


Школа Инженерная школа ядерных технологий
 Направление подготовки 01.03.02 «Прикладная математика и информатика»
 Отделение школы (НОЦ) Отделение экспериментальной физики

БАКАЛАВРСКАЯ РАБОТА

Тема работы
Формирование научных групп при анализе извлеченных данных из наукометрической базы Scopus

УДК 004.65:001.891:330.43

Студент

Группа	ФИО	Подпись	Дата
0В71	Рахмонкулов Салохиддин Мухсумович		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОЭФ ИЯТШ	Крицкий О. Л.	Кандидат ф-м. наук, доцент		

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Спицына Любовь Юрьевна	Кандидат экономических наук		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент ООД	Мезенцева Ирина Леонидовна	-		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Руководитель ООП 01.03.02 «Прикладная математика и информатика»	Крицкий Олег Леонидович	Кандидат ф-м. наук, доцент		

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

Школа Инженерная школа ядерных технологий
 Направление подготовки 01.03.02 «Прикладная математика и информатика»
 Отделение школы (НОЦ) Отделение экспериментальной физики

УТВЕРЖДАЮ:
 Руководитель ООП
26.04.2021 Крицкий О.Л.
 (Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ на выполнение выпускной квалификационной работы

В форме:

Бакалаврской работы

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

Студенту:

Группа	ФИО
0В71	Рахмонкулову Салохиддину Мухсумовичу

Тема работы:

Формирование научных групп при анализе извлеченных данных из наукометрической базы Scopus

Утверждена приказом директора (дата, номер)

Срок сдачи студентом выполненной работы:

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<p>Исходные данные к работе</p> <p><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></p>	<p>Данные о публикациях и авторах, извлеченные из наукометрической базы данных Scopus.</p>
<p>Перечень подлежащих исследованию, проектированию и разработке вопросов</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"> 1. Изучить основы технологии big data, научиться работать с большими данными. 2. Изучить алгоритмы парсинга данных. 3. Провести анализ данных наукометрической базы данных Scopus в области «Economic, Econometric and Finance», выделить группы авторов. 4. Проанализировать найденные кластеры, построить их визуальное представление, выделить кластеры. Определить структуру

	сотрудничества между исследователями в области «Экономики, Эконометрики и Финансов» (ECON) 5. Анализ полученных результатов.
Перечень графического материала <i>(с точным указанием обязательных чертежей)</i>	
Консультанты по разделам выпускной квалификационной работы <i>(с указанием разделов)</i>	
Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Спицына Любовь Юрьевна
Социальная ответственность	Мезенцева Ирина Леонидовна

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	
---	--

Задание выдал руководитель / консультант (при наличии):

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОЭФ ИЯТШ	Крицкий Олег Леонидович	Кандидат ф-м. наук, доцент		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
0В71	Рахмонкулов Салохиддин Мухсумович		

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

Студенту:

Группа	ФИО
	Рахмонкулову Салохиддину Мухсумовичу

Школа	ИЯТШ	Отделение школы (НОЦ)	Отделение экспериментальной физики
Уровень образования	Бакалавриат	Направление/специальность	01.03.02 «Прикладная математика и информатика»

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

1. <i>Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i>	Бюджет проекта – не более 65958,95 руб., в т.ч. затраты по оплате труда – не более 41197 руб.
2. <i>Нормы и нормативы расходования ресурсов</i>	30% премии; 20% надбавки; 16% накладные расходы; 30% районный коэффициент.
3. <i>Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	Отчисления во внебюджетные фонды – 30%.

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. <i>Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения</i>	1. Потенциальные потребители результатов исследования; 2. Анализ конкурентных технических решений; 3. SWOT – анализ
2. <i>Планирование и формирование бюджета научных исследований</i>	1. Структура работ в рамках научного исследования; 2. Определение трудоемкости выполнения работ и разработка графика проведения научного исследования; 3. Бюджет научно - технического исследования (НТИ).
4. <i>Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования</i>	1. Определение интегрального показателя ресурсоэффективности разработки.

Перечень графического материала (с точным указанием обязательных чертежей):

1. <i>Оценка конкурентоспособности технических решений</i> 2. <i>Матрица SWOT</i> 3. <i>Альтернативы проведения НИ</i> 4. <i>График проведения и бюджет НИ</i> 5. <i>Оценка ресурсной, финансовой и экономической эффективности НИ</i>	
--	--

Дата выдачи задания для раздела по линейному графику

--	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ОСГН ШБИП ТПУ	Спицына Любовь Юрьевна	к.э.н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
0B71	Рахмонкулов Салохиддин Мухсумович		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
0B71	Рахмонкулов Салохиддин Мухсумович

Школа	ИЯТШ	Отделение (НОЦ)	Экспериментальной физики
Уровень образования	Бакалавриат	Направление/специальность	01.03.02 «Прикладная математика и информатика»

Тема ВКР:

Формирование научных групп при анализе извлеченных данных из наукометрической базы Scopus	
Исходные данные к разделу «Социальная ответственность»:	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	<i>Объектом исследования являются данные о публикациях и авторах, извлеченные из наукометрической базы данных Scopus. Данная работа направлена на определение структур сотрудничества между исследователями в области «Экономики, Эконометрики и Финансов» (ECON) и выявлении учреждений, направленных на данную тематику.</i>
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. Правовые и организационные вопросы обеспечения безопасности: – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны.	– Трудовой кодекс Российской Федерации от 30.12.2001 N 197-ФЗ (ред. от 09.03.2021). – ГОСТ 12.2.032-78 Рабочее место при выполнении работ сидя.
2. Производственная безопасность: 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	Вредные факторы: – отклонение параметров микроклимата в помещении; – недостаточная освещённость рабочей зоны; – повышенный уровень шума на рабочем месте. Опасные факторы: – опасность поражения электрическим током.
3. Экологическая безопасность:	– Атмосфера: загрязнение воздуха выбросами с электростанций;

	– Литосфера и гидросфера: наличие отходов (бумага, картриджи, компьютеры и т. д.).
4. Безопасность в чрезвычайных ситуациях:	Возможные ЧС: – природного характера: повышение уровня воды, град; – техногенного характера: обрушение зданий, пожары. Наиболее типичная ЧС: пожар.

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

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Мезенцева Ирина Леонидовна	-		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
0В71	Рахмонкулов Салохиддин Мухсумович		

Реферат

Выпускная квалификационная выполнена на 63 страницах, содержит 20 таблиц, 11 рисунков, 15 источников, 2 приложения.

Ключевые слова: кластерный анализ; Elsevier Api; взвешенный ориентированный граф; алгоритм PageRank; Экономика, Эконометрика и Финансы.

Объект исследования: данные о публикациях и авторах, извлеченные из наукометрической базы данных Scopus.

Цель работы: Формирование научных групп при анализе извлеченных данных из наукометрической базы Scopus.

Методы проведения работы: теоретические (изучение литературы, обзор методов и моделей анализа) и практические (применение методов для построения модели).

В результате исследования: проведен анализ найденных кластеров, построено их визуальное представление, выделены кластеры, содержащие самое большое количество учёных.

Бакалаврская работа написана в Microsoft Word 2015. Для получения данных использовался скрипт, написанный на языке программирования Python. Для хранения полученных данных использовалась реляционная система управления базами данных PostgreSQL. Для формирования графов и обработки их использовалась библиотека igraph (язык программирования R). Для отрисовки графа использовалась программа Gephi.

Для достижения поставленной цели были сформулированы следующие задачи:

1. Извлечь данные из библиографической базы данных Scopus данные за 2011-2020 гг. в области «Economic, Econometric and Finance» (ECON).
2. Выделить группы авторов. Проанализировать найденные кластеры, построить их визуальное представление, выделить кластеры.
3. Определить структуру сотрудничества между исследователями в области «Экономики, Эконометрики и Финансов» (ECON)
4. Провести анализ полученных результатов.

Оглавление

Введение.....	11
Обзор литературы	13
1. Теоретическая часть.....	15
1.1. CiteScore	15
1.2. Реляционная база данных.....	16
1.1. Application Programming Interface.....	17
1.2. Кластерный анализ.....	17
1.3. Облако тегов	19
1.4. Алгоритм PageRank.....	20
2. Практическая часть	22
2.1. Разработка структуры базы данных	22
2.2. Выгрузка данных.....	22
2.3. Разведочный анализ данных	23
2.4. Построение облака тегов.....	27
2.5. Построение графа.....	29
2.6. Выделение кластеров учёных и их анализ	30
3. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ.....	35
3.1 Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения	35
3.1.1 Потенциальные потребители результатов исследования	35
3.1.2 Анализ конкурентных технических решений.....	37
1.1.1. SWOT-анализ	38
3.2 Планирование научно-исследовательских работ	38
3.2.1 Структура работ в рамках научного исследования.....	38
3.2.2 Определение трудоемкости выполнения работ и разработка графика проведения научного исследования	39
3.3 Бюджет научно-технического исследования (НТИ)	42
3.3.1 Расчёт материальных затрат НТИ.....	42

3.3.2	Расчёт затрат на специально оборудование для научных (экспериментальных) работ	43
3.3.3	Основная заработная плата.....	43
3.3.4	Отчисления во внебюджетные фонды (страховые отчисления) 44	
3.3.5	Накладные расходы	45
3.3.6	Формирование бюджета затрат НТИ.....	46
3.4	Определение ресурсной эффективности исследования	46
3.5	Выводы по разделу	47
4.	Социальная ответственность.....	48
4.1.	Введение	48
4.2.	Правовые и организационные вопросы обеспечения безопасности 48	
4.3.	Производственная безопасность.....	49
4.4.	Анализ опасных и вредных производственных факторов.....	49
4.4.1.	Отклонение параметров микроклимата в помещении.....	49
4.4.2.	Недостаточная освещённость рабочей зоны.....	51
4.4.3.	Повышенный уровень шума на рабочем месте	52
4.4.4.	Опасность поражения электрическим током.....	53
4.4.5.	Обоснование мероприятий по снижению уровней воздействия опасных и вредных факторов на исследователя (работающего)	54
4.5.	Экологическая безопасность.....	54
4.6.	Безопасность в чрезвычайных ситуациях.....	55
4.7.	Выводы по разделу.....	56
	Заключение	57
	Список используемых источников.....	58
	Приложение А	60
	Приложение Б.....	63

Введение

Объектом исследования являются данные о публикациях и авторах, извлеченные из наукометрической базы данных Scopus, в области «Economic, Econometric and Finance».

Цель работы – провести анализ данных и выделить группы авторов, работающих в одной предметной области. Изучить основы технологии big data, научиться работать с большими данными и обрабатывать их, изучить алгоритмы парсинга данных.

В результате исследования проведен анализ найденных кластеров, построено их визуальное представление, выделены кластеры, содержащие самое большое количество учёных. Дана интерпретация полученных результатов и сделаны соответствующие выводы.

Теперь, как никогда ранее научное сообщество располагает огромным количеством материала. Согласно некоторым исследованиям, объём производства научных публикаций удваивается каждые девять лет.

Благодаря быстрому развитию информационно-коммуникационных технологий (ИКТ), теперь можно получить доступ к большей части научной литературы через интернет, что позволяет более эффективно передавать знания. Однако этот процесс также ведёт к усилению конкуренции между исследователями и увеличению требований к качеству материала. Эти изменения побуждают исследователей, особенно тех, кто находится на ранних этапах своей карьеры, сотрудничать с другими исследователями из своего или другого учреждения. Динамичность такого партнёрства делает анализ сети научного сотрудничества предметом большого интереса.

В основном для оценки автора используются индикаторы, полученные через поисковые системы, такие как Google Scholar, Scopus и т.д., которые включают количество и качество публикаций на основе журнальных и авторских показателей. Тем не менее, эти критерии не учитывают взаимосвязь между авторами. Хотя глобализация науки в современном мире обусловлена

сотрудничеством между исследователями и это стало большой проблемой анализа эффективности работы конкретных авторов.

Данная работа направлена на определение структур сотрудничества между исследователями в области «Экономики, Эконометрики и Финансов» (ECON) и выявлении учреждений, направленных на данную тематику. В качестве основного критерия используется сотрудничество в публикациях. Доступ к данным осуществляется с помощью интерфейса Elsevier API.

Обзор литературы

1. Бин Анналин, Су Кеннет. [1] — Хотя название говорит о больших данных, большая часть книги рассказывает о data science и её основах, а также об алгоритмах и их описании.
2. Мартышкин А. И., Бершадская Е. Г. [6] — В статье рассматривается процесс подготовки к исследованию неструктурированных данных, полученных из социальных сетей, с использованием технологии Big Data, под термином которой принято понимать любые наборы данных, достаточно большие и сложные для того, чтобы их можно было обработать традиционными средствами работы с данными. Представлено описание web-интерфейса программной системы. Описано назначение и функционал разработанной системы с детализацией по вкладкам программы.
3. Казиев Г. З., Курдюков В. В. [5] — В работе приводится описание методов кластеризации BIGDATA, представлены ключевые понятия технологии Больших данных: основные характеристики, модели, методы, этапы технологии, сферы применения технологии. Проанализирован современный рынок использования технологии BIGDATA. Выполнен сравнительный анализ суперкомпьютеров. В рамках проведённых исследований разработана функциональная схема технологии, рассмотрены методы кластеризации больших объёмов данных.
4. Юре Лесковец, Ананд Раджараман, Джеффри Д. Ульман. [2] — В книге описываются алгоритмы, которые использовались для решения задач добычи данных и могут быть с успехом применены даже к очень большим наборам данных. Изложение начинается с рассмотрения технологии MapReduce – важного средства распараллеливания алгоритмов. Излагаются алгоритмы хэширования с учетом близости и потоковой обработки данных, которые поступают слишком быстро для тщательного анализа. В последующих главах рассматривается идея показателя PageRank, нахождение частых предметных наборов и кластеризация.

5. Сенько А. [3] — Основное внимание уделено решениям Microsoft Azure и AWS. Рассматриваются все этапы работы — получение данных, подготовленных для обработки в облаке, использование облачных хранилищ, облачных инструментов анализа данных. Особое внимание уделено службам SAAS, продемонстрированы преимущества облачных технологий по сравнению с решениями, развернутыми на выделенных серверах или в виртуальных машинах.
6. Зайцева Е. А. [7] — В статье рассмотрена возможность применения аналитической платформы IBM InfoSphere BigInsights, реализующей вышеприведенные технологии, в сочетании с IBM i2, которая позволяет работать с нереляционными графовыми базами данных, с целью построения социального профиля клиента на основе открытых источников информации.
7. Elsevier [4] — Предоставлена документация по работе с API Scopus. Пошаговые технические руководства, которые помогут начать работу с API Elsevier. Каждое руководство знакомит с подходом, обычно используемым для конкретного варианта использования, объясняя, какие вызовы API необходимо выполнить в какой точке приложения и для какой цели.

1. Теоретическая часть

Выборка статей осуществляется из журналов, которые имеют предметную область «Economic, Econometric and Finance» и показатель CiteScore > 0.5. Данный показатель помогает отсеять журналы, имеющие низкую цитируемость статей. Для анализа используются статьи за последние 10 лет (2011-2021).

Для полного анализа необходима следующая информация:

- информация о журнале, в котором была опубликована статья;
- информация о статье;
- информация об авторе;
- информация о принадлежности автора к университету или другой организации.

1.1. CiteScore

CiteScore — количественная оценка уровня цитируемости мест публикации, например журналов. [4]

Расчет рейтинга CiteScore основывается на среднем количестве цитирований одного документа. Рейтинг CiteScore — это количество цитирований за год документов, опубликованных в журнале за предыдущие три года, деленное на количество документов, проиндексированных в Scopus, которые были опубликованы за эти же три года. [4]

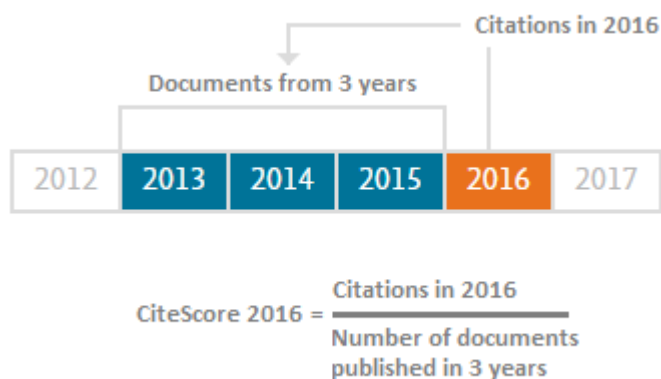


Рисунок 1. Расчёт CiteScore за 2016 г.

Например, рейтинг CiteScore за 2016 г. рассчитывается как количество цитирований в 2016 г. документов, которые были опубликованы в 2013, 2014 и 2015 г.г., деленное на количество проиндексированных в Scopus документов, которые были опубликованы в 2013, 2014 и 2015 г.г. [4]

CiteScore использует 3-летнее временное окно. Согласно данным рецензируемой библиометрической литературы, 3 года — это наилучший компромисс для базы данных с широким охватом, позволяющий включить показательную долю цитирований для всех дисциплин, но при этом отражающий относительно свежие данные. [4]

1.2. Реляционная база данных

Для данного проекта было принято решение использовать реляционную базу данных, т.к.

Среди популярных РСУБД пал выбор на PostgreSQL – СУБД с открытым исходным кодом. Она поддерживает большую часть стандарта SQL, позволяет проводить быстрый анализ большого количество информации, имеет в наличии множество дополнительных типов данных и удобный механизм работы с XML, JSON форматами, а также предлагает множество полезных функций:

- complex queries (сложные запросы);
- foreign keys (внешние ключи);
- triggers (триггеры);
- materialized views (изменяемые представления);

- transactional integrity (транзакционная целостность);
- multiversion concurrency control (многоверсионность);
- indexes (индексы).

Создание индексов позволяет ускорить выборку и обработку данных средствами СУБД. Например, для данной работы скорость выполнения запросов увеличилась до 3 раз. Для больших таблиц это существенное ускорение.

1.1. Application Programming Interface

API (программный интерфейс приложения, интерфейс прикладного программирования) (англ. application programming interface, API [эй-пи-ай]) — интерфейс прикладного программирования, который определяет взаимодействие между несколькими программными посредниками. API определяет виды вызовов или запросов, которые могут быть сделаны, как их сделать, форматы данных, которые следует использовать, соглашения, которым следует следовать, и т. д. API может быть полностью настраиваемым, специфичным для компонента, или он может быть разработан на основе отраслевого стандарта для обеспечения совместимости. Благодаря скрытию информации API обеспечивают модульное программирование, что позволяет пользователям использовать интерфейс независимо от реализации.

При использовании в контексте веб-разработки API обычно определяется как набор спецификаций, таких как сообщения запроса протокола передачи гипертекста (HTTP), наряду с определением структуры ответных сообщений, обычно в расширяемом языке разметки (XML) или объектной нотации JavaScript (JSON) формат.

Доступ к базе данных Scopus осуществляется с помощью Elsevier API, с примерами запросов (request) и ответов (response) можно ознакомиться по в приложении А.

1.2. Кластерный анализ

Кластерный анализ – это анализ, позволяющий получить разбиение большого объема данных на классы или группы (от англ. cluster = класс) согласно некоторому критерию или их совокупности. При этом считается, что отсутствует дополнительная информация о характере исходных данных, т.е. они анализируются «как есть». Кластерный анализ позволяет:

- 1) провести классификацию объектов с учетом признаков, отражающих их сущность;
- 2) проверить предположения о наличии некоторой структуры в совокупности этих объектов;
- 3) построить новые классификации для слабоизученных явлений, когда необходимо установить наличие связей внутри совокупности и структурировать их.

Для проведения классификации данных используют понятие метрики или расстояния.

Метрикой называется функция, отображающая некоторое метрическое пространство в пространство действительных чисел и обладающая следующими свойствами (аксиомами метрики):

- 1) $p(x, y) = 0 \Leftrightarrow x = y$ (аксиома тождества)
- 2) $p(x, y) = p(y, x)$ (аксиома симметрии)
- 3) $p(x, z) \leq p(x, y) + p(y, z)$ (аксиома треугольника)

В теории кластерного анализа используются следующие метрики для измерения расстояния между отдельными точками:

- 1) евклидово расстояние;
- 2) взвешенное евклидово расстояние;
- 3) хеммингово расстояние;
- 4) расстояние Махаланобиса;
- 5) расстояние Минковского.

В ряде алгоритмов наряду с расстояниями между векторами используются расстояниями между кластерами и объединениями кластеров:

- 1) расстояние между кластерами по принципу «ближнего соседа»;
- 2) расстояние между кластерами по принципу «дальнего соседа»;
- 3) расстояние между центрами тяжести групп;
- 4) расстояние между кластерами по принципу «средней связи»;
- 5) обобщённое расстояние по Колмогорову.

1.3. Облако тегов

Это визуальное представление списка ключевых слов (или тегов, также называемых метками, ярлыками, и т. п.). Ключевые слова чаще всего представляют собой отдельные слова или небольшие словосочетания, и важность каждого ключевого слова обозначается размером шрифта или цветом. Такое представление удобно для быстрого восприятия часто используемых терминов и для распределения ключевых слов по популярности относительно друг друга.

Для построения облаков тегов пользуются кластеризацией, в частности методом k-средних. Для отображения облака тегов нужно разбить все ключевые слова на группы (кластеры) по частоте их появления и изменять размер слов для каждой группы отдельно.

Облако текста или облако слов — это визуализация частоты слов в тексте в виде взвешенного списка.

Облака словосочетаний. Развивая принципы облаков текста, облака словосочетаний дают более сфокусированный вид текста или набора текстов. Полученное облако содержит слова, которые часто используются в сочетании с исходным словом. Эти словосочетания представляют частоту, выделяемую размером шрифта, наряду с силой связи, выделяемой яркостью. Это позволяет интерактивно изучать язык.

Реализация облака тегов также включает синтаксический анализ фильтрацию ненужных тегов, таких как предлоги, местоимения, числа и знаки препинания.



Рисунок 2. Набор слов из списка 1000 избранных статей Википедии, упорядоченный по количеству просмотров.

1.4. Алгоритм PageRank

PageRank – это алгоритм ссылочного ранжирования, разработанный для определения относительной важности объекта, связанного с сетью объектов. Ссылочное ранжирование – это тип сетевого анализа, определяющий ассоциации (связи) между объектами.

Рассмотрим ориентированный взвешенный граф. Пусть у графа n вершин. Каждой паре вершин соответствует некоторый вес (вероятность перехода).

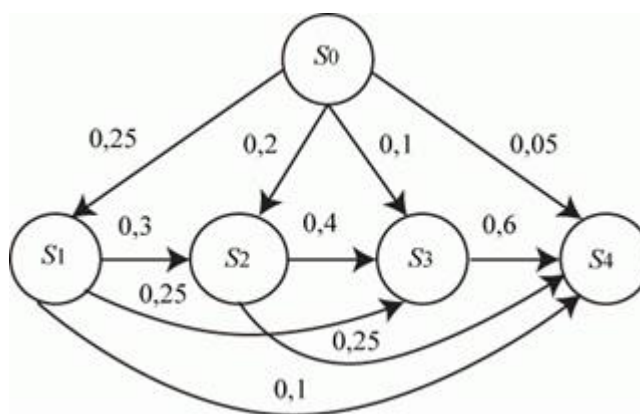


Рисунок 3. Ориентированный взвешенный граф.

Запишем уравнение Колмогорова-Чэпмена:

$$p^T(t + 1) = p^T(t)P, \quad (2.6.1)$$

где P – переходная матрица.

Предположим, что существует предел

$$\lim_{t \rightarrow \infty} (p^t) = v$$

Тогда:

$$v^T = v^T P. (2.6.2)$$

Вектор v называют стационарным распределением вероятности.

Именно из соотношения (2.6.2) и было предложено искать вектор ранжирования web-страниц в модели Брина-Пейджа.

Так как v является пределом, то стационарное распределение можно вычислить методом простых итераций (МПИ).

Для того чтобы перейти с одной web-страницы на другую, пользователь должен с определенной вероятностью выбрать, по какой именно ссылке перейти. В случае если у документа несколько исходящих ссылок, то будем считать, что пользователь переходит по каждой из них с одинаковой вероятностью. Кроме того, есть еще коэффициент телепортации, который показывает нам, с какой вероятностью пользователь может с текущего документа переместиться на другую страницу, не обязательно связанную напрямую со страницей, в которой мы находимся в данный момент. Пусть пользователь телепортируется с вероятностью δ , тогда, уравнение (2.6.1) примет вид

$$p(t + 1) = (1 - \delta) P^T p(t) + \delta \left(\frac{1}{n} \dots \frac{1}{n} \right)^T$$

При $0 < \delta < 1$ уравнение гарантированно имеет единственное решение. На практике для вычисления PageRank обычно используют уравнение с $\delta = 0,15$.

2. Практическая часть

2.1. Разработка структуры базы данных

Для эффективного анализа данных из сети было решено организовать свою базу данных. Для этого необходимо выбрать систему управления базами данных. Была выбрана реляционная система управления базами данных PostgreSQL. При разработке были выделены необходимые элементы, возвращаемые Elsevier API. Визуализация структуры БД представлена на следующем рисунке.

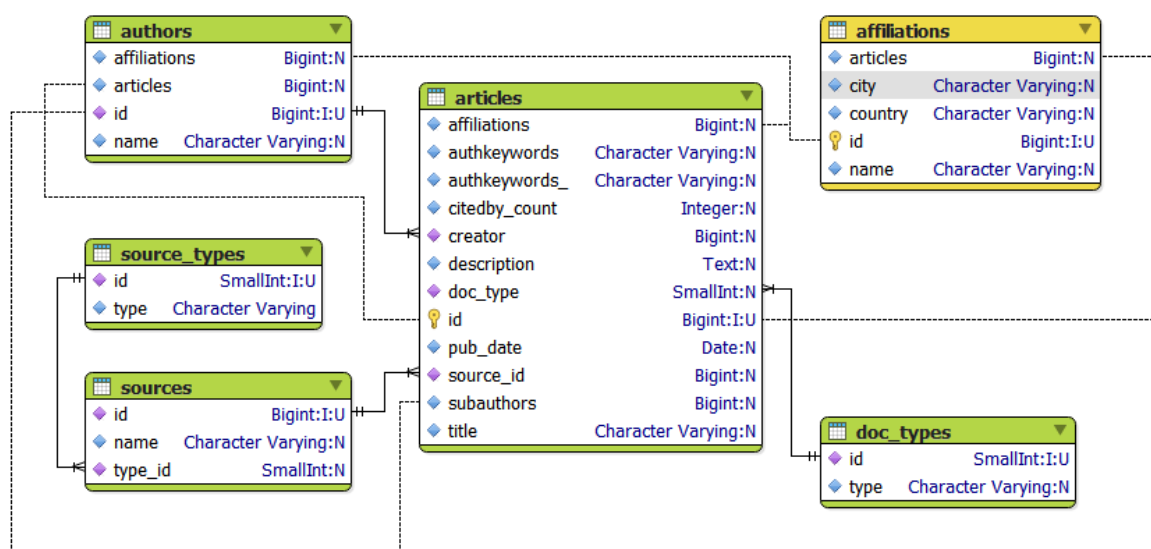


Рисунок 4. Структура БД

Центральным элементом данной структуры БД является таблица статей (articles), из неё исходят таблицы с информацией о научных организациях (affiliations), авторах работ (authors), типов документов (doc_types) и информации о научных журналах (sources).

2.2. Выгрузка данных

В данном разделе представлен метод выгрузки научных публикаций для последующего анализа. Выгрузка данных один из важнейших этапов данного исследования, так как от качества и количества данных зависит качество

работы. Данную методику можно применять для выгрузки любого допустимого по объёму хранилища данных с использованием различных скриптов. Прежде, чем выгружать данные, необходимо было понять какие журналы и статьи интересны для анализа. Для этого, из всего списка индексируемых журналов отберем те, что имеют CiteScore > 0,5. Кроме того, нас интересует конкретная область «Economic, Econometric and Finance», а также период с 2011 по 2020 годы.

Для выгрузки данных был использован автоматизированный скрипт, написанный на языке программирования Python, использующий Elsevier API. Ссылка на исходный код парсера указана в приложении Б.

Скрипт проходится по всем журналам с CiteScore > 0.5 в области «Economic, Econometric and Finance» и извлекает статьи за 10-летний период:

1. Для каждого журнала проверяем наличие интересующих нас публикаций.
2. При наличии таковых извлекаем для каждой найденной статьи следующую информацию:
 - Название статьи, описание, ключевые слова авторов, тип источника и т.д.
 - Имена авторов и их принадлежность к университету или другой организации.
 - Информацию об организации, к которой принадлежат авторы

При этом перезаписи или повторной записи в БД не будет, если какая-либо информация уже есть в базе данных.

Далее, анализируя эти данные можно установить различные зависимости, первичную информацию о статьях и т.д.

По итогу выгрузки в базе данных находились данные о 364886 статьях, из которых 19000 содержали ошибки. Статьи, содержащие неточности, не были включены в выборку для анализа.

2.3. Разведочный анализ данных

Исходя из полученных данных, были построены таблица и диаграмма распределения количества публикаций относительно года.

Таблица 1. Количество публикаций в год

Год	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Кол-во	26094	28172	30836	33420	33943	34690	36408	37210	39764	29729	43

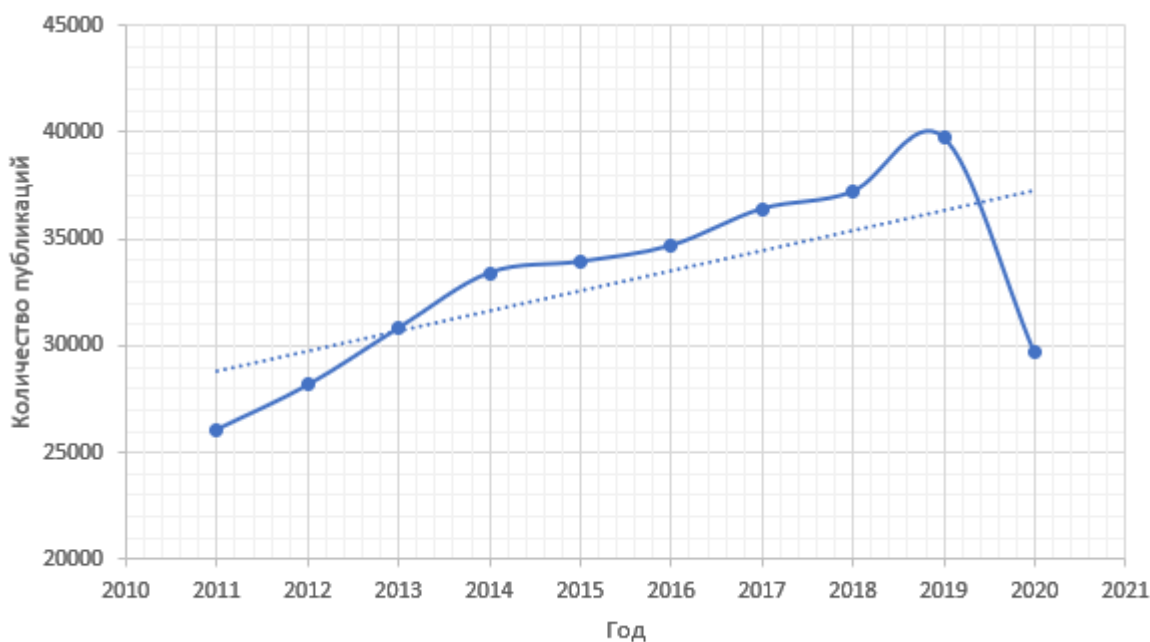


Рисунок 5. Количество публикаций в год

Можно заметить, что количество публикаций растёт с каждым годом. В 2014 и 2019 были небольшие «скачки». 2020 год на момент парсинга данных был ещё не завершён, поэтому количество публикаций за этот год существенно меньше. Но можно заметить, что 75% от количества научных работ предыдущего года уже достигнуто. Также в базе находятся статьи датированные 2021 годом.

Рассмотрим диаграмму и картограмму распределения количества публикаций по странам (рисунок 6). Лидером по количеству публикаций является США. Стоит отметить, что учитывается принадлежность статьи к научной организации, которая в свою очередь находится в определённой стране. Но при написании научной работы могут принимать участие исследователи из различных стран.

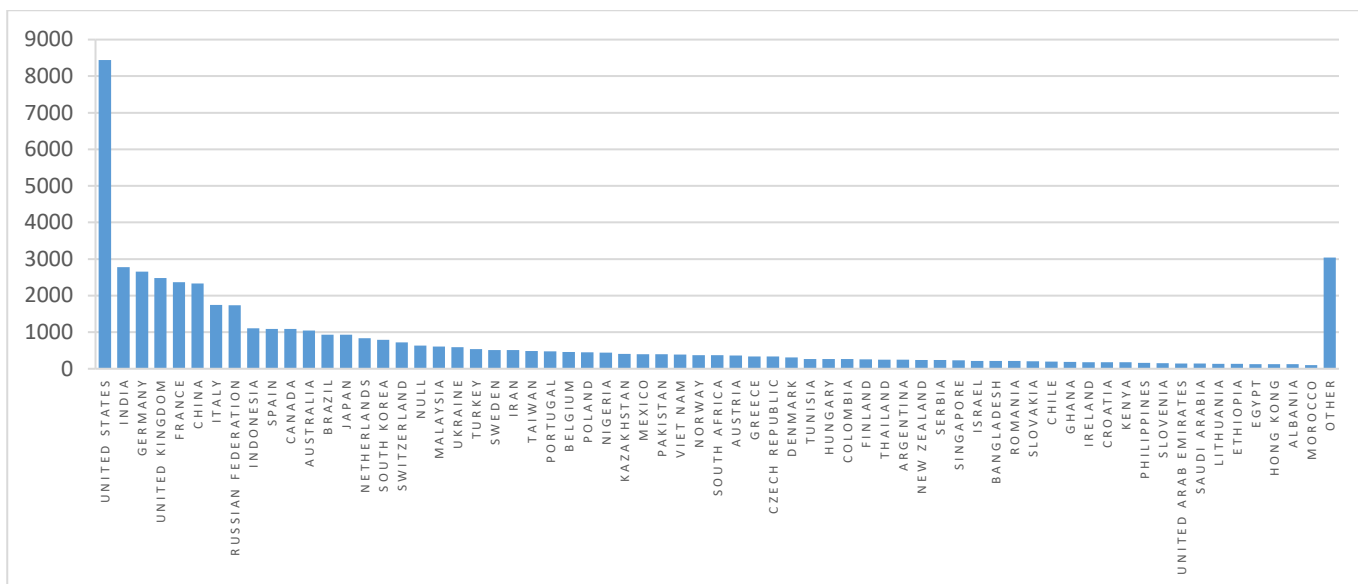


Рисунок 6. Диаграмма распределения количества публикаций по странам

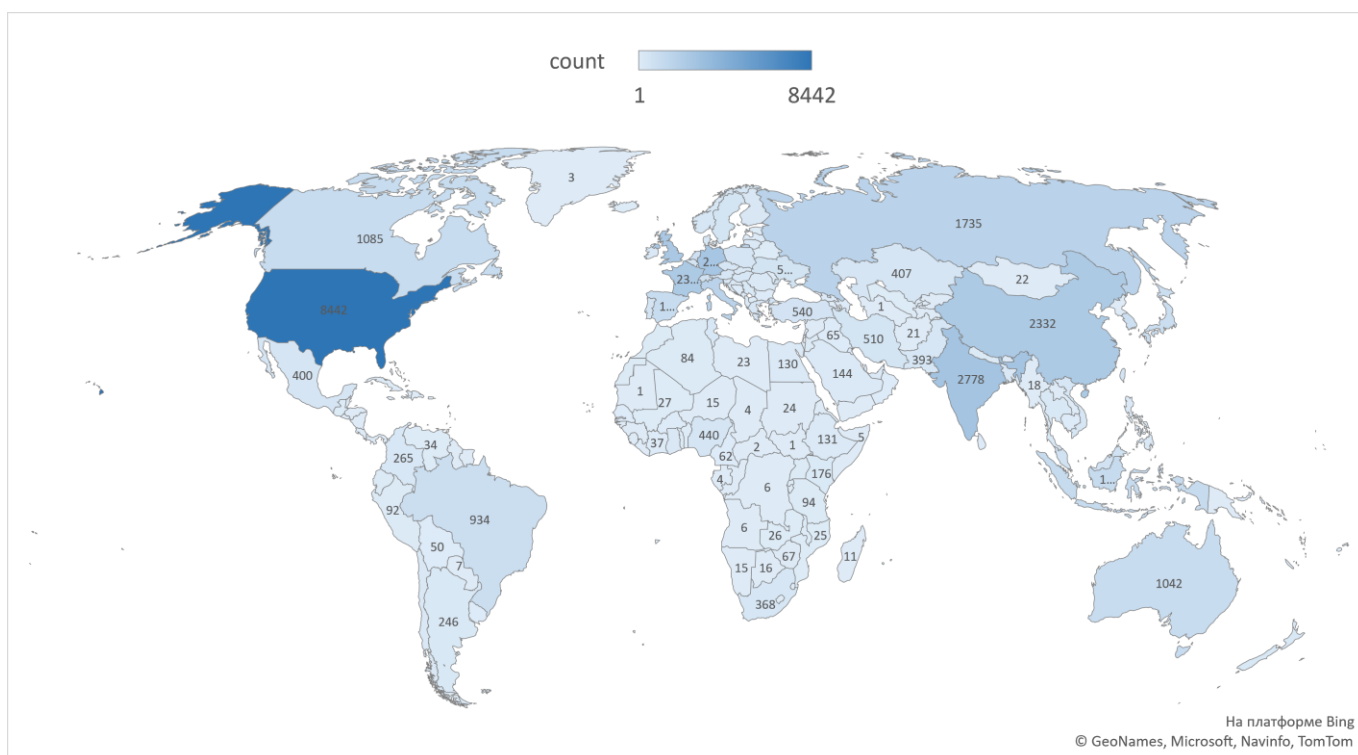


Рисунок 7. Картограмма распределения количества публикаций по странам

Рассмотрим рейтинг научных организаций и вузов по количеству публикаций, больше тысячи (таблица 2). В данной таблице наблюдается аналогичная ситуация, как и с картограммой и гистограммой, а именно большинство научных организаций из США.

Таблица 2. Топ вузов (организаций) по количеству публикаций (больше 1000)

National Bureau of Economic Research	2757
London School of Economics and Political Science	2548
The World Bank, USA	1763
University of Oxford	1628
National Research University Higher School of Economics	1551
University of New South Wales (UNSW) Australia	1472
Columbia University in the City of New York	1467
Centre for Economic Policy Research, London	1459
The University of Sydney	1449
Harvard University	1435
Institut Zur Zukunft Der Arbeit	1419
Monash University	1397
Cornell University	1369
Stanford University	1352
International Monetary Fund	1338
University of California, Berkeley	1321
Peking University	1264
UCL	1254
National University of Singapore	1229
KU Leuven	1225
Universiteit van Amsterdam	1218
Renmin University of China	1212
The Australian National University	1205
University of Cambridge	1179
University of Melbourne	1177
Universiti Utara Malaysia	1155
University of Michigan, Ann Arbor	1129
Yale University	1125
Michigan State University	1113
University of Groningen	1108
Kazan Federal University	1092
Università Bocconi	1087
Tilburg University	1086
ETH Zürich	1084
Alma Mater Studiorum Università di Bologna	1082
Münchener Gesellschaft zur Förderung der Wirtschaftswissenschaft - CESifo GmbH	1080
Universiti Kebangsaan Malaysia	1080
University of Illinois at Urbana-Champaign	1067
Massachusetts Institute of Technology	1056
Duke University	1037
Vrije Universiteit Amsterdam	1030
Universität Göttingen	1028
University of Toronto	1022
Tsinghua University	1007

2.4. Построение облака тегов

Для построения облака тегов использовались ключевые слова всех статей (authkeywords из таблицы articles). Для составления облака слов и сочетаний ключевые слова были переведены в нижний регистр, удалены стоп-слова (артиккли, союзы и так далее). Далее ключевые слова были обработаны с помощью скрипта с использованием библиотеки WorldCloud на языке Python.

Было построено два вида облака тегов: облако отдельных слов и облако отдельных слов и словосочетаний.

Среди словосочетаний в топ вошли:

- economic growth
- monetary policy
- model
- exchange rate
- system
- corporate governance
- china
- panel data
- financial crisis

Соответствующее облако представлено на рисунке ниже.

будет добавлять единицу к весу ребра. Таким образом получим взвешенный ориентированный граф, который будем анализировать далее.

2.6. Выделение кластеров учёных и их анализ

Для создания, анализа и выделения кластеров авторов статей была использована библиотека `igraph` языка R (данная библиотека есть и для языков Python и C, а также для математического пакета Wolfram Mathematica).

С помощью различных алгоритмов, встроенных в библиотеку `igraph`, были выделены связанные подгруппы графа. Так как исходный граф содержит порядка 307 тысяч вершин и более 371 тысяча рёбер, сначала были выделены подгруппы исходного графа, затем был выбран самый крупный кластер и уже он был проанализирован. Кластеризация и выделения самого крупного кластера было выполнено с помощью жадного алгоритма `fast_greedy` (`cluster_fast_greedy` в библиотеке `igraph` для языка R).

Графическое представление даже подграфа с трудом поддаётся анализу, так как в нём более 16 тысяч вершин и более 25 тысяч рёбер. Сам кластер также состоял из меньших кластеров. Каждый подкластер выделен своим цветом. Варьируя гиперпараметры можно уменьшать количество кластеров. Так был получен граф с 7 кластерами. Но они сильно связаны между собой, из-за чего получается следующая картина.

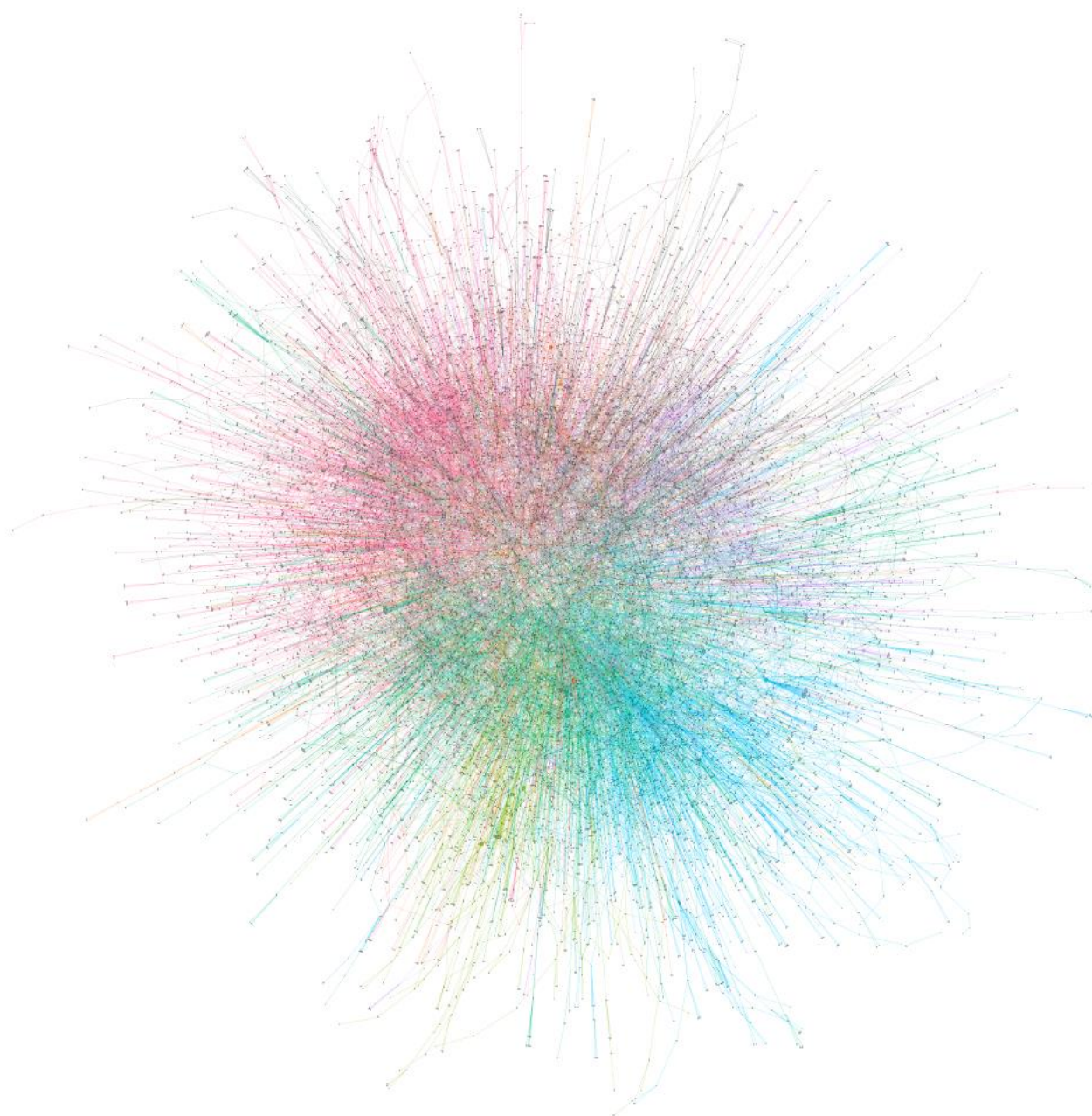


Рисунок 10. Изображение самого большого кластера с выделением
меньших кластеров

Детальные изображения в формате PDF доступны по ссылке в приложении Б в формате PDF. Разница в файлах заключается в раскладке и отрисовке.

Исходя из полученной картины, можно заметить, что на протяжении 10 лет среди авторов статей образовались группы, а внутри групп они

предпочитают плодотворно работать с многими другими авторами, и лишь некоторые выделяются, представляя на рисунке некие «лучи» из центра.

В процессе анализа, сформированный граф был разделен на 7 классов (сообществ). Лидеры были определены с помощью алгоритма определения авторитарности PageRank. Лидеры сообществ представлены в таблице 3.

Таблица 3. Лидеры сообществ и их PageRank

Name	PageRank	Modularity Class
Acemoglu D. n11021	0,002369971	0
Avraham R. n10538	0,002084327	0
Batrancea I. n8631	0,001228035	0
Bender S. n16098	0,001989232	1
Ambec S. n3174	0,001843845	1
Nichita A. n6710	0,001242115	1
Angelini V. n1803	8,697E-04	2
Card D. n12594	0,002130552	3
Amialchuk A. n8070	0,001409403	3
Batrancea L. n6708	0,001380223	3
Aghion P. n3256	0,002187200	4
Andersson C. n5153	0,001884082	4
Yildiz E. n14484	0,002255115	5
Andersson T. n3037	0,001887304	5
Lanoie P. n454	0,001715791	5
Kogler C. n7810	0,001280864	6
Al-Ubaydli O. n11727	0,001097667	6

На следующем рисунке представлены сообщества, собранные вместе.

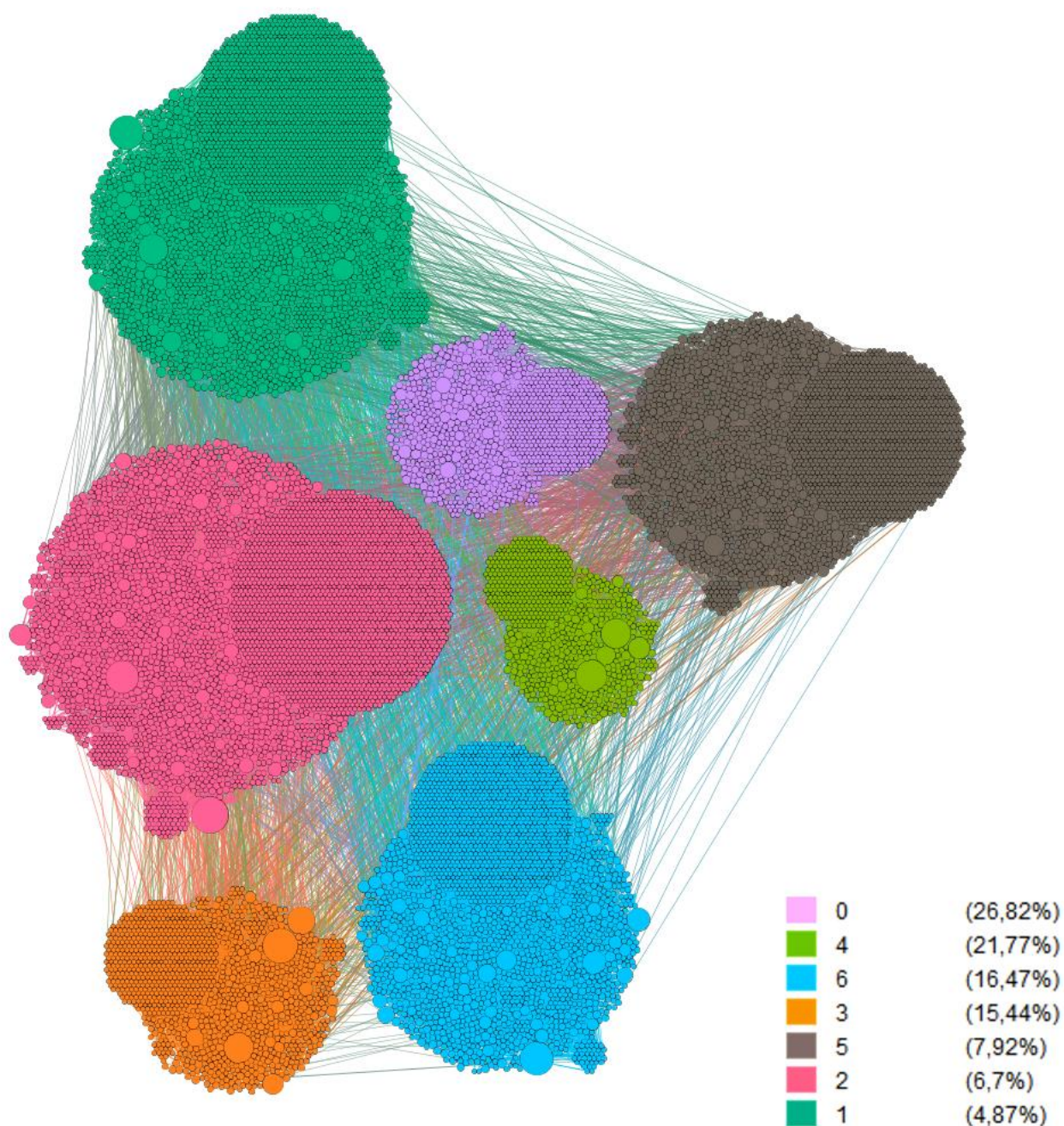


Рисунок 11. Выделенные сообщества

На данном рисунке можно визуальнo определить размеры сообществ, кроме того, на рисунке представлены их (сообществ) обозначения. Величина каждой вершины отражает размер PageRank автора. Можно заметить, что в каждом сообществе есть крупные узлы – так называемые «авторитетные» исследователи. Можно заметить, что в каждом сообществе есть некие круги, состоящие из узлов примерно одинакового размера. Это наименее «авторитетные» исследователи. Размеры таких кругов (количество учёных в

них) сопоставимы относительно размеров самих сообществ, это позволяет сказать, что сообщества сформированы достаточно хорошо.

Кроме того, на рисунке 11 хорошо видны сильные взаимосвязи между различными сообществами. Исследователи неограниченны своим сообществом и плотно сотрудничают с другими учёными. Стоит учитывать, что это данные за 2011-2021 годы, что довольно большой промежуток. Для более детального анализа лучше рассматривать промежутки меньшей длины, например, 1 год. Это касается не только формирования групп учёных и их взаимосвязей, но и анализа направленности работ, ключевых слов и выделения наиболее авторитарных исследователей.

3. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ

3.1 Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения

3.1.1 Потенциальные потребители результатов исследования

Для анализа потребителей результатов исследования необходимо рассмотреть целевой рынок и провести его сегментирование.

Целевой рынок – сегменты рынка, на котором будет продаваться в будущем разработка. В свою очередь, сегмент рынка – это особым образом выделенная часть рынка, группы потребителей, обладающих определенными общими признаками.

Сегментирование – это разделение покупателей на однородные группы, для каждой из которых может потребоваться определенный товар (услуга).

Предоставляемая услуга – исследование данных о публикациях и авторах методами кластерного анализа, в результате которых предоставляется анализ найденных кластеров, их визуальное представление, выделение самого большого кластера и анализ авторитарности каждого учёного. Целевым рынком являются компании, имеющие библиографические данные о публикациях и авторах.

Потребителями разработанной технологии могут быть:

- научные организации, производящие индексацию статей;
- интернет ресурсы рассчитывающие рейтинги журналов (Elsevier, Semantic Scholar и другие);
- научное сообщество (университеты и организации).

Основными конкурентами являются g-индекс и h-индекс. h-индекс, или индекс Хирша — наукометрический показатель, являющийся количественной характеристикой продуктивности учёного, группы учёных, научной

организации или страны в целом, основанной на количестве публикаций и количестве цитирований этих публикаций. g-индекс — это индекс для измерения научной продуктивности, рассчитываемый на основе распределения цитирований, полученных публикациями учёного. Пользователями данных конкурентных решений являются следующие наукометрические базы данных: Scopus, Web of Science (WoS), Google Scholar, Microsoft Academic, PubMed, Semantic Scholar, eLIBRARY.ru, ИСТИНА МГУ и др.

Таблица 4. Карта сегментированного рынка.

Факторы	Использование разрабатываемого проекта		
	Научные организации	Библиографические базы данных	Научное сообщество
Затраты компьютерных ресурсов	1	2	1
Цена	3	0	1
Надежность	2	3	2
Удобство в использовании	2	3	2
Смысловая нагрузка	3	3	3

Карта сегментирования рынка показывает, какие ниши не заняты конкурентами и где уровень конкуренции низок. Оценка степени значимости факторов разработки осуществляется по четырёхпольной шкале – от 0 до 3. Нулевой уровень означает, что фактор не имеет значения для данного сегмента потребителей. Максимальное значение равно 3, оно показывает, что данный фактор имеет особое значение для сегмента. Например, цена со значением 3 свидетельствует о том, что этот фактор наиболее значим и потребитель заинтересован в минимально возможной цене.

Из карты сегментирования рынка, представленной в таблице 4, можно определить соответствие существующих разработок и требования потребителей, приобретающих данный продукт. Для научных организаций и научного сообщества требования схожи между собой. Соответствия выявленные при составлении карты сегментирования рынка показывают, что существующие разработки нуждаются в доработке.

Новая разработка будет востребованной при повышении уровня надежности, смысловой нагрузки и удобства использования. Так же

разработанный продукт должен иметь минимальную цену, вследствие чего разработка станет доступной таким потребителям, как научные организации.

3.1.2 Анализ конкурентных технических решений

На данный момент большинство интернет ресурсов, связанных с индексацией научных работ, используют очень простые для расчёта методики, которые несут малую смысловую нагрузку, основывающуюся лишь на количестве публикаций и рейтинге журнала в котором данная работа напечатана.

Бф – разработка, представленная в данной работе

Бк1 – h-индекс

Бк2 – g-индекс

Таблица 5. Оценочная карта для сравнения конкурентных технических решений

Критерии оценки	Вес критерия	Баллы			Конкурентоспособность		
		Бф	Бк1	Бк2	Кф	Кк1	Кк2
1	2	3	4	5	6	7	8
Технические критерии оценки ресурсоэффективности							
1. Затраты компьютерных ресурсов для расчёта	0,1	5	4	3	0,5	0,4	0,3
2. Скорость получения результатов	0,05	2	4	5	0,1	0,2	0,25
3. Смысловая нагрузка критерия	0,5	5	2	2	2,5	1	1
4. Удобство расчёта	0,1	2	4	4	0,2	0,4	0,4
Экономические критерии оценки эффективности							
1. Конкурентоспособность	0,05	5	4	4	0,25	0,2	0,2
2. Уровень проникновения на рынок	0,1	1	5	4	0,1	0,5	0,4
3. Затраты на разработку	0,1	1	3	3	0,1	0,3	0,3
Итого	1				3,75	3	2,85

Конкурентоспособность $3,75/3 = 1,25$ и $3,75/2,85 = 1,32$.

Конкурентные решения не используют сложные взаимосвязи между научными трудами. Кроме того, разрабатываемый метод анализа лучше показывает значимость учёного, исходя из его взаимосвязей с другими учёными, что конкурентные решения не могут.

1.1.1. SWOT-анализ

Разработанная для данного исследования матрица SWOT представлена в таблице 6.

Таблица 6. Матрица SWOT

	Сильные стороны научно-исследовательского проекта: 1. Полученные результаты показывают влияние и значимость учёного, исходя из его взаимодействий с другими учёными	Слабые стороны научно-исследовательского проекта: 1. Большой объём исследовательских данных
Возможности: 1. Применение полученных результатов при ранжировании учёных и научных организаций	Разработанную модель можно использовать для ранжирования учёных и научных организаций.	Использование наиболее быстрых решений, дающих достаточный результат.
Угрозы: 1. Отсутствие спроса на разработанную модель	В случае отсутствия спроса или появления новых конкурентов в данной области исследования проект может оказаться не рентабельным.	Поиск высококвалифицированных специалистов для увеличения скорости обработки.

Результатом анализа стало следующее: мы можем смело выделить сильные стороны разработки, слабые, её возможности и угрозы для потенциального клиента.

3.2 Планирование научно-исследовательских работ

3.2.1 Структура работ в рамках научного исследования

Трудоемкость выполнения ВКР оценивается экспертным путем в человеко-днях и носит вероятностный характер, т.к. зависит от множества трудно учитываемых факторов.

Для выполнения научно-исследовательской работы формируется рабочая группа, в состав которой могут входить:

- 1) руководитель проекта (Р);
- 2) бакалавр (Б).

Примерный порядок составления этапов и работ, распределение исполнителей по данным видам работ приведен в таблице 7.

Таблица 7. Комплекс работ по разработке проекта

Основные этапы	№	Содержание работ	Должность исполнителя
----------------	---	------------------	-----------------------

Разработка технического задания	1	Составление и утверждение научного задания	Бакалавр, Руководитель
Выбор направления исследований	2	Подбор и изучение материалов по теме	Бакалавр
	3	Выбор направления исследований	Бакалавр, Руководитель
	4	Анализ исходных данных	Бакалавр
	5	Выбор метода выполнения	Бакалавр, Руководитель
	6	Календарное планирование работ по теме	Бакалавр, Руководитель
Теоретические и экспериментальные исследования	7	Разработка программного продукта	Бакалавр
	8	Проведение расчётов и построение модели	Бакалавр
Обобщение и оценка результатов	9	Анализ результатов работы	Бакалавр, Руководитель

В результате был составлен перечень этапов работ в рамках проведения научного исследования.

3.2.2 Определение трудоемкости выполнения работ и разработка графика проведения научного исследования

Трудовые затраты в большинстве случаев образуют основную часть стоимости разработки, поэтому важным моментом является определение трудоемкости работ каждого из участников научного исследования.

Трудоемкость выполнения научного исследования оценивается экспертным путем в человеко-днях и носит вероятностный характер, т.к. зависит от множества трудно учитываемых факторов. Для определения ожидаемого (среднего) значения трудоемкости используется следующая формула:

$$t_{ож\ i} = \frac{3t_{min\ i} + 2t_{max\ i}}{5}$$

где $t_{ож\ i}$ – ожидаемая трудоемкость выполнения i -й работы, человеко-дни;

$t_{min\ i}$ – минимально возможная трудоемкость выполнения заданной i -й работы, человеко-дни;

$t_{max\ i}$ – максимально возможная трудоемкость выполнения заданной i -й работы, человеко-дни;

Рассчитаем значение ожидаемой трудоемкости работы.

Установление длительности работ в рабочих днях осуществляется по формуле:

$$T_{pi} = \frac{t_{ож i}}{Ч_i}$$

где T_{pi} - продолжительность одной работы, раб. дн.;

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

При выполнении дипломных работ студенты в основном становятся участниками сравнительно небольших по объему научных тем. Поэтому наиболее удобным и наглядным является построение ленточного графика проведения научных работ в форме диаграммы Ганта.

Для удобства построения графика, длительность каждого из этапов работ из рабочих дней следует перевести в календарные дни. Для этого необходимо воспользоваться формулой:

$$T_k = T_{pi} * k_{кал}$$

где T_k – продолжительность выполнения i -й работы в календарных днях;

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

$k_{кал}$ – коэффициент календарности, предназначен для перевода рабочего времени в календарное.

Коэффициент календарности определяется по формуле:

$$k_{кал} = \frac{T_{кал}}{T_{кал} - T_{вых} - T_{пр}}$$

где $T_{кал}$ – количество календарных дней в году;

$T_{вых}$ – количество выходных дней в году;

$T_{пр}$ – количество праздничных дней в году.

Рассчитанные значения в календарных днях по каждой работе необходимо округлить до целого числа.

Вычислим коэффициент календарности:

$$k_{\text{кал}} = \frac{365}{365 - 52 - 14} = 1,22$$

Расчитанные временные показатели были сведены в таблицу 8, представленную ниже.

Таблица 8. Временные показатели осуществления комплекса работ

№ работы	Продолжительность работ			Исполнители	T _{pi} , человеко-дни		T _{ki} , человеко-дни	
	t _{mini} , человеко-дни	t _{maxi} , человеко-дни	t _{ожи} , человеко-дни		Р	Б	Р	Б
	1	1	3					
2	14	18	16	Б	0	16	0	20
3	3	5	2	Б, Р	1	1	1	1
4	7	12	9	Б	0	9	0	11
5	3	6	4	Б, Р	2	2	2	2
6	2	4	2	Б, Р	1	1	1	1
7	7	14	7	Б	0	7	0	9
8	14	18	16	Б	0	16	0	20
9	5	7	6	Б, Р	3	3	4	4
				Итого:	8	56	9	69

На основе таблицы 8 составлен календарный план-график, показывающий продолжительность выполнения работ ВКР. В результате планирования графика, продолжительность работ равна двум месяцам (табл. 9).

Таблица 9. Календарный план-график выполнения работ

№ работы	Наименование работы	Исполнители	дни	Продолжительность выполнения работ, дни																
				март			апрель			май										
				10	20	30	10	20	30	10	20	30	10							
1	Составление и утверждение научного задания	Бакалавр, Руководитель	1	■																
2	Подбор и изучение материалов по теме	Бакалавр	20	■	■	■														
3	Выбор направления исследований	Бакалавр, Руководитель	1			■														
4	Анализ исходных данных	Бакалавр	11			■														
5	Выбор метода выполнения	Бакалавр, Руководитель	2				■													
6	Календарное планирование работ по теме	Бакалавр, Руководитель	1				■													
7	Разработка программного продукта	Бакалавр	9					■	■	■										
8	Проведение расчётов и построение модели	Бакалавр	20								■	■	■	■	■					
9	Анализ результатов работы	Бакалавр, Руководитель	4																	■

3.3 Бюджет научно-технического исследования (НТИ)

При планировании бюджета НТИ должно быть обеспечено полное и достоверное отражение всех видов расходов, связанных с его выполнением. В процессе формирования бюджета НТИ используется следующая группировка затрат по статьям:

- материальные затраты НТИ;
- основная заработная плата исполнителей темы;
- отчисления во внебюджетные фонды (страховые отчисления);
- накладные расходы.

3.3.1 Расчёт материальных затрат НТИ

Данная статья отражает стоимость всех материалов, используемых при разработке проекта, включая расходы на их приобретение и доставку. Транспортные расходы принимаются в пределах 15-25% от стоимости материалов. В материальные затраты, помимо вышеуказанных, включаются дополнительно затраты на канцелярские принадлежности, диски, картриджи и т.п. Однако их учет ведется в данной статье только в том случае, если в научной организации их не включают в расходы на использование оборудования или накладные расходы.

Материальные затраты, необходимые для данной разработки, заносятся в таблицу 10.

Таблица 10. Материальные затраты

Наименование	Единица измерения	Количество	Цена за ед., руб.	Затраты на материалы (Зм), руб.
Бумага	Пачка	1	250	250
Картридж для принтера	Шт	1	4000	4000
Канцелярские принадлежности	Шт	1	250	250
Итого				4500

Итого материальные затраты составляют 4500 рублей.

3.3.2 Расчёт затрат на специально оборудование для научных (экспериментальных) работ

В данную статью включают все затраты, связанные с приобретением специального оборудования (приборов, контрольно-измерительной аппаратуры, стендов, устройств и механизмов), необходимого для проведения работ по конкретной теме. Определение стоимости спецоборудования производится по действующим прейскурантам, а в ряде случаев по договорной цене.

Таблица 11. Расчет затрат на приобретение спецоборудования для научных работ

№	Наименование оборудования	Кол-во единиц оборудования	Цена единицы оборудования, тыс.руб.	Общая стоимость оборудования, тыс.руб
1	Ноутбук	1	40000	40000
Итого:				40000

Всё используемое ПО предоставляется по бесплатной подписке. Ноутбук приобретается на 4 года (48 месяцев) и амортизационное начисление (линейным методом) за время работы над проектом (69 дней) составит 1916,66 рублей.

3.3.3 Основная заработная плата

В состав основной заработной платы включается премия, выплачиваемая ежемесячно из фонда заработной платы в размере 20 – 30 % от тарифа или оклада.

Основная заработная плата ($Z_{осн}$) руководителя (лаборанта, инженера) от предприятия (при наличии руководителя от предприятия) рассчитывается по следующей формуле:

$$Z_{осн} = Z_{дн} * T_p$$

где $Z_{осн}$ – основная заработная плата одного работника;

$Z_{дн}$ – продолжительность работ, выполняемых научнотехническим работником, раб. дн.;

T_p – среднедневная заработная плата работника, руб.

Среднедневная заработная плата рассчитывается по формуле:

$$Z_{\text{дн}} = \frac{Z_{\text{м}} * M}{F_{\text{д}}}$$

где $Z_{\text{м}}$ – месячный должностной оклад работника, руб.;

M – количество месяцев работы без отпуска в течение года;

$F_{\text{д}}$ – действительный годовой фонд рабочего времени научно-технического персонала, раб. дн.

Месячный должностной оклад работника:

$$Z_{\text{м}} = Z_{\text{тс}} * (1 + k_{\text{пр}} + k_{\text{д}}) * k_{\text{р}}$$

где $Z_{\text{тс}}$ – заработная плата по тарифной ставке, руб.;

$k_{\text{пр}}$ – премиальный коэффициент, равный 0,3 (т.е. 30% от $Z_{\text{тс}}$);

$k_{\text{д}}$ – коэффициент доплат и надбавок составляет примерно 0,2 – 0,5 (в НИИ и на промышленных предприятиях – за расширение сфер обслуживания, за профессиональное мастерство, за вредные условия: 15-20 % от $Z_{\text{тс}}$);

$k_{\text{р}}$ – районный коэффициент, равный 1,3 (для г. Томска).

Результат расчетов заработных плат представлен в таблице 12.

Таблица 12. Расчёт основной заработной платы

Исполнители	$Z_{\text{тс}}$, руб	$k_{\text{пр}}$	$k_{\text{д}}$	$Z_{\text{м}}$, руб	$Z_{\text{дн}}$, руб	T_p , дни	$Z_{\text{осн}}$, руб
Руководитель	33300	0,3	0,2	64935	2691	8	21528
Бакалавр	9893	0	0	12861	533	56	29848
ИТОГО							51376

Итого основная заработная плата составит 56376 рублей.

3.3.4 Отчисления во внебюджетные фонды (страховые отчисления)

Отчисления во внебюджетные фонды являются обязательными по установленным законодательством Российской Федерации нормам органам

государственного социального страхования (ФСС), пенсионного фонда (ПФ) и медицинского страхования (ФФОМС) от затрат на оплату труда работников.

Величина отчислений во внебюджетные фонды определяется исходя из следующей формулы:

$$З_{внеб} = k_{внеб} * (З_{осн} + З_{доп})$$

где $k_{внеб}$ - коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.).

На 2014 г. в соответствии с Федеральным закона от 24.07.2009 №212-ФЗ установлен размер страховых взносов равный 30%.

Отчисления во внебюджетные фонды представлены в таблице 13.

Таблица 13. Отчисления во внебюджетные фонды

Исполнители	Основная ЗП, руб
Руководитель	21528
Бакалавр	29848
Коэффициент отчислений во внебюджетные фонды	0,3
ИТОГО	15412,8

Отчисления во внебюджетные фонды составят 15412,8 рублей.

3.3.5 Накладные расходы

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

$$З_{накл} = k_{пр} * (З_{осн} + З_{внеб}) + \text{Затраты на электроэнергию}$$

где $k_{пр}$ - коэффициент, учитывающий накладные расходы.

Величину коэффициента накладных расходов можно взять в размере 16%.

Затраты на электроэнергию = $0,440 * 47 * 6 * 5,8 = 719,67$ р.

$$З_{накл} = 52361,4 * 0,16 + 719,67 = 9097,55$$

3.3.6 Формирование бюджета затрат НИИ

Расчитанная величина затрат научно-исследовательской работы (темы) является основой для формирования бюджета затрат проекта, который при формировании договора с заказчиком защищается научной организацией в качестве нижнего предела затрат на разработку научно-технической продукции. Определение бюджета затрат на научно-исследовательский проект приведен в таблице 14.

Таблица 14. Расчет бюджета затрат НИИ

Наименование статьи	Сумма, руб.
1. Материальные затраты НИИ	4500
2. Затраты на специальное оборудование для научных (экспериментальных) работ	41916,6
3. Затраты по основной заработной плате исполнителей темы	51376
4. Отчисления во внебюджетные фонды	15412,8
5. Накладные расходы	9097,55
6. Бюджет затрат НИИ	122302,95

Общие затраты НИИ составят 122302,95 рублей.

3.4 Определение ресурсной эффективности исследования

Расчет интегрального показателя ресурсоэффективности приведен в виде таблицы 15.

Под Исполнением 1 имеется ввиду разработка представленная в данной ВКР. Под Исполнением 2 подразумевается расчёт с помощью h-индекса.

Таблица 15. Сравнительная оценка характеристик вариантов исполнения проекта

Критерии	Вес критерия	Исп. 1	Исп. 2
1. Затраты компьютерных ресурсов	0,2	5	4
2. Удобство в эксплуатации	0,2	5	5
3. Смысловая нагрузка	0,35	5	3
4. Скорость расчётов	0,2	3	5
5. Надёжность	0,05	5	5
Итого	1		

$$I_{p-исп1} = 5 * 0,2 + 5 * 0,2 + 5 * 0,35 + 3 * 0,2 + 5 * 0,05 = 5,6$$

$$I_{p-исп2} = 4 * 0,2 + 5 * 0,2 + 3 * 0,35 + 5 * 0,2 + 5 * 0,05 = 4,9$$

Исходя из полученных результатов сравнительной оценки показателя ресурсоэффективности, можно сделать вывод, что научно-техническое исследование будет ресурсоэффективно при первом варианте исполнения проекта, т.е. при высокой производительности и смысловой нагрузке проекта.

3.5 Выводы по разделу

В процессе выполнения части работы по финансовому менеджменту, ресурсоэффективности и ресурсосбережению был проведен анализ разрабатываемого исследования.

Проведено планирование НИР, а именно: определена структура и календарный план работы, трудоемкость, составлена ленточная диаграмма Ганта, и определен бюджет научно-исследовательской работы. В ходе планирования научно-исследовательских работ определён перечень работ, выполняемый рабочей группой. В данном случае рабочая группа состоит из двух человек: руководитель и инженер. Результаты соответствуют требованиям ВКР по срокам и иным параметрам.

Представленная в данной ВКР разработка имеет перспективы быть востребованной, если улучшить показатели скорости вычислений и удобства эксплуатации. И может занять место на ряду с показателем индекса Хирша.

Бюджет научно-технического исследования составил 65958,95 рубля. Бюджет НТИ состоит из затрат на разработку (41197 рублей), отчислений во внебюджетные фонды (11164,4 рубля), материальных затрат (4500 рублей) и накладных расходов (9097,55 рубля).

4. Социальная ответственность

4.1. Введение

Объектом исследования данной ВКР являются данные о публикациях и авторах, извлеченные из наукометрической базы данных Scopus. Данная работа направлена на определение структур сотрудничества между исследователями в области «Экономики, Эконометрики и Финансов» (ECON) и выявлении учреждений, направленных на данную тематику. В результате исследования проводится анализ найденных кластеров, строится их визуальное представление, выделяются кластеры, содержащие самое большое количество учёных.

Так как работа производится непосредственно на персональном компьютере (ПК), то в данном разделе целесообразно рассмотреть вопросы анализа опасных и вредных факторов при работе с ПК, влияния этих факторов на окружающую среду и мероприятий по её защите.

Предметом исследования является рабочая зона офисного сотрудника, включая письменный стол, персональный компьютер, клавиатуру, компьютерную мышь и стул, а также помещение в котором эта рабочая зона находится.

4.2. Правовые и организационные вопросы обеспечения безопасности

Согласно Ст. 108 ТК РФ режим офисного работника в течении рабочего дня не должен превышать 8 часов с перерывом на обед от 30-ти минут до 2-х часов. Необходимо придерживаться правильного режима труда и отдыха, чтобы избежать вредного воздействия компьютера на здоровье человека.

Рабочее место должно быть организовано с учетом требований ГОСТ 12.2.032-78 «Система стандартов безопасности труда (ССБТ). Рабочее место при выполнении работ сидя. Общие эргономические требования» и СанПиН 1.2.3685-21 «Гигиенические нормативы и требования к обеспечению безопасности и безвредности для человека факторов среды обитания».

4.3. Производственная безопасность.

Для идентификации потенциальных факторов был использован ГОСТ 12.0.003-2015 «Опасные и вредные производственные факторы. Классификация». При работе с ПК пользователь подвергается воздействию опасных и вредных производственных факторов, представленных в таблице 16.

Таблица 16. Возможные опасные и вредные факторы

Факторы (ГОСТ 12.0.003-2015)	Этапы работ			Нормативные документы
	Разработка	Изготовление	Эксплуатация	
1. Отклонение параметров микроклимата в помещении	+	+		СанПиН 1.2.3685-21 «Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания» [9]
2. Недостаточная освещённость рабочей зоны	+	+	+	СП 52.13330.2016 «Естественное и искусственное освещение» [14]
3. Повышенный уровень шума на рабочем месте	+	+		СН 2.2.4/ 2.1.8.562-96 «Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки» [15]
5. Опасность поражения электрическим током	+	+	+	ГОСТ Р 58698-2019 «Защита от поражения электрическим током» [13]

4.4. Анализ опасных и вредных производственных факторов

4.4.1. Отклонение параметров микроклимата в помещении

Микроклимат производственных помещений – это климат внутренней среды помещений, который определяется действующими на организм человека сочетаниями температур воздуха и поверхностей, относительной влажности воздуха, скорости движения воздуха и интенсивности теплового излучения. Показатели микроклимата должны обеспечивать сохранение теплового баланса человека с окружающей средой и поддержание оптимального или допустимого теплового состояния организма.

Оптимальные микроклиматические при воздействии на человека в течение рабочей смены обеспечивают сохранение теплового состояния организма и не вызывают отклонений в состоянии здоровья. Допустимые микроклиматические условия могут приводить к незначительным

дискомфортным тепловым ощущениям. Возможно, временное (в течение рабочей смены) снижение работоспособности, без нарушения здоровья.

Нормы оптимальных и допустимых показателей устанавливает СанПиН 1.2.3685-21. Он регулирует множество параметров, среди которых: температуру воздуха, температуру поверхностей конструкции, относительную влажность воздуха, скорость движения воздуха и интенсивность теплового облучения. Допустимые величины параметров микроклимата на рабочих местах в помещениях оцениваются в зависимости от категории работ по уровню энергозатрат организма. Работа, производимая сидя и сопровождаемая незначительным физическим напряжением, относится к категории Ia – работа с интенсивностью энергозатрат до 139 Вт. Допустимые нормы микроклимата приведены в таблице 17.

Таблица 17. Допустимые нормы микроклимата в рабочей зоне производственных помещений

Период года	Категория работ по уровню энергопотребления, Вт	Температура воздуха, °С		Температура воздуха поверхностей	Относительная влажность воздуха	Скорость движения воздуха	
		Диапазон ниже оптимальных величин	Диапазон выше оптимальных величин			Для диапазона температур воздуха ниже оптимальных величин, не более	Для диапазона температур воздуха выше оптимальных величин, не более
Холодный	Ia (до 139)	20,0 - 21,9	24,1 - 25,0	19,0 - 26,0	15 - 75	0,1	0,1
Тёплый	Ia (до 139)	21,0 - 22,9	25,1 - 28,0	20,0 - 29,0	15 - 75	0,1	0,2

В производственных помещениях, где допускаемые нормативные величины локального микроклимата поддерживать не представляется возможным, необходимо проводить мероприятия по защите работников от возможного перегревания и охлаждения. Это достигается разными способами: использование систем местного кондиционирования воздуха; регламентацией периодов работы в неблагоприятном локальном микроклимате и отдыха в помещении с микроклиматом, нормализующим тепловое состояние; уменьшение длительности рабочей смены и др.

4.4.2. Недостаточная освещённость рабочей зоны

Свет является естественным условием жизни человека. Верно, спроектированное и выполненное освещение обуславливает высокую степень работоспособности, оказывает положительное психологическое воздействие на человека и содействует увеличению производительности труда. На рабочей поверхности обязаны отсутствовать резкие тени, которые создают неравномерное рассредоточивание поверхностей с различной яркостью в поле зрения, искажает размеры и формы объектов различия, в итоге увеличивается утомляемость персонала и понижается производительность труда.

Свет влияет на физиологическое состояние человека, правильно организованное освещение стимулирует протекание процессов высшей нервной деятельности и повышает работоспособность. При недостаточном освещении человек работает менее продуктивно, быстро устает, растет вероятность ошибочных действий, что может привести к травматизму.

Нормы оптимальных и допустимых показателей устанавливает СанПиН 1.2.3685-21.

Таблица 18. Нормы оптимальных и допустимых значений освещенности.

Рабочая поверхность и плоскость нормирования КЕО и освещённости и высота плоскости над полом, м	Естественное освещение		Совмещённое освещение		Искусственное освещение				
	КЕО, %		КЕО, %		Освещённость, лк			Объединённый показатель дискомфорта, UGR, не более	Коэффициент пульсации освещённости, Кп, %, не более
	при верхнем или комбинированном освещении	при боковом освещении	при верхнем или комбинированном освещении	при боковом освещении	при комбинированном освещении	при общем освещении	все		
Г-0,8	3,0	1,0	1,8	0,6	400	200	300	21	15

К средствам нормализации освещенности производственных помещений рабочих мест относятся:

- источники света;

- осветительные приборы;
- световые проемы;
- светозащитные устройства.

4.4.3. Повышенный уровень шума на рабочем месте

Основной источник создаваемого шума в помещении – это другие электрические машины.

Повышенный уровень шума может привести к хронической бессоннице, сердечным заболеваниям, нарушениям слуха, повышению в организме гормонов стресса, снижению иммунитета, неврозам.

Может возникнуть шумовая болезнь, которая далеко не всегда поддаётся лечению.

Предельно допустимый уровень (ПДУ) шума – это степень фактора, который при ежедневной (кроме выходных дней) работе, но не более 40 часов в неделю в течение всего рабочего стажа, не должен вызывать заболеваний или же отклонений в состоянии здоровья и самочувствия. Соблюдение ПДУ шума не исключает нарушения здоровья и самочувствия у сверхчувствительных лиц.

Допустимые значения уровня шума ограничены СанПиН 1.2.3685-21.

Таблица 19. Допустимые значения уровня шума.

Для источников постоянного шума									Для источников непостоянного шума		
Уровни звукового давления, дБ, в октавных полосах со среднегеометрическими частотами, Гц									Уровни звука L(A), дБА	Эквивалентные уровни звука, L(Aэв.), дБА	Максимальные уровни звука, L(Aмакс), дБА
31,5	63	125	250	500	1000	2000	4000	8000			
79	63	52	45	39	35	32	30	28	40	40	55

При значениях выше допустимого уровня необходимо предусмотреть средства коллективной защиты (СКЗ).

Средства коллективной защиты:

- устранение причин шума или значительное его ослабление в источнике возникновения;
- изоляция источников шума от окружающей среды средствами звуко- и виброизоляции, звуко- и вибропоглощения;
- использование средств, снижающих шум и вибрацию на пути их передачи;

4.4.4. Опасность поражения электрическим током

Электробезопасность подразумевает под собой систему мероприятий, технических и организационных, направленных на защиту людей от опасного воздействия электрического тока, статического электричества и электромагнитного поля. Значения вышеперечисленных факторов регулируются ГОСТ Р 58698-2019.

Таблица 20. Пороги напряжения прикосновения для реагирования.

Характер реагирования	Пороги напряжения
Реакция испуга	2 В переменный ток
	8 В постоянный ток
Мышечная реакция	20 В переменный ток
	40 В постоянный ток

Меры предосторожности для основной защиты от повреждения электрическим током:

- использование защитных ограждений или оболочек;
- размещение опасных для жизни и здоровья человека участков электропроводов и приборов вне зоны досягаемости рукой;
- ограничение напряжения или питание должно осуществляться от безопасного источника питания;
- автоматическое отключение питания (защитное устройство, которое будет отключать систему, питающую электрическое оборудование или установку в случае замыкания);

Меры защиты:

- посредством системы безопасного сверхнизкого напряжения (БСНН) и защитного сверхнизкого напряжения (ЗСНН);

4.4.5. Обоснование мероприятий по снижению уровней воздействия опасных и вредных факторов на исследователя (работающего)

1. Перед началом работы следует убедиться в отсутствии свешивающихся со стола или висящих под столом проводов электропитания, в целостности вилки и провода электропитания, в отсутствии видимых повреждений аппаратуры и рабочей мебели, в отсутствии повреждений и наличии заземления приэкранного фильтра.

2. При отклонении от нормы предоставить обогреватель, вентилятор или увлажнитель воздуха в зависимости от требуемых условий работы.

3. При отклонении от нормы предоставить дополнительные источники света (например, настольные лампы, точечные светильники и т.п.) в зависимости от требуемых условий работы.

4. Монитор компьютера служит источником ЭМП – вредного фактора, который отрицательно влияет на здоровье работника при продолжительной непрерывной работе и приводит к снижению работоспособности. Поэтому во избежание негативного влияния на здоровье необходимо делать перерывы при работе с ЭВМ и проводить специализированные комплексы упражнений для глаз.

4.5. Экологическая безопасность

При выполнении ВКР основными отходами являются: бумага и люминесцентные лампы. Для утилизации бумаги существуют пункты переработки бумаги или специальные мусорные контейнеры. Так же существуют пункты приема для утилизации перегоревших люминесцентных ламп. Объект исследования не оказывает влияния на окружающую среду и не наносит ей ущерб.

4.6. Безопасность в чрезвычайных ситуациях

Возможными чрезвычайными ситуациями при разработке, производстве или эксплуатации ПК являются: повышение уровня воды, град, обрушение зданий и пожары.

Наиболее типичной чрезвычайной ситуацией, которая может произойти на рабочем месте исследователя при написании ВКР, является пожар.

Основными причинами возникновения пожара считаются:

- использование неисправных электроприборов;
- перегрузка сети;
- курение не в специальных местах;
- оставление электрических приборов без присмотра на долгое время.

Для устранения оснований появления и локализации пожаров в помещении лаборатории должны проводиться надлежащие мероприятия:

- a) использование только исправного и надежного оборудования;
- b) осуществление периодических инструктажей по пожарной безопасности;
- c) отключение электрического оборудования, освещения и электропитания при предполагаемом отсутствии обслуживающего персонала или по завершении работ;
- d) курение в строго отведённом месте;
- e) содержание путей и проходов для эвакуации людей в свободном состоянии.

Здание должно отвечать требованиям пожарной безопасности, а именно: наличие охранно-пожарной сигнализации, плана эвакуации, порошковых или углекислотных огнетушителей с поверенным клеймом, табличек с указанием направления к запасному (эвакуационному) выходу

Потушить пожар можно с помощью ручных углекислотных огнетушителей, установленных в помещениях из расчета один огнетушитель на 40-50 м². Но, помимо этого, помещение должно быть оборудовано пожарными сигнализациями и средствами связи. При возгорании необходимо сообщить в службу пожарной охраны по телефону 101 или 112, сообщить адрес и место возникновения пожара.

4.7. Выводы по разделу

Рабочее помещение, где была разработана ВКР, соблюдены все нормы безопасности. Действие вредных и опасных факторов сведено к минимуму. Само помещение и рабочее место удовлетворяет всем требованиям.

Действие вредных и опасных факторов сведено к минимуму, т.е. микроклимат, освещение и электробезопасность соответствуют требованиям, предъявленным в соответствующих нормативных документах. Не стоит забывать, что монитор компьютера служит источником вредного фактора и отрицательно влияет на здоровье офисного сотрудника. Во избежание этого, нужно делать перерывы в работе и проводить специальные комплексы упражнений для разминки тела.

Заключение

1. В данной работе было проведено исследование данных Scopus за 2011-2021 годы. Разработан интерфейс для работы со Scopus API.
2. Интерфейс реализован на языке Python, данные были выгружены в реляционную базу данных PostgreSQL, анализ произведён при помощи языка R, Python и программы для анализа и визуализации данных Gephi.
3. Получена информация о 364886 статьях из области ECON, информация о 338963 авторах и их принадлежности к университетам или научным организациям. Произведен первичный анализ полученных данных.
4. Все авторы были разбиты на 87597 кластеров, которые основываются на взаимосвязи авторов через совместное написание работ.

Список используемых источников

1. Бин Анналин, Су Кеннет. Теоретический минимум по Big Data. Всё, что нужно знать о больших данных. — СПб.: Питер, 2019. — 208 с.: ил. — (Серия «Библиотека программиста»).
2. Юре Лесковец, Ананд Раджараман, Джеффри Д. Ульман. Анализ больших наборов данных. / Пер. с англ. Слинкин А. А. – М.: ДМК Пресс, 2016. – 498 с.: ил.
3. Сенько А. Работа с BigData в облаках. Обработка и хранение данных с примерами из Microsoft Azure. — СПб.: Питер, 2019. — 448 с.: ил. — (Серия «Для профессионалов»).
4. Elsevier, 2020. Scopus database API interface. Режим доступа: <https://dev.elsevier.com/> — свободный. (дата обращения: 27.05.21)
5. Казиев Г. З., Курдюков В. В. Модели и методы кластеризации big data для их анализа и обработки. Современный взгляд на будущее науки: приоритетные направления и инструменты развития // Сборник научных статей по итогам международной научно-практической конференции 2017. Издательство: Общество с ограниченной ответственностью "Редакционно-издательский центр "КУЛЬТ-ИНФОРМ-ПРЕСС" (Санкт-Петербург)
6. Мартышкин А. И., Бершадская Е. Г. Программная система сбора неструктурированной информации для построения социального профиля человека // XXI век: итоги прошлого и проблемы настоящего плюс.
7. Зайцева Е. А. Построение социального профиля клиента для персонализации банковского предложения на основе открытых источников информации с применением технологий больших данных. // Вестник университета № 6
8. PAGERANK: алгоритм ранжирования WEB-страниц. Режим доступа: <https://habr.com/ru/post/533096/> — свободный. (дата обращения: 27.05.21)
9. СанПиН 1.2.3685-21 «Гигиенические нормативы и требования к обеспечению безопасности и безвредности для человека факторов среды обитания»

10. Трудовой кодекс Российской Федерации от 30.12.2001 N 197-ФЗ (ред. от 27.12.2018)
11. ГОСТ 12.2.032-78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования от 01.03.1986: дата введения 01.01.1979
12. ГОСТ 12.0.003-2015 «Опасные и вредные производственные факторы. Классификация».
13. ГОСТ Р 58698-2019 (МЭК 61140:2016) Защита от поражения электрическим током. Общие положения для электроустановок и электрооборудования
14. СП 52.13330.2016 «Естественное и искусственное освещение»
15. СН 2.2.4/ 2.1.8.562-96 «Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки»

Приложение А

URL запроса:

[https://api.elsevier.com/content/search/scopus?query=all\(gene\)&apiKey=7f59af901d2d86f78a1fd60c1bf9426a&count=1&view=COMPLETE](https://api.elsevier.com/content/search/scopus?query=all(gene)&apiKey=7f59af901d2d86f78a1fd60c1bf9426a&count=1&view=COMPLETE)

Результат (формат XML):

```
1. <search-results xmlns="http://www.w3.org/2005/Atom"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:opensearch="http://a9.com/-/spec/opensearch/1.1/"
  xmlns:prism="http://prismstandard.org/namespaces/basic/2.0/"
  xmlns:atom="http://www.w3.org/2005/Atom">
2.   <opensearch:totalResults>7095352</opensearch:totalResults>
3.   <opensearch:startIndex>0</opensearch:startIndex>
4.   <opensearch:itemsPerPage>1</opensearch:itemsPerPage>
5.   <opensearch:Query role="request" searchTerms="all(gene)" startPage="0"/>
6.   <link ref="self"
  href="https://api.elsevier.com/content/search/scopus?start=0&count=1&query=a
  ll%28gene%29&apiKey=7f59af901d2d86f78a1fd60c1bf9426a&view=complete"
  type="application/xml"/>
7.   <link ref="first"
  href="https://api.elsevier.com/content/search/scopus?start=0&count=1&query=a
  ll%28gene%29&apiKey=7f59af901d2d86f78a1fd60c1bf9426a&view=complete"
  type="application/xml"/>
8.   <link ref="next"
  href="https://api.elsevier.com/content/search/scopus?start=1&count=1&query=a
  ll%28gene%29&apiKey=7f59af901d2d86f78a1fd60c1bf9426a&view=complete"
  type="application/xml"/>
9.   <link ref="last"
  href="https://api.elsevier.com/content/search/scopus?start=4999&count=1&quer
  y=all%28gene%29&apiKey=7f59af901d2d86f78a1fd60c1bf9426a&view=complete"
  type="application/xml"/>
10.  <entry>
11.    <link ref="self"
  href="https://api.elsevier.com/content/abstract/scopus_id/0000742471"/>
12.    <link ref="author-affiliation"
  href="https://api.elsevier.com/content/abstract/scopus_id/0000742471?field=a
  uthor,affiliation"/>
13.    <link ref="scopus"
  href="https://www.scopus.com/inward/record.uri?partnerID=HzOxMe3b&scp=000074
  2471&origin=inward"/>
14.    <link ref="scopus-citedby"
  href="https://www.scopus.com/inward/citedby.uri?partnerID=HzOxMe3b&scp=00007
  42471&origin=inward"/>
15.    <prism:url>https://api.elsevier.com/content/abstract/scopus_id/0000742471</p
  rism:url>
16.    <dc:identifier>SCOPUS_ID:0000742471</dc:identifier>
17.    <eid>2-s2.0-0000742471</eid>
18.    <dc:title>A global regulatory gene controls secondary
  metabolites, motility, and pathogenicity factors in erwinia
  amylovora</dc:title>
19.    <dc:creator>Mukherjee A.</dc:creator>
20.    <prism:publicationName>Acta
  Horticulturae</prism:publicationName>
21.    <prism:issn>05677572</prism:issn>
22.    <prism:volume>411</prism:volume>
23.    <prism:issueIdentifier>1</prism:issueIdentifier>
24.    <prism:pageRange>237-247</prism:pageRange>
25.    <prism:coverDate>2096-04-01</prism:coverDate>
```

26. <prism:coverDisplayDate>1 April 2096</prism:coverDisplayDate>
27. <prism:doi>10.17660/actahortic.1996.411.48</prism:doi>
28. <dc:description>We previously reported that a low copy plasmid carrying *Erwinia carotovora* *rsmA* suppressed the production of extracellular polysaccharide and protease, motility, the ability to cause hypersensitive reaction in tobacco leaves and the production of fire blight symptoms in apple shoots by *E. amylovora* strain E9. We document here that *rsmA* and flanking sequences have been conserved in *E. amylovora* strains. Northern analysis using the *rsmA* DNA of *E. carotovora* as the probe also revealed that *E. amylovora* wild-type strains produced *rsmA* transcripts of ca. 310 bases. By screening a library of E9 chromosomal DNA, we have isolated several plasmids that carry sequences homologous to the *rsmA* DNA of *E. carotovora* subsp. *carotovora* strain 71. One such plasmid suppresses the synthesis of the cell density (quorum) sensing signal, A/(3-oxohexanoyl)-L-homoserine lactone, production of extracellular enzymes, motility and pathogenicity of *E. carotovora* subsp. *carotovora* strains. We conclude that (i) *rsmA* is present in *E. amylovora*; (ii) the gene is expressed in this bacterium under normal cultural conditions; and (iii) *rsmA* controls the production of various secondary metabolites as well as pathogenicity factors.</dc:description>
29. <citedby-count>4</citedby-count>
30. <affiliation>
31. <affiliation-url>https://api.elsevier.com/content/affiliation/affiliation_id/60006173</affiliation-url>
32. <afid>60006173</afid>
33. <affilname>University of Missouri</affilname>
34. <affiliation-city>Columbia</affiliation-city>
35. <affiliation-country>United States</affiliation-country>
36. </affiliation>
37. <prism:aggregationType>Book Series</prism:aggregationType>
38. <subtype>cp</subtype>
39. <subtypeDescription>Conference Paper</subtypeDescription>
40. <author-count limit="100" total="5">5</author-count>
41. <author seq="1">
42. <author-url>https://api.elsevier.com/content/author/author_id/55453750900</author-url>
43. <authid>55453750900</authid>
44. <authname>Mukherjee A.</authname>
45. <surname>Mukherjee</surname>
46. <given-name>A.</given-name>
47. <initials>A.</initials>
48. <afid>60006173</afid>
49. </author>
50. <author seq="2">
51. <author-url>https://api.elsevier.com/content/author/author_id/7402595360</author-url>
52. <authid>7402595360</authid>
53. <authname>Cui Y.</authname>
54. <surname>Cui</surname>
55. <given-name>Y.</given-name>
56. <initials>Y.</initials>
57. <afid>60006173</afid>
58. </author>
59. <author seq="3">
60. <author-url>https://api.elsevier.com/content/author/author_id/57192560565</author-url>
61. <authid>57192560565</authid>
62. <authname>Liu Y.</authname>
63. <surname>Liu</surname>
64. <given-name>Y.</given-name>
65. <initials>Y.</initials>

```
66.         <afid>60006173</afid>
67.         </author>
68.         <author seq="4">
69.         <author-
url>https://api.elsevier.com/content/author/author_id/6602636444</author-
url>
70.         <authid>6602636444</authid>
71.         <authname>Korsi Dumenyo C.</authname>
72.         <surname>Korsi Dumenyo</surname>
73.         <given-name>C.</given-name>
74.         <initials>C.</initials>
75.         <afid>60006173</afid>
76.         </author>
77.         <author seq="5">
78.         <author-
url>https://api.elsevier.com/content/author/author_id/7403038808</author-
url>
79.         <authid>7403038808</authid>
80.         <authname>Chatterjee A.K.</authname>
81.         <surname>Chatterjee</surname>
82.         <given-name>A. K.</given-name>
83.         <initials>A.K.</initials>
84.         <afid>60006173</afid>
85.         </author>
86.         <source-id>26209</source-id>
87.         <fund-no>undefined</fund-no>
88.         <openaccess>0</openaccess>
89.         <openaccessFlag>>false</openaccessFlag>
90.     </entry>
91. </search-results>
```

Приложение Б

Ссылка на визуализацию самого большого кластера: <https://clck.ru/VPgnp>

Ссылка на исходный код: <https://clck.ru/VSmKH>