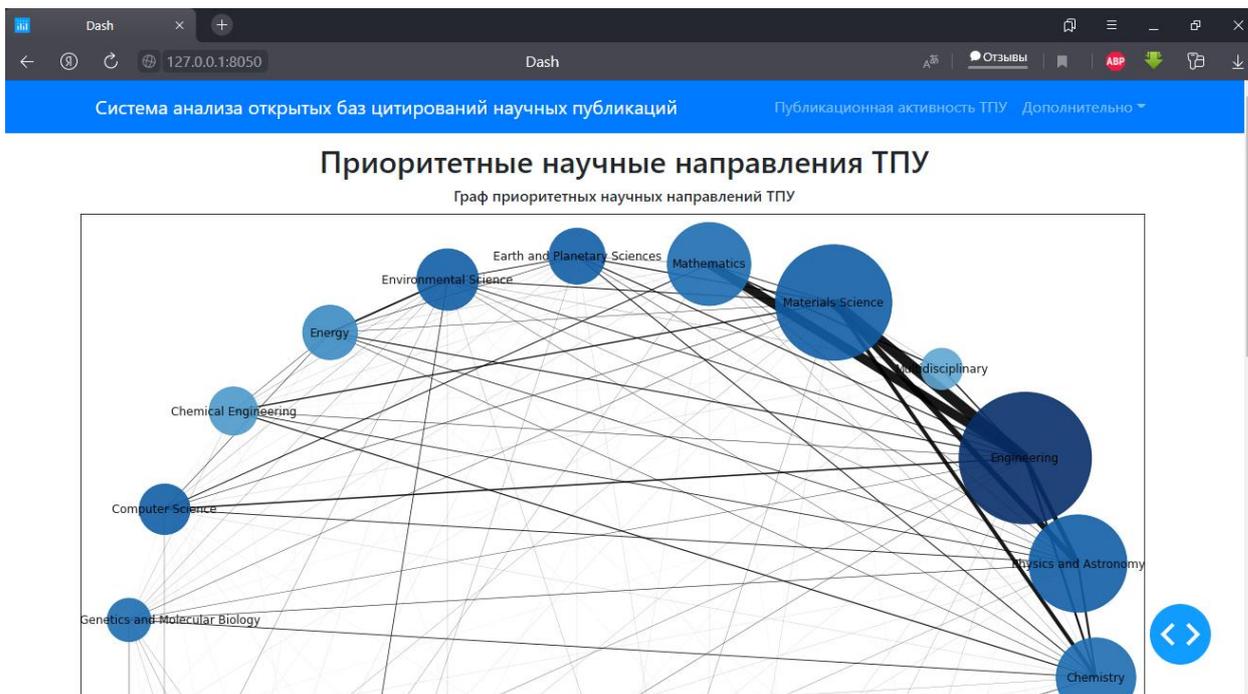


Рисунок 22 – Страница «Приоритетные направления»

Также на данной странице размещена диаграмма приоритетных научных направлений, позволяющая количественно оценить какое количество авторов публикуется в каждой научной области (рисунок 23).

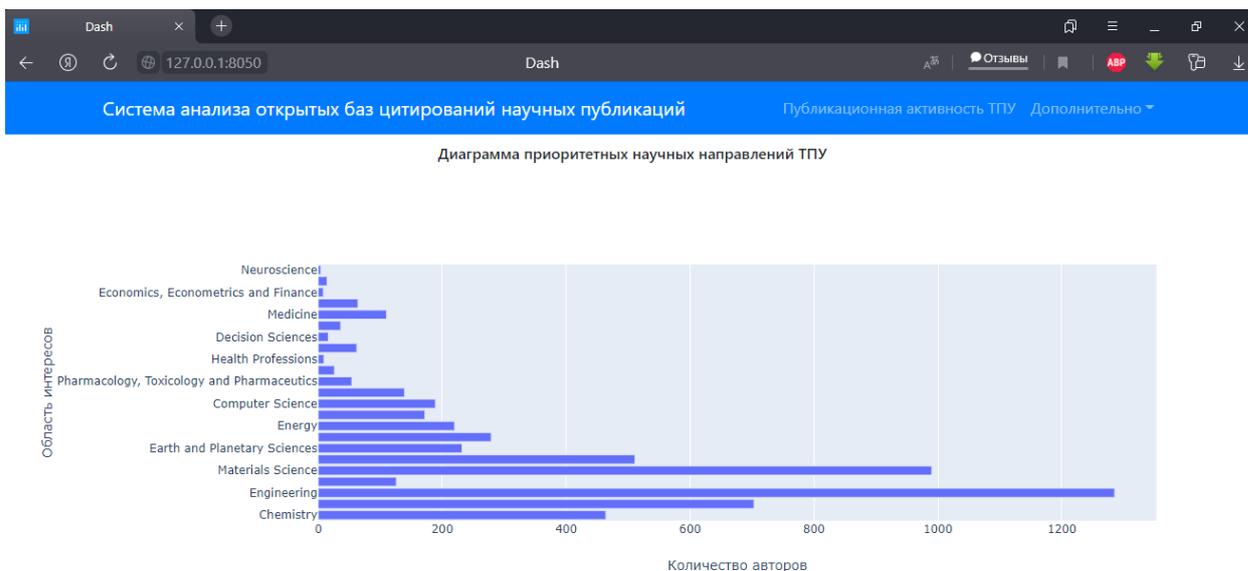


Рисунок 23 – Страница «Приоритетные направления»

На рисунке 24 представлена страница «Карта журналов», на которой отмечены местоположения журналов, в которых публикуются авторы ТПУ.

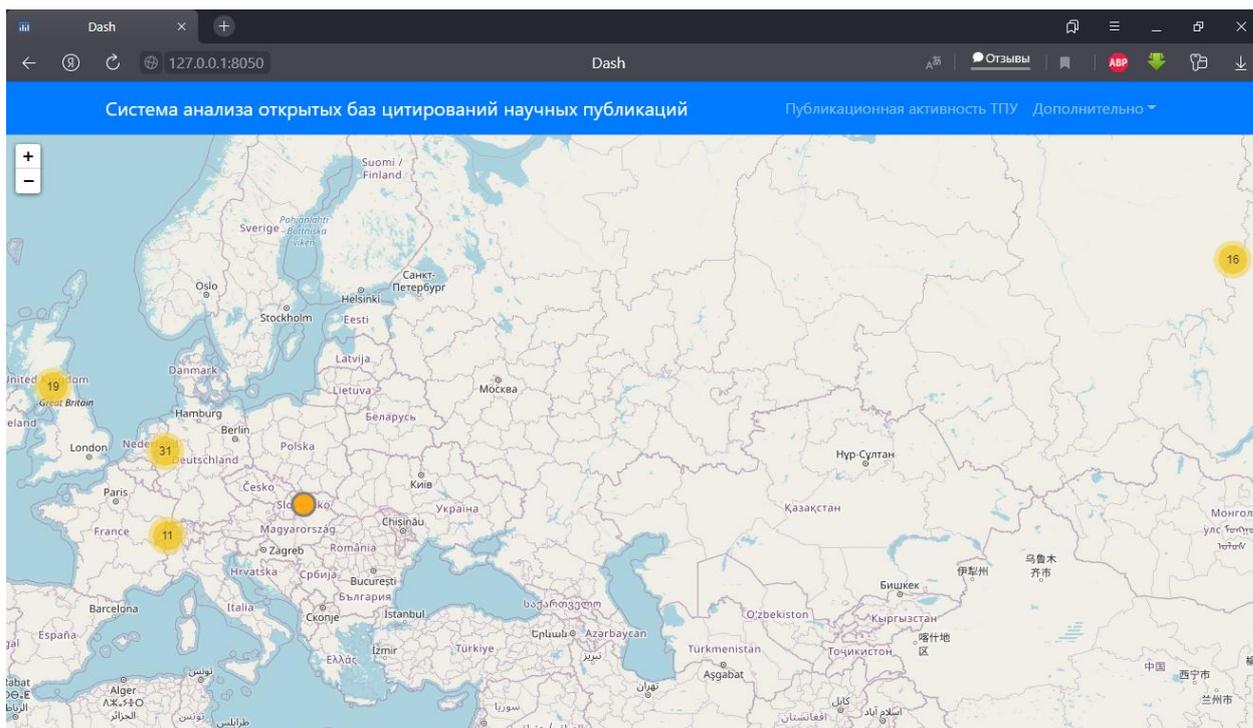


Рисунок 24 – Страница «Карта журналов»

Помимо обычного просмотра графиков и диаграмм, благодаря библиотеке plotly, можно использовать интерактивные элементы для управления просмотром графического элемента и получения более узкой выборки. На рисунке 25 представлены элементы для управления графиком и столбчатой диаграммой:

1. Download plot as a png – скачать сюжет в формате png;
2. Zoom, zoom in, zoom out – масштабирование, увеличить масштаб, уменьшить масштаб;
3. Pan – панорамирование;
4. Box select, Lasso select – инструмент прямоугольник, инструмент лассо;
5. Autoscale – автоматическое масштабирование;
6. Reset axes – сброс осей;
7. Toggle spike lines – переключение линий шипов;
8. Show closest data on hover, compare data on hover – показать ближайшие данные при наведении, сравнить данные при наведении.

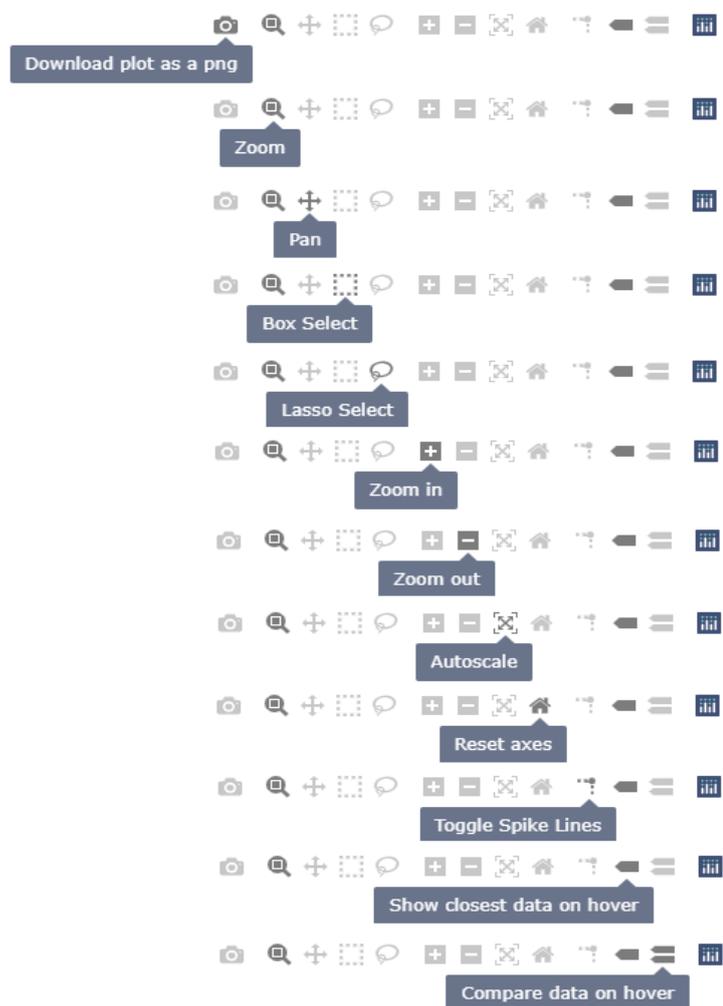


Рисунок 25 – Интерактивные элементы для управления просмотром графика и столбчатой диаграммы

Для круговой диаграммы возможно использование только двух интерактивных элементов (рисунок 26):

1. Download plot as a png – скачать сюжет в формате png;
2. Toggle show closest data on hover – переключить отображение ближайших данных при наведении курсора.



Рисунок 26 – Интерактивные элементы для управления просмотром круговой диаграммы

Помимо этого, на графиках с легендой есть возможность отключать ряды данных, что позволяет пользователю рассмотреть только интересующие

его данные и скомбинировать их на свое усмотрение. Пример данной функции представлен на рисунке 27.

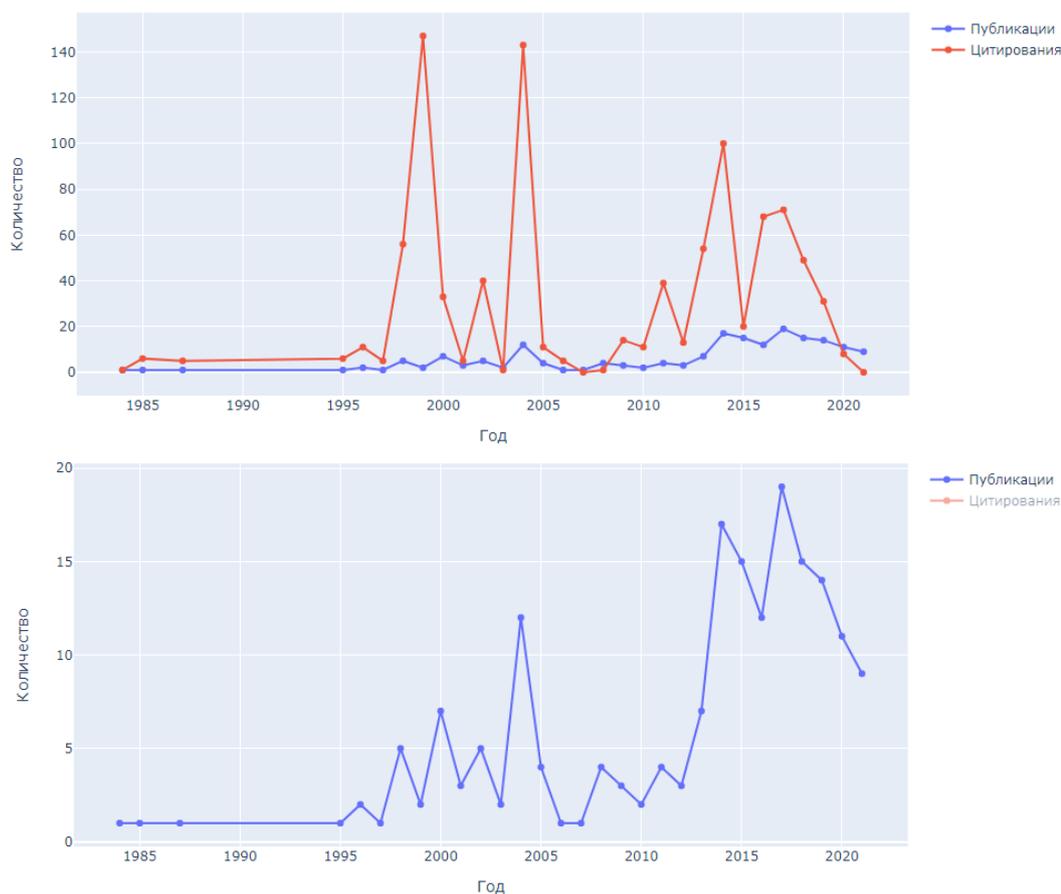


Рисунок 27 – Пример функции отключения ряда данных

### 3.6 Реализация функции интеграции с инфраструктурой научного сообщества

Для формирования статистики по отдельным подразделениям научного сообщества необходимо иметь возможность создавать объединения авторов и просматривать статистику не только по всей научной организации или отдельному автору, но и по созданным объединениям авторов.

В систему была добавлена возможность создавать объединения. На рисунке 28 представлена форма создания научного сообщества, для этого необходимо ввести название сообщества и выбрать авторов из выпадающего списка. После нажатия на кнопку «Создать» научное сообщество появится в

списке научных сообществ ТПУ, расположенном под формой на странице научных сообществ.

Создать научное сообщество

Геоинформационные системы и технологии

Abdiyaynov A. Adischev Y.N.

Создать

Рисунок 28 – Форма создания научного сообщества

Для проверки формирования статистических данных по сообществам были созданы объединения из авторов Томского политехнического университета. На рисунке 29 представлен список тестовых объединений авторов.

Номер	Наименование	Авторы	
1	Отделение химической инженерии	[Slepchenko G.B.]	Удалить
2	Исследовательская школа химических и биомедицинских технологий	[Di Martino A., Zinoviev A., Novikov V., Pershina A.]	Удалить
3	Центр промышленной томографии	[Janz J.]	Удалить
4	Научно-образовательный центр Н.М.Кизнера	[Kuznetsova A.S., Kuznetsova A.S., Vakalova T.V., Sharafiev S.M., Sergeev N.P.]	Удалить
5	Отделение автоматизации и робототехники	[Muravyov S.V., Konkin D.A.]	Удалить
6	Научная лаборатория высокоинтенсивной имплантации ионов	[Ryabchikov A.I., Dektyarev S.V., Ivanova A.I., Korneva O.S., Sivin D.O., Ivanova A.]	Удалить
7	Исследовательская школа физики высокоэнергетических процессов	[Shevelov A.E., Glushkov P.A., Belova A., Petrova A., Petrov M.]	Удалить
8	Научно-образовательный центр И.Н.Бутова	[Kuznetsov G.V., Kostoreva Z.A., Krainov D.A.]	Удалить
9	Отделение иностранных языков	[Kotorova E., Markova N.A., Aksenova N.V., Iuzhakova M.A.]	Удалить
10	Научно-производственная лаборатория "Импульсно-пучковых, электроразрядных и плазменных технологий"	[Iseberlinova A.A., Lapteva O., Remnev G.]	Удалить
11	Отделение нефтегазового дела	[Fadeeva S.V., Tsibulnikova M.R., Orlova Y.N.]	Удалить
12	Отделение ядерно-топливного цикла	[Pavlov V., Bespalov V.I., Vidyayev D., Lutsenko Y., Vergun A., Dorofeeva L., Stavitskaya K., Karengin A., Novoselov I.]	Удалить
13	Отдел развития публикационной активности	[Ryabkina A.]	Удалить
14	Отделение электронной инженерии	[Dikman E.Y., Bolotina I., Baranov P.]	Удалить
15	Отделение математики и информатики	[Mikhailchuk A.A., Arefyev K.]	Удалить
16	Научно-производственная лаборатория "Чистая вода"	[Ivanova M.P.]	Удалить
17	Отделение электроэнергетики и электротехники	[Izaev Y.N.]	Удалить
18	Инженерная школа энергетики	[Kropotova S.]	Удалить
19	Отделение общетехнических дисциплин	[Tomlin A.K.]	Удалить
20	Отделение геологии	[Orekhov A.N., Sokolov B.R.]	Удалить
21	Отделение экспериментальной физики	[Erofeeva G.V.]	Удалить
25	Юргинский технологический институт (филиал) ТПУ	[Kryukov A.V.]	Удалить
26	Отделение материаловедения	[Kuvshinov K.A., Petkar P., Lisitsyn V.M.]	Удалить
28	Международная научно-образовательная лаборатория неразрушающего контроля	[Koneva D.]	Удалить
29	Военный учебный центр	[Lenskii V.]	Удалить
30	Отдел студенческих объединений	[Gamov D.]	Удалить

Рисунок 29 – Список объединений авторов Томского политехнического университета

По полученным объединениям также строятся следующие статистики:

1. Статистика публикационной активности по объединению. Данная статистика хранит в себе общее число публикаций, общее число цитирований и среднее значение цитируемости по всему объединению за каждый год публикационной активности объединения.
2. Статистика по числу публикаций типа «статья» по объединению. Данная статистика хранит в себе процент числа публикаций, имеющих тип «статья», от общего числа публикаций на все

объединение за каждый год публикационной активности объединения.

3. Статистика по публикациям в различных источниках по отдельным объединениям. Данная статистика хранит в себе все объединения, журналы, где они публикуются, и число статей, опубликованных объединением в данном журнале за все время публикационной активности объединения.

### **3.7 Реализация функции построения графа популярных научных областей**

По каждому автору с базы Scopus извлекаются его области интересов. Для визуализации графа популярных научных областей после выгрузки данных с базы Scopus по каждой полученной области интересов подсчитываем число авторов, публикующихся в данной области. Помимо этого, для каждой пары областей интересов подсчитываем число общих авторов, интересующихся этими областями. Для получения описанных данных был составлен следующий запрос к базе данных:

```
SELECT sa1.id as sa1_id, sa1.name, sa2.id as sa2_id, sa2.name,  
(SELECT count(*) filter (where count_authors > 1 ) FROM (SELECT  
count(author_subject_areas.id) as count_authors, author_id FROM  
author_subject_areas where subject_area_id IN (sa1.id, sa2.id) group by  
author_subject_areas.author_id) sub)  
FROM subject_areas sa1 CROSS JOIN subject_areas sa2  
WHERE sa1.id != sa2.id  
GROUP BY sa1.id, sa1.name, sa2.id, sa2.name
```

Визуализированный граф популярных научных областей представлен на рисунке 30. Диаметр узла графа зависит от количества авторов в данной научной области, а ширина ребра зависит от числа общих авторов каждой пары областей интересов.

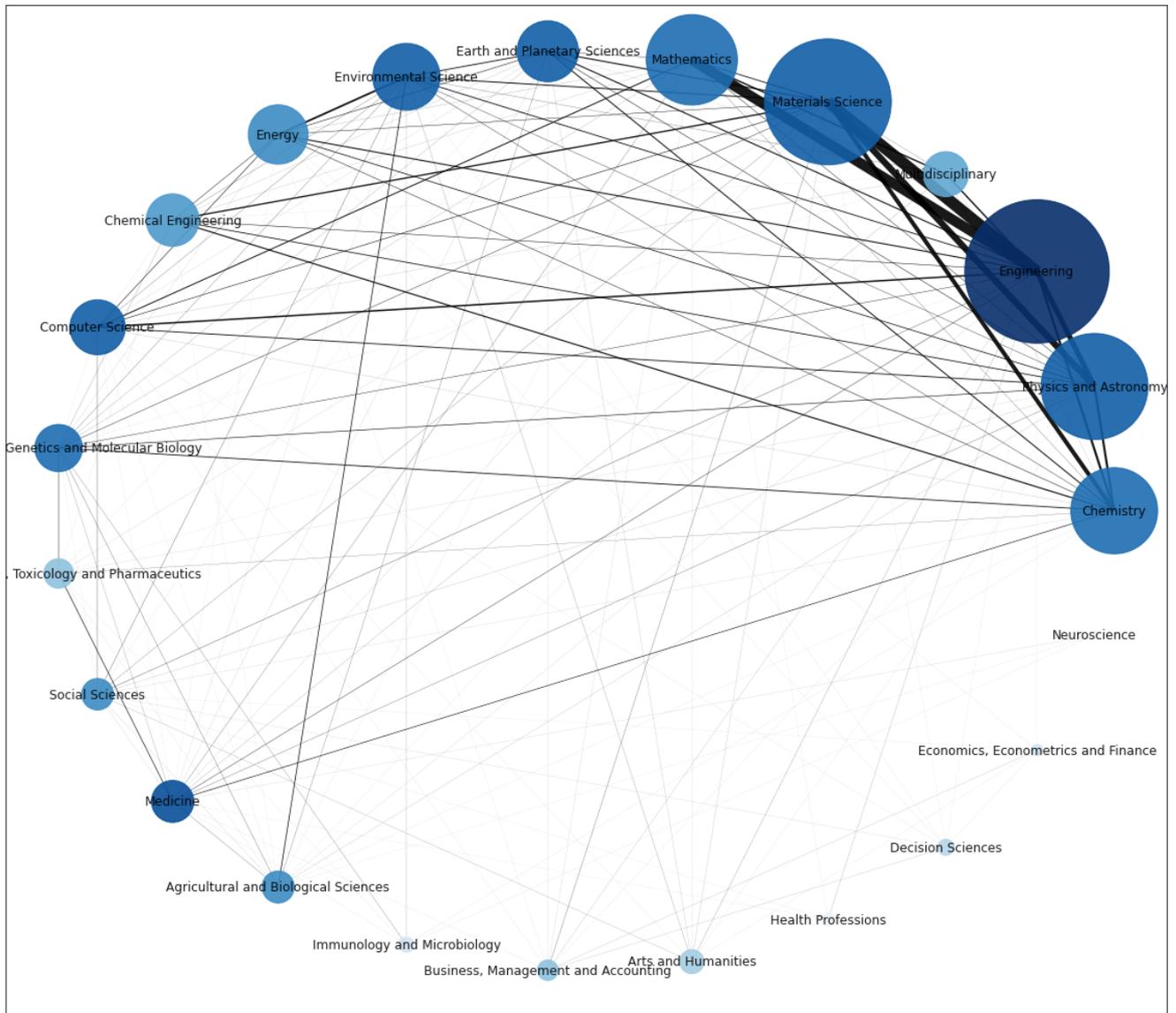


Рисунок 30 – Граф популярных научных областей

Построенный граф позволяет визуально оценить приоритетные научные направления, для количественной оценки строится столбчатая диаграмма сравнения научных областей по количеству авторов (рисунок 31).

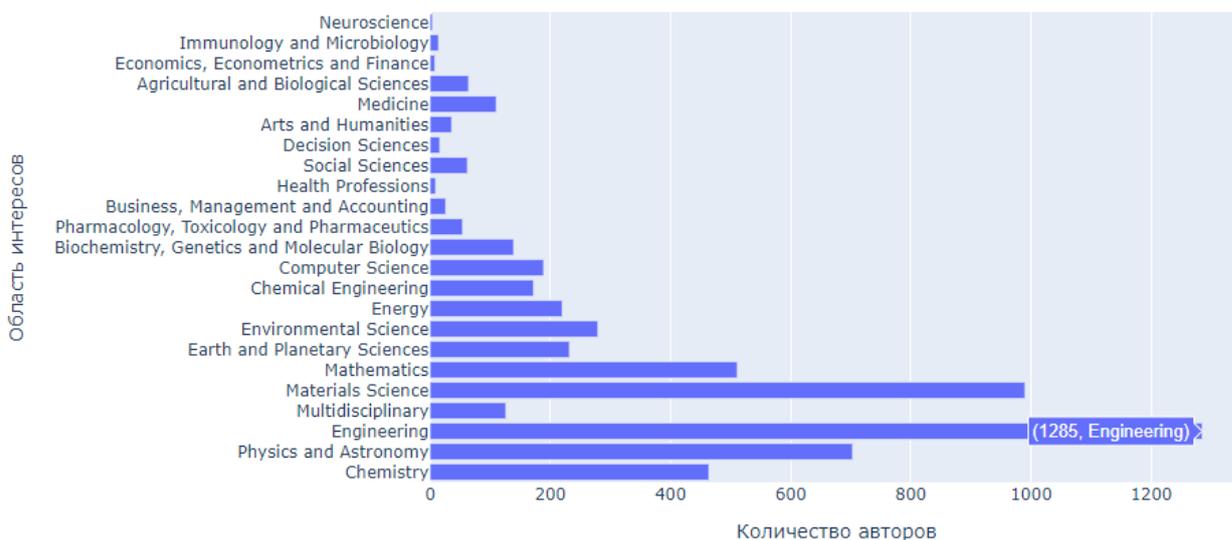


Рисунок 31 – Диаграмма количества авторов, публикующихся в научных областях

### 3.8 Реализация функции построения карты журналов с публикационной активностью

По всем публикациям, аффилированным с научной организацией, с библиографической базы Scopus выгружается информация о том, в каких журналах были опубликованы данные публикации. В рамках разрабатываемой системы было принято решение реализовать скрипт для визуализации географического местоположения журналов, в которых публикуются авторы научной организацией, с отображением количества публикации в журнале.

Данные о том, какой стране принадлежит тот или иной журнал, были получены с помощью сервиса Scimago Journal & Country Rank. Для построения карты использовалась библиотека Folium, которая позволила на основе полученных данных реализовать интерактивную карту журналов с указанием числа публикаций научной организации в этом журнале. Для установки маркеров с журналами были сохранены географические координаты (широта и долгота) каждой страны, располагающей издательства. Для создания карты запускается Python-скрипт и сохраняет карту в виде html-страницы.

На рисунке 32 представлен фрагмент описанной карты, построенной по данным о публикационной активности Томского политехнического университета.

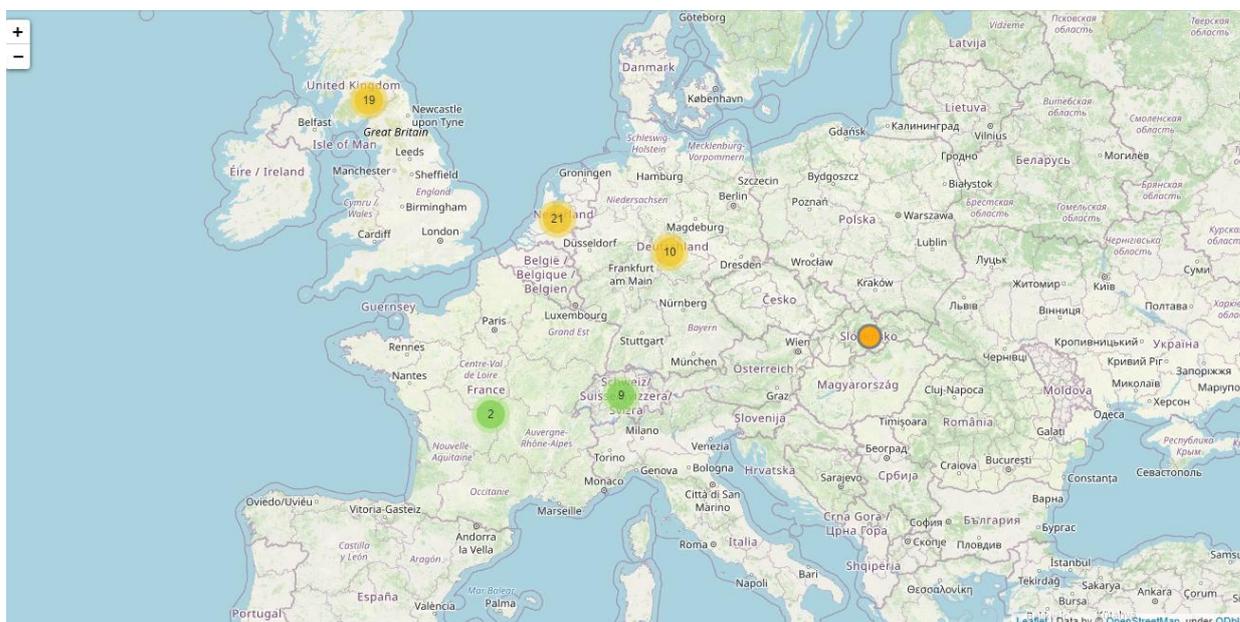


Рисунок 32 – Фрагмент карты публикационной активности организации в различных журналах

При увеличении масштаба карты маркер журнала увеличивается, при выборе маркера какой-либо страны появляются маркеры с наименованием журнала и количеством публикаций в нем. Маркер журнала имеет цвет соответственно числу публикаций научной организации в нем (зеленый – меньше 50, оранжевый – больше 50, но меньше 200 публикаций, а красный – более 200 публикаций). На рисунке 33 представлен фрагмент карты журналов, в которых публикуются авторы ТПУ, в Германии.

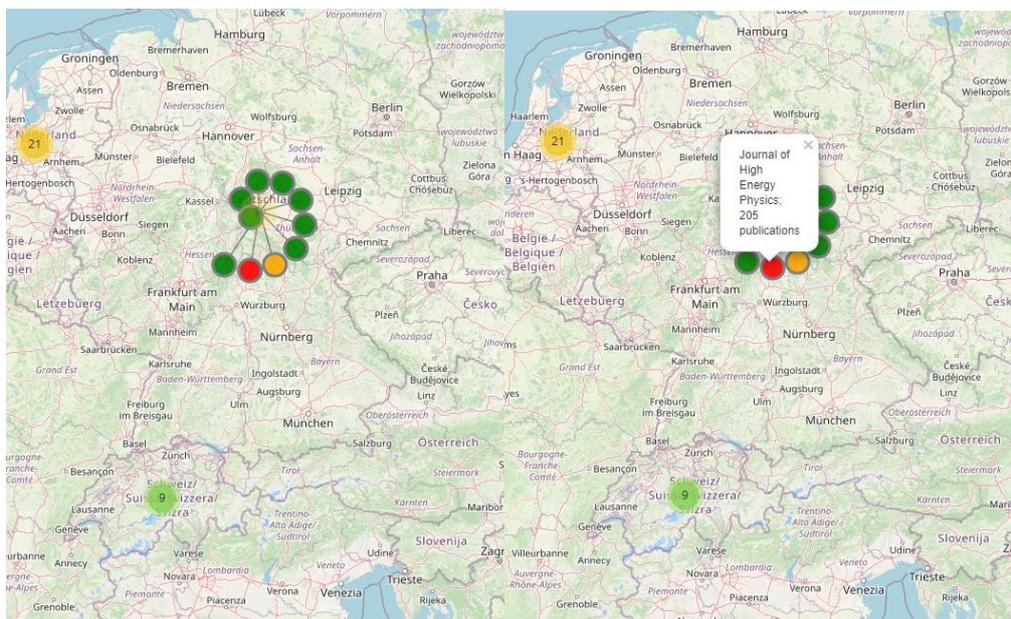


Рисунок 33 – Фрагмент карты публикационной активности организации в различных журналах с отображением маркеров

#### **4. Апробация системы анализа открытых данных баз цитирований научных публикаций**

##### **4.1. Апробация функций анализа научной активности организации**

Публикационная активность является важнейшим критерием оценки эффективности деятельности научных организаций, сообществ, авторов и научных журналов. В данной работе рассмотрим, по каким показателям и данным можно оценить публикационную активность научно-исследовательской деятельности Томского политехнического университета. Все эти данные отображены на веб-странице публикационной активности ТПУ.

На рисунке 34 представлена круговая диаграмма, на которой отображено процентное соотношение общего количества публикаций авторов университета. Согласно данной диаграмме, наибольшее число авторов имеют в среднем меньше 10 публикаций за время своей научной деятельности.

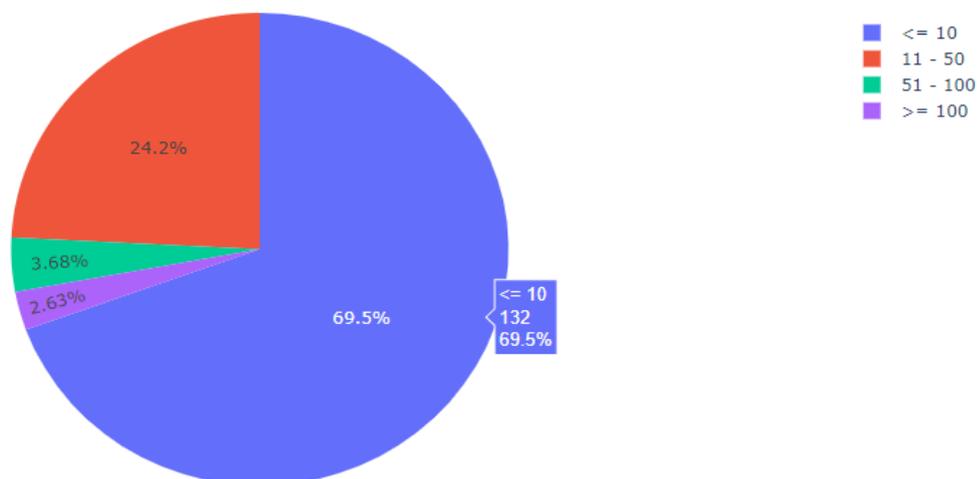


Рисунок 34 – Диаграмма статистики авторов университета по количеству публикаций

Процентное соотношение общего количества цитирований авторов университета отображено на круговой диаграмме на рисунке 35. Согласно полученным данным, наибольшее количество авторов имеют наименьшее количество цитирований или даже имеют нулевое значение данного показателя.

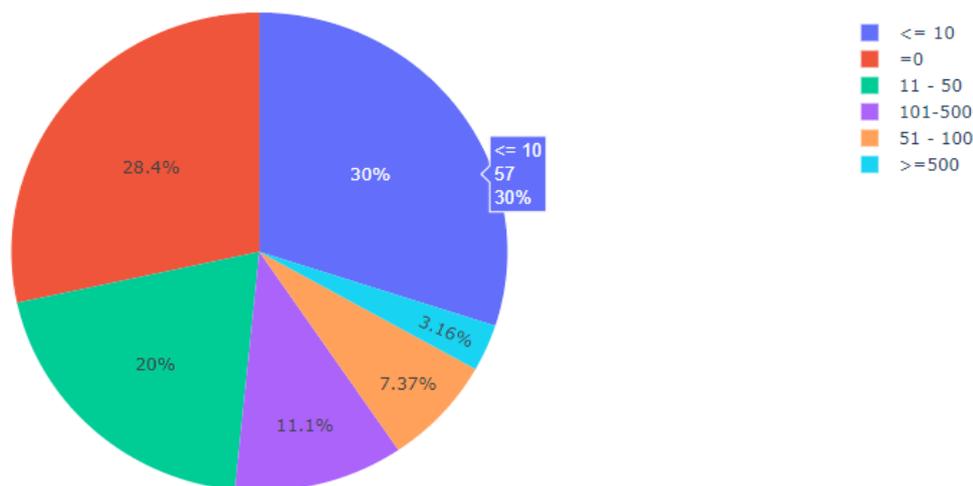


Рисунок 35 – Диаграмма статистики авторов университета по количеству цитирований

Преобладание авторов с числом публикаций меньше 10 и публикаций с числом цитирований меньше 10 можно объяснить тем, что в качестве исследуемой научной организации выбран университет, имеющий в списках

своих авторов большое число студентов, не продолжающих свою научную деятельность после завершения обучения.

Таким образом, при использовании описанных диаграмм научная организация может давать поверхностную оценку результативности научной деятельности всех своих авторов. Для более подробного анализа необходимо анализировать деятельность отдельных авторов или групп авторов.

Для отслеживания и количественной оценки авторов с наибольших числом публикаций за свою научную деятельность университета были построены столбчатые диаграммы. На рисунке 36 представлена диаграмма авторов университета с наибольшим количеством публикаций (более 100), а на рисунке 37 диаграмма авторов с количеством публикаций от 50 до 100.

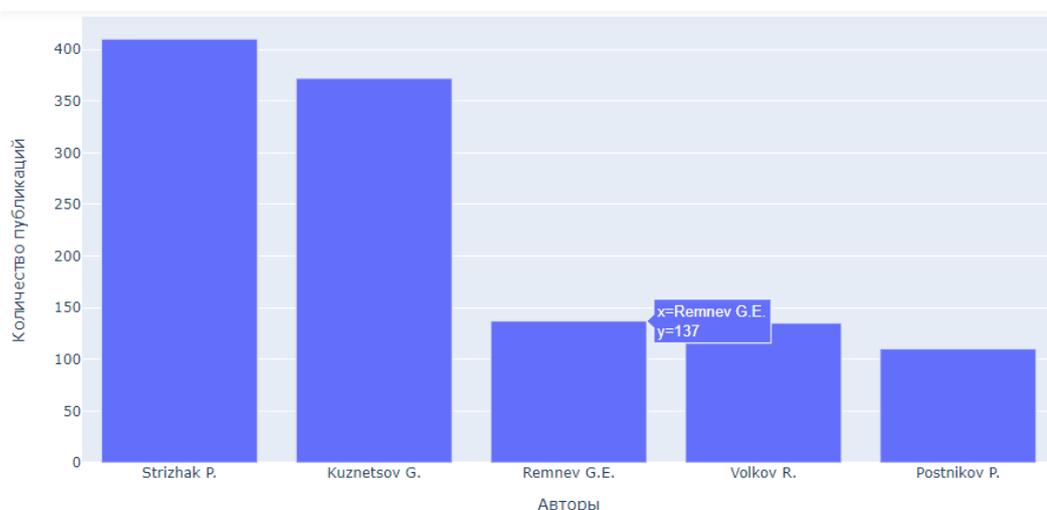


Рисунок 36 – Авторы университета с наибольшим количеством публикаций (более 100)

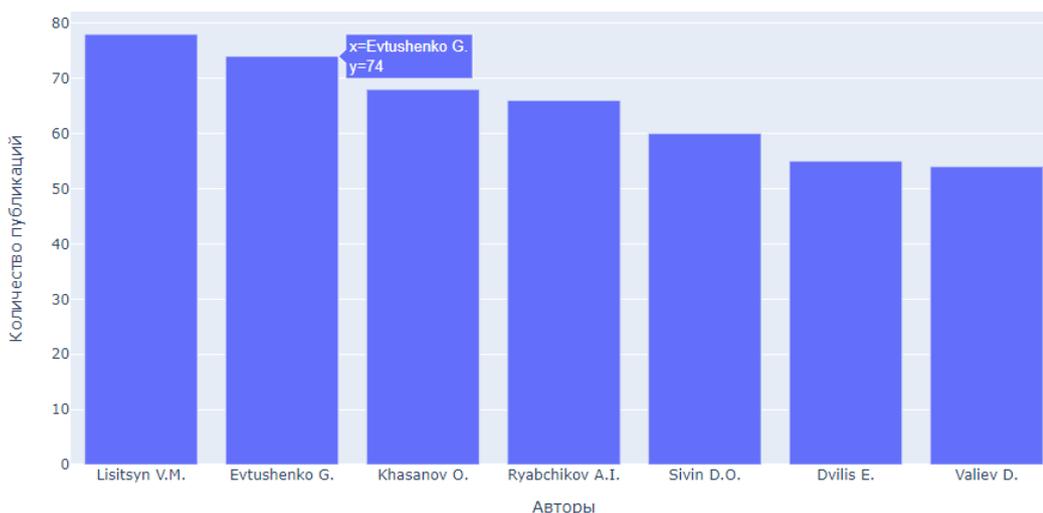


Рисунок 37 – Авторы университета с количеством публикаций от 50 до 100

Аналогично были построены диаграммы авторов с наибольшим количеством цитирований (более 500) на рисунке 38 и диаграмма авторов университета с количеством публикаций от 100 до 500 на рисунке 39.

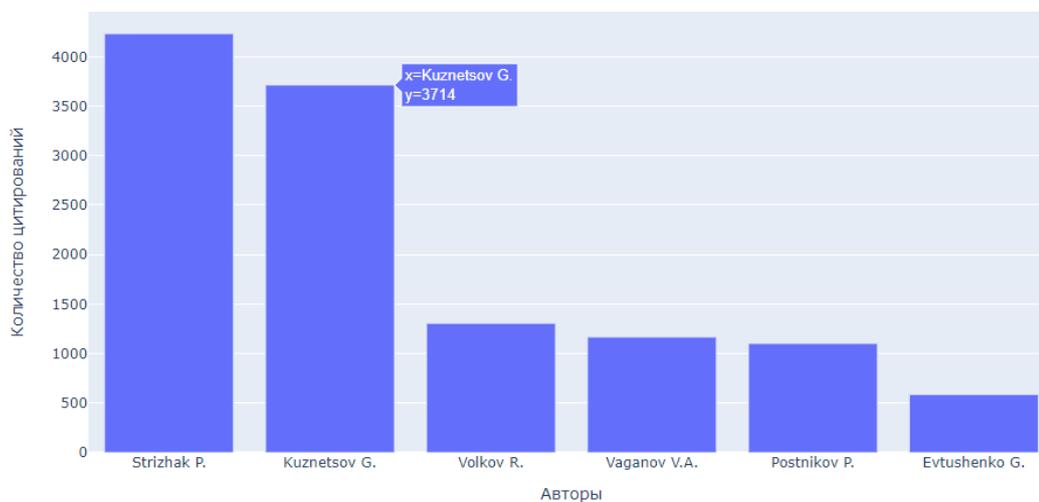


Рисунок 38 – Авторы университета с наибольшим количеством цитирований (более 500)

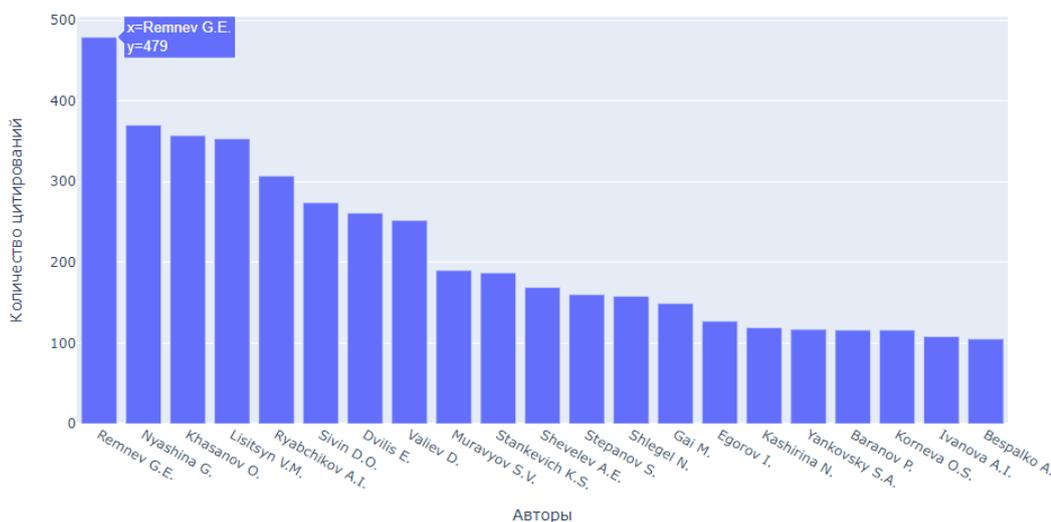


Рисунок 39 – Авторы университета с количеством публикаций от 100 до 500

Исходя из анализа сформированных диаграмм, можно выделить совпадения в авторах с наибольшим числом цитирований и наибольшим числом публикаций, что повышает показатели результативности данных авторов.

Статистика по типу публикаций хранит в себе информацию о процентном соотношении типов полученных публикаций. Публикации

разбиты на 5 типов: статья, глава, документ конференции, обзор и другое, если тип не указан в полученной публикации. На рисунке 40 представлена интерактивная круговая диаграмма, показывающая процентное и количественное соотношение публикаций по их типу. Из диаграммы видно, что на данный момент по имеющимся данным наибольшее количество составляют публикации типа статья.

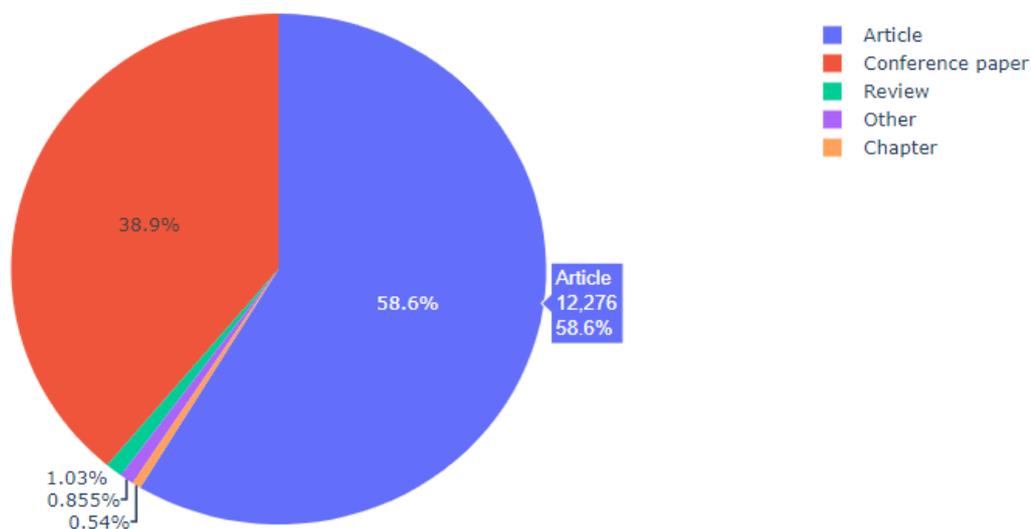


Рисунок 40 – Диаграмма статистики публикаций по типу

Статистика по типу издания хранит в себе информацию о процентном соотношении типов полученных изданий. Издания делятся на журналы и материалы конференции. На рисунке 41 представлена интерактивная круговая диаграмма, показывающая процентное и количественное соотношение изданий по их типу. Наибольшее количество изданий – это журналы.

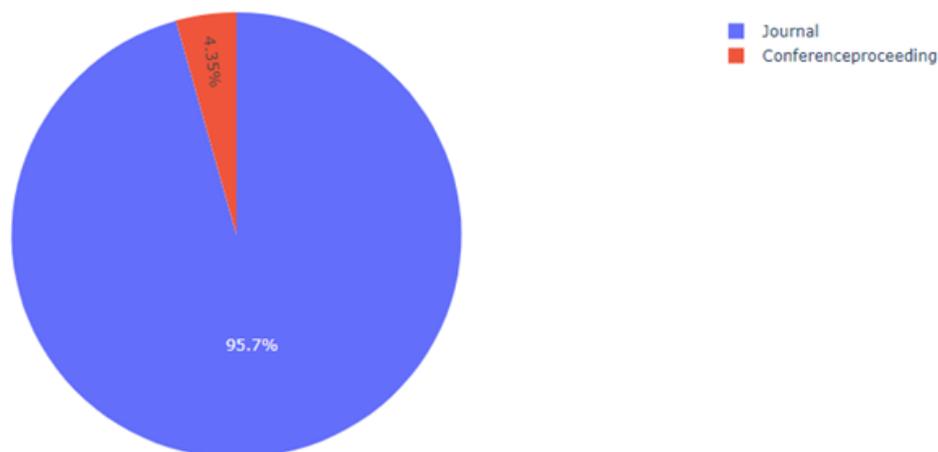


Рисунок 41 – Диаграмма статистики изданий по типу

По каждому автору с базы Scopus извлекаются значения индекса Хирша, что позволяет проводить анализ этого значения авторов в разных научных областях. Индекс Хирша призван учитывать не только количество статей и цитирований, но и распределение статей по количеству полученных цитирований.

На рисунке 42 представлена столбчатая диаграмма авторов с наибольшим индексом Хирша в научной области «Computer Science». Данная диаграмма может быть полезна для нахождения сотрудников с наибольшей эффективностью научной деятельности в рамках определенной научной деятельности, может использоваться отдельным автором для поиска партнеров.

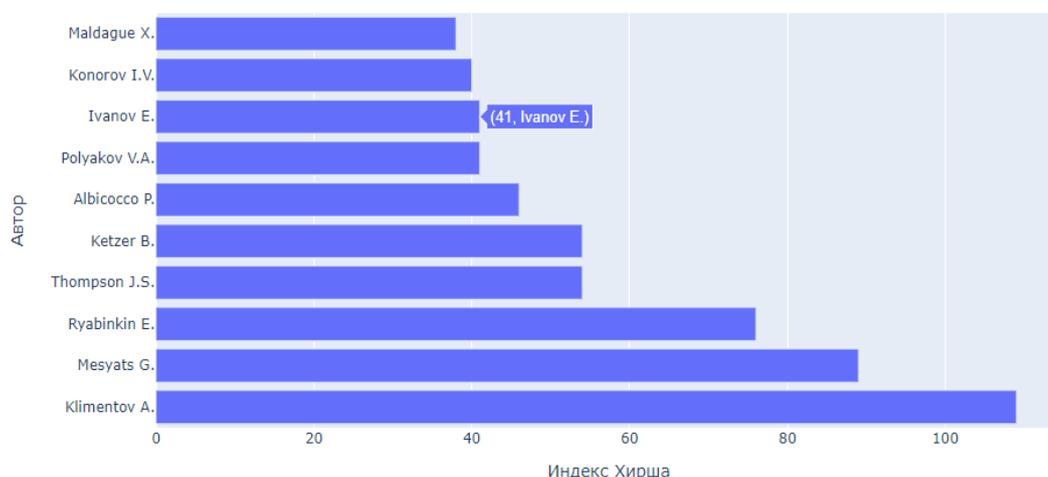


Рисунок 42 – Сравнение индекса Хирша авторов в научной области «Computer Science»

## 4.2. Апробация функций анализа научной активности журнала

Публикационная активность журнала представлена данными: название, страна, индекс Хирша, график изменения индекса SJR по годам.

Все эти данные отображены на веб-странице издания. Также на карте журналов с публикационной активностью можно увидеть число публикаций научной организации (рисунок 32).

SJR – это мера научного влияния журналов, которая учитывает как количество цитат, полученных журналом, так и важность или престиж журналов, где такие цитаты исходят от него, измеряет научное влияние средней статьи в журнале, выражает, насколько центральной для глобальной научной дискуссии является средняя статья журнала.

Таким образом, можно анализировать то, какую меру научного влияния имеют журналы, в которых наиболее часто публикуются авторы научной организации, отдельного научного сообщества внутри организации или отдельные авторы организации.

По каждому журналу строится график изменения индекса SJR по годам, на рисунке 43 представлен пример графика для российского журнала «Russian Engineering Research», а на рисунке 44 – для немецкого журнала «Electrical Engineering».

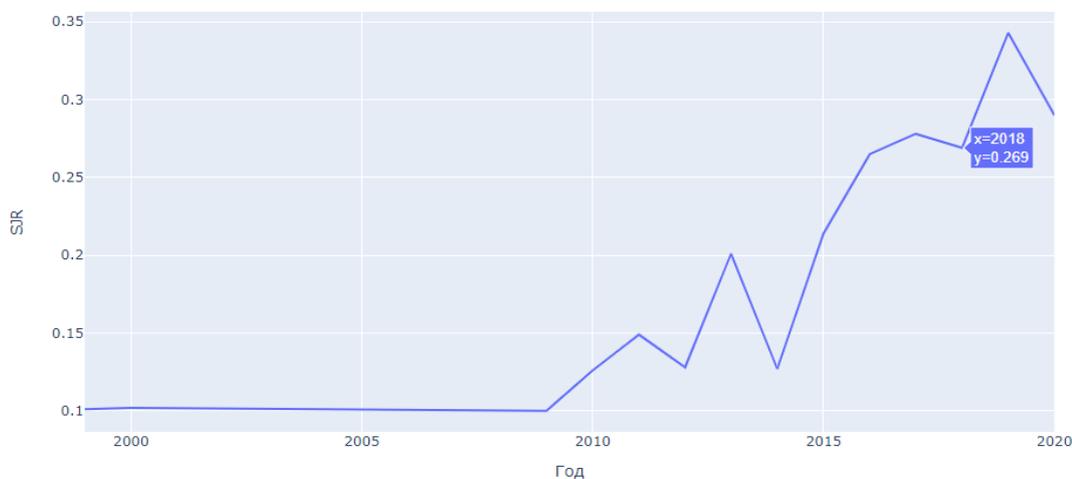


Рисунок 43 – Изменение индекса SJR по годам журнала «Russian Engineering Research»

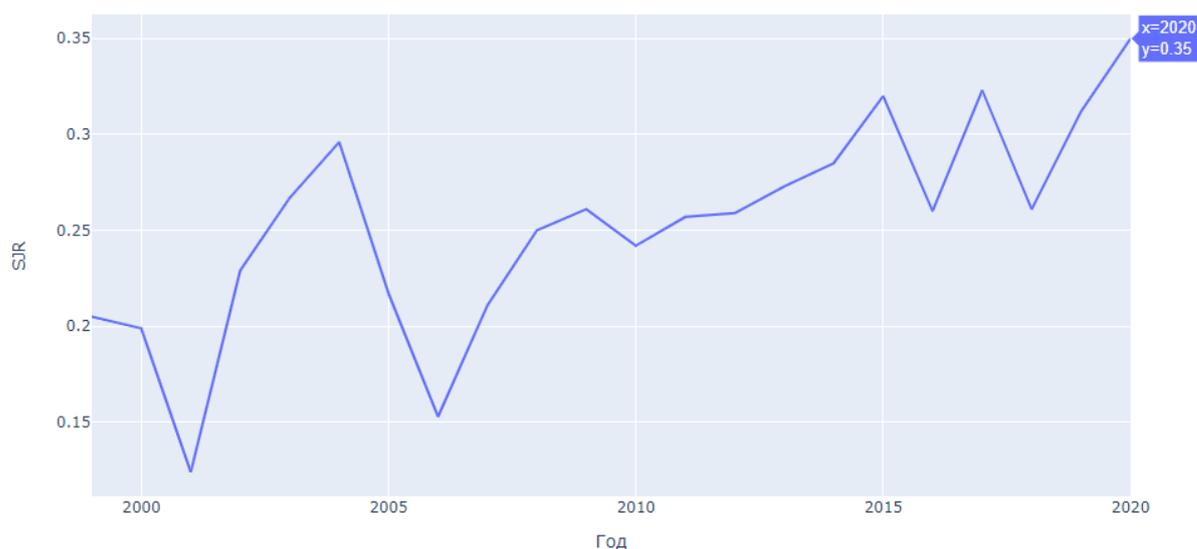


Рисунок 44 – Изменение индекса SJR по годам журнала «Electrical Engineering»

#### 4.3. Апробация функций анализа научной активности научного сообщества

Проводя исследование построенных графиков и диаграмм на примере Томского политехнического университета, было выявлено, что оценка публикационной активности по всему списку авторов, связанных с университетом, не дает полного понимания состояния научной деятельности внутри организации. Для решения данной проблемы реализован модуль интеграции с инфраструктурой научной организации, позволяющий объединять авторов в сообщества внутри организации и получать статистику уже по множеству авторов, принадлежащих сообществу.

На рисунках 45, 46 и 47 представлены статистики по созданному сообществу из 3х сотрудников Томского политехнического университета. Диаграмма на рисунке 45 позволяет отметить сотрудников, имеющих наибольший вклад в развитие научной деятельности сообщества.

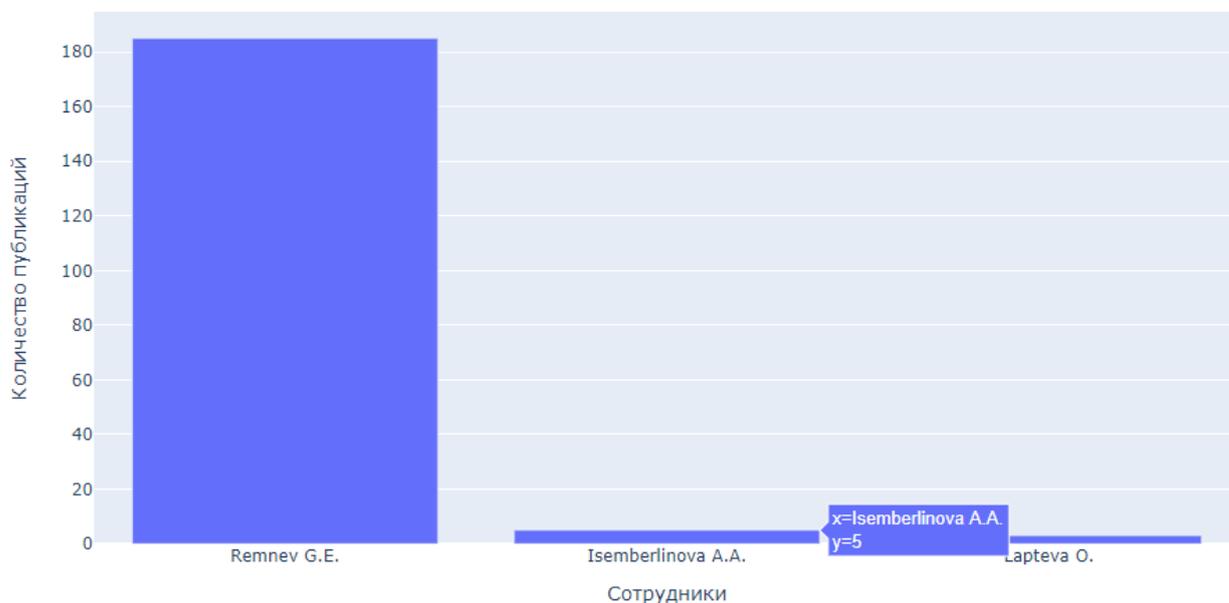


Рисунок 45 – Диаграмма сравнения общего количества публикаций сотрудников

На рисунке 46 представлен график состояния публикационной активности сообщества на протяжении всего времени активности. Таким образом, можно отследить периоды, когда научная деятельность сообщества требует поддержки для поддержания или повышения уровня результативности.

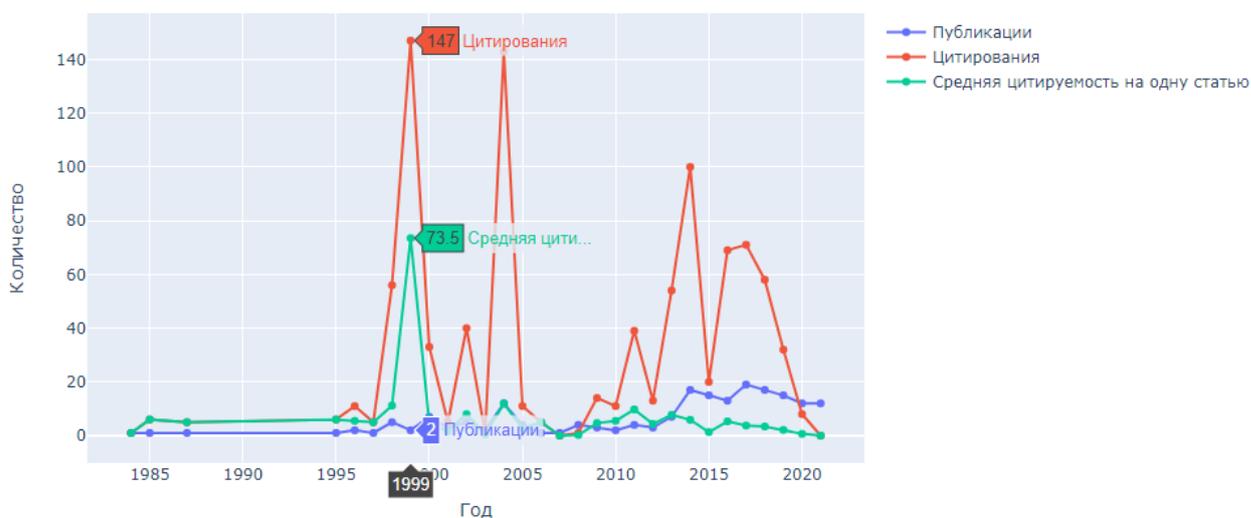


Рисунок 46 – График публикационной активности сообщества

На рисунке 47 представлена диаграмма публикационной активности научного сообщества в журналах, которые высоко котируются в

международной научной среде и индексируются крупнейшими мировыми реферативными базами.



Рисунок 47 – Диаграмма публикационной активности сообщества в журналах (количество публикаций более 5)

#### 4.4. Апробация функций анализа научной активности автора

В ситуациях, когда статистики по научному сообществу недостаточно, имеется возможность провести анализ научной деятельности по отдельному автору. Данный функционал продемонстрирован на примере сотрудника ТПУ.

График публикационной активности на рисунке 48 показывает то, каким образом меняется число публикаций и число цитирований автора на протяжении его научной деятельности. Можно отметить, что на период с 2010 до 2015 года число публикаций в год от автора выросло в 2 раза, наиболее цитируемые публикации пришлись на 1999 год. Также на графике отмечена средняя цитируемость на одну статью, что позволяет оценить наиболее влиятельные статьи автора.

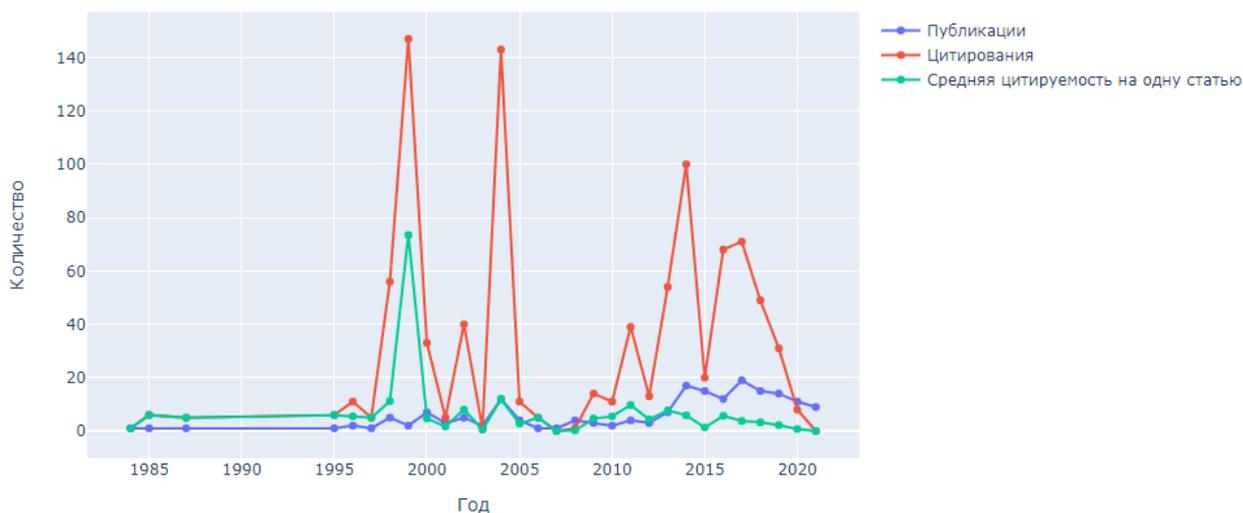


Рисунок 48 – График публикационной активности автора

На рисунке 49 представлен фрагмент таблицы с публикациями автора, которая содержит следующую информацию по публикациям: дата публикации, название публикации, тип публикации и количество цитирований. Такое представление публикаций автора позволяет посмотреть наиболее цитируемые публикации и по дате, отраженной на графике, посмотреть темы наиболее влиятельных из них.

Дата публикации	Название публикации	Тип публикации	Кол-во цитирований
15.12.2016	Flexible hard Al-Si-N films for high temperature operation	Article	28
05.08.2017	Atmospheric pressure plasma assisted immobilization of hyaluronic acid on tissue engineering PLA-based scaffolds and its effect on primary human macrophages	Article	26
01.04.2011	Optical properties of GaAs films deposited via pulsed ion ablation	Article	19
01.01.2016	Effect of energy on the formation of flexible hard Al-Si-N films prepared by magnetron sputtering	Article	16
24.10.2013	Pulse plasma-chemical synthesis of ultradispersed powders of titanium and silicon oxide	Article	15
01.01.2014	Progress in plasma focus research at the Kurchatov Institute	Conference Paper	15
15.08.2014	Formation of nanoscale carbon structures in the surface layer of metals under the impact of high intensity ion beam	Article	15
20.08.2013	A high-repetition rate pulsed electron accelerator	Article	13
01.01.2014	Characteristics of pulsed plasma-chemical synthesis of silicon dioxide nanoparticles	Article	13
15.12.2011	Plasma-chemical synthesis of composite nanodispersed oxides	Article	12
01.12.2012	A high-repetition rate pulsed electron accelerator	Conference Paper	12
01.01.2013	The Astra repetitive-pulse electron accelerator	Article	12
01.01.2014	Physicomechanical properties of the surface of a zirconium alloy modified by a pulsed ion beam	Article	12

Рисунок 49 – Фрагмент таблицы с наиболее цитируемыми публикациями автора

Так как у автора большое количество публикаций, на диаграмме представлены журналы, в которых количество публикаций автора более 5 (рисунок 50).



Рисунок 50 – Диаграмма публикационной активности автора в журналах

Помимо этого, автор входит в список с наибольшим количеством публикаций (рисунок 36) и в список с количеством цитирований от 100 до

500 (рисунок 39). Также была построена диаграмма сравнения индекса Хирша автора Remnev G.E. с авторами ТПУ (рисунок 51).

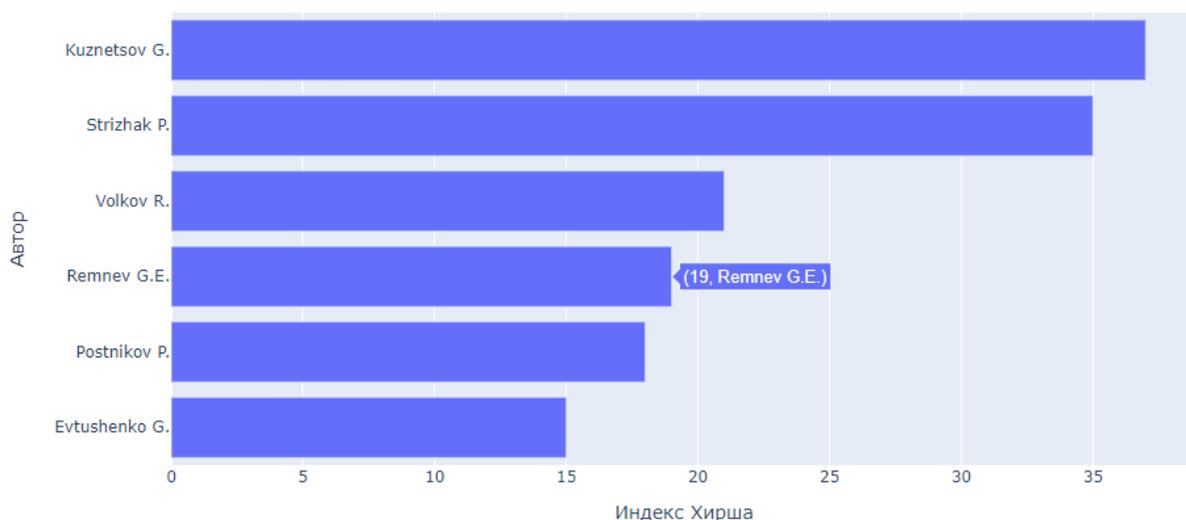


Рисунок 51 – Диаграмма сравнения индекса Хирша автора с авторами ТПУ

Данные диаграммы показывают, что автор вносит наибольший вклад в научную деятельность организации, а его научная деятельность результативна.

Помимо визуализации статистик был построен граф поиска партнеров по данному автору. Ширина ребра зависит от количества общих публикаций между автором и его соавтором. На рисунке 52 представлен граф поиска партнеров по соавторам автора.





Рисунок 53 – Граф партнеров автора (число общих публикаций более 10)

Таким образом, данные графики и диаграммы позволяют оценить продуктивность и качество научной деятельности автора, а граф показывает, что автор имеет большое число научных сотрудничеств, что говорит о его востребованности в научных кругах.

#### 4.5. Заключение по разделу

В процессе выполнения работы была спроектирована, разработана и апробирована система обеспечения информационной базы процессов принятия управленческих решений для развития научной деятельности.

В качестве научной организации для апробации работы системы был выбран Томский политехнический университет, это позволило провести оценку использования системы на реальной научной организации. В ходе

исследования данных по университету были выявлены такие особенности, как большое число авторов с числом публикаций, не превышающим 10, что говорит о специфике организации. На примере университета подчеркнута необходимость модуля интеграции с инфраструктурой организации, позволяющего исследовать научную активность не всех авторов, а наиболее интересных, активно занимающихся научной деятельностью. Таким образом, разработанная система позволяет производить оценку деятельности всей организации, сообществ авторов или отдельных авторов внутри нее, предоставляет информационное обеспечение для стратегического планирования и просмотра результатов принятых управленческих решений в виде статистических данных. Данную систему в представленном виде можно использовать научными организациями, поскольку она дает возможность изучить все основные наукометрические показатели, при необходимости возможно расширить предоставляемые статистические данные и источники данных. В сравнении с мировыми известными аналогами разработанная система имеет преимущества для российских научных организаций, предоставляя возможность сбора данных с любой открытой библиографической базы, имеющей API.

## **5. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение**

Разработка системы анализа открытых данных баз цитирований научных публикаций в рамках выполнения магистерской диссертации предполагает ряд затрат на исследование и разработку. Целью данного раздела является планирование работ, затрат на разработку и анализ сроков окупаемости разрабатываемой системы.

### **5.1. Предпроектный анализ**

#### **5.1.1. Потенциальные потребители результатов исследования**

Цитируемость научных публикаций российских ученых за последние 10 лет выросла вдвое, но остается ниже общемирового уровня по данным Института статистических исследований и экономики знаний (ИСИЭЗ) НИУ ВШЭ [17]. Это приводит к необходимости принятия управленческих решений в сфере развития научной деятельности российских организаций.

Разрабатываемая система направлена на предоставление конечному пользователю информационной базы для поддержки процессов принятия решений в области управления научной деятельностью. В качестве информационной базы предоставляются статистики по текущей публикационной активности авторов в виде интерактивных графов, графиков, диаграмм и карт науки. Основными заказчиками данной системы будут выступать научные организации, научные сообщества и возможно отдельные авторы. Помимо возможности производить оценку научной деятельности авторов, система предоставляет функционал мониторинга за наукометрическими показателями журналов. Таким образом, отдельной категорией заказчиков могут быть представители различных издательств.

### **5.1.2. Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения**

Для проведения анализа конкурентных технических решений было выбрано 2 из наиболее популярных продуктов:

- Интегрированная модульная платформа SciVal (K1): система позволяет анализировать массив публикаций, проиндексированных в Scopus, сопоставлять результативность по 230 странам, 12 600 исследовательским организациям, а также отдельным авторам.

- Инструмент для оценки научной деятельности InCites (K2): модуль позволяет анализировать массив публикаций Web of Science, сопоставлять научную продуктивность разных стран, организаций, авторов, выявлять публикационные тренды в разных областях науки, оценивать результативность финансирующих исследования фондов.

В качестве критериев для проведения анализа конкурентных решений выбраны:

- Охват различных баз источников данных: критерий показывает на наличие разных (в том числе и российских) библиографических баз в качестве источников данных или возможность их подключения.

- Интеграция с инфраструктурой организации: критерий определяет возможность воспроизведения в системы инфраструктуры организации и проведения анализа по отдельным составляющим инфраструктуры.

- Экспорт данных в формат Excel: критерий формируется, исходя из наличия возможности экспортировать статистические данные в отчеты формата Excel.

- Функциональные возможности: под критерием подразумевается объем предоставляемых функциональных возможностей для проведения анализа научной деятельности.

- Стоимость: чем ниже стоимость разработки, тем выше ее

конкурентоспособность по данному критерию.

– Мощность клиентской базы: критерий показывает количество клиентов, пользующихся разработкой.

– Техническая поддержка: по критерию можно понять доступность технической поддержки системы в случае, если у клиентов возникнут вопросы или пожелания по улучшению.

Таблица 1 – Оценочная карта для сравнения конкурентных технических решений (разработок)

Критерии оценки	Вес критерия	Баллы			Конкурентоспособность		
		Бф	Бк1	Бк2	Кф	Кк1	Кк2
1	2	3	4	5	6	7	8
<b>Технические критерии оценки ресурсоэффективности</b>							
1. Охват различных баз источников данных	0.25	3	2	2	0.75	0.5	0.5
2. Интеграция с инфраструктурой организации	0.15	5	4	3	0.75	0.6	0.45
3. Экспорт данных в формат excel	0.05	5	5	5	0.25	0.25	0.25
4. Функциональные возможности	0.25	3	5	5	0.75	1.25	1.25
<b>Экономические критерии оценки эффективности</b>							
5. Стоимость	0.15	5	2	2	0.75	0.3	0.3
6. Мощность клиентской базы	0.05	2	5	5	0.1	0.25	0.25
7. Техническая поддержка	0.1	5	4	4	0.5	0.4	0.4
<b>Итого</b>	1				3.85	3.55	3.4

Исходя из выбранных критериев и их весов, можно отметить, что наиболее важными при анализе считаются: охват различных баз источников данных, функциональные возможности, стоимость и интеграция с

инфраструктурой сообщества. Приведенный анализ конкурентоспособности показал, что разрабатываемый модуль обладает высокой конкурентоспособностью, поскольку имеет высокие показатели по наиболее важным для решения поставленной задачи факторам.

### 5.1.3. FAST-анализ

В рамках данной магистерской диссертации в качестве объекта FAST-анализа выступает система анализа данных открытых баз цитирований. Главной функцией системы является обеспечение информационной базы для процесса принятия управленческих решений в области развития научной деятельности организации. В состав системы входят следующие модули: модуль сбора данных с библиографических баз, модуль формирования статистик, модуль визуализации статистики, модуль интеграции с инфраструктурой научной организации и модуль экспорта статистики. В таблице 2 представлены все функции, выполняемые данными модулями.

Таблица 2 – Классификация функций, выполняемых объектом исследования

Наименование модуля	Выполняемая функция	Ранг функции		
		Главная	Основная	Вспомогательная
Модуль сбора данных с библиографических баз	Обеспечение системы исходными данными		×	
Модуль формирования статистики	Формирование статистики на основе исходных данных		×	
Модуль визуализации статистик	Визуализация сформированных статистических данных	×		
Модуль интеграции с инфраструктурой научной организации	Воспроизведение инфраструктуры организации внутри системы и отображение статистики по отдельным элементам инфраструктуры			×

Модуль экспорта статистики	Экспорт статистики в отчеты формата Excel			×
----------------------------	---	--	--	---

В таблице 3 представлена матрица смежности для всех выделенных функций. В основу данного метода положено расчетно-экспертное определение значимости каждой функции.

Таблица 3 – Матрица смежности

	Ф1	Ф2	Ф3	Ф4	Ф5
Ф1	=	>	=	>	>
Ф2	<	=	<	>	>
Ф3	=	>	=	>	>
Ф4	<	<	<	=	>
Ф5	<	<	<	<	=

Следующим этапом является преобразованием матрицы смежности в матрицу количественных соотношений функций. Матрица количественных соотношений представлена в таблице 4.

Таблица 4 – Матрица количественных соотношений функций

	Ф1	Ф2	Ф3	Ф4	Ф5	Итого	Относительная значимость
Ф1	1	1.5	1	1.5	1.5	6.5	0.26
Ф2	0.5	1	0.5	1.5	1.5	5	0.2
Ф3	1	1.5	1	1.5	1.5	6.5	0.26
Ф4	0.5	0.5	0.5	1	1.5	4	0.16
Ф5	0.5	0.5	0.5	0.5	1	3	0.12
$\Sigma =$						25	1

Таблица 5 представляет собой расчет себестоимости всех перечисленных выше функций. В ситуации разработки программного обеспечения себестоимость включает в себя расходы на выплату заработной

платы разработчикам в зависимости от трудоемкости разрабатываемых функций.

Таблица 5 – Определение стоимости функций, выполняемых объектом исследования

Наименование модуля	Выполняемая функция	Трудоемкость модуля, нормо-ч	Заработная плата, руб.	Себестоимость, руб.	Относительная себестоимость
Модуль сбора данных с библиографических баз	Обеспечение системы исходными данными (Ф1)	504	47 880	47 880	0.41
Модуль формирования статистики	Формирование статистики на основе исходных данных (Ф2)	200	19 000	19 000	0.16
Модуль визуализации статистики	Визуализация сформированных статистических данных (Ф3)	504	47 880	47 880	0.41
Модуль интеграции с инфраструктурой научной организации	Воспроизведение инфраструктуры организации внутри системы и отображение статистики по отдельным элементам инфраструктуры (Ф4)	16	1 520	1 520	0.01
Модуль экспорта статистики	Экспорт статистики в отчеты формата Excel (Ф5)	16	1 520	1 520	0.01
$\Sigma =$				117 800	1

Информация об относительной себестоимости и относительной значимости выделенных функций используется для построения функционально-стоимостной диаграммы, изображенной на рисунке 54.

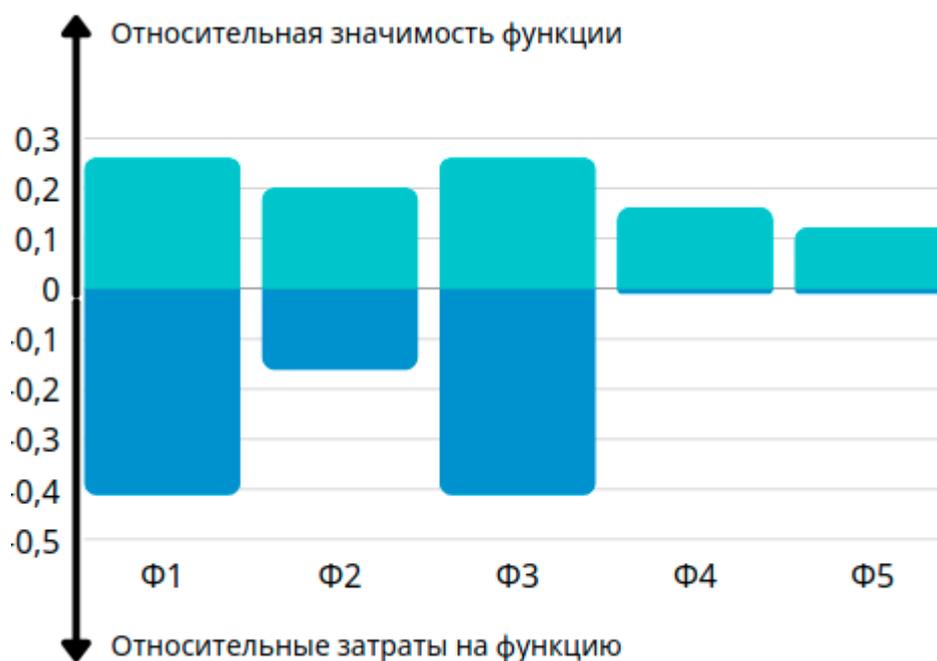


Рисунок 54 – Функционально-стоимостная диаграмма

Согласно полученной диаграмме, можно отметить небольшой дисбаланс по функциям 1 и 3. Для ликвидации данных пропорций можно переработать и упростить архитектуру системы, найти библиотеки для ускорения процесса разработки, повысить квалификацию разработчиков.

#### 5.1.4. Диаграмма Исикава

С помощью диаграммы Исикавы можно найти причину рассматриваемой в рамках работы проблемы. В данной работе в качестве проблемы рассматривается неэффективная научная деятельность организации. Данная проблема возникает в ситуациях низкого числа новых публикаций или же высокого числа публикаций с низким индексом цитирования. Диаграмма Исикавы по изучаемой проблеме представлена на рисунке 55.

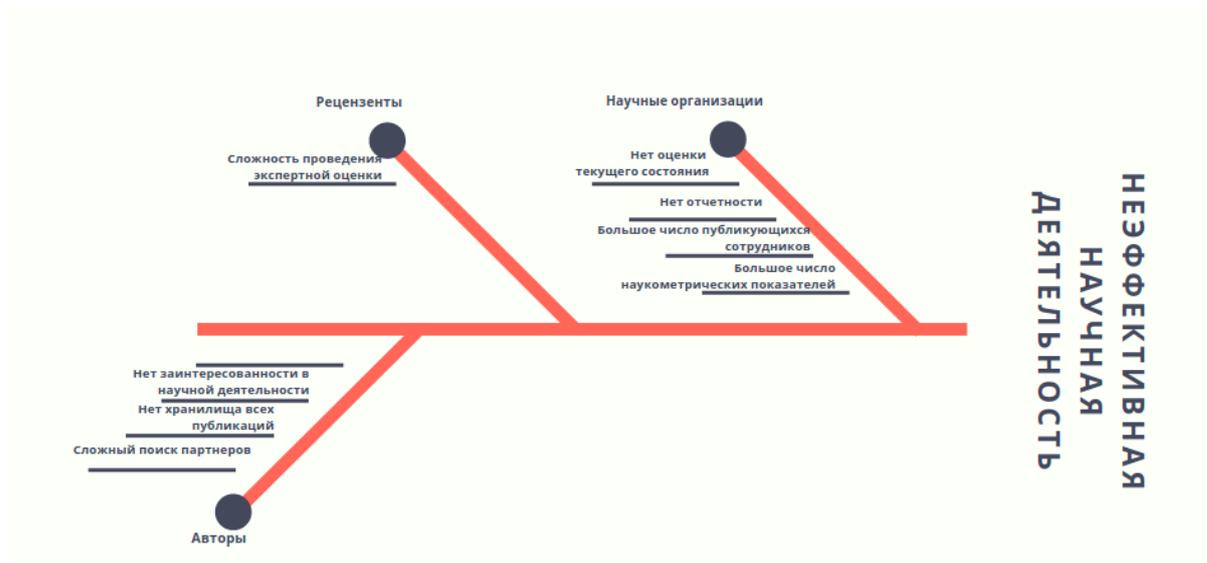


Рисунок 55 – Диаграмма Исикава

Анализ выявленных групп причин показал, что для решения поставленной проблемы необходимо предоставление инструментов оценки и формирования отчетности по научной деятельности организации, опирающихся на информацию обо всех публикациях и наукометрических показателях. В рамках данной диссертации описывается проектирование и разработка системы для выполнения поставленных задач.

### 5.1.5. SWOT-анализ

SWOT-анализ представляет собой комплексный анализ научно-исследовательского проекта и применяется для исследования внешней и внутренней среды проекта. Полученная SWOT-матрица представлена в таблице 6.

Таблица 6 – SWOT-анализ

		Внутренние факторы	
		<p><b>Сильные стороны:</b></p> <p>С1. Возможность расширения системы дополнительными модулями для интеграции с другими базами цитирований</p> <p>С2. Интеграция с инфраструктурой организации</p> <p>С3. Наличие экспорта отчетов</p>	<p><b>Слабые стороны:</b></p> <p>Сл1. Недостаток человеческих ресурсов и ресурсов времени</p> <p>Сл2. Отсутствие в команде разработчики специалистов-дизайнеров</p> <p>Сл3. Длительный процесс сбора данных</p>

Внешние факторы	<p>Возможности:</p> <p>В1. Появление единой международной библиографической базы</p> <p>В2. Появление новых наукометрических показателей</p> <p>В3. Создание расширения для построения карт науки</p> <p>В4. Появление новых библиографических баз данных</p>	<p>С1В4. С добавлением в систему интеграции с все большим числом библиографических баз увеличивается число потенциальных клиентов из разных стран.</p> <p>В3С2. При появлении удобного инструмента для построения карты науки, появляется возможность внедрения его в модуль интеграции с инфраструктурой организации</p>	<p>В1Сл1. Ограниченность ресурсов не позволяет реализовать интеграцию сразу со всеми популярными библиографическими базами, однако, при создании единой международной базы такая проблема исчезнет.</p> <p>В2Сл2. При появлении большого числа новых наукометрических показателей возникнет задача удобного отображение статистики по всем показателям, для чего потребуется специалист в сфере дизайна приложений.</p>
	<p>Угрозы:</p> <p>У1. Конкуренция с аналогами</p> <p>У2. Ограничения со стороны библиографических баз на использование API</p> <p>У3. Организации-заказчики с не стабильным Интернет-соединением</p>	<p>У1С1. Среди аналогов мало систем, имеющих в качестве источников несколько библиографических баз. Поэтому интеграции с несколькими базами повысят конкурентоспособность системы.</p> <p>У3С3. В ситуации, когда организация-заказчик не имеет постоянного выхода в сеть Интернет, есть возможность экспортировать статистику в отчеты формата Excel.</p>	<p>У1Сл1Сл2. Привлечение в команду разработки высококвалифицированных специалистов.</p> <p>У2Сл3. Реализация универсального и оптимизированного модуля интеграции с API библиографических баз.</p> <p>У4Сл3. Реализация парсера, позволяющего отказаться от использования API библиографических баз</p>

В целом, SWOT-анализ показал: перспективным улучшением системы является увеличение числа библиографических баз, используемых в качестве источников данных. Помимо этого, необходимо либо реализация универсального модуля интеграции с API данных баз, либо реализация парсера, позволяющая отказаться от использования внешних API.

В таблице 7 приведена интерактивная матрица проекта, показывающая взаимосвязи выделенных областей матрицы SWOT-анализа.

Таблица 7 – Интерактивная матрица проекта

		Сильные стороны			Слабые стороны		
		С1	С2	С3	Сл1	Сл2	Сл3
Возможности	В1				+		
	В2					+	
	В3		+				
	В4	+					
Угрозы	У1	+			+		
	У2						+
	У3			+			

## 5.2. Планирование научно-исследовательских работ

### 5.2.1. Разработка графика проведения научного исследования

Первоначально необходимо определить перечень работ, сроки их выполнения и занятость исполнителей. Практической пользой данного этапа работ является эффективное использование временных и человеческих ресурсов при выполнении работ.

Исполнителями в рамках рассматриваемой диссертации являются научный руководитель (НР) и 2 непосредственных исполнителя инженера (И1, И2).

Для расчета продолжительности этапов работ выбран опытно-статистический метод. При этом выбран экспертный способ расчетов, так как нет данных по аналоговым проектам.

Для определения трудоемкости выполнения работ на основе экспертной оценки ожидаемой трудоемкости выполнения каждой работы необходимо рассчитать длительность работ в рабочих и календарных днях для каждого из вариантов исполнения работ по следующим формулам:

$$t_{ож i} = \frac{3t_i + 2t_i}{5}, \quad (1)$$

где  $t_{ож i}$  – ожидаемая трудоемкость выполнения  $i$ -ой работы чел.-дн.;  
 $t_i$  – минимально возможная трудоемкость выполнения заданной  $i$ -ой работы, чел.-дн.;

$t_i$  – максимально возможная трудоемкость выполнения заданной  $i$ -ой работы (пессимистическая оценка: в предположении наиболее неблагоприятного стечения обстоятельств), чел.-дн.

$$T_{pi} = \frac{t_{ож i}}{Ч_i}, \quad (2)$$

где  $T_{pi}$  – продолжительность одной работы, раб. дн.;  
 $t_{ож i}$  – ожидаемая трудоемкость выполнения одной работы, чел.-дн.;

$Ч_i$  – численность исполнителей, выполняющих одновременно одну и ту же работу на данном этапе, чел.

$$k_{кал} = \frac{T_{кал}}{T_{кал} - T_{вых} - T_{пр}}, \quad (3)$$

где  $k_{кал}$  – коэффициент календарности;  
 $T_{кал}$  – количество календарных дней в году;  
 $T_{вых}$  – количество выходных дней в году;  
 $T_{пр}$  – количество праздничных дней в году.

$$T_{ki} = T_{pi} * k_{кал}, \quad (4)$$

где  $T_{ki}$  – продолжительность выполнения  $i$ -й работы в календарных днях;

$T_{pi}$  – продолжительность выполнения  $i$ -й работы в рабочих днях;

При проведении данных расчетов учитывался тот факт, что согласно производственному календарю для 6-ти дневной рабочей недели в 2021 году 365 календарных дней, которые включают в себя 299 рабочих дней и 66 выходных/праздничных дней. Для инженера расчет производился по 5-ти дневной рабочей недели, для которой в 2021 году 365 календарных дней,

которые включают в себя 247 рабочих дней и 118 выходных/праздничных дней.

Отсюда следует, что коэффициент календарности для 6-ти дневной рабочей недели равен следующему значению:

$$k_{\text{кал}} = \frac{T_{\text{кал}}}{T_{\text{кал}} - T_{\text{вых}} - T_{\text{пр}}} = \frac{365}{365 - 66} = 1,22$$

Для 5-ти дневной рабочей недели равен следующему значению:

$$k_{\text{кал}} = \frac{T_{\text{кал}}}{T_{\text{кал}} - T_{\text{вых}} - T_{\text{пр}}} = \frac{365}{365 - 118} = 1,48$$

В таблице 8 представлены все этапы работы над диссертацией и результаты расчетов трудозатрат.

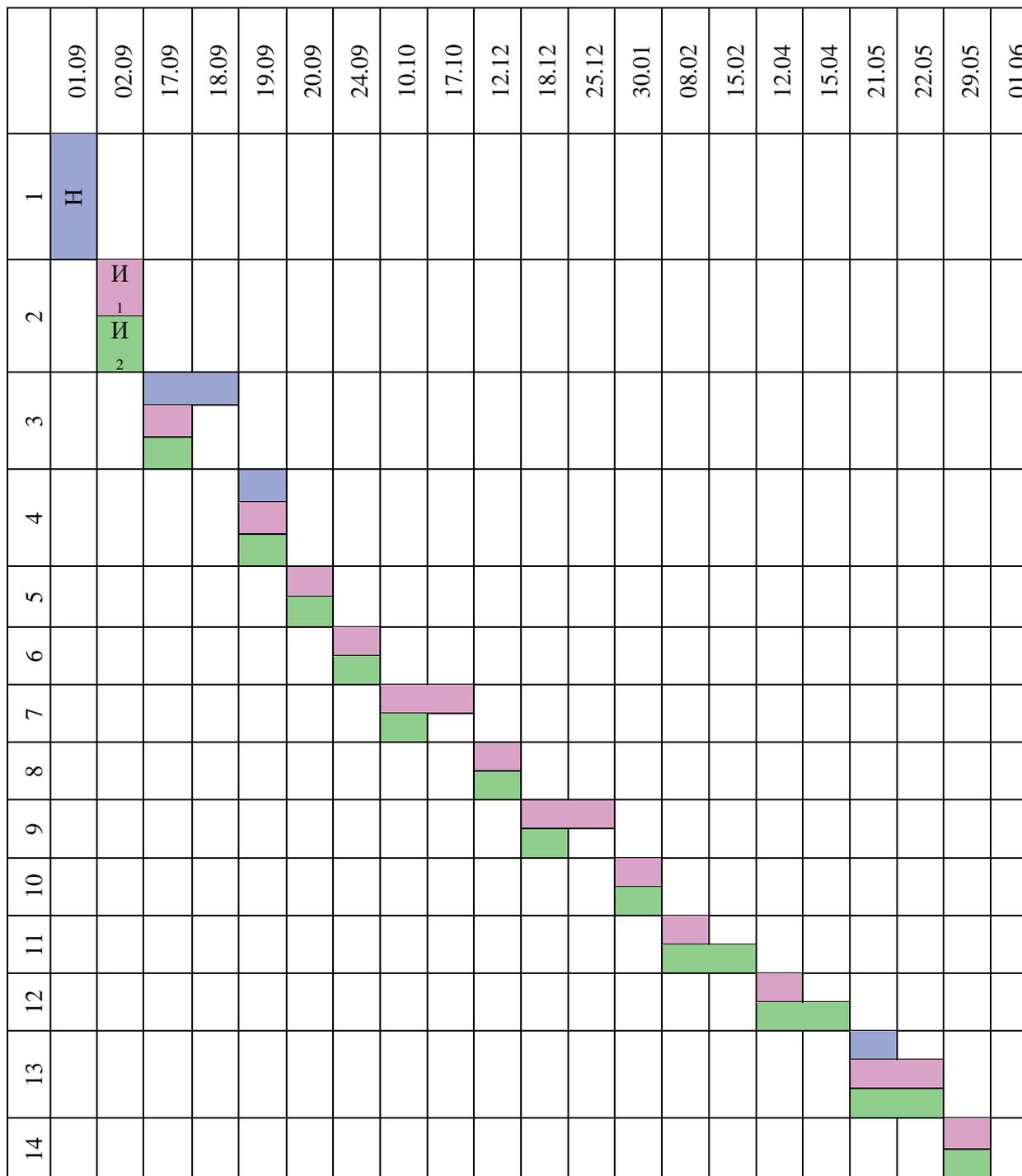
Таблица 8 – Трудозатраты на выполнение диссертации

	Наименование работы	Исполнители	Трудоемкость, чел.-дни			Длительность, дни	
			$t_{\min}$	$t_{\max}$	$t_{\text{ож}}$	$T_p$	$T_k$
	Постановка целей и задач	НР	1	2	1.4	1	1
	Анализ предметной области	И1, И2	19	23	20.6	10	15
	Составление и утверждение технического задания	НР	1	2	1.4	1	1
		НР, И1, И2	1	2	1.4	1	1
	Планирование работ и сроков их выполнения	НР, И1, И2	1	2	1.4	1	1
	Определение инструментов для разработки	И1, И2	5	9	6.6	3	4
	Проектирование архитектуры системы	И1, И2	15	30	21	11	16
	Реализация модуля интеграции со Scopus API	И1	30	50	8	38	56
		И1, И2	7	15	10.2	5	7
	Изучение основных наукометрических показателей	И1, И2	7	9	7.8	4	6

	Формирование статистики для оценки эффективности научной деятельности	И1	20	30	24	24	36	
		И1, И2	7	5	0.2	5	7	
10	Проектирование пользовательского интерфейса	И1, И2	10	5	2	6	9	
11	Реализация модуля визуализации статистических данных	И2	30	50	38	38	56	
		И1, И2	7	15	10.2	5	7	
12	Анализ результатов работы системе на примере научной деятельности ТПУ	И2	20	30	24	24	36	
		И1, И2	3	7	4.6	2	3	
13	Оформление отчетной документации	И1, И2	9	13	10.6	5	7	
		НР, И1, И2	1	2	1.4	1	1	
14	Подготовка к защите диссертации	И1, И2	2	7	4	2	3	
Итого	И1						123	179
	И2						123	179
	НР						5	5

На основе полученных значений трудозатрат исполнителей строится календарный план-график в виде диаграммы Ганта для отображения всех этапов работ и занятости исполнителей на каждом этапе. Данная диаграмма представляет собою ленточный график, на котором каждая из проведенных работ отображается в виде отрезка с заданным интервалом времени выполнения и характеризует каждый вид работы отдельно.

Таблица 9 – Линейный график работ



Из диаграммы Гантта видно, что между исполнителями И1 и И2 работы распределены равномерно по объему. В первой половине календарного графика большую часть работ выполняет исполнитель И1, занимающийся разработкой серверной части, во второй половине календарного графика большую часть работ выполняет исполнитель И2, занимающийся разработкой клиентской части. Наиболее длительными являются этапы проектирования и разработки. Планируемый срок завершения работ 01.06.21.

### **5.3. Бюджет научно-технической разработки**

В состав затрат на создание проекты включается величина всех расходов, необходимых для реализации всего комплекса работ. Расчет ведется по следующим статьям затрат:

- материалы;
- амортизационные отчисления;
- заработная плата;
- отчисления во внебюджетные фонды;
- прочие (накладные расходы) расходы.

Рассмотрим подробнее и определим затраты по этим статьям ниже.

#### **5.3.1. Расчет затрат на материалы**

К данной статье расходов относится стоимость материалов, покупных изделий, полуфабрикатов и других материальных ценностей, расходуемых непосредственно в процессе выполнения работ над объектом проектирования. Цена материальных ресурсов определяется по соответствующим ценникам или договорам поставки.

Приближенно они оцениваются в процентах к отпускной цене закупаемых материалов, как правило, это 5%. Перечень приобретенных материалов и их стоимость приведены в таблице 10.

Таблица 10 – Затраты на материалы

Наименование материалов	Цена за ед., руб.	Кол-во	Сумма, руб.
Бумага для принтера формата А4	262	1 уп	262
Картридж для принтера	1100	1 шт.	1100
<b>Итого</b>			1362

### 5.3.2. Расчет затрат на специальное оборудование

В данную статью затрат входят суммы, необходимые на обеспечение амортизации используемого оборудования.

При подготовке и написании работы использовались 2 персональных компьютера стандартной комплектации стоимостью 100000 рублей каждый. Срок полезного использования для офисных машин – 3 года, планируется использовать персональные компьютеры для постановки вычислительных экспериментов и написания ВКР в течение 8 месяцев. Тогда:

- норма амортизации:

$$A_n = \frac{1}{n} \cdot 100\% = \frac{1}{3} \cdot 100\% = 33,33\%, \quad (5)$$

- годовые амортизационные отчисления:

$$A_g = (2 \cdot 100000) \cdot 0,33 = 66000 \text{ руб.}, \quad (6)$$

- ежемесячные амортизационные отчисления:

$$A_m = \frac{66000}{12} = 5500 \text{ руб.}, \quad (7)$$

- итоговая сумма амортизации основных средств:

$$A = 5500 \cdot 8 = 44000 \text{ руб.} \quad (8)$$

Итоговая сумма затрат на амортизацию составила 44000 руб.

### 5.3.3. Расчет заработной платы

Проведем расчет заработной платы исполнителей проекта: два инженера и научного руководителя.

Месячный оклад руководителя с должностью доцент и кандидат технических наук, учитывая индексацию, составляет 35111,5 руб., для

студента (инженера) был взят оклад ассистента без научной степени – 22695 руб.

Месячный должностной оклад исполнителя рассчитывается как:

$$Z_m = Z_{ок} \cdot k_p, \quad (9)$$

где  $Z_{ок}$  – размер оклада,  $k_p = 1,3$  – районный коэффициент (значение взято для г. Томска).

Среднедневная заработная плата рассчитывается как:

$$Z_{дн} = \frac{Z_m \cdot M}{F_d}, \quad (10)$$

где  $Z_m$  – месячный должностной оклад работника;  $M$  – количество месяцев работы без отпуска в течение года: для инженера (пятидневная рабочая неделя)  $M = 11,2$  мес., для научного руководителя (шестидневная рабочая неделя)  $M = 10,4$  мес.;  $F_d$  – действительный годовой фонд рабочего времени научно-технического персонала: для инженера  $F_d = 223$  раб. дня, для научного руководителя  $F_d = 243$  раб. дня.

Тогда основная заработная плата от предприятия может быть рассчитана следующим образом:

$$Z_{осн} = Z_{дн} \cdot T_p, \quad (11)$$

где  $Z_{дн}$  – среднедневная заработная плата,  $T_p$  – продолжительность работ

Результаты расчетов заработной платы исполнителей приведены в таблице 11.

Таблица 11 – Расчет основной заработной платы

Исполнитель	З <sub>ок</sub> , руб.	k <sub>р</sub>	З <sub>м</sub> , руб.	З <sub>дн</sub> , руб.	T <sub>р</sub> , раб. дни	З <sub>осн</sub> , руб.
Инженер 1	22695,00	1,3	29503,50	1481,79	123	182260,17
Инженер 2	22695,00	1,3	29503,50	1481,79	123	182260,17
Научный руководитель	35111,50	1,3	45644,95	1953,53	5	9767,65
Итого						374287,99

#### 5.3.4. Дополнительная заработная плата исполнителей темы

В данную статью расходов входит заработная плата, начисленная рабочим и служащим не за фактически выполненные работы или проработанное время, а в соответствии с действующим законодательством, в том числе оплата очередных отпусков рабочих, времени, связанного с выполнением государственных и общественных обязанностей.

Зная основную заработную плату, можно рассчитать дополнительную заработную плату в размере 12 % от основной:

$$Z_{\text{доп}} = k_{\text{доп}} \cdot Z_{\text{осн}}, \quad (12)$$

где  $k_{\text{доп}} = 0,12$  – коэффициент дополнительной заработной платы,  $Z_{\text{осн}}$  – основная заработная плата.

Результаты расчетов дополнительной заработной платы приведены в таблице 12.

Таблица 12 – Расчет дополнительной заработной платы

Исполнитель	k <sub>доп</sub>	З <sub>осн</sub> , руб.	З <sub>доп</sub> , руб.
Инженер 1	0,12	182260,17	21871,22
Инженер 2		182260,17	21871,22
Научный руководитель		9767,65	1172,19
Итого			44914,56

#### 5.3.5. Отчисления во внебюджетные фонды

В данную статью затрат включаются отчисления в органы социального страхования, пенсионные фонды и отчисления медицинского страхования.

Отчисления во внебюджетные фонды рассчитываются как:

$$Z_{\text{внеб}} = k_{\text{внеб}} \cdot (Z_{\text{осн}} + Z_{\text{доп}}), \quad (13)$$

где  $k_{\text{внеб}} = 0,271$  – коэффициент отчислений во внебюджетные фонды (принимается пониженная ставка для учреждений, осуществляющих образовательную и научную деятельность),  $Z_{\text{осн}}$  – основная заработная плата,  $Z_{\text{доп}}$  – дополнительная заработная плата.

Результаты расчетов отчислений во внебюджетные фонды приведены в таблице 13.

Таблица 13 – Расчет отчислений во внебюджетные фонды

Исполнитель	$k_{\text{внеб}}$	$Z_{\text{доп}}$ , руб.	$Z_{\text{осн}}$ , руб.	$Z_{\text{внеб}}$ , руб.
Инженер 1	0,271	21871,22	182260,17	55319,61
Инженер 2		21871,22	182260,17	55319,61
Научный руководитель		1172,19	9767,65	2964,70
Итого				113603,92

### 5.3.6. Накладные расходы

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов: печать и ксерокопирование материалов, оплата услуг связи, электроэнергии и т.д. Их величина определяется согласно следующей формуле:

$$Z_{\text{накл}} = k_{\text{накл}} \cdot (\text{сумма статей расходов}), \quad (14)$$

где  $k_{\text{накл}}$  – коэффициент накладных расходов, принятый за 16 %.

Результат расчета накладных расходов приведен в таблице 14.

Таблица 14 – Расчет накладных расходов

Статьи затрат	Сумма, руб.
Материальные затраты	1362
Амортизацию	44000
Основная заработная плата	374287,99
Дополнительная заработная плата	44914,56
Отчисления во внебюджетные фонды	113603,92
Накладные расходы	92506,95

### 5.3.7. Бюджет затрат

Рассчитанные величины затрат научно-исследовательской работы являются основой для формирования бюджета затрат проекта. Результаты составления итогового бюджета разработки представлены в таблице 15.

Таблица 15 – Бюджет затрат на разработку

Наименование	Сумма, руб.	Удельный вес, %
Материальные затраты	1362	0,002
Затраты на амортизацию	44000	6,56
Затраты на основную заработную плату	374287,99	55,81
Затраты на дополнительную заработную плату	44914,56	6,70
Затраты на отчисления во внебюджетные фонды	113603,92	16,94
Накладные расходы	92506,95	13,79
Общий бюджет	670675,42	100

### 5.4. Определение ресурсной, финансовой, бюджетной, социальной и экономической эффективности исследования

В данном подразделе сравнивается разработка, описанная в данной работе, со своим прямым конкурентом – Интегрированная модульная платформа SciVal (аналог 1), описанной в подразделе 5.1.2, так как данный аналог решает ту же самую задачу.

### 5.4.1. Определение финансовой эффективности исследования

Интегральный финансовый показатель разработки для  $i$ -го варианта исполнения рассчитывается как:

$$I_{\Phi}^{\text{исп.}i} = \frac{\Phi_i}{\Phi_{\text{max}}}, \quad (15)$$

где  $\Phi_i$  – стоимость  $i$ -го варианта исполнения,  $\Phi_{\text{max}}$  – максимальная стоимость исполнения из всех вариантов.

Реализация системы описанной в данной магистерской диссертации займет меньше времени по причине меньшего объема работы (система анализа открытых данных баз цитирований научных публикаций использует данные только в рамках одной организации – ТПУ), поэтому примем, что для разработки системы потребуется в два раза меньше времени.

Тогда:

$$I_{\Phi}^p = \frac{\Phi_p}{\Phi_{\text{max}}} = \frac{\Phi_p}{\Phi_p} = 1, \quad (16)$$

$$I_{\Phi}^1 = \frac{\Phi_1}{\Phi_{\text{max}}} = \frac{\Phi_1}{\Phi_p} = \frac{(2 \cdot 1) \cdot \Phi_p}{\Phi_p} = 2, \quad (17)$$

### 5.4.2. Определение показателя ресурсоэффективности

Интегральный показатель ресурсоэффективности для  $i$ -го варианта исполнения рассчитывается как:

$$I_p^{\text{исп.}i} = \sum_{j=1}^n a_j \cdot b_j, \quad (18)$$

где  $a_j$  – весовой коэффициент  $j$ -го параметра сравнения,  $b_j$  – оценка  $j$ -го параметра сравнения по пятибалльной шкале,  $n$  – число параметров сравнения.

В качестве параметров сравнения были взяты критерии конкурентоспособности из подраздела 5.1.2 с теми же весовыми коэффициентами. Результаты определения показателя ресурсоэффективности для разработки и аналоги приведены в таблице 16.

Таблица 16 – Показатель ресурсоэффективности для разработки и аналога

Параметр	Весовой коэффициент параметра	Объект исследования	
		Разработка	Аналог 1
1. Охват различных баз источников данных	0.25	2	2
2. Интеграция с инфраструктурой организации	0.15	3	3
3. Экспорт данных в формат excel	0.05	4	2
4. Функциональные возможности	0.25	3	4
5. Стоимость	0.15	4	2
6. Мощность клиентской базы	0.05	2	4
7. Техническая поддержка	0.1	4	3
Итого	1	3,00	2,85

### 5.4.3. Интегральный показатель эффективности

Интегральный показатель эффективности  $i$ -го варианта исполнения вычисляется как:

$$I_{\text{исп.}i} = \frac{I_{\text{р.}i}^{\text{исп.}i}}{I_{\text{ф.}i}^{\text{исп.}i}} \quad (19)$$

Для сравнения  $i$ -го и  $j$ -го вариантов исполнения разработки используется сравнительная эффективность проекта:

$$\mathcal{E}_{\text{ср}} = \frac{I_{\text{исп.}i}}{I_{\text{исп.}j}} \quad (20)$$

Взяв за  $i$ -ый вариант разработку, описанную в данной работе, а за  $j$ -ый – прямого конкурента, можно сделать вывод о целесообразности предлагаемого варианта: если  $\mathcal{E}_{\text{ср}} > 1$ , то предлагаемая разработка более целесообразна. В таблице 17 приведены все рассчитанные значения интегральных показателей и сравнительная эффективность вариантов исполнения.

Таблица 17 – Сравнение интегральных показателей

Показатель	Разработка	Аналог 1
Интегральный финансовый показатель разработки	2	1
Интегральный показатель ресурсоэффективности	3,00	2,85

разработки		
Интегральный показатель эффективности	6,00	2,85
Сравнительная эффективность вариантов исполнения	2,1	

Таким образом, значение сравнительной эффективности показывает большую целесообразность разрабатываемой системы.

### **5.5. Заключение по разделу**

В ходе выполнения магистерской диссертации была разработана система для информационного обеспечения процессов управления научной деятельностью.

В ходе работы над разделом был проведен предпроектный анализ, включающий анализ конкурентных решений и SWOT-анализ. Анализ конкурентных решений показал преимущество разрабатываемого подхода по сравнению с аналогами. SWOT-анализ позволил определить потенциальные пути улучшения разработки: перспективными можно считать улучшение предлагаемого в данной работе подхода путем применения новых методов машинного обучения и понижение порога вхождения для рядовых пользователей.

Также в ходе инициации проекта был составлен календарный план проекта, проведена оценка трудоемкости работ. График-план проекта был представлен в виде диаграммы Ганта.

Был сформирован бюджет разработки, включающий материальные затраты, затраты на амортизацию, основную заработную плату исполнителям, дополнительную заработную плату, отчисления во внебюджетные фонды и накладные расходы. Общий бюджет разработки составил 670675,42 руб.

Было проведено сравнение эффективности выполнения для данной разработки и ее прямого аналога на основании сравнения значений интегральных показателей эффективности. Предлагаемый в данной работе подход оказался более целесообразным.



## **6. Социальная ответственность**

Целью данной работы является разработка веб-системы для проведения оценки и анализа научной деятельности организации. В качестве основных конечных пользователей системы будут выступать сотрудники научных организаций, занимающиеся управленческой деятельностью по развитию научной активности организации. Результатом использования разрабатываемой системы является информационное обеспечение процессов управления научной деятельностью с помощью инструментов мониторинга базовых показателей научной деятельности.

Разработка и ее дальнейшее использование системы подразумевает работу за рабочим столом, оборудованным персональным компьютером с доступом к сети Интернет. В данном разделе будут рассмотрены возможные вредные факторы при проектировании, разработке и эксплуатации системы конечными пользователями.

### **6.1. Правовые и организационные вопросы обеспечения безопасности**

Основу законодательства, регламентирующего правовые гарантии в области производственной безопасности, составляют Трудовой кодекс РФ, а также Федеральные законы «Об обязательном социальном страховании от несчастных случаев на производстве и профессиональных заболеваний» (1998 г.), «О техническом регулировании» (2002 г.) и др.

Поскольку разработка и дальнейшая эксплуатация системы предполагает работу за персональным компьютером в положении сидя, для изучения требований к рабочему месту необходимо обратиться к стандартам ГОСТ 12.2.032–78 «ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования» [18] и ГОСТ 21889–76 «Кресло человека-оператора. Общие эргономические требования» [19]. Основными

требованиями к конструкции оборудования и рабочего места при работе в положении сидя, исходя из вышеперечисленных документов, являются:

1. Конструкцией рабочего места должно быть обеспечено выполнение трудовых операций в пределах зоны досягаемости моторного поля;
2. Оптимальное положение рабочего места достигается регулированием высоты рабочей поверхности, сидения и пространства для ног в зависимости от роста человека;
3. Перечень требований к размещению органов управления;
4. Перечень требований к размещению средств отображения информации.

В СанПиН 2.2.2/2.4.1340-03 содержатся правила и нормы, описывающие организацию работы с персональным компьютером. Согласно СанПиН 2.2.2/2.4.1340-03 и п. 3.2.2.4 Перечня вредных и опасных факторов излучение от компьютера, если работник проводит за ним более 50 % рабочего времени, является вредным фактором условий труда [20]. Основные эргономические требования к помещению, в котором выполняется работа с использованием персонального компьютера:

1. Окна в помещениях, где работают с компьютерами, должны быть ориентированы на север и северо-восток. Оконные проемы необходимо оборудовать регулируемыми жалюзи, внешними козырьками и пр.
2. По истечении каждого часа работы проводится влажная уборка и проветривание.
3. Площадь рабочего места для сотрудника, работающего с использованием компьютера на базе плоских дискретных экранов, составляет не менее 4,5 м<sup>2</sup>
4. Уровни шума на рабочих местах не должны превышать предельно допустимых значений, установленных для данных видов работ в

соответствии с действующими санитарно-эпидемиологическими нормативами.

5. При размещении рабочих мест с персональными компьютерами расстояние между рабочими столами с видеомониторами, должно быть не менее 2,0 м, а расстояние между боковыми поверхностями видеомониторов - не менее 1,2 м.

Согласно Трудовому кодексу РФ режим рабочего времени должен предусматривать продолжительность рабочей недели, работу с ненормированным рабочим днем для отдельных категорий работников, продолжительность ежедневной работы (смены), в том числе неполного рабочего дня (смены), время начала и окончания работы, время перерывов в работе, число смен в сутки, чередование рабочих и нерабочих дней, которые устанавливаются правилами внутреннего трудового распорядка в соответствии с трудовым законодательством и иными нормативными правовыми актами, содержащими нормы трудового права, коллективным договором, соглашениями, а для работников, режим рабочего времени которых отличается от общих правил, установленных у данного работодателя, - трудовым договором[21].

## **6.2. Производственная безопасность**

### **6.2.1. Анализ вредных и опасных факторов, которые могут возникнуть при эксплуатации объекта исследования**

Вредным называется производственный фактор, воздействие которого на работающего в определенных условиях приводит к заболеванию или снижению работоспособности. При изменении уровня и времени воздействия вредные производственные факторы могут стать опасными. Опасными считаются производственные факторы, воздействие которых на работающего в конкретных условиях может привести к травмам, а также другим

внезапным резким ухудшениям здоровья. Рассмотрим возможные вредные и опасные факторы, возникающие при работе с персональным компьютером.

Таблица 18 – Возможные опасные и вредные факторы

Факторы	Этапы работ		Нормативные документы
	Разработка	Эксплуатация	
Отклонение показателей микроклимата	+	+	СанПиН 2.2.2/2.4.1340-03, СанПиН 2.2.4.548-96
Недостаточная освещенность рабочей зоны	+	+	СанПиН 2.2.1/2.1.1.1278-03, СанПиН 2.2.2/2.4.1340-03, СП 52.13330.2011, СП 52.13330.2016
Повышенный уровень электромагнитных излучений	+	+	СанПиН 2.2.2/2.4.1340-03, СанПиН 2.2.4.1191-03, СанПиН 2.2.4/2.1.8.055-96, ГОСТ Р 12.1.019-2009
Повышенный уровень шума на рабочем месте	+	+	СанПиН 2.2.2/2.4.1340-03, ГОСТ 12.1.003-2014, СН 2.2.4/2.1.8.562-96

### **6.2.2. Отклонение показателей микроклимата рабочего помещения**

Микроклимат производственных помещений это условия внутренней среды помещений, которые определяются действующими на организм человека сочетаниями температуры, влажности, скорости движения воздуха и теплового излучения; комплекс физических факторов, оказывающих влияние на теплообмен человека с окружающей средой, на тепловое состояние человека и определяющих самочувствие, работоспособность, здоровье и производительность труда. Требования к микроклимату производственных помещений регламентируются санитарными нормами и правилами СанПиН 2.2.2548-96 «Гигиенические требования к микроклимату производственных помещений» [22]. Данные правила предназначены для предотвращения неблагоприятного воздействия микроклимата рабочих мест,

производственных помещений на самочувствие, функциональное состояние, работоспособность и здоровье человека. В таблице 19 приведены оптимальные и допустимые параметры микроклимата производственного помещения для категории работ 1а, к которой относят деятельность программиста.

Таблица 19 – Оптимальные и допустимые параметры микроклимата

Период года	Вид параметров	Температура воздуха, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Теплый	Оптимальные	22-24	40-60	0,1
	Допустимые	20-25	15-75	0,1
Холодный	Оптимальные	23-25	40-60	0,1
	Допустимые	21-28	15-75	0,1

### 6.2.3. Недостаточная освещенность рабочего помещения

Физическая величина, численно равная световому потоку, падающему на единицу площади освещаемой поверхности, называется освещенностью. Плохое освещение негативно воздействует на зрение, приводит к быстрому утомлению, снижает работоспособность, вызывает дискомфорт, является причиной головной боли и бессонницы. Естественное и искусственное освещение в помещениях регламентируется СанПиН 2.2.1/2.1.1.1278-03 «Гигиенические требования к естественному, искусственному и совмещенному освещению жилых и общественных зданий» в зависимости от характера зрительной работы, системы и вида освещения, фона, контраста объекта с фоном. Осветительные установки, независимо от используемых источников света и световых приборов, должны обеспечивать нормативные требования к общему искусственному освещению. В помещениях с персональным компьютером нормируемый показатель искусственного

освещения согласно СанПиН 2.2.1/2.1.1.1278–03 равен 300 Лк. Нормируемые показатели естественного, искусственного и совмещенного освещения в соответствии с СанПиНом 2.2.1/2.1.1.1278-03 указаны в таблице 20[23].

Таблица 20 – Нормируемые показатели естественного, искусственного и совмещенного освещения

Помещения	Рабочая поверхность и плоскость нормирования КЕО и освещенности (Г-горизонтальная, В-вертикальная) и высота плоскости над полом, м	Естественное освещение		Совмещенное освещение	
		КЕО $e_n$ , %		КЕО $e_n$ , %	
		При верхнем или комбинированном освещении	При боковом освещении	При верхнем или комбинированном освещении	При боковом освещении
Кабинеты, рабочие комнаты, офисы, представительства	$\Gamma - 0,8$	3,0	1,0	1,8	0,6
Помещения	Искусственное освещение				
	Освещенность, лк				
	При комбинированном освещении		При общем освещении	Показатель дискомфорта, М, не более	Коэффициент пульсации освещенности, Кп, %, не более
	Всего	От общего			
Кабинеты, рабочие комнаты, офисы, представительства	400	200	300	401	15

Произведем оценку соответствия освещенности рабочего места инженера стандартам. Для этого определим, соответствует ли световой поток

от имеющегося количества и мощности ламп требуемому [25]. Размеры рабочего помещения, следующие: длина  $A = 3$  м, ширина  $B = 3$  м, высота  $H = 3$  м. В помещении находится одно окно (источник естественного освещения) и один потолочный светильник.

Общий световой поток определяется по формуле:

$$F_{\text{общ}} = \frac{E_H * S * z_1 * z_2}{n}, \quad (21)$$

где  $E_H$  – нормированная освещенность (в данном случае – 300 Лк);  $S$  – площадь помещения;  $z_1$  – коэффициент запаса, учитывающий старение ламп и загрязнение светильников ( $z_1 = 1,3$ );  $z_2$  – коэффициент, учитывающий неравномерность освещения помещения (согласно СНиП 23-05-95[7]  $z_2 = 1,1$ );  $n$  – коэффициент использования светового потока (определяется в соответствии с индексом помещения).

Индекс помещения определяется по формуле:

$$I = \frac{S}{h * (A + B)}, \quad (22)$$

где  $h$  – высота подвеса, м;  $A$  – ширина помещения, м;  $B$  – длина помещения, м.

$$I = \frac{9}{(3 - 0,3) * 6} = 0,6$$

Цвета потолка, стен и пола имеют коэффициенты отражения 70%, 50% и 30% соответственно. Согласно таблице значений коэффициента использования светового потока светильников используемого типа для данных коэффициентов отражения значение коэффициента использования  $n = 0,3$ . Из полученных данных определяется значение светового потока  $F$ . Сделаем расчет для нормативного значения освещенности, равного 300 Лк.

$$F_{\text{общ}} = \frac{300 * 9 * 1,3 * 1,1}{0,3} = 12870 \text{ Лм.}$$

В рабочем помещении используются компактные люминесцентные лампы со световым потоком около 1300 Лм. Определим необходимое количество ламп по формуле:

$$N = \frac{F_{\text{общ}}}{F_{\text{л}}}. \quad (23)$$

Таким образом, для соответствия требованиям стандартов освещенности необходимое число ламп составляет:

$$N = \frac{12870}{1300} = 9,8 \approx 10$$

Люстра в рабочем помещении пяти рожковая, в ее конструкции предусмотрена установка 6 ламп. Таким образом, для соблюдения рекомендуемого уровня освещенности необходимо использовать дополнительные источники освещения, например, настольную лампу.

#### 6.2.4. Повышенный уровень электромагнитных излучений

Электромагнитное излучение – это электромагнитные волны, возбуждаемые различными излучающими объектами, заряженными частицами, атомами, молекулами, антеннами и пр. Электромагнитное излучение может вызвать расстройства нервной системы, снижение иммунитета, расстройства сердечно-сосудистой системы и представляет собой особую опасность для беременных. При работе с персональным компьютером необходимо предусматривать защиту от излучения системного блока. СанПиН 2.2.4/2.1.8.055-96 регулирует максимально допустимые уровни напряженности электрического и магнитного полей, плотности потока энергии электромагнитного поля [26]. Соответствующие значения приведены в таблице 21.

Таблица 21 – Предельно допустимые уровни энергетических экспозиций электромагнитного поля

Диапазоны частот	По электрической составляющей, (В/м) <sup>2</sup> ч	По магнитной составляющей, (А/м) <sup>2</sup> ч	По плотности потока энергии, (мкВт/м) <sup>2</sup> ч
30 кГц – 3 МГц	20000	200	-
3 – 30 МГц	7000	-	-

30 – 50 МГц	800	0,72	-
50 – 300 МГц	800	-	-
300 МГц – 300 ГГц	-	-	200

### 6.2.5. Повышенный уровень шума

Предельно допустимый уровень (ПДУ) шума – это уровень фактора, который при ежедневной работе не должен вызывать заболеваний или отклонений в состоянии здоровья, обнаруживаемых современными методами исследований в процессе работы или в отдаленные сроки жизни настоящего и последующих поколений. Длительное воздействие шумов снижает производительность труда и приводит к ухудшению слуха, головным болям, к различным нарушениям деятельности нервной системы и изменению сосудистого давления. Источниками шума при работе с персональным компьютером являются:

- Кулер на процессоре;
- Кулеры на корпусе;
- Система охлаждения видеокарты;
- Жесткие диски;
- CD-ROM.

По санитарным нормам, допустимым уровнем шума, который не наносит вреда слуху даже при длительном воздействии на слуховой аппарат, принято считать: 55 децибел (дБ) в дневное время и 40 децибел (дБ) ночью. Уровень шума исправного современного компьютера находится в пределах от 35 до 50 дБА. Если в компьютере установлен плохо сбалансированный вентилятор, то он, особенно на первых минутах после включения, может достигать 55 дБА и более. В таблице 22 представлены предельно допустимые уровни звука и эквивалентные уровни звука на рабочих местах с учетом напряженности и тяжести трудовой деятельности согласно СН 2.2.4/2.1.8.562-96 [10].

Таблица 22 – Предельно допустимые уровни звука и эквивалентные уровни звука на рабочих местах для трудовой деятельности разных категорий тяжести и напряженности в дБА

Категория напряженности и трудового процесса	Категория тяжести трудового процесса				
	Легкая физическая нагрузка	Средняя физическая нагрузка	Тяжелый труд 1 степени	Тяжелый труд 2 степени	Тяжелый труд 3 степени
Напряженность легкой степени	80	80	75	75	75
Напряженность средней степени	70	70	65	65	65
Напряженный труд 1 степени	60	60	-	-	-
Напряженный труд 2 степени	50	50	-	-	-

На рисунке 56 представлено измерение уровня шума на рабочем месте. Средний уровень шума на рабочем месте составляет 45 дБА, из чего можно сделать вывод, что уровень шума соответствует норме.



Рисунок 56 – Измерение уровня шума на рабочем месте

Воздействие перечисленных факторов уменьшает работоспособность человека, а длительное и систематическое воздействие может приводить к профессиональному заболеванию.

### 6.3. Экологическая безопасность

В настоящее время проблема экологической безопасности и охраны окружающей среды является приоритетной. Это стало поводом для принятия жестких законов, ограничивающих обычную утилизацию компьютерной техники. В большей мере это обуславливается тем, что в производстве такой техники используется множество различных материалов, которые способны нанести непоправимый вред окружающей среде и, соответственно, здоровью человека.

В соответствии с постановлением правительства юридическим лицам запрещено самостоятельно утилизировать компьютерную технику. Для этого необходимо найти специальную компанию, которая занимается утилизацией

в частном порядке. В нормативном документе СанПиН 2.2.2/2.4.1340-03 даются следующие общие рекомендации по снижению опасности для окружающей среды, исходящей от компьютерной техники [28]:

- применять оборудование, соответствующее санитарным нормам и стандартам экологической безопасности;
- применять расходные материалы с высоким коэффициентом использования и возможностью их полной или частичной регенерации;
- отходы в виде компьютерного лома утилизировать;
- использовать экономичные режимы работы оборудования.

В ходе выполнения данной магистерской диссертации и дальнейшем использовании системы отсутствуют выбросы каких-либо вредных веществ в атмосферу и гидросферу, следовательно, загрязнение воздуха и воды не происходит. Однако, люминесцентные лампы, применяющиеся для искусственного освещения рабочих мест, требуют особой утилизации, т.к. в них присутствует от 10 до 70 мг ртути, которая относится к чрезвычайно-опасным химическим веществам и может стать причиной отравления живых существ, а также загрязнения атмосферы, гидросферы и литосферы. Сроки службы таких ламп составляют около 5-ти лет, после чего их необходимо сдавать на переработку в специальных пунктах приема.

В ходе выполнения работы также создается бытовой мусор (канцелярские, пищевые отходы), который должен быть утилизирован в соответствии с определенным классом опасности или переработан, чтобы не оказывать негативное влияние на состояние литосферы.

#### **6.4. Безопасность в чрезвычайных ситуациях**

К возможным чрезвычайным ситуациям на данном рабочем месте выделяют внезапное обрушение здания, аварии на коммунальных системах жизнеобеспечения населения, пожар, угроза пандемии.

С учетом специфики работы и наличием вычислительной техники в помещении наиболее вероятно возникновение пожара, под которым понимается вышедший из-под контроля процесс горения, обусловленный возгоранием вычислительной техники и угрожающий жизни и здоровью работников.

Возникновение пожара может произойти по нескольким факторам:

- возникновением короткого замыкания в электропроводке вследствие неисправности самой проводки или электросоединений и электрораспределительных щитов;
- возгоранием устройств вычислительной аппаратуры вследствие нарушения изоляции или неисправности самой аппаратуры;
- возгоранием мебели или пола по причине нарушения правил пожарной безопасности, а также неправильного использования дополнительных бытовых электроприборов и электроустановок;
- возгоранием устройств искусственного освещения.

В связи с этим, согласно ГОСТ 12.1.004-91 «Пожарная безопасность. Общие требования», при работе с компьютером необходимо соблюдать следующие нормы пожарной безопасности [29]:

- для предохранения сети от перегрузок запрещается одновременно подключать к сети количество потребителей, превышающих допустимую нагрузку;
- работы за компьютером проводить только при исправном состоянии оборудования, электропроводки;
- иметь средства для тушения пожара (огнетушитель);
- установить количество, размеры и соответствующее конструктивное исполнение эвакуационных путей и выходов;
- обеспечить возможность беспрепятственного движения людей по эвакуационным путям.

Для предотвращения пожара помещение с вычислительной техникой должно быть оборудовано первичными средствами пожаротушения: углекислотным огнетушителем типа ОУ-2 или ОУ-5 [20]. При появлении пожара, любой, увидевший пожар должен: незамедлительно заявить о данном в пожарную службу по телефонному номеру 01 или 112, заявить о происшествии и соблюдать покой.

Пожар может нанести не только вред здоровью, но и материальный ущерб. Применимо к выполняемой работе в случае пожара могут быть уничтожены бумажные документы и\или электронные носители информации.

Для защиты информации рекомендуется использовать облачные хранилища данных для данных и документов. Для исходных кодов программ рекомендуется использовать системы контроля версий.

#### **6.5. Выводы по разделу**

В данном разделе были рассмотрены ключевые вопросы безопасности при разработке и эксплуатации реализуемой в рамках магистерской диссертации системы. Используя все приведенные правила и нормы, касающиеся работы с персональным компьютером, исследование и дальнейшая эксплуатация разрабатываемой системы не приведет к ухудшению здоровья работника и окружающей его среды.

В рамках рассмотрения соблюдения требуемых норм при разработке для защиты от вредных факторов выявлена необходимость дополнительного источника освещения. Остальные рассматриваемые факторы соответствуют нормам.

Рассматривая вопросы экологического влияния разработки и использования системы, была выявлена необходимость решения вопроса правильной утилизации использованного сырья (ламп, электрических приборов, канцелярских отходов).

При изучении возможных чрезвычайных ситуаций возникновение пожара выделено как одна из наиболее вероятных чрезвычайных ситуаций, рассмотрены способы предотвращения пожара и основные нормы и правила пожарной безопасности.

## Список используемых источников

1. Мальцева А.А., Барсукова Н.Е. Управление развитием научных организаций с применением сбалансированной системы показателей [Электронный ресурс], URL: <https://vspu2019.ipu.ru/proceedings/1768.pdf> – Свободный доступ (дата обращения: 05.01.2020).
2. Публикационная активность и индексы научного цитирования: в вопросах и ответах / [авт.-сост. Гниденко Е. В.]; Витеб. гос. ордена Дружбы народов мед. ун-т, Библиотека. — Витебск: [б. и.], 2020. — 72 с.: ил. — Библиогр.: с. 64—67.
3. Основные наукометрические показатели [Электронный ресурс], URL: [http://bibl.rusoil.net/jirbis2/images/phocagallery/library/nauk\\_pokasateli.pdf](http://bibl.rusoil.net/jirbis2/images/phocagallery/library/nauk_pokasateli.pdf) – Свободный доступ (дата обращения: 25.01.2020).
4. SciVal: наукометрический инструмент для ученых и аналитиков от Elsevier [Электронный ресурс]: Управление научных исследований СПбГТИ (ТУ). URL: <http://science.spb.ru/csnews/item/3091-scival> – Свободный доступ (дата обращения: 07.03.2020).
5. InCites – аналитический инструмент для оценки научной деятельности ученого, организации, страны [Электронный ресурс]: МИСиС Национальный исследовательский технологический университет. URL: <http://lib.misis.ru/inc.html> – Свободный доступ (дата обращения: 10.03.2020).
6. Хохлов А.Р., Информационно-аналитическая система Наука-МГУ («Истина»): методическое пособие. Москва: Изд-во МГУ, 2012. (дата обращения: 27.03.2020).
7. Информационно-аналитическая система "Наука-МГУ" [Электронный ресурс]: Московский Государственный университет имени М.В.Ломоносова. URL: <https://www.msu.ru/science/ris/> – Свободный доступ (дата обращения: 12.04.2020).

8. ИСТИНА в науке и образовании [Электронный ресурс]: OSP – Гид по технологиям цифровой трансформации. URL: <https://www.osp.ru/os/2016/03/13050261> – Свободный доступ (дата обращения: 22.04.2020).
9. Карта российской науки [Электронный ресурс]: Википедия. URL: [https://ru.wikipedia.org/wiki/Карта\\_российской\\_науки](https://ru.wikipedia.org/wiki/Карта_российской_науки) – Свободный доступ (дата обращения: 03.05.2020).
10. Networkx: программный пакет python для изучения сложных сетей [Электронный ресурс]. URL: <http://espressocode.top/networkx-python-software-package-study-complex-networks/> – Свободный доступ (дата обращения: 05.02.2021).
11. Визуализация данных с python-фреймворком dash [Электронный ресурс]: Python School. URL: <https://python-school.ru/dash-basics/> – Свободный доступ (дата обращения: 16.03.2021).
12. API Interface Specification [Электронный ресурс] // Elsevier Developers - URL: [https://dev.elsevier.com/api\\_docs.html](https://dev.elsevier.com/api_docs.html) – Свободный доступ (дата обращения: 24.03.2021).
13. Трубников В.С., Туральчук К.А. Проектирование системы сбора, анализа и визуализации наукометрических данных [Электронный ресурс]: CyberLeninka – 2015. – URL: <https://cyberleninka.ru/article/n/proektirovanie-sistemy-sbora-analiza-i-vizualizatsii-naukometricheskih-dannyh> – Свободный доступ (дата обращения: 08.03.2021).
14. Laravel Documentation [Электронный ресурс]: Laravel – URL: <https://laravel.com/docs/8.x> – Свободный доступ (дата обращения: 23.03.2021).
15. Москалева О.В. Использование наукометрических показателей для оценки научной деятельности [Электронный ресурс]: CyberLeninka – 2013. URL: <https://cyberleninka.ru/article/n/ispolzovanie>

- naukometricheskih-pokazateley-dlya-otsenki-nauchnoy-deyatelnosti –  
Свободный доступ (дата обращения: 02.04.2021).
- 16.Laravel Excel Documentation [Электронный ресурс]: Laravel Excel. URL:  
<https://docs.laravel-excel.com/3.1/getting-started/> – Свободный доступ  
(дата обращения: 14.04.2021).
- 17.Число публикаций российских ученых растет, но их редко цитируют.  
[Электронный ресурс]. Режим доступа – URL:  
<https://nauka.tass.ru/nauka/10164209>, – Свободный доступ (дата  
обращения: 22.05.2021).
- 18.ГОСТ 12.2.032–78 «ССБТ. Рабочее место при выполнении работ сидя.  
Общие эргономические требования»
- 19.ГОСТ 21889–76 «Кресло человека-оператора. Общие эргономические  
требования»
- 20.СанПиН 2.2.2/2.4.1340-03. Гигиенические требования к персональным  
электронно-вычислительным машинам и организация работы
- 21.Трудовой кодекс Российской Федерации [Электронный ресурс] федер.  
закон от 30.12.2001 № 197-ФЗ (ред. от 09.03.2021) – Доступ из  
справочно правовой системы «КонсультантПлюс»
- 22.СанПиН 2.2.2548-96 «Гигиенические требования к микроклимату  
производственных помещений»
- 23.СанПиН 2.2.1/2.1.1.1278-03 «Гигиенические требования к  
естественному, искусственному и совмещенному освещению жилых и  
общественных зданий»
- 24.СНиП 23-05-95 "Естественное и искусственное освещение"
- 25.Расчет освещенности помещения: формулы и методы. [Электронный  
ресурс]. Режим доступа – URL: [https://elektrika-svoimi-  
rukami.com/raschet-osveshheniya/raschet-osveshheniya](https://elektrika-svoimirukami.com/raschet-osveshheniya/raschet-osveshheniya), свободный. Дата  
обращения 02.05.2021г

26. СанПиН 2.2.4/2.1.8.055-96 «Электромагнитные излучения радиочастотного диапазона (ЭМИ РЧ)»
27. СН 2.2.4/2.1.8.562-96 «Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки»
28. ГОСТ 12.1.004-91 «Система стандартов безопасности труда. Пожарная безопасность. Общие требования»
29. ГОСТ Р 51057-2001 Техника пожарная. Огнетушители переносные. Общие технические требования. Методы испытаний

## Приложение А

(справочное)

### Chapter 1

#### Analytical review of scientific and technical documentation

Студенты:

Группа	ФИО	Подпись	Дата
8ИМ92	Непомнящих Дарья Сергеевна		
8ИМ92	Сутягина Анна Андреевна		

Руководитель ВКР:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Савельев Алексей Олегович	к.т.н.		

Консультант-лингвист отделения иностранных языков ШБИП:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИЯ ШБИП	Коротченко Татьяна Валериевна	к.ф.н.		

## **1. Analytical review**

### **1.1. Scientific activities in the modern world**

Scientific publications as the main product of the activity of scientists characterize the state of the science work organization and the country as a whole. Modern science is characterized by increasing number of publications, increasing number of authors in almost every scientific field.

Several decades ago, it was impossible to imagine encountering the scientific paper written by more than four co-authors. The same can be said about the research carried out by several labs or scientific centers. Today, such situation is rather common. Actually, this fact can be explained by the rise in interdisciplinary studies, approaches, and research. Under present conditions interdisciplinary scientific work can be confined to one company (institute) or be transnational and well-structured. The number of such transnational research work increased from 8% to 19% during 15 years (1988 – 2013) [1].

There is one certain limitation of the above-described tendency, precisely, it makes much more difficult to implement managerial decision concerning scientific activity. In this case, it is necessary to analyze large amounts of data on already conducted research and predict the situation in the scientific fields for the future.

### **1.2. Scientometric Indicators**

Scientometrics is a new scientific field joining science and technology with information science and expanding numerous mathematical, statistical, data mining techniques, and procedures to measure and quantify scientific information [2]. There are a large number of bibliographic indicators that somehow characterize the effectiveness of scientific activity. Via that, ratings and statistics for information support during the evaluation of the state of scientific activity have become possible.

Data sources of bibliometric/scientometric research and technology are open bibliographies and bibliographic databases [3]. The most well-known information

and analytical systems operating in developed countries are Web of Science and Scopus.

Let us consider the most commonly used scientometric indicators for tracking the scientific activities of authors:

- The Hirsch index (h-index) is considered to be a quantitative feature which shows the efficiency of a scientist with regard to his or her number of articles and scientific works, as well as the number of the citations to his or her works. For a scientist to have h index it is necessary to publish h papers. The main purpose of the Hirsch Index was to guarantee a more adequate evaluation of scientists' efficiency. It was an alternative to a simple enumeration of scientists' publications, as it considered not only the number of the published works, but also the number of citations. The Hirsch index is calculated automatically using special applications in the abstract databases Scopus, Web of Science, RSCI.

- The citation index is a measure of the "importance" of a scientific work. The importance of work calculated on the basis of the number of references made to this to the work in other sources, i.e. the number of references to the work made by other researchers in the other works and publications;

- Average citation shows the average number of citations per publication.

Scholars distinguish a number of widely-applied scientometric indicators to identify and evaluate scientists' efficiency based on their publications. They are as follows:

- The IF (impact factor) of the journal is termed as the dependence of the number of references which is obtained by the journal during the year and the articles which were published in this journal in the previous two years, to the number of articles published in this journal in the same two previous years.

- The IF is usually calculated in the following platforms: Web of Science (analytical module Journal Citation Reports), RSCI. To find out a

conventional IF, one should know the number of publications for 3 years, including a 2-year citation window.

– The IF is intended to estimate the frequency with which the average cited article of the journal is cited. In addition, it may also be used as a factor that shows the "demand" and "recognition" of the journal in the scientific world. Obviously that the higher the rating of the journal, the higher the rating of the author of the article published in this journal.

– SJR (SCImago Journal Ranking) is a rating of journals with the help of which it is possible to consider the total number of citations and weighted citation rates by year. Also, it involves the qualitative indicators, for example, the credibility of references. It is worth noting that this system does not greatly differ from that of traditional IF, however, it provides a lot of opportunities and can be applicable for a great number of journals due to its openness.

– SNIP (Source Normalized Impact per Paper) is a new enhanced system to evaluate efficiency of a scientist. It considers the number and intensity of citation in a certain scientific sphere. It makes it possible to carry out the comparative analysis of publications referring to various scientific fields. The following indicator characteristics are distinguished: references made in the current year to articles published during the previous three years. Publication window equals 3 years, Citation window equals 1 year, Document types are the same for all stages of the indicator calculation. A special definition of the "individual field of science" is introduced for the journal, or "journal environment": all articles published in the current year (in any edition) that at least once cited issues of the journal published in the last ten years. To determine the citation potential (this is the average number of positions, the average "length" of the lists of cited literature in the articles of the "environment"), the average number of references in the articles that make up the "journal environment" is calculated. But only those links that: a) lead to articles published during the previous three

years are taken into account; b) lead to the articles available in the database for which the calculation is made.

– The immediacy index is an indicator of the number of references to the journal's publications received in the year of publication. It reflects how quickly the articles published in the journal become known in the scientific world. It is calculated as the ratio of the number of references received by the journal in a certain year to the articles published in the same year, to the total number of articles published in the journal for that year.

As noted by domestic and foreign analysts, bibliometric analysis using information from databases has its advantages and disadvantages. It is not right that these indicators completely exclude expert evaluation, but they are an excellent information base for expert evaluation.

### **1.3. Evaluation of the effectiveness of scientific activities**

The development of bibliometrics as a field is strongly linked to the creation of the Science Citation Index (SCI) by Eugene Garfield in 1961. As a supplemental property, it enabled scientific literature to be analyzed quantitatively. Since the 1960s, the SCI and other similar databases, now included in the online product Web of Science, have been applied in a large number of studies covering many different fields. The option for citation analysis has been a crucial cause for this popularity. In the database, all the references of the indexed articles are registered. Based on this, each article can be ascribed a citation count showing how many times it has been cited by later papers registered in the database. In the early 2000s, competing databases were introduced which also include citation statistics, most importantly the Scopus database and Google Scholar. The coverage of the scientific and scholarly literature varies across these databases, and the results of citation studies are thus dependent upon the particular characteristics of the databases and their coverage [4].

Let us consider the existing modern systems of evaluation of scientific activity.

## 1. Integrated modular SciVal platform

SciVal is considered to be a modular platform which makes it possible to show and evaluate the findings within various fields of science. The platform allows more than 4.600 organizations to be involved. In addition, these organization are from 220 countries. The main obvious advantage of the platform is that it allows organizations to optimize strategic investment. Also, it provides them with an opportunity to define the future steps in scientific works and choose the required staff and partners to perform the tasks.

The SciVal is developed on the basis of the data taken from Scopus – the world's largest database of peer-reviewed literature, which includes more than 20 thousand publications and 30 thousand books from more than 5000 international publishers. It is a well-known fact that this database is intended to make the results of scientific work more obvious, as well as to help a scientist to compare and track the findings obtained by any company or organization, country or scientific group/community.

Basic tools SciVal is intended to:

- visualize the scientific findings, define the level of competence of this or that organization;
- reveal the strong aspects and multidisciplinary field of study, make visual reports, carry out in-depth analyses by combining the areas of research and metrics to achieve the goals;
- specify and analyze the current and possible opportunities carry out joint scientific work based on the publications and citations.

SciVal grants the opportunity to develop individual research areas which can be further studied and monitored. In accordance with these areas it is possible to specify focus topics, analyze scientific teams in order to establish unique centers of excellence.

SciVal also contains an advanced Trends module. It makes it possible to determine the most relevant areas within a certain subject. In addition, it provides

scientist to track the most successful research teams, including the country, within the certain research area.

## 2. InCites Research Assessment Tool

InCites is another scientometric tool which is also intended to estimate research findings, however, it can be easily adjusted to the needs of a customer. In addition, it helps to monitor research performance and compare it with other organizations worldwide.

InCites is based on the Web of Science citation Index data as a source and allows for both a broad analysis of the organization's place in the research process, and making sample data slices that demonstrate certain aspects of scientific performance.

The InCites scientometric tool will easily help you identify the most productive joint projects. Scientometric data on countries will help determine where scientific activity is developing most rapidly, and identify opportunities for new cooperation.

InCites allows you to analyze the productivity of institutions and compare the activities of the organization and its employees with the activities of colleagues in Russia and around the world.

All employees and students of St. Petersburg State University have the opportunity to use InCites from all computers of St. Petersburg State University and by usernames and passwords from home computers [3].

InCites provides the ability to create both standard and custom reports with a specified set of parameters. The following features are available in the main menu item "Research performance profiles":

- Overview and summary metrics;
- Productivity and Researcher output;
- Collaboration and research networks;
- Specialization and field strengths;
- Trends and time series analysis;

- Impact and citation ranking.

All data for compiling Research performance profiles is updated quarterly.

### 3. Analytical system "Research in view»

At the end of February 2011, Thomson Reuters released a new analytical product-the "Research in view" system, as well as "InCites" focused on the organization. Technologically, "Research in view" differs from the "InCites" system: "InCites" is a web application, "Research in view" is considered to be an integrated system that is installed on the organization's server, but is supported by the "Web of knowledge" servers via a client server.

"Research in view" makes it possible not only to analyze the scientific performance of the organization itself or its individual employees (authors of publications), but also to introduce the organizational structure of the institution into the analysis. Thus, it is possible to analyze the publication activity of the laboratory, department, faculty, department and other structural divisions of the organization. "Research in view" implements almost the same features as "InCites", but the organization and its individual divisions can enter new (external) data on publications in addition to those that the system borrows from bibliographic sources located on the "Web of knowledge" platform.

Thus, the organization can analyze the effectiveness of its scientific activities, the activities of its scientific departments and its employees, based on a cumulative array of data, some of which is taken from the "Web of knowledge", and the other part is entered into the system by the organization itself and reflects other types of publications – monographs, scientific reports, patents, dissertations, textbooks, etc.

### 4. Information and analytical system Nauka-MSU (ISTINA)

The ISTINA system is designed for permanent collection and systematization, storage and analysis of scientometric information in universities and scientific organizations for the purpose of preparing and making managerial decisions.

In order to automate the accounting of published publications, the scientific information system "Nauka-MSU" (previously called "Truth") was put into operation (in pilot mode) in May 2012. This system allows employees to enter their own published scientific publications, as well as other achievements in the scientific and pedagogical field [4].

Objectives of the system:

- collection, systematization and storage of information describing the results of the activities of scientific organizations;
- statistical and semantic analysis of data, issuing information at the request of various categories of users: individual scientists, scientific secretaries, heads of departments.

Basic principles:

- Distributed content: all information is entered in a decentralized manner, by specific registered users;
- Publicity: all entered information is open and accessible to all Internet users;
- Verifiability: entered publications in highly rated journals are manually checked (when determining incentive allowances), to account for citations, publications are automatically checked in WOS (when calculating scientometric indicators for re-election).

To systematize and analyze the publications introduced by MSU employees in the "Science-MSU" system, a special toolkit was created with generalized information on publications among departments and fields of science.

The ISTINA system allows you to add bibliographic records from external databases in semi-automatic mode, such as:

- Web of Knowledge / Web of Science;
- ResearcherID.com;
- Scopus;
- eLibrary.ru;

- PubMed;
- MathSciNet and other databases of mathematical articles;
- Websites of publishing houses (Elsevier, Springer, etc.)

## 5. Map of Russian Science

The Map of Russian Science is an information system designed to regularly automatically update information about scientists and organizations (including their performance indicators), perform statistical analysis of research activity, and provide a basis for creating analytical materials on the state of the Russian research and development sector. The information system "Map of Russian Science" uses only publicly available data, officially obtained from the following sources:

- Scientific Electronic Library LLC (scientific publications included in the Russian Science Citation Index);
- Thomson Scientific Inc. (scientific publications indexed in the Web of Science database and an array of data on foreign patents issued to Russian organizations);
- Federal State Budgetary Institution "Federal Institute of Industrial Property" (information on patents for inventions, utility models and industrial designs);
- Federal State Budgetary Institution of Science "Russian Book Chamber" (information on monographs, textbooks for universities, textbooks for universities and collections of scientific works).

From the very beginning, it is an ambiguous and criticized project. The purpose of the system is to optimize the control of scientific activities in Russia by the state through the means of scientists themselves.

According to the statement of the Council for Science under the Ministry of Education and Science of the Russian Federation of 31.01.2017 on the competition of scientific projects carried out within the framework of the state task in the universities subordinate to the Ministry of Education and Science, as a result of the evaluation of the activities of working with the information system, the

unsatisfactory quality of the tool for four years of its existence was established. The Council called on the Ministry of Education and Science of the Russian Federation not to use the "Map of Russian Science" for any purpose, but to use a list of generally accepted databases for various fields.

In general, the following conclusions can be drawn from all of the above:

1. In general, all products are online services for displaying a variety of scientometric indicators for user-selected sets of entities of authors, organizations, countries, etc. They are equipped with built-in data visualization tools that facilitate and speed up analysis, as well as allow you to export tables with indicators.

2. The value of SciVal and InCites is due to the commercial nature of the activities of Thomson Reuters and Elsevier in fact, these are analytical add-ons for the Web of Science (WoS) and Scopus, which are placed in separate expensive products. At the same time, in WoS and Scopus themselves, the capabilities for analytics and data export are significantly limited.

3. In SciVal with InCites, there are many strict limits on data uploads, so for large-scale research, you still need to purchase uploads, connect the API, or manually pump raw data in pieces of the allowed size from WoS and Scopus.

4. Another drawback of modern SciVal and InCites tools is that they are based on foreign publications and they practically do not take into account Russian journals.

5. With regard to Russian tools, the ISTINA system is aimed at generating reports and presenting data in various formats, as well as analyzing statistical indicators by division and by topic for MSU employees.

#### **1.4. Development of scientific activities in organization**

The need for new approaches to the management of scientific organizations is discussed in the Strategy of Scientific and Technological Development of the Russian Federation. The management of scientific activity includes the organization, planning, management and coordination of human and material

resources to achieve certain results in research in terms of the composition and scope of work, cost, time and quality.

An analytical review allows us to conclude that based on the data from bibliographic databases, you can build such analytical tools that will help you make management decisions in the field of scientific activity development more quickly and efficiently. Such tools should organize data from all bibliographic databases so that the end user can rely on the results of the work of intelligent analytical tools and maps of science to make an expert assessment. This makes it possible to use the method of making managerial decisions based on the analysis of big data to find the most effective way to develop scientific activities. This approach ensures the validity and objectivity of the decisions made and, as a rule, qualitatively improves management activities.

#### **List of sources used**

1. Fostering Integrity in Research [Electronic resource] // The National Academic Press - 2017. -URL: <https://www.nap.edu/read/21896/chapter/6>
2. Tazeem Zainab, Zahid Wani. Advancement and Application of Scientometric Indicators for Evaluation of Research Content [Electronic resource] // ResearchGate - URL: [https://www.researchgate.net/publication/318432935\\_Advancement\\_and\\_Application\\_of\\_Scientometric\\_Indicators\\_for\\_Evaluation\\_of\\_Research\\_Content](https://www.researchgate.net/publication/318432935_Advancement_and_Application_of_Scientometric_Indicators_for_Evaluation_of_Research_Content)
3. Wolfgang Glänzel. Bibliometrics as a research field: A course on theory and application of bibliometric indicators [Electronic resource] // ResearchGate - 2003. -URL: [https://www.researchgate.net/publication/242406991\\_Bibliometrics\\_as\\_a\\_research\\_field\\_A\\_course\\_on\\_theory\\_and\\_application\\_of\\_bibliometric\\_indicators](https://www.researchgate.net/publication/242406991_Bibliometrics_as_a_research_field_A_course_on_theory_and_application_of_bibliometric_indicators)
4. Dag W. Aksnes, Liv Langfeldt, Paul Wouters. Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories

[Electronic resource]// SAGE Journals - URL:  
<https://journals.sagepub.com/doi/full/10.1177/2158244019829575>

## Приложение Б

```
// Статистика по числу публикаций типа «статья» по университету
SELECT round((count(publications.id) FILTER (WHERE
publications.type::text = 'Article'::text)::double precision
count(publications.id)::double precision * 100::double precision)::numeric,
2) AS article_percent,
date_part('year'::text, publications.date_of_publication) AS year
FROM publications
GROUP BY (date_part('year'::text, publications.date_of_publication))
ORDER BY (date_part('year'::text, publications.date_of_publication));

// Статистика публикационной активности по объединению авторов
SELECT author_unions.id as author_union_id,
author_unions.title as author_union_title,
count(publications.id) as count_publications,
sum(publications.cited_count) as count_citations,
ROUND((sum(publications.cited_count)/count(publications.id)::float)
::decimal,2) as average_citation_rate,
extract(year FROM publications.date_of_publication) as year
FROM author_unions
LEFT JOIN author_union_authors ON
author_union_authors.author_union_id = author_unions.id
LEFT JOIN authors ON authors.id = author_union_authors.author_id
LEFT JOIN publication_authors ON publication_authors.author_id =
authors.id
LEFT JOIN publications ON publication_authors.publication_id =
publications.id
WHERE affiliation_id IS NOT NULL
GROUP BY author_unions.id,
author_unions.title,
```

```

    extract(year FROM publications.date_of_publication)
    ORDER BY extract(year FROM publications.date_of_publication)
// Статистика по числу публикаций типа «статья» по отдельным авторам
    SELECT authors.id AS author_id,
    authors.name AS author_name,
    round((count(publications.id) FILTER (WHERE publications.type::text =
'Article'::text)::double precision / count(publications.id)::double precision *
100::double precision)::numeric, 2) AS article_percent,
    date_part('year'::text, publications.date_of_publication) AS year
    FROM authors
    LEFT JOIN publication_authors ON publication_authors.author_id =
authors.id
    LEFT JOIN publications ON publication_authors.publication_id =
publications.id
    WHERE authors.affiliation_id IS NOT NULL
    GROUP BY (date_part('year'::text, publications.date_of_publication)),
    authors.id,
    authors.name
    ORDER BY (date_part('year'::text, publications.date_of_publication));
// Статистика по числу публикаций типа «статья» по отдельным авторам
    SELECT authors.id AS author_id,
    authors.name AS author_name,
    count(publications.id) AS count_publications,
    sum(publications.cited_count) AS count_citations,
    date_part('year'::text, publications.date_of_publication) AS year
    FROM authors
    LEFT JOIN publication_authors ON publication_authors.author_id =
authors.id

```

```

LEFT JOIN publications ON publication_authors.publication_id =
publications.id
WHERE authors.affiliation_id IS NOT NULL
GROUP BY (date_part('year'::text, publications.date_of_publication)),
authors.name,
authors.id
ORDER BY (date_part('year'::text, publications.date_of_publication));
// Статистика по публикациям в различных источниках по отдельным
авторам
SELECT
    authors.id as author_id,
    authors.name as author_name,
    COUNT(publications.id) as count_publications,
    publications.source_id,
    sources.title as source_title,
    sources.country as source_country,
    sources.latitude as source_latitude,
    sources.longitude as source_longitude
FROM authors
LEFT JOIN publication_authors on publication_authors.author_id =
authors.id
LEFT JOIN publications on publications.id =
publication_authors.publication_id
LEFT JOIN sources on sources.id = publications.source_id
WHERE affiliation_id IS NOT NULL
GROUP BY publications.source_id,
    sources.title,
    authors.id,
    authors.name,

```

*sources.country,*  
*sources.latitude,*  
*sources.longitude*  
**ORDER BY** *country;*

