

Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Инженерная школа ядерных технологий
Направление подготовки: Прикладная математика и информатика
Отделение школы (НОЦ): Математики и информатики

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Эффективный анализ больших данных, извлеченных из наукометрической базы Scopus

УДК 004.62:001.891(048.84):519.237

Студент

Группа	ФИО	Подпись	Дата
0ВМ61	Неретин А.А.		

Руководитель

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

КОНСУЛЬТАНТЫ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Меньшикова Е.В.	Кандидат философских наук		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор	Федорчук Ю.М.	Доктор технических наук		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
	Трифонов А.Ю.	Профессор ф- м. наук		

Томск – 2018 г

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
0ВМ61	Неретин Артём Александрович

Инженерная школа ядерных технологий			
Уровень образования	Магистр	Направление/специальность	Прикладная математика и информатика

Тема дипломной работы: «Эффективный анализ больших данных, извлеченных из наукометрической базы Scopus»

Исходные данные к разделу «Социальная ответственность»:

1. Целью данной работы является формирование научных групп на основе анализа данных, полученных из базы данных Scopus.
2. Описание рабочего места на предмет возникновения:
 - вредных проявлений факторов производственной среды (необходимо обеспечить оптимальные, в крайнем случае, допустимые значения метеоусловий на рабочем месте, обеспечить комфортную освещенность рабочего места, уменьшить до допустимых пределов шум от персональной ЭВМ, вентиляции, обеспечить безопасные значения электромагнитных полей от персонального компьютера);
 - опасных проявлений факторов производственной среды (в связи с присутствием электричества для питания энергоблока персонального компьютера и освещенности аудитории необходимо предусмотреть средства коллективной и индивидуальной защиты от электро-, пожаро- и взрывоопасности).

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Анализ выявленных вредных факторов проектируемой производственной среды в следующей последовательности:
 - указывается воздействие фактора на организм человека;
 - приводятся данные по оптимальным и допустимым значениям микроклимата на рабочем месте, перечисляются методы обеспечения этих значений; приводится расчет освещенности на рабочем месте;
 - приводятся данные по реальным значениям шума на рабочем месте, разрабатываются мероприятия по защите персонала от шума, при этом приводятся значения предельно-допустимого уровня, средства коллективной защиты (СКЗ), средства индивидуальной защиты (СИЗ);
 - приводятся данные по реальным значениям электромагнитных полей на рабочем месте от персонального компьютера, перечисляются СКЗ и СИЗ;
 - приводятся допустимые нормы с необходимой размерностью (с ссылкой на соответствующий нормативно-технический документ);
 - предлагаемые средства защиты (сначала коллективной защиты, затем – индивидуальные защитные средства).

<p>2. Анализ выявленных опасных факторов проектируемой производённой среды в следующей последовательности</p> <ul style="list-style-type: none"> – приводятся данные по значениям напряжения используемого оборудования, классификация помещения по электробезопасности, допустимые безопасные для человека значения напряжения, тока и заземления (в т.ч. статическое электричество, молниезащита - источники); перечисляются СКЗ и СИЗ; – приводится классификация пожароопасности помещений, указывается класс пожароопасности вашего помещения, перечисляются средства пожаробнаружения и принцип их работы, средства пожаротушения, принцип работы, назначение (какие пожары можно тушить, какие – нет), маркировка; пожаровзрывобезопасность (причины, профилактические мероприятия).
<p>3. Охрана окружающей среды:</p> <ul style="list-style-type: none"> – анализ воздействия при работе на ПЭВМ на атмосферу, гидросферу, литосферу; – наличие отходов (бумага, картриджи, компьютеры и т. д.); – методы утилизации отходов.
<p>4. Защита в чрезвычайных ситуациях:</p> <ul style="list-style-type: none"> – выявление типичных аварийных ситуаций, причин их возникновения; – разрабатываются превентивные меры по предупреждению ЧС; – разработка мер по повышению устойчивости объекта к данной ЧС; – разработка действий в результате возникшей ЧС и мер по ликвидации её последствий; – ЧС природная – сильный мороз; – ЧС техногенная – предполагаемая диверсия.
<p>5. Правовые и организационные вопросы обеспечения безопасности:</p> <ul style="list-style-type: none"> – специальные (характерные для проектируемой рабочей зоны) правовые нормы трудового законодательства: СанПиН 2.2.2.542-96; СанПин 2.2.2.542-96; СанПиН 2.2.2/2.4.1340-03; СНиП-23-05-95; Сан.ПиН 2.2.2. 542 – 96; ГОСТ 12.1.036-96;ГОСТ 12.1.012-96; ГОСТ 12.1.004-76; ГОСТ 12.1.010-76; ГОСТ 12.1.013-78.
<p>Перечень графического материала:</p> <ol style="list-style-type: none"> 1) План эвакуации; 2) План размещения светильников на потолке рабочего помещения.

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

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор	Федорчук Юрий Митрофанович	Доктор технических наук		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
ОВМ61	Неретин Артём Александрович		

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

Студенту:

Группа	ФИО
ОВМ61	Неретину Артёму Александровичу

Школа	ИЯТШ		
Уровень образования	Магистр	Направление/специальность	Прикладная математика и информатика

**Тема дипломной работы: «Эффективный анализ больших данных,
извлеченных из наукометрической базы Scopus»**

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:	
1. <i>Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i>	1. <i>Стоимость расходных материалов</i> 2. <i>Стоимость расхода электроэнергии</i> 3. <i>Норматив заработной платы</i>
2. <i>Нормы и нормативы расходования ресурсов</i>	1. <i>Тариф на электроэнергию</i> 2. <i>Коэффициенты для расчета заработной платы</i>
3. <i>Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	1. <i>Отчисления во внебюджетные фонды (27,1%)</i> 2. <i>Расчет дополнительной заработной платы (12%)</i>
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. <i>Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения</i>	1. <i>Потенциальные потребители результатов исследования;</i> 2. <i>Анализ конкурентных технических решений;</i> 3. <i>SWOT – анализ.</i>
2. <i>Планирование и формирование бюджета научных исследований</i>	1. <i>Структура работ в рамках научного исследования;</i> 2. <i>Определение трудоемкости выполнения работ и разработка графика проведения научного исследования;</i> 3. <i>Бюджет научно - технического исследования (нти).</i>
3. <i>Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования</i>	1. <i>Определение интегрального финансового показателя разработки;</i> 2. <i>Определение интегрального показателя ресурсоэффективности разработки;</i> 3. <i>Определение интегрального показателя эффективности</i>
Перечень графического материала (с точным указанием обязательных чертежей):	

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

Дата выдачи задания для раздела по линейному графику

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Меньшикова Екатерина Валентиновна	Кандидат философских наук		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
0ВМ61	Неретин Артём Александрович		

РЕФЕРАТ

Выпускная квалификационная работа содержит 93 страницы, 13 рисунков, 27 таблиц, 13 источников, 4 приложения + 2 справочных приложения.

Ключевые слова: Elsevier, Scopus API, PostgreSQL, PHP, кластерный анализ, экономика, эконометрика, финансы, базы данных, walktrap, R language.

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.

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

Степень внедрения: низкая, на стадии разработки.

Область применения: научное сообщество, сервисы, производящие сбор, хранение, индексацию и распространение научных работ.

Значимость работы выражается в том, что по результатам исследования возможно выявить мировых лидеров в научном сообществе и устанавливать мировые стандарты, основываясь на полученном результате.

Обзор литературы

T.A. Brooks [1] – в работе рассматриваются различные показатели авторов в зависимости от уровня цитирования, и проведён анализ их эффективности.

Bárbara S. Lancho-Barrantes, Vicente P. Guerrero-Bote, Félix Moya-Aneón [2] – в данной работе произведён анализ современного научного сообщества и зависимость количества цитирований от области публикаций, также предоставлен примитивный алгоритм анализа выборки по авторам.

Cristopher J. Date [3] – на основе стандартов предоставленных в данном источнике была произведена разработка схемы базы данных, позволяющая выполнять эффективное хранение и выгрузку данных.

Asadi, S., Dahlan, H. M. [7] – данная статья включает в себя пример анализа данных по статьям из области «Telematics and Informatics». В ней производятся разносторонние исследования, позволяющие выделить лидеров области, основываясь на данных, предоставляемых Scopus.

Иванов Г.С., Welling L., Thomson L. [4,5] – описаны основы программирования для работы PHP в связке SQL подобными базами данных, также представлены наиболее эффективные методы работы с POST и GET методами, позволяющие выполнять быструю выгрузку и хранение данных.

Elsevier [8] – на основе документации, предоставленной Elsevier, построен интерфейс для выгрузки данных при помощи API Scopus.

Uddin S., Hossain L., Rasmussen K. [9] – в данной статье произведён анализ научного сообщества, основанный на публикациях и взаимосвязях между авторами.

Schlattmann S., 2017 [10] – произведено выделение взаимосвязей между авторами и построение сети сотрудничества при помощи методов кластерного анализа.

Baghdadi, H. S., Ranaivo-Malan, B. [11] – описан алгоритм фильтрации данных, предоставляемых API Scopus, для избавления от повторяющихся данных. Описан метод идентификации автора, для случая, когда один автор имеет несколько записей профиля в базе.

Can˜as-Guerrero, I., Mazarron, F., Pou-Merina, A., Calleja-Perucho, C., Diaz Rubio, G., 2013 [12] – содержит пример анализа статей из области “agronomy” для статей, полученных из интернет источников.

Szymanski, B., De La Rosa, J., Krishnamoorthy, M [13] – в данной статье предоставлен алгоритм “правильного” подсчёта количества цитирований для статей, находящихся в базе данных Scopus.

Оглавление	
Введение	12
Глава 1. Теоретическая часть.....	13
1.1. Методика расчёта <i>CiteScore</i>	13
1.2. Реляционная база данных PostgreSQL	14
1.3. GET и POST запросы.....	15
1.4. Наукометрические базы данных.....	15
1.5. Выгрузка данных	18
1.6. Кластерный анализ.....	20
1.7. Облако тегов	21
Глава 2. Практическая часть.....	23
2.1. Разработка и реализация структуры базы данных.....	23
2.2. Выгрузка данных с использованием автоматизированного скрипта	24
2.3. Первичный количественный анализ	26
2.4. Построение облака тегов	27
2.5. Выделение кластеров ученых	28
2.6. Построение облака тегов имён авторов	32
2.7. Анализ кластеров	32
Глава 3. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	34
3.1. Потенциальные потребители результатов исследования	34
3.2. Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения.....	34
3.3. Диаграмма Исикава.....	36
3.4. SWOT-анализ	37
3.5. Оценка готовности проекта к коммерциализации	38
3.6. Методы коммерциализации результатов научно-технического исследования.....	40
3.5. Инициация проекта	40
3.5.1 Цели и результат проекта.	40
3.5.2 Организационная структура проекта.	41
3.5.3 Ограничения и допущения проекта.	42

3.6. Планирование управления научно-техническим проектом	42
3.6.1. План проекта.....	42
3.6.3. Бюджет научного исследования	46
3.6.4. Организационная структура проекта	47
3.6.5. Расчёт электроэнергии	47
3.6.6. Накладные расходы.....	47
3.6.7. Бюджет научного исследования	48
3.6.8. Реестр рисков проекта.....	49
3.7. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования	50
3.7.1 Оценка сравнительной эффективности исследования	50
Глава 4. Социальная ответственность.	52
1. Социальная ответственность	52
4.1. Описание рабочего места.....	52
4.2. Анализ выявленных вредных факторов проектируемой производственной среды.....	53
4.3. Микроклимат в помещении.....	55
4.4. Освещенность	56
4.5. Электромагнитные поля.....	60
4.6. Шум.....	61
4.7. Организация рабочего места оператора ПЭВМ	62
4.8. Анализ опасных производственных факторов	65
4.9. Электробезопасность	65
4.10. Защита в чрезвычайных ситуациях.....	68
4.11. Пожарная безопасность	69
4.12. Экологическая безопасность	70
4.13. Научно техническая документация	73
4.14. Выводы и рекомендации	74
Список литературы	76
Приложение А. Пример данных о статье (XML формат)	78
Приложение Б. Метод класса отправки запросов	79
Приложение В. Метод получения данных по номеру журнала	80
Приложение Г	82

Приложение Д	83
Приложение Е	84

Введение

Теперь, как никогда ранее научное сообщество располагает огромным количеством материала. Согласно некоторым исследованиям, объём производства научных публикаций удваивается каждые девять лет [6]. Благодаря быстрому развитию информационно-коммуникационных технологий (ИКТ), теперь можно получить доступ к большей части научной литературы через интернет [7], 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 Scopus Database API [8].

Глава 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 15 лет (2003-2018).

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. Методика расчёта *CiteScore*

Для расчёта *CiteScore* используется простая методика, разработанная издательским домом Elsevier, который выпускает около четверти всех научных статей в научном мире.

CiteScore – это общее количество цитирований документов за календарный год, которые были опубликованы в журналах за три предыдущих года, разделенное на количество документов, проиндексированных в Scopus, которые также были опубликованы за три этих года.

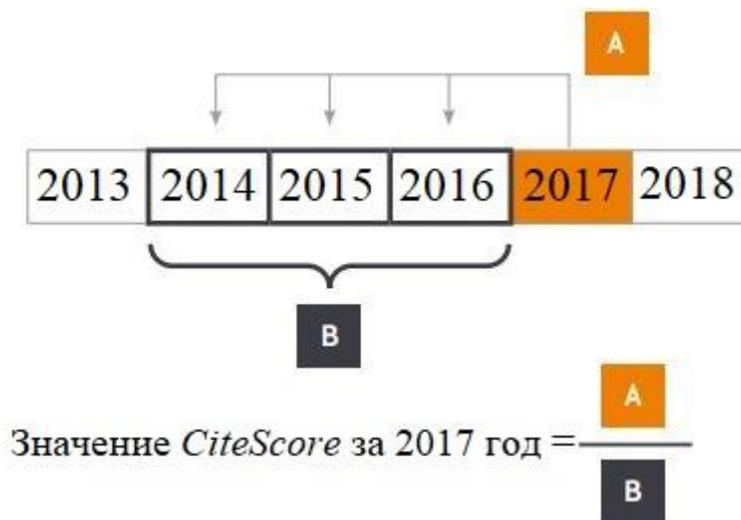


Рисунок 1 - Пример расчёта CiteScore для журнала за 2017 год

Для знаменателя используется трехлетнее окно, которое является оптимальным при работе с большим количеством данных. Оно позволяет получить наиболее полные и актуальные данные [2].

1.2. Реляционная база данных PostgreSQL

Для хранения и выборки данных была выбрана СУБД PostgreSQL, которая является идеальным бесплатным решением для достижения поставленных целей.

PostgreSQL — СУБД с открытым исходным кодом, основой которой является код, написанный в Беркли. Она поддерживает большую часть стандарта SQL, позволяет проводить быстрый анализ большого количества 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.

1.3. GET и POST запросы

POST и GET методы используются для отправки запросов и получения ответов между клиентом и сервером.

На раннем этапе развития сети Интернет существовали запросы, использующие GET метод. Они представляют собой передачу данных непосредственно в адресной строке браузера, в нашем случае используется PHP класс **Curl**, которые позволяет отправлять запросы непосредственно из скрипта.

Основным недостатком GET-запросов является ограниченность передаваемых данных. На стороне сервера строка запроса ограничивается максимальным значением, которое задаётся в настройках сервера. Например, если максимальный размер запроса может составлять 1024 символа, то входные данные обрежутся, в следствии чего часть информации потеряется. Вторым существенным ограничением является возможность передачи строго определенных наборов символов.

POST метод был разработан для удаления данных недостатков GET метода, он позволяет передавать большие объёмы данных, в основном данные запросы используются для загрузки файлов на сервер, отличным примером использования POST метода является загрузка личной фотографии в социальные сети.

1.4. Наукометрические базы данных

В настоящее время большинство исследователей имеют доступ к высококачественным междисциплинарным научным базам данных. Некоторые из них являются базами данных открытого доступа (например,

Google Scholar), в то время как к другим получают доступ благодаря подпискам, сделанным государственными или частными организациями, для которых работают эти исследователи (например, Web of Science, Scopus]). Аналогичным образом, существуют базы данных по конкретным областям или областям, таким как медицина, которые включают ссылки на биомедицинскую литературу из MEDLINE, журналов по науке о жизни и онлайн-книг. В различных исследованиях были проанализированы характеристики научных баз данных. Основные источники информации, такие как научные базы данных (например, Scopus), поисковые системы (например, Google Scholar) и социальные сети (например, academia.edu или ResearchGate), анализируются Асади [7]. Широкое обсуждение было посвящено преимуществам и недостаткам этих источников информации. Например, важным преимуществом Google Scholar является то, что информация периодически обновляется с использованием существующей информации в Интернете. Однако Google Scholar включает только профили исследователей, которые добровольно вписали себя, с многочисленными ошибками в связывании документов с авторами, так что многие профили исследователей имеют публикации, которые действительно не соответствуют этому автору. Это неудобство частично устраняется с помощью уникальных и глобальных идентификаторов, таких как Open Researcher и идентификатор участника (ORCID), который присваивает каждому автору уникальный идентификатор, хотя такая информация должна быть включена или связана вручную. Эти недостатки устранены в литературных базах данных, источники данных которых являются редакционными, профессиональными ассоциациями. Это касается Web of Science и Scopus, чья информация считается более надежной, хотя она включает в себя меньшее количество документов, чем другие системы, такие как Google Scholar. Чтобы проверить надежность и точность этих баз данных, в разных исследованиях были проанализированы типичные ошибки в информации, содержащейся в ней. Например, среди наиболее распространенных ошибок Scopus является наличие нескольких

идентификаторов авторизации для одного и того же человека из-за их присутствия в разных исследовательских институтах или того факта, что они использовали разные форматы подписей в разных публикациях. Таким образом, достоверность любого библиометрического исследования на основе этих данных ограничена. В настоящем исследовании эти неточности были проверены и были максимально сведены к минимуму благодаря использованию инструментов для уточнения текста и сверки данных. Ожидается, что надежность информации, содержащейся в Scopus, будет увеличиваться в будущем, и в результате будет улучшена точность результатов, скомпилированных с помощью представленного здесь метода.

Все эти базы данных часто ориентированы на поиск отдельных исследователей и документов по конкретной теме исследования. Однако его использование для извлечения и анализа совокупных данных не представляло интереса до последних лет. К счастью, некоторые из этих баз недавно включили инструменты, которые позволяют выполнять поиск и собирать данные для разработки исследований другого характера. Это относится к интерфейсу API базы данных Scopus, который доступен для общественности и позволяет получать необработанные данные из базы данных Scopus на основе разных критериев. Более конкретно, документы в Scopus классифицируются по четырем широким областям: экономика, эконометрика и финансы, науки о жизни, физические науки, науки о здоровье и социальные науки и гуманитарные науки. Эти общие предметы дополнительно подразделяются на 27 основных тематических категорий, а также на более чем 300 конкретных предметных категорий, хотя данный документ может быть одновременно включен в разные категории. Для анализа мы выбрали только одну предметную область: экономика, эконометрика и финансы.

1.5. Выгрузка данных

Elsevier API в основном работает с GET методом, поэтому для выгрузки данных требуется разработка интерфейса для отправки GET и парсинга ответа с дальнейшей загрузкой данных в базу данных.

Разработка осуществляется с использованием языков HTML, PHP и Python [5]. Для анализа данных используется язык Python.

Для работы с данными были созданы следующие классы:

- класс работы с базой данных (class_db.php);
- класс для парсинга и обработки данных (class_parser.php);
- класс для отправки запросов GET методом (class_request.php).

Пример выгружаемых данных представлен в приложение А. Для наглядности данные приведены к формату XML. В разработанном интерфейсе работа осуществляется с ответом сервера в формате JSON.

Пример метода класса приведён в приложение Б.

Перед запуском скрипта был произведён выбор журналов по следующим критериям:

- SiteScore > 0.5 или неиндексированный;
- Направленность журнала на область Economic, Econometric and Finance;
- Наличие статей в базе данных Scopus с датой публикации не ранее чем 2008 год.

Упрощенный алгоритм работы скрипта представлен на рисунке 2.

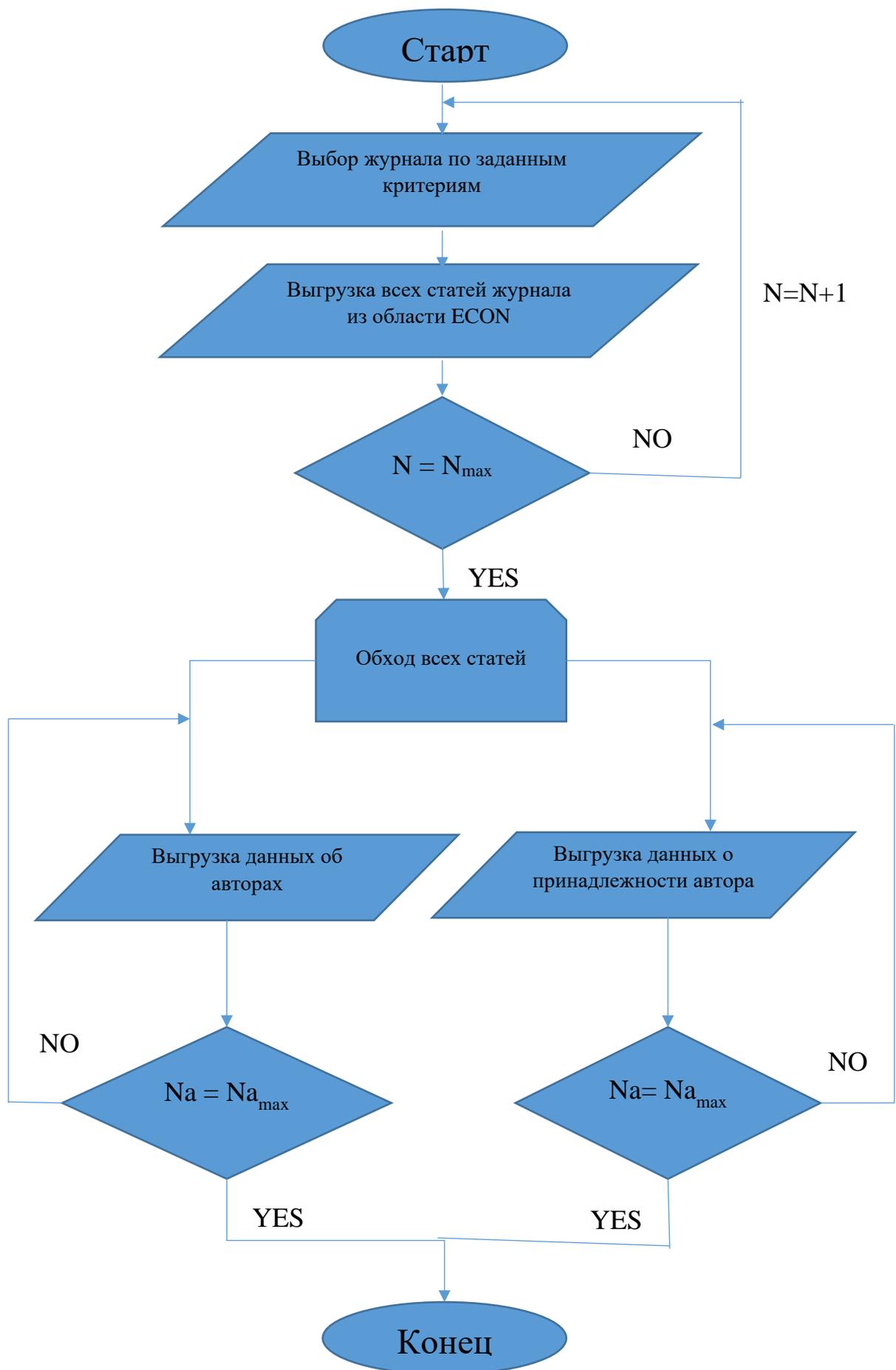


Рисунок 2 - Упрощенная схема работы скрипта выгрузки данных через API

1.6. 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.

Методы кластерного анализа подразделяются на:

- 1) агломеративные (объединяющие);
- 2) дивизимные (разделяющие);
- 3) итеративные.

Первые последовательно объединяют отдельные объекты в кластеры, вторые, наоборот, расчленяют кластеры на объекты. Третьи объединяют первые два. Их особенностью является формирование кластеров, исходя из условий разбиения (так называемых параметров), которые могут быть изменены в процессе работы алгоритма для улучшения качества разбиения.

Итеративные методы обычно используются для классификации больших объемов информации.

Для проведения классификации данных используют понятие метрики или расстояния.

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

1) $\rho(X,Y) \geq 0$, 2) $\rho(X,Y) = \rho(Y,X)$, 3) $\rho(X,Y) = 0 \Leftrightarrow X = Y$, 4) $\rho(X,Y) \leq \rho(X,Z) + \rho(Z,Y)$

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.

1.7. Облако тегов

Это визуальное представление списка ключевых слов (или тегов, также называемых метками, ярлыками, и т. п.). Обычно используется для представления неформатированного текста. 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 воспользуемся 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.

Принцип работы алгоритма для построения облака тегов представлен на рисунке 3.

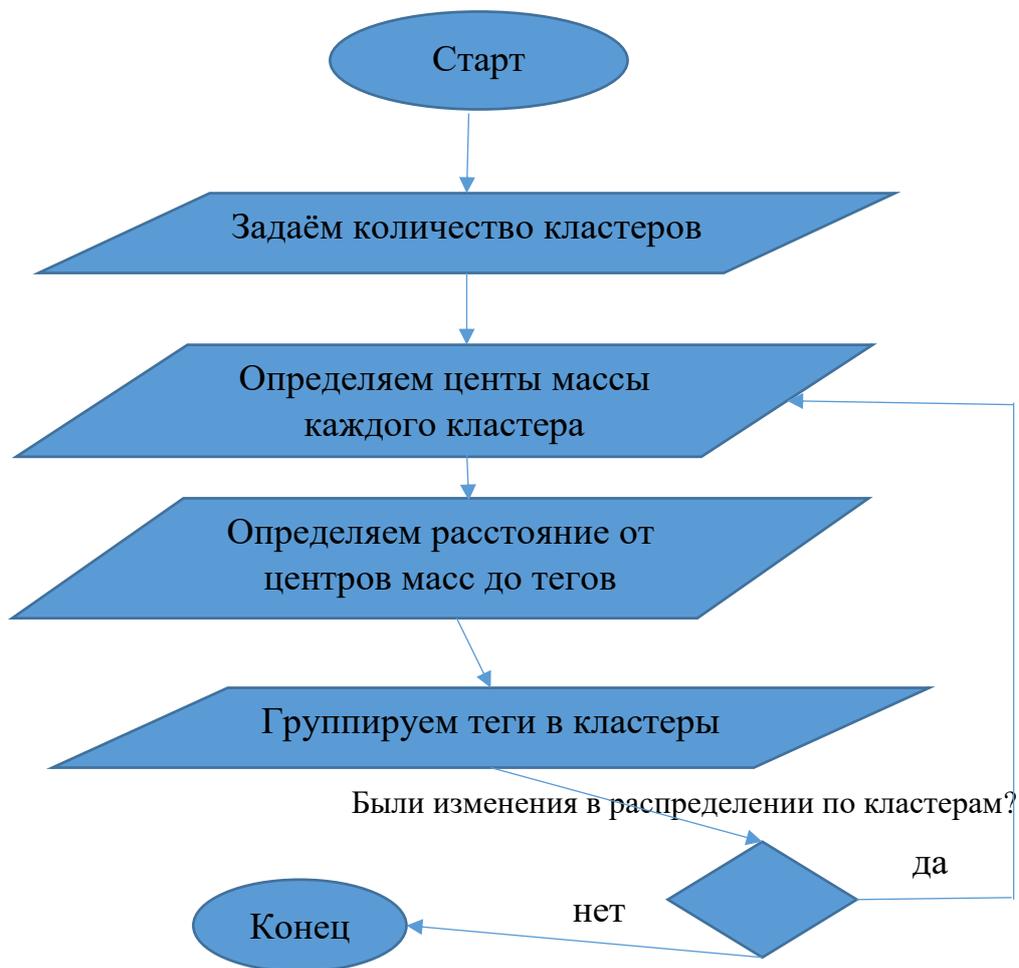


Рисунок 3. Принцип работы алгоритма построения облака тегов

Глава 2. Практическая часть

2.1. Разработка и реализация структуры базы данных

Необходимо разработать структуру данных для того, чтобы хранить требуемую для анализа информацию. Для решения данной задачи была выбрана реляционная база данных (БД), основывающаяся на SQL синтаксисе. Разработанная структура БД имеет вид, представленный на рисунке 4.

Для визуализации структуры базы данных использован интернет ресурс dbdesigner.net.

Поля в таблицах имеют следующее соответствие с полями в базе данных Scopus.

- `affiliations.id = affiliation.afid;`
- `sources.id = source.record_id;`
- `authors.id = author.dc:identifier(SCOPUS_ID);`
- `articles.id = articles.dc:identifie(SCOPUS_ID).`

Для составления облака тегов была создана функция, производящая обработку столбца `AuthorKeyWords`, таблицы `articles`. Данная функция подсчитывает статистику по количеству и частоте использования различных ключевых слов автора.

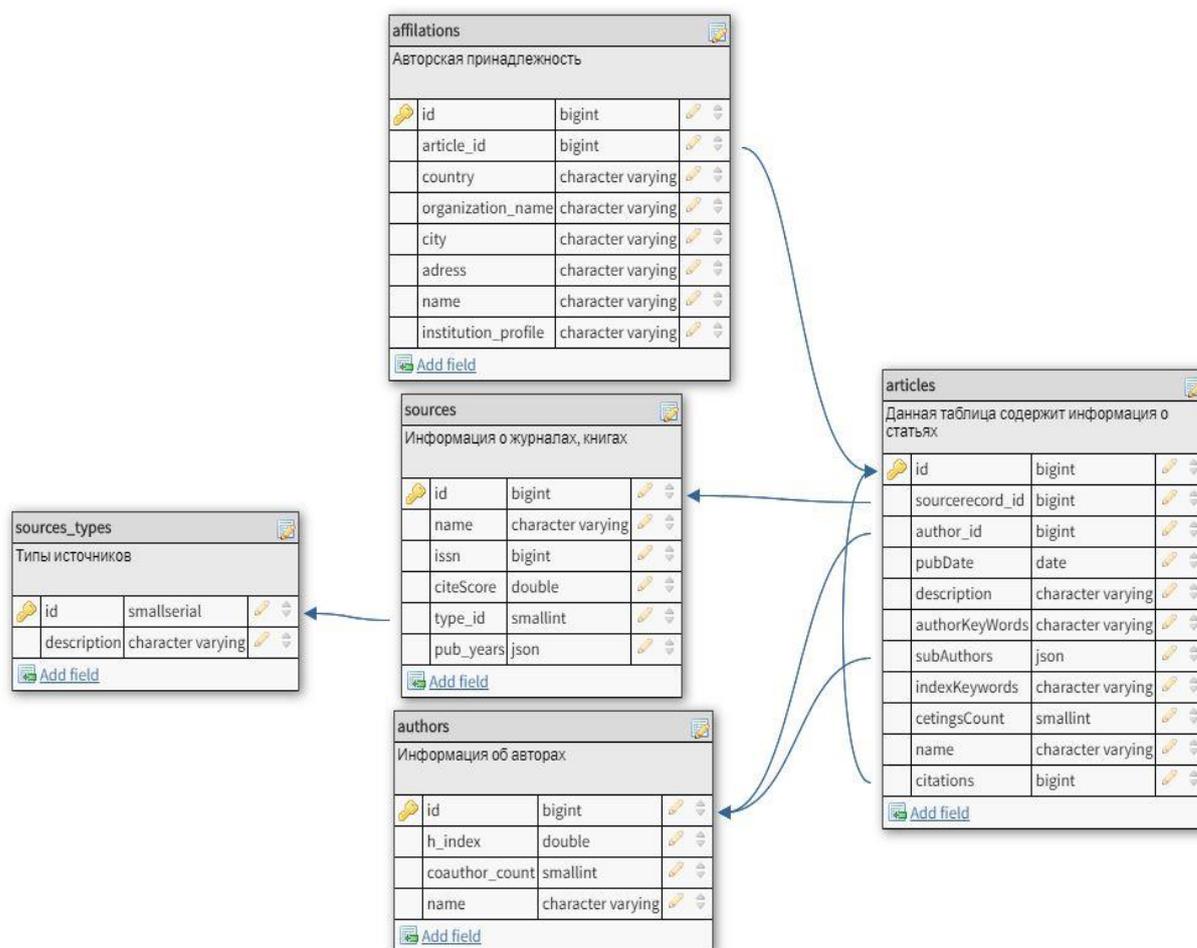


Рисунок 4 - Структура БД для хранения данных

2.2. Выгрузка данных с использованием автоматизированного скрипта

В этом разделе представлен метод выгрузки большого объема данных из любого академического или исследовательского учреждения с целью определения различных библиографических показателей отдельных или групп исследователей, а также сетей сотрудничества этих исследователей с точки зрения публикаций. Этот метод, который может представлять большой интерес для исследователей, научно-исследовательских институтов и финансирующих агентств, использует возможности, предоставляемые интерфейсом API базы данных Scopus, для автоматизации поиска рукописей, опубликованных авторами и организациями. Scopus Elsevier также является самой большой базой данных рецензируемой литературы в различных

научных областях. В нашем случае мы производим выгрузку по области публикаций.

Для нашей работы мы создали скрипт автоматического извлечения информации из базы данных Scopus. Действия этого скрипта, которые позволяют собирать все данные, необходимые для разных анализов, можно разделить на три следующих этапа:

- 1) Получить данные статей: на этом этапе скрипт извлекает всю информацию из публикаций авторов, которые публиковали свои работы, в определенных журналах, которые мы выбрали для анализа. Для выполнения этой задачи информация в каждой из статей извлекается, а «authorid» (уникальный идентификатор автора в Scopus) будет храниться в базе данных скрипта, если он ранее не был зарегистрирован.
- 2) Получить данные авторов. Из списка исследователей, чья принадлежность совпадает с желаемой, скрипт извлекает и сохраняет в своей базе данных информацию, которую Scopus предоставляет для каждого из этих авторов, включая зарегистрированные публикации, даты этих публикаций, историю принадлежности, h-индекс, количество ссылок и т. д.
- 3) Получить сети сотрудничества. Индивидуальная информация о каждом исследователе используется для установления совместных отношений, основанных на совместных публикациях. Более конкретно, на этом этапе скрипт применяет итеративный процесс для каждого из авторов учреждения, затем получает информацию, такую как название учреждения, город, страна, количество соавторов, текущая принадлежность этих сотрудников и т. д.

2.3. Первичный количественный анализ

В данном разделе представлен первичный анализ информации о статьях за 2003-2018 годы (без учёта *SiteScore*). В ходе импорта была получена информация о 653908 статьях, которые написаны представителями различных стран.

Стоит отметить, что при написании статьи автор может иметь принадлежность сразу к нескольким организациям, находящимся в разных городах и даже разных странах.

Таблица 1. количество публикаций за каждый год

Год	2003	2004	2005	2006	2007	2008	2009	2010
Количество документов	20656	21212	26149	29248	33445	38270	41667	45472

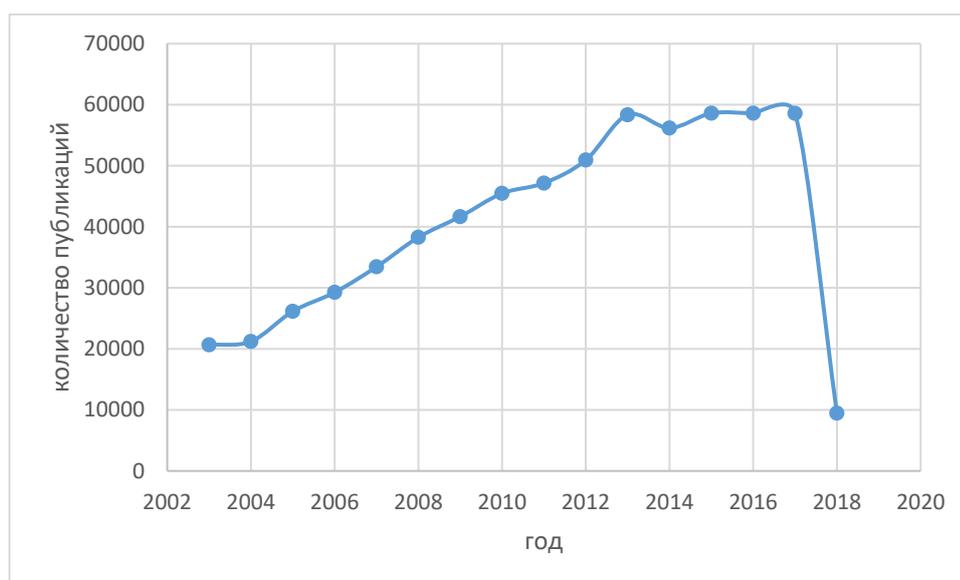


Рисунок 5. Количество публикаций в области ECON за 2003-2018г.

По графику видно, что с 2003 по 2014 годы количество публикаций росло линейно, а с 2015 по 2017 прекратило свой рост и сохраняет количество научных работ около 60000 в год.

Таблица 2. Количество публикаций в зависимости от страны

Страна	США	Неопределенная	Великобритания	Германия	Россия
Количество публикаций	179391	100769	71343	38921	8456

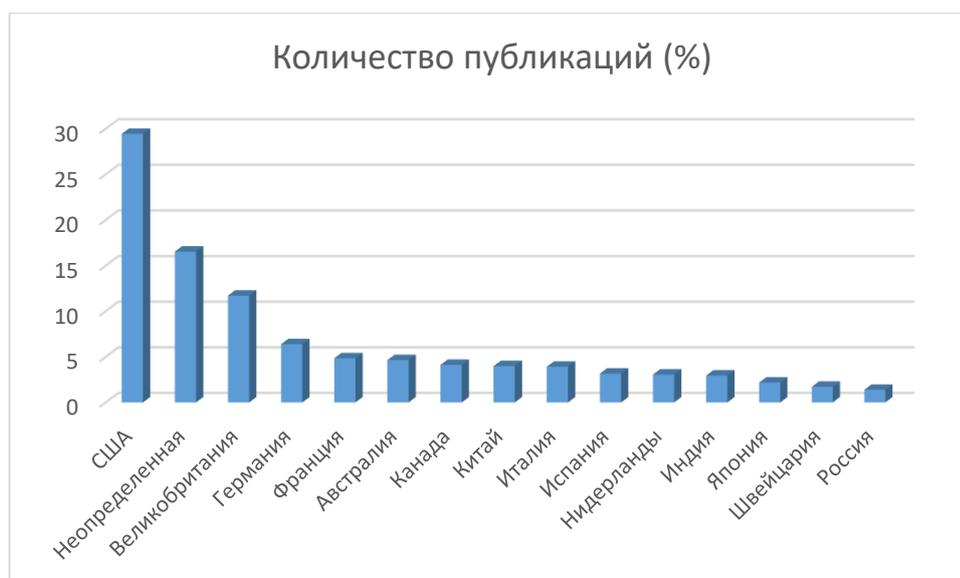


Рисунок 6. Количество публикаций в области ECON за 2003-2018г. в зависимости от страны (все работы, представленные в базе)

2.4. Построение облака тегов

Для построения облака тегов использовались ключевые слова всех статей, выгруженных из выбранных журналов.

В Топ используемых ключевых слов вошли:

- China (3157);
- Innovation (3106);
- Economic growth (2965);
- Entrepreneurship (2877);
- Perfomance (2602).

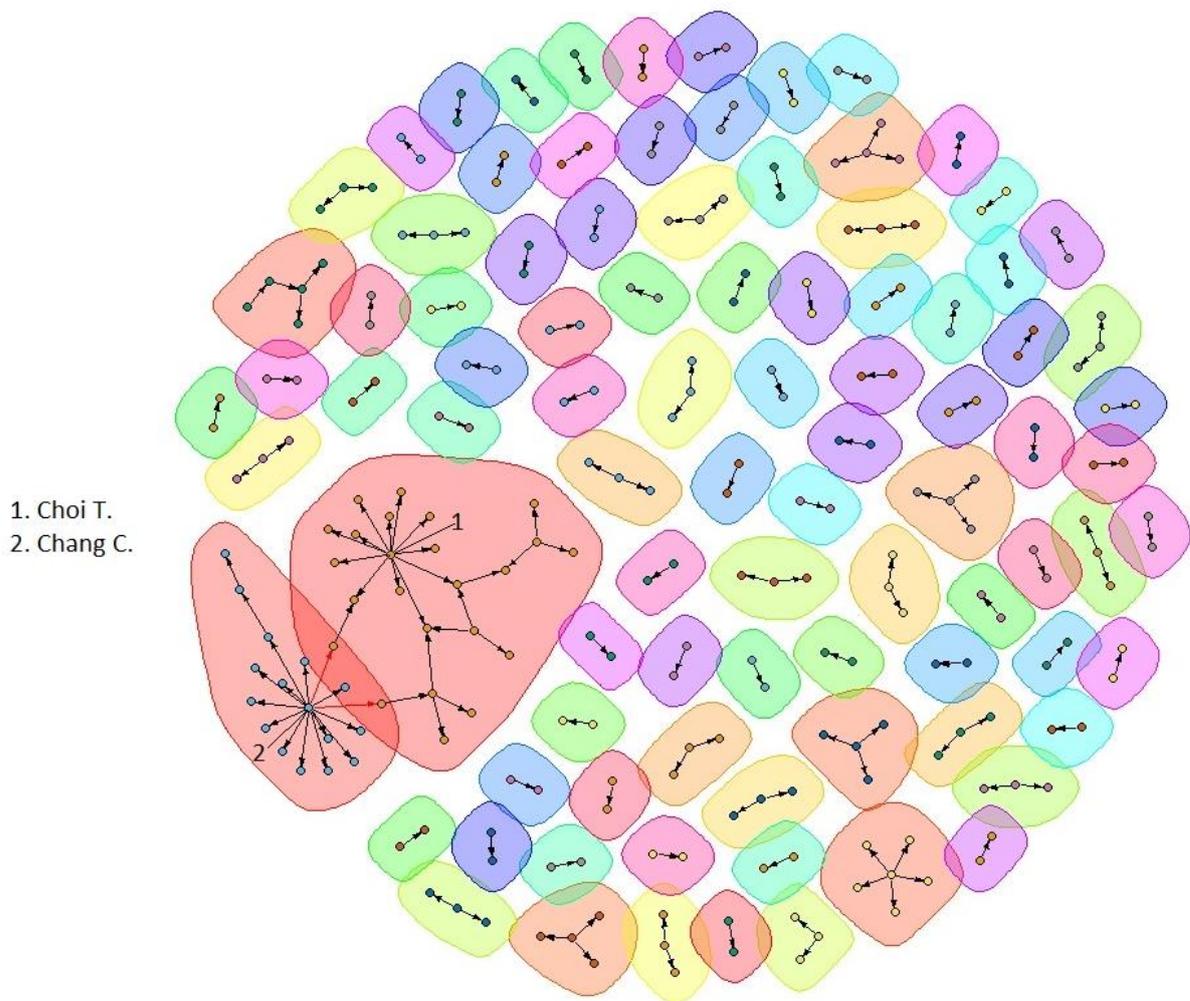


Рисунок 8. Пример распределения авторов на кластеры

На рисунке 8 представлены 92 кластера, такое количество получилось из-за малого количества данных, добавленных для анализа. Из графа видно, что по данной выборке данных мы имеем 2 взаимосвязанных кластера учёных, лидерами по количеству цитирований в которых являются Choi T. и Chang C.

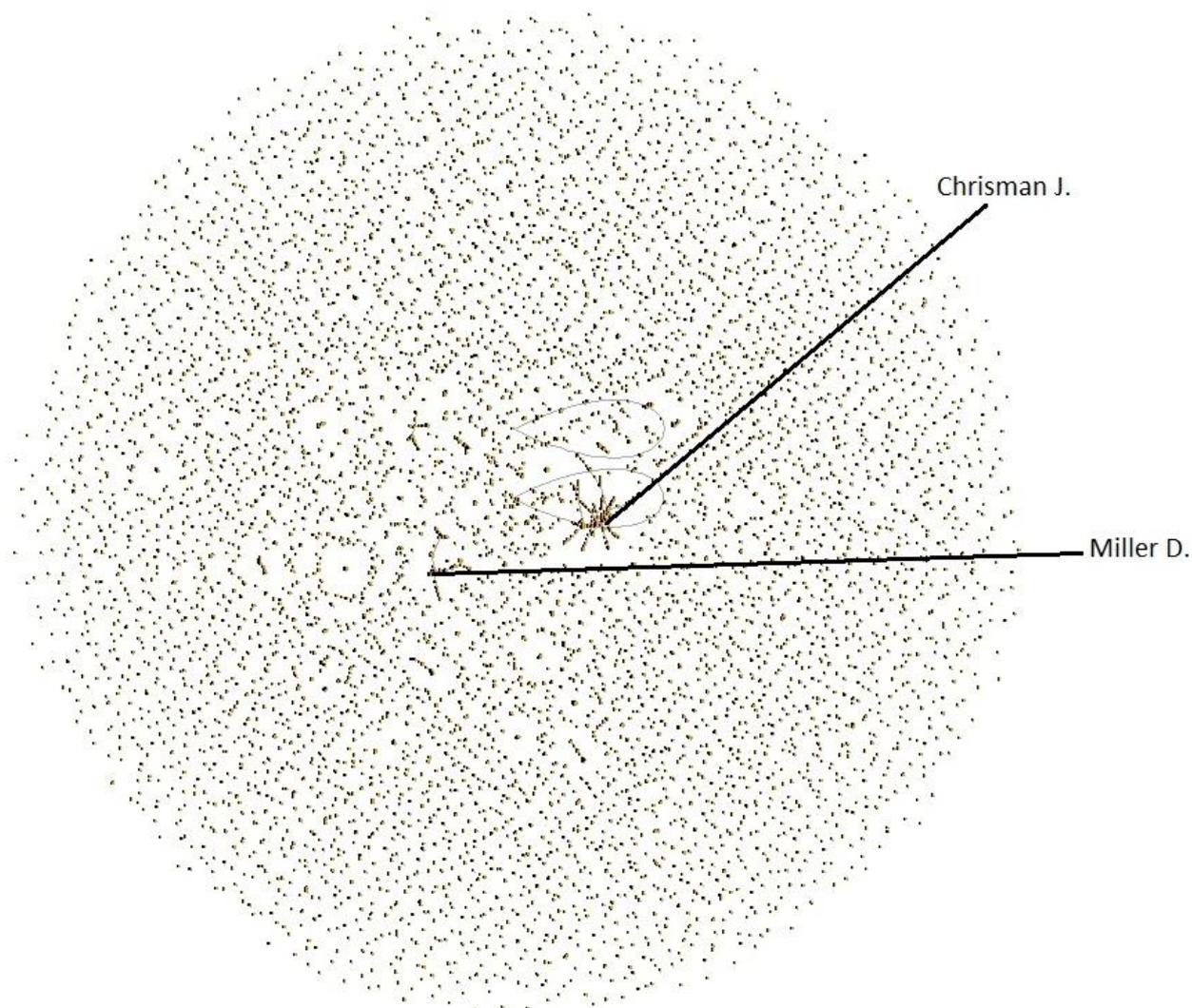


Рисунок 9. Взаимосвязи 5000 авторов без выделения кластеров.



Рисунок 10. Граф взаимосвязей для самого большого кластера выборки

На рисунке 9 представлены граф всех взаимосвязей, присутствующих в выборке, а также лидеры по цитированиям из данных групп. Цветовое представление невозможно в связи с большим количеством взаимосвязей и конечного числа найденных кластеров (около 2500).

На рисунке 10 представлен граф взаимосвязей для самого большого кластера присутствующего во всей выборке.

Таблица 3. Статистика самых больших кластеров

Лидер кластера	Количество авторов в кластере
Chrisman J.	400
Miller D.	241

2.6. Построение облака тегов имён авторов

В данном разделе мы строим облако тегов для самого большого кластера, найденного в пункте 2.6. Основными данными для него являются имена авторов, размер имени изменяется от количества цитирований статей предоставляемого автора. Будем рассматривать только авторов с количеством цитирований больше 200.

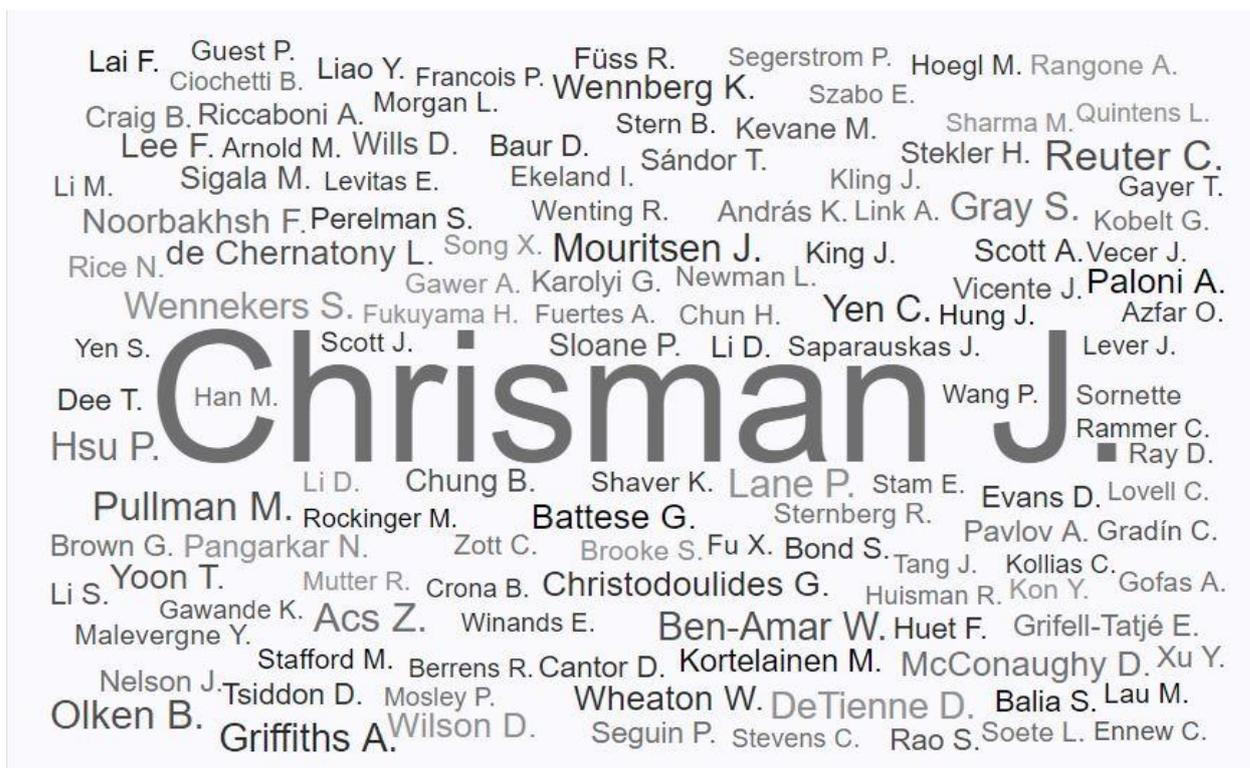


Рисунок 11. Облако тегов для самого большого кластера

2.7. Анализ кластеров

В данной пункте мы проведём количественный анализ кластеров, включающих в себя более одной статьи, имеющие не нулевое количество цитирований. В полученной выборке удалось выделить 57 кластеров, которые удовлетворяют поставленным условиям, все эти кластеры включают в себя 50 и более авторов.

Наибольшее количество людей, входящих в кластер, равняется 400, а наименьшее в свою очередь - 50. Статистика по анализу выделенных кластеров представлена в таблицах 4,5.

Таблица 4. Статистика анализируемых кластеров

Количество групп	57
Среднее число авторов	106
Среднее число цитирований (на всю группу)	3125
Число авторов в 57 анализируемых группах	6043
Вероятность попадания автора в группу анализа на основе количества людей	26.6%
Вероятность попадания автора в группу анализа на основе количества анализируемых кластеров*	12.5%
Средний географический охват	1.67
Среднее число статей	9

* для вычисления данного пункта использовались только группы, в которые попало более 10 человек, в которых не нулевое число цитирований и более одной опубликованной статьи на весь кластер.

Таблица 5. Статистика кластера с наибольшим количеством авторов

Число авторов	400
Число цитирований	13012
Вероятность попадания автора в группу на основе количества людей	1.76%
Вероятность попадания автора в группу анализа на основе количества анализируемых кластеров*	0,025%
Географический охват	4
Число статей	25

* для вычисления данного пункта использовались только группы, в которые попало более 10 человек, в которых не нулевое число цитирований и более одной опубликованной статьи на весь кластер.

Глава 3. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

3.1. Потенциальные потребители результатов исследования

Потребителями разработанной технологии могут быть:

- научные организации, производящие индексацию статей;
- интернет ресурсы рассчитывающие рейтинги журналов (Elsevier);
- научное сообщество (университеты);

3.2. Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения

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.

B_{ϕ} – разработка, представленная в диссертации

B_{k1} – h-индекс

B_{k2} – g-индекс

Таблица 6. Оценочная карта для сравнения с h-индексом и g-индексом

Критерии оценки	Вес критерия	Баллы			Конкурентоспособность		
		B_{ϕ}	B_{k1}	B_{k2}	K_{ϕ}	K_{k1}	K_{k2}
1	2	3	4	5	6	7	8
Технические критерии оценки ресурсоэффективности							
1. Удобство расчёта	0.1	3	4	4	0.3	0.4	0.4
2. Затраты компьютерных ресурсов для расчёта	0.05	2	5	3	0.1	0.25	0.15
3. Скорость получения результатов	0.1	2	4	5	0.2	0.4	0.5
4. Смысловая нагрузка критерия	0.5	5	2	2	2.5	1	1
Экономические критерии оценки эффективности							
1. Конкурентоспособность продукта	0.05	5	4	4	0.25	0.2	0.2

2. Уровень проникновения на рынок	0.1	1	5	5	0.1	0.5	0.5
3. Затраты на разработку	0.1	5	1	1	0.5	0.1	0.1
Итого	1				3.95	2.85	2.85

Конкурентоспособность = $3.95/2.85 = 1.39$.

Конкуренты уязвимы перед нашей разработкой, потому что они используют устаревшие разработки для индексации статей.

Представленная разработка производит расчёт по большему количеству данных и учитывает множество факторов, которые влияют на научные труды, что является основным конкурентным преимуществом.

3.3. Диаграмма Исикава



Рисунок 12. Диаграмма Исикава

3.4. SWOT-анализ

Таблица 7. Интерактивная матрица проекта

Сильные стороны проекта						
Возможности проекта		C1	C2	C3	C4	C5
	B1	+	+	+	0	0
	B2	+	+	+	+	0
	B3	+	-	0	0	0
	B4	+	-	0	-	0
	B5	-	-	+	+	0

Возможные направления реализации проекта:

- B2B1C1C2C3C4;
- B3C1;
- B5C3C4.

Таблица 8. SWOT-анализ

	<p>Сильные стороны научно-исследовательского проекта:</p> <p>C1. Низкая стоимость проекта и высокая смысловая нагрузка разрабатываемого критерия.</p> <p>C2. Наличие доступа к наукометрической базе Scopus через VPN Томского Политехнического.</p> <p>C3. Отсутствие высоких требований к оборудованию.</p> <p>C4. Наличие бюджетного финансирования.</p> <p>C5. Опытные программисты, имеющие широкий опыт в сфере анализа данных.</p> <p>C6. Широкий охват предметной области.</p>	<p>Слабые стороны научно-исследовательского проекта:</p> <p>Сл1. Отсутствие прототипа научной разработки.</p> <p>Сл2. Узкая направленность на определенную структуру данных.</p> <p>Сл3. Отсутствие полного доступа к требуемым данным.</p>
<p>Возможности:</p> <p>B1. Использование инновационной инфраструктуры ТПУ.</p> <p>B2. Использование инновационных разработок, предоставляемы ТПУ.</p>	<p>- Поиск программистов на базе ТПУ.</p> <p>- Расширение научной базы для совершенствования критерия.</p> <p>- Установление связи со Scopus при помощи их контракта с ТПУ.</p>	<p>- Произвести разработку прототипа программы с использованием ресурсов подконтрольных ТПУ.</p> <p>- Проведение рекламных акций, которые позволят популяризировать данное решение.</p>

<p>В3. Появление дополнительного спроса на новый продукт</p> <p>В4. Бесплатный доступ к требуемому программному обеспечению, посредством ELMS ТПУ.</p> <p>В5. Представление необоснованности использования конкурентных решений.</p>		<p>- Показ преимуществ над решениями, используемыми научным сообществом на данном этапе.</p>
<p>Угрозы:</p> <p>У1. Незаинтересованность научного сообщества в новых, более полных критериях.</p> <p>У2. Появление новых разработок, учитывающих более широкий спектр факторов и взаимосвязей.</p> <p>У3. Ограничение уровня доступа к данным для ТПУ.</p> <p>У4. Введения дополнительных государственных требований к оценки научных работ.</p> <p>У5. Несвоевременное финансовое обеспечение научного исследования со стороны государства и научного сообщества.</p>	<p>- Демонстрация эффективности разработки заинтересованным сторонам и малые требования к бюджету.</p> <p>- Проведение консультаций с государственными представителями, для выяснения точных критериев, необходимых при оценки научных работ.</p>	<p>- Поиск инвесторов в различных кругах, а не только в научных.</p> <p>- Презентация экономической эффективности данного решения.</p>

3.5. Оценка готовности проекта к коммерциализации

Таблица 9. Бланк оценки степени готовности научного проекта к коммерциализации

№ п/п	Наименование	Степень проработанности научного проекта	Уровень имеющихся знаний у разработчика
1.	Определен имеющийся научно-технический задел	5	5
2.	Определены перспективные направления коммерциализации научно-технического задела	3	3

3.	Определены отрасли и технологии (товары, услуги) для предложения на рынке	5	4
4.	Определена товарная форма научно-технического задела для представления на рынок	4	2
5.	Определены авторы и осуществлена охрана их прав	3	5
6.	Проведена оценка стоимости интеллектуальной собственности	2	3
7.	Проведены маркетинговые исследования рынков сбыта	3	2
8.	Разработан бизнес-план коммерциализации научной разработки	2	1
9.	Определены пути продвижения научной разработки на рынок	4	4
10.	Разработана стратегия (форма) реализации научной разработки	1	3
11.	Проработаны вопросы международного сотрудничества и выхода на зарубежный рынок	2	3
12.	Проработаны вопросы использования услуг инфраструктуры поддержки, получения льгот	1	2
13.	Проработаны вопросы финансирования коммерциализации научной разработки	5	2
14.	Имеется команда для коммерциализации научной разработки	3	4
15.	Проработан механизм реализации научного проекта	5	4
	ИТОГО БАЛЛОВ	48	47

Перспективность данной разработки выше среднего.

Для её реализации разработчику требуются высокие знания в программировании и команда опытных программистов, которые помогут оптимизировать и ускорить все требуемые расчёты. Для оплаты труда разработчиков требуются небольшие инвестиции.

3.6. Методы коммерциализации результатов научно-технического исследования.

Для коммерциализации результатов научно-технического исследования идеальным решением будет «Торговля патентными лицензиями», потому что данная разработка требует получения патента, который в дальнейшем может передаваться таким научным организациям как Scopus.

В связи с малым количеством бюджетного финансирования продажа патентных лицензий поможет оплатить труд наёмных разработчиков, что приведёт к быстрейшему завершению проекта.

3.5. Инициация проекта

Таблица 10. Заинтересованные стороны проекта

Заинтересованные стороны проекта	Ожидания заинтересованных сторон
ТПУ	Эффективный критерий оценки характеристик научных публикаций.
Scopus	Быстрый, не затратный, несущий высокую смысловую нагрузку критерий, позволяющий сгруппировать научные работы.

3.5.1 Цели и результат проекта.

Таблица 11. Цели и результат проекта

Цели проекта:	Разработка эффективного критерия оценки характеристик научных публикаций.
Ожидаемые результаты проекта:	Программное обеспечение, позволяющие производить анализ большого количество научных работ.
Критерии приемки результата проекта:	Ёмкость критерия, основанная на различных характеристиках статьи.
	Требование:

Требования к результату проекта:	Низкое потребление ресурсов.
	Высокая скорость расчёта.
	Высокая смысловая нагрузка критерия.

3.5.2 Организационная структура проекта.

Таблица 12. Организационная структура проекта

№ п/п	ФИО, основное место работы, должность	Роль в проекте	Функции	Трудо-затраты, час.
1	Неретин А.А., ТПУ, магистр	Разработка backend (алгоритмы, парсеры) программного продукта.	Эксперт проекта, основной разработчик	954
2	Крицкий О.Л., ТПУ, доцент	Консультации по анализу большого количества данных.	Руководитель проекта	204
ИТОГО:				1156

3.5.3 Ограничения и допущения проекта.

Таблица 13. Ограничения проекта

Фактор	Ограничения/ допущения
3.1. Бюджет проекта	120 т.р.
3.1.1. Источник финансирования	ТПУ
3.2. Сроки проекта:	6 месяцев
3.2.1. Дата утверждения плана управления проектом	21.12.2017
3.2.2. Дата завершения проекта	01.06.2018

3.6. Планирование управления научно-техническим проектом

3.6.1. План проекта

Таблица 14. План проекта

Код работы (из ИСР)	Название	Длительность, дни	Дата начала работ	Дата окончания работ	Состав участников (ФИО ответственных исполнителей)
1	Выбор темы ВКР	3	21.12.2017	25.12.2017	Неретин, Крицкий
2	Получение технического задания	4	26.12.2017	30.12.2017	Неретин
3	Подбор материала, его анализ и обобщение	1	01.01.2018	10.01.2018	Неретин
4	Выбор метода выполнения работы	5	11.01.2018	16.01.2018	Неретин
5	Календарное планирование работ по теме	8	17.01.2018	27.01.2018	Неретин, Крицкий
6	Первичный анализ данных	12	28.01.2018	13.02.2018	Неретин
7	Разработка программного продукта	46	14.02.2018	14.04.2018	Неретин

8	Тестирование и выявление недочётов	14	15.04.2018	05.05.2018	Неретин, Крицкий
9	Исправление найденных багов, доработка модели	11	06.05.2018	21.05.2018	Неретин
10	Составление отчёта о проделанной работе с полным анализом результатов	11	22.05.2018	05.06.2018	Неретин
11	Защита дипломного проекта	1	06.06.2018	07.06.2018	Неретин
И т о г о:		116			

3.6.3. Бюджет научного исследования

Специальное оборудование для научных (экспериментальных) работ.

Таблица 16. Расчет затрат по статье «Спецоборудование для научных работ»

№ п/п	Наименование оборудования	Кол-во единиц оборудования	Цена единицы оборудования, тыс.руб.	Общая стоимость оборудования, тыс.руб.
1.	Компьютер	1	42000	42000
2.	МФУ	1	8000	8000
3.	Монитор	1	10000	10000
4.	Комплектующие	1	2000	2000

Таблица 17. Баланс рабочего времени

Показатели рабочего времени	Руководитель	Инженер
Календарное число дней	365	365
Количество нерабочих дней		
- выходные дни	48	48
- праздничные дни	14	14
Потери рабочего времени		
- отпуск	52	52
- невыходы по болезни		
Действительный годовой фонд рабочего времени	251	251

Таблица 6. Расчет основной заработной платы

Исполнители	З _б , руб.	k _р	З _м , руб	З _{дн} , руб.	T _р , раб. дн.	З _{осн} , руб.
Руководитель	33130	1.3	43069	1785	25	44625
Инженер	9893	1.3	12860	533	116	61828

Отчисления на социальные нужды

Отчисления во внебюджетные фонды для руководителя: $27,1\% \cdot (44625) = 12093$ (р)

Отчисления во внебюджетные фонды для инженера: $27,1\% \cdot (61828) = 16755$ (р)

3.6.4. Организационная структура проекта



Рисунок 13. Проектная структура проекта

3.6.5. Расчёт электроэнергии

Потребление:

- Системный блок ~ 400 Вт/ч;
- Монитор ~ 40 Вт/ч;

Затраты на электричество = $0,440 \text{ кВт} \cdot 116 \text{ дней} \cdot 6 \text{ часов/день} \cdot 5,8 \text{ р/кВтч} = 1776 \text{ р.}$

3.6.6. Накладные расходы

$$C_{\text{накл}} = k_{\text{накл}} \cdot (Z_{\text{осн}} + Z_{\text{доп}})$$

$k_{\text{накл}} = 0,8$ (80 % от основной заработной платы работников).

$C_{\text{накл}} = 85163 \text{ р.}$

3.6.7. Бюджет научного исследования

Таблица 7. Группировка затрат по статьям

Вид работ	Статьи					Итого плановая себестоимость
	Техническое оборудование	Электроэнергия	Основная заработная плата	Отчисления на социальные нужды	Накладные расходы	
1. Разработка и реализация проекта	62000	1776	106453	28848	85163	284240

3.6.8. Реестр рисков проекта

Таблица 20. Реестр рисков

№	Риск	Потенциальное воздействие	Вероятность наступления (1-5)	Влияние риска (1-5)	Уровень риска*	Способы смягчения риска	Условия наступления
1	DOS атака	Внедрение вредоносного кода в проект	1	4	Средний	Повышение уровня безопасности	Атака конкурентов
2	Отсутствие соединения с сетью интернет	Отсутствие доступа к требуемым данным	2	5	Высокий	Наличие нескольких провайдеров	Проблемы на линии передач
3	Перегорание комплектующих	Беспрерывная работа оборудования	1	5	Высокий	Наличие комплектующих на замену	Износ оборудования

3.7. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования

3.7.1 Оценка сравнительной эффективности исследования

Интегральный финансовый показатель разработки: 1.

Таблица 21. Сравнительная оценка характеристик вариантов исполнения проекта

Критерии \ ПО	Весовой коэффициент параметра	Текущий проект	Аналог 1
1. Способствует росту производительности труда пользователя	0,1	5	3
2. Удобство в эксплуатации	0,15	5	5
3. Смысловая нагрузка	0,15	5	2
4. Скорость расчёта	0,20	2	5
5. Надежность	0,25	4	2
6. Область охвата	0,15	5	1
ИТОГО	1		

$$I_{\text{П}} = 0.1 \cdot 5 + 0.15 \cdot 5 + 0.15 \cdot 5 + 0.20 \cdot 2 + 0.25 \cdot 4 + 0.15 \cdot 1 = 3.55$$

$$\text{Аналог 1} = 0.1 \cdot 3 + 0.15 \cdot 5 + 0.15 \cdot 2 + 0.2 \cdot 5 + 0.25 \cdot 2 + 0.15 \cdot 1 = 3$$

$$\text{Сравнительная эффективность проекта} = 3.55/3 = 1.18$$

Таблица 22. Сравнительная эффективность разработки

№ п/п	Показатели	Аналог	Разработка
1	Интегральный финансовый показатель разработки	1	1
2	Интегральный показатель ресурсоэффективности разработки	3	3.55
3	Интегральный показатель эффективности	3	3.55
4	Сравнительная эффективность вариантов исполнения	1.18	

Вывод: сравнение значений интегральных показателей эффективности позволило выбрать наиболее эффективный вариант решения проекта. С позиций технической и финансовой ресурсоэффективности можно сделать вывод о том, что научно – техническое решение, разрабатываемое нами, является более эффективным чем предоставленные на рынке аналоги.

Глава 4. Социальная ответственность.

1. Социальная ответственность

4.1. Описание рабочего места

Охрана труда и окружающей среды, а также обеспечения безопасности рабочих в чрезвычайных ситуациях являются важными компонентами организации рабочего процесса в настоящее время.

В современной жизни компьютер широко применяется в жизни человека: и дома, и в офисе, и в магазине, и в производстве, и даже в бытовой технике. Другими словами, компьютеры прочно вошли в повседневную жизнь людей и их использование постоянно увеличивается.

Несоблюдение требований безопасности приводит к тому, что при работе за компьютером сотрудник может ощущать дискомфорт: возникают головные боли и резь в глазах, появляются усталость и раздражительность. У некоторых людей нарушается сон, аппетит, ухудшается зрение, начинают болеть руки, шея, поясница и тому подобное. При ненормированной работе возможно нервное истощение.

Внедрение компьютерных технологий принципиально изменило характер труда и требования к организации и охране труда. Работники, используют персональные ЭВМ (ПЭВМ) практически во всех отраслях деятельности, как в производстве, так и в научно-исследовательских работах. Одновременно с этим, работая с компьютером, работник подвергается вредному воздействию, что может привести к различным профессиональным заболеваниям.

В рамках данного раздела необходимо:

- выявить и изучить вредные и опасные производственных факторы при работе с ПЭВМ и определить способы их устранения;
- оценить условия труда;

- определение способов снижения действия вредных факторов до безопасных пределов или, по возможности, полного их исключения;
- рассмотреть вопросы техники безопасности, безопасности в чрезвычайных ситуациях и охраны окружающей среды.

Объектом исследования является рабочее место (РМ) и помещение, в котором оно находится.

Характеристика помещения, где был разработана магистерская работа: ширина, составляет $b = 5$ м, длина комнаты $a = 6$ м, высота $h = 3,5$ м. Тогда площадь помещения будет составлять $S = ab = 30$ м², объем равен $V = abh = 105$ м³. Также в нем присутствует одно окно, через которое осуществляется вентиляция помещения, с параметрами: ширина 1,5 м, высота 2 м. Количество РМ, $n = 2$. В помещении используется комбинированное освещение – искусственное (люминесцентные лампы типа ЛБ) и естественное (свет из окна). В зимнее время помещение отапливается. Электроснабжение сети переменного напряжения 220В. Помещение без повышенной опасности в отношении поражения человека электрическим током по ГОСТ 12.1.013-78. Компьютер, расположенный на рабочей поверхности высотой 0,77м., обладает следующими характеристиками: процессор AMD A8, оперативная память 8 ГБ, система Microsoft Windows 8.1, частота процессора – 2,00 ГГц, PnP 15,6-и дюймовый монитор с разрешением 1366 на 768 точек и частотой 60 Гц.

4.2. Анализ выявленных вредных факторов проектируемой производственной среды

Пространство или производственная среда, в которой осуществляется трудовая деятельность человека, называется производственным помещением. В этом помещении человек находится определенное время, выполняет определенную работу и на его безопасность и самочувствие, соответственно, оказывается существенное влияние окружающей среды. Поэтому в

производственном помещении должны быть обеспечены и соблюдены нормативные санитарно-технические условия.

Согласно санитарным нормам СанПин 2.2.2.542-96, на каждого работника должно быть выделено не менее 6 м² площади и не менее 20 м³ объема[16]. При высоте помещения – 3,5 м, расстояние от глаз до клавиатуры, экрана и документов было 500 – 600 мм, клавиатура находилась на высоте 740 – 790 мм от уровня пола, верхний край дисплея находился на уровне прямого взгляда, а нижний – не ниже 40° от прямого взгляда.

Найдем площадь и объем исследуемого помещения на одного человека. Площадь всего помещения $S = 30 \text{ м}^2$. Количество рабочих мест $n = 2$. Тогда на одного работающего приходится:

$$\frac{S}{n} = \frac{30}{2} = 15(\text{м}^2).$$

Объём помещения на одного человека составит:

$$V = \frac{S \cdot h}{n} = \frac{105}{2} = 52.5(\text{м}^3).$$

Таким образом, из выше приведенных расчетов, можно сделать вывод, что по занимаемой площади и по занимаемому объёму помещение полностью удовлетворяет нормативным требованиям.

Выбор типа производственного помещения определяется производственным процессом, анализируя санитарно-технические условия которого (площадь и объём производственных помещений, освещённость, микроклимат, вентиляция, шумы, излучения), проводится оценка вредных факторов и сравнение их с требованиями нормативных документов.

Таблица 23 – Основные элементы производственного процесса, формирующие вредные факторы

Наименование видов работ и параметров производственного процесса	Факторы (ГОСТ 12.0.003-74 ССБТ)	Нормативные документы
--	---------------------------------	-----------------------

Оператор ПК	Отклонение показателей микроклимата	ГОСТ 12.1.005–88
Оператор ПК	Недостаточная освещенность	СанПиН 2.2.1/2.1.1.1278-03

4.3. Микроклимат в помещении

Микроклимат производственных помещений – это климат внутренней среды помещений, который определяется действующими на организм сочетаниями температуры, влажности и скорости движения воздуха, а также температуры окружающих поверхностей. Показатели микроклимата должны обеспечивать сохранение теплового баланса человека с окружающей средой и поддержание оптимального или допустимого теплового состояния организма человека [16].

Нормы оптимальных и допустимых метеорологических условий устанавливает СанПиН 2.2.4.548–96. Эти нормы устанавливаются в зависимости от времени года, характера трудового процесса и характера производственного помещения. Все категории работ разграничиваются на основе интенсивности энергозатрат организма в ккал/ч (Вт). Работа, выполняемая математиком-экономистом, производимая сидя и сопровождающаяся незначительным физическим напряжением, относится к категории Ia – работа с интенсивностью энергозатрат до 120 ккал/ч (до 139 Вт). Для данной категории допустимые нормы микроклимата помещения представлены в таблице 24.

Таблица 24 – Допустимые нормы микроклимата в рабочей зоне производственных помещений

Сезон года	Категория тяжести	Температура, С°	Относительная влажность, %	Скорость движения воздуха, м/сек
------------	-------------------	-----------------	----------------------------	----------------------------------

		Фактическое значение	Допустимое значение	Фактическое значение	Допустимое значение	Фактическое значение	Допустимое значение
Холодный	Ia	(21÷23)	(20÷25)	55	(15÷75)	0,1	0,1
Теплый	Ia	(22÷24)	(21÷28)	55	(15÷75)	0,1	0,1

Анализируя, данные таблицы 28 и состояние рабочей комнаты, микроклимат которой поддерживается на оптимальном уровне системой водяного центрального отопления и естественной вентиляцией, можно сделать вывод, что параметры микроклимата производственного помещения соответствуют нормам.

4.4. Освещенность

Свет является естественным условием жизни человека. Правильно спроектированное и выполненное освещение обеспечивает высокий уровень работоспособности, оказывает положительное психологическое действие на человека и способствует повышению производительности труда. На рабочей поверхности должны отсутствовать резкие тени, которые создают неравномерное распределение поверхностей с различной яркостью в поле зрения, искажает размеры и формы объектов различия, в результате повышается утомляемость и снижается производительность труда.

Существует три вида освещения: естественное – за счёт солнечного излучения, искусственное – за счёт источников искусственного света и совмещенное – освещение, включающее в себя как естественное, так и искусственное освещения.

Оценка освещенности рабочей зоны проводится в соответствии с СанПиН 2.2.2/2.4.1.1340-03.

В данном рабочем помещении используется комбинированное освещение — искусственное и естественное. Искусственное освещение создается люминесцентными лампами типа ЛД.

Расчёт общего равномерного искусственного освещения горизонтальной рабочей поверхности выполняется методом коэффициента светового потока, учитывающим световой поток, отражённый от потолка и стен. Длина помещения $a = 6$ м, ширина $b = 5$ м, высота $H = 2,8$ м. Высота рабочей поверхности над полом $h_p = 0,75$ м. Интегральным критерием оптимальности расположения светильников является величина λ , которая для люминесцентных светильников с защитной решёткой лежит в диапазоне 1,1–1,3.

Выбираем лампу дневного света ЛД-40, световой поток которой равен $\Phi_{ЛД} = 2300$ Лм.

Выбираем светильники с люминесцентными лампами типа ОДОР-2-40. Этот светильник имеет две лампы мощностью 40 Вт каждая, длина светильника равна 1227 мм, ширина – 265 мм.

На первом этапе определим значение индекса освещенности i .

$$i = \frac{S}{(a + b)h}, \quad (21)$$

где S – площадь помещения;

h – расчетная высота подвеса светильника, м;

a и b – длина и ширина помещения, м.

Высота светильника над рабочей поверхностью h

$$h = H - h_p - h_c = 2,8 - 0,75 - 0,3 = 1,55 \text{ м}, \quad (22)$$

где H - высота помещения, м;

h_p - высота рабочей поверхности, м;

В результате проведенных расчетов, индекс освещенности i равен

$$i = \frac{S}{(a+b)h} = \frac{30}{(5+6) \cdot 1,55} = 1,7.$$

Расстояние между соседними светильниками или рядами определяется по формуле:

$$L = \lambda \cdot h = 1,1 \cdot 1,55 = 1,6 \text{ м}$$

Число рядов светильников в помещении:

$$Nb = \frac{5}{L} = \frac{4}{1,6} = 3,1 \approx 3$$

Число светильников в ряду:

$$Na = \frac{a}{L} = \frac{6}{1,6} = 3,75 \approx 4$$

Общее число светильников:

$$N = Na \cdot Nb = 4 \cdot 3 = 12$$

Учитывая, что в каждом светильнике установлено две лампы, общее число ламп в помещении $N = 24$.

Расстояние от крайних светильников или рядов до стены определяется по формуле:

$$l1 = \frac{L}{3} = \frac{1,6}{3} = 0,53 \text{ м}$$

$$l2 = \frac{L}{3} = \frac{0,3}{3} = 0,1 \text{ м}$$

Размещаем светильники в три ряда. План помещения и размещения светильников с люминесцентными лампами представлен в приложении Г.

Световой поток лампы определяется по формуле:

$$\Phi = \frac{E_n \cdot S \cdot K_z \cdot Z}{N \cdot \eta}, \quad (23)$$

где E_n – нормируемая минимальная освещённость по СНиП 23-05-95, лк;

S – площадь освещаемого помещения, м²;

K_z – коэффициент запаса, учитывающий загрязнение светильника (источника света, светотехнической арматуры, стен и пр., т. е. отражающих поверхностей), наличие в атмосфере цеха дыма, пыли;

Z – коэффициент неравномерности освещения, отношение E_{cp}/E_{min} . Для люминесцентных ламп при расчётах берётся равным 1,1;

N – число ламп в помещении;

η – коэффициент использования светового потока.

Данное помещение относится к типу помещения со средним выделением пыли, в связи с этим $K_z = 1,5$; состояние потолка – свежепобеленный, поэтому значение коэффициента отражения потолка $\rho_n = 70$; состояние стен – побеленные бетонные стены, поэтому значение коэффициента отражения стен $\rho_c = 50$. Коэффициент использования светового потока, показывающий какая часть светового потока ламп попадает на рабочую поверхность, для светильников типа ОДОР с люминесцентными лампами при $\rho_n = 70\%$, $\rho_c = 50\%$ и индексе помещения $i = 1,5$ равен $\eta = 0,47$.

Нормируемая минимальная освещенность при использовании ЭВМ и одновременной работе с документами должна быть равна 600лк.

$$\Phi = \frac{E_n \cdot S \cdot K_z \cdot Z}{N \cdot \eta} = \frac{600 \cdot 24 \cdot 1,5 \cdot 1,1}{24 \cdot 0,47} = 2106 \text{ Лм}$$

Для люминесцентных ламп с мощностью 40Вт и напряжением сети 220В, стандартный световой поток ЛД равен 2300 Лм.

$$-10\% \leq \frac{\Phi_{\text{ЛД}} - \Phi_{\text{л.расч}}}{\Phi_{\text{ЛД}}} \cdot 100\% \leq 20\%$$

$$\frac{2300 - 2106}{2300} \cdot 100\% = 8,43\%$$

$$-10\% \leq 8,43\% \leq 20\%$$

Таким образом необходимый световой поток светильника не выходит за пределы требуемого диапазона.

4.5. Электромагнитные поля

Известно, что ПЭВМ являются источником электромагнитного поля (ЭМП) радиочастотного диапазона. При длительном постоянном воздействии, которого на организм человека наблюдаются нарушения сердечно-сосудистой, дыхательной и нервной систем, появляется утомляемость, ухудшение самочувствия, гипотония, также характерна головная боль, изменение проводимости сердечной мышцы. Тепловое воздействие ЭМП характеризуется повышением температуры тела, локальным избирательным нагревом тканей, органов, клеток вследствие перехода ЭМП в тепловую энергию.

При работе с компьютером допустимые уровни электромагнитных полей (ЭМП) нормируются СанПиН 2.2.2/2.4.1340-03:

1. Напряженность электромагнитного поля на расстоянии 50 см вокруг ПЭВМ по электрической составляющей должна быть не более:

- В диапазоне частот 5Гц-2кГц - 25В/м;
- В диапазоне частот 2кГц/400кГц - 2,5В/м.

2. Плотность магнитного потока должна быть не более:

- В диапазоне частот 5Гц-2кГц - 250нТл;
- В диапазоне частот 2кГц/400кГц - 25нТл.

3. Напряжённость электростатического поля должна быть 15 кВ/м;

4. Электростатический потенциал экрана видеомонитора – 500 В.

Среди средств защиты от ЭМП выделяют следующие [16]:

1. Организационные мероприятия – выбор рациональных режимов работы оборудования, ограничение места и времени нахождения персонала в зоне воздействия ЭМП, т.е. защита расстоянием и временем;

2. Инженерно-технические мероприятия, включающие рациональное размещение оборудования, использование средств, ограничивающих

поступление электромагнитной энергии (поглотители мощности, экранирование и т.п.);

3. Лечебно-профилактические мероприятия в целях предупреждения, ранней диагностики и лечения здоровья персонала;

4. Средства индивидуальной защиты, к которым относятся защитные очки, щитки, шлемы, защитная одежда и др.

В данном случае воздействие ЭМП происходит только от монитора компьютера. Исходя из паспортных данных компьютера и монитора, они соответствуют нормам ТСО-99, ТСО-03.

Ионизирующее излучение вызывает в организме цепочку обратимых и необратимых последствий. Ионизирующая радиация при воздействии на организм человека может вызвать два вида эффектов: детерминированные пороговые эффекты (лучевая болезнь, лучевой ожог, лучевая катаракта, лучевое бесплодие и др.) и стохастические (вероятностные) беспороговые эффекты (злокачественные опухоли, лейкозы, наследственные болезни).

Оценка уровней ионизирующих излучений проводится при работе компьютерами, оснащенными мониторами с электроннолучевой трубкой. В данном случае работа велась за компьютером, снабженным монитором с жидкокристаллическим дисплеем, поэтому оценка параметров по данному пункту раздела не проводилась.

4.6. Шум

Производственным шумом называется неблагоприятно воспринимаемый звук на рабочих местах, на участках или на территориях предприятий, который возникает во время производственного процесса.

Средства индивидуальной защиты органов слуха работающих установлены ГОСТ 12.4.011-75; это противозумные шлемофоны (шлемы), наушники, заглушки, вкладыши. Они эффективно защищают организм от раздражающего действия шума, предупреждая возникновение различных

функциональных нарушений и расстройств, если правильно подобраны и систематически используются.

Предельно допустимый уровень (ПДУ) шума - это уровень 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 отклонений в состоянии здоровья, обнаруживаемых современными методами исследований в процессе работы или в отдаленные сроки жизни настоящего и последующих поколений. Соблюдение ПДУ шума не исключает нарушения здоровья у сверхчувствительных лиц.

Допустимый уровень шума ограничен ГОСТ 12.1.003-83 и СанПиН 2.2.4/2.1.8.10-32-2002. Максимальный уровень звука постоянного шума на рабочих местах не должно превышать 80 дБА. Уровень шума на рабочем месте математиков-программистов и операторов видеоматериалов не должен превышать 50дБА, а в залах обработки информации на вычислительных машинах - 65дБА.

Методы и средства коллективной защиты включают в себя: изменение направленности излучения шума; рациональную планировку предприятий и производственных помещений; акустическую обработку помещений; применение звукоизоляции; ремонт оборудования; замена шумящих деталей на бесшумные аналоги.

Если на рабочем месте превышает допустимый уровень шума, то оптимальным решением для борьбы с этой проблемой будет вынос системного блока компьютера в отдельное звукоизолированное помещение. Для изоляции данного помещения можно использовать звукоизолирующие панели или звукоизолирующие мембраны.

4.7. Организация рабочего места оператора ПЭВМ

Рабочее место – это часть рабочей зоны, где постоянно или временно пребывает работник в процессе трудовой деятельности. Оно должно, прежде всего, обеспечивать возможность удобного выполнения работ, учитывая её

физическую тяжесть и технологические особенности, а также включать в себя пространство, необходимое для передвижения в ней работающего.

Невыполнение этих требований может привести к получению работником производственной травмы или развития у него профессионального заболевания. Рабочее место при выполнении работ в положении сидя должно соответствовать требованиям ГОСТ 12.2.032-78.

В соответствии с СанПиНом 2.2.2/2.4.1340-03, где даны общие требования к организации и оборудованию рабочих мест с ПЭВМ, при работе инженера за столом, конструкция стола и стула должна обеспечивать оптимальное положение тела работающего. Параметры рабочего места при работе с ПЭВМ, а также с нормативной и технической документацией приведены в таблице 25.

Таблица 25 – Параметры рабочего места при работе

Параметры	Значение параметра	Реальные значения
Высота рабочей поверхности стола	От 600 до 800, мм	770 мм
Высота клавиатуры	600-700, мм	630 мм
Удаленность клавиатуры	Не менее 80, мм	85 мм
Удаленность экрана монитора	500-700, мм	650 мм
Высота сидения	400-500, мм	470 мм
Угол наклона монитора	0-30, град.	10 мм
Наклон подставки ног	0-20, град.	0 мм

Параметры рабочего стола удовлетворяют нормативным требованиям. Кроме того, не рекомендуется располагать компьютеры вблизи друг от друга в целях уменьшения действия переменного электрического поля. Рабочие места с компьютерами должны размещаться так, чтобы расстояние от экрана одного монитора до тыла другого было не менее 2,0 м, а расстояние между боковыми поверхностями мониторов - не менее 1,2 м.

Нормативные параметры для мониторов при работе с ПЭВМ указаны в таблице 26.

Таблица 26 – Допустимые визуальные параметры устройств отображения информации

Параметры	Допустимые значения
Яркость белого поля	Не менее 35 кд/м ²
Неравномерность яркости рабочего поля	Не более $\pm 20\%$
Контрастность (для монохромного режима)	Не менее 3:1
Временная нестабильность изображения (непреднамеренное изменение во времени яркости изображения на экране дисплея)	Не должна фиксироваться

При оформлении помещения большое значение имеет цветовое решение. Психофизическое воздействие цвета - первый и наиболее важный фактор, учитываемый при выборе цветового решения. Учитывая характер работ, следует выбирать неяркие, малоконтрастные оттенки, которые не рассеивали бы внимание в рабочей зоне. Так как работа требует спокойствия и сосредоточенности, предпочтительно использовать оттенки «холодных» цветов. В данном случае рабочее помещение окрашено в серо-белые тона, что является оптимальным и нейтральным цветом для работы.

Продолжительность непрерывной работы с компьютером без регламентированного перерыва не должна превышать 2-х часов. Это связано с тем, что на человека при работе с компьютером оказывают влияние опасные и вредные производственные факторы, а также наступает общее утомление, что негативно сказывается на здоровье и самочувствии человека.

Санитарные правила и нормы содержат ряд комплексов упражнений, которые способствует снятию локального утомления. По содержанию они различны и предназначаются для конкретного воздействия на ту или иную группу мышц или систему организма в зависимости от самочувствия и ощущения усталости.

4.8. Анализ опасных производственных факторов

Основные опасные факторы, которые могут возникнуть при работе с компьютером, представлены в таблице 27.

Таблица 27 – Основные элементы производственного процесса, формирующие опасные факторы

Наименование видов работ и параметров производственного процесса	Факторы (ГОСТ 12.0.003-74 ССБТ)	Нормативные документы
Оператор ПК	Электрический ток	ГОСТ 12.1.002–84

4.9. Электробезопасность

Электробезопасность представляет собой систему организационных и технических мероприятий и средств, обеспечивающих защиту людей от вредного и опасного воздействия электрического тока, электрической дуги, электромагнитного поля и статистического электричества.

Электроустановки классифицируют по напряжению: с номинальным напряжением до 1000 В (помещения без повышенной опасности), до 1000 В с присутствием агрессивной среды (помещения с повышенной опасностью) и свыше 1000 В (помещения особо опасные).

В отношении опасности поражения людей электрическим током различают:

1. Помещения без повышенной опасности, в которых отсутствуют условия, создающие повышенную или особую опасность.
2. Помещения с повышенной опасностью, которые характеризуются наличием в них одного из следующих условий, создающих повышенную опасность: сырость, токопроводящая пыль, токопроводящие полы

(металлические, земляные, железобетонные, кирпичные и т.п.), высокая температура, возможность одновременного прикосновения человека к имеющим соединение с землей металлоконструкциям, технологическим аппаратам, с одной стороны, и к металлическим корпусам электрооборудования - с другой.

3. Особо опасные помещения, которые характеризуются наличием оборудования свыше 1000 В и одного из следующих условий, создающих особую опасность: особой сырости, химически активной или органической среды, одновременно двух или более условий повышенной опасности. Территории размещения наружных электроустановок в отношении опасности поражения людей электрическим током приравниваются к особо опасным помещениям.

Помещение, где была разработана магистерская работа, принадлежит к категории помещений без повышенной опасности по степени вероятности поражения электрическим током, вследствие этого к оборудованию предъявляются следующие требования:

- экран монитора должен находиться на расстоянии не менее 50 см от пользователя (расстояния от источника);
- применение приэкранных фильтров, специальных экранов.

Защитное заземление — это преднамеренное электрическое соединение с землей или ее эквивалентом металлических нетоковедущих частей, которые могут оказаться под напряжением.

Сопротивление заземления — основной показатель заземляющего устройства, определяющий его способность выполнять свои функции и определяющий его качество в целом.

Сопротивление заземления зависит от площади электрического контакта заземлителя (заземляющих электродов) с грунтом (“стекание” тока) и удельного электрического сопротивления грунта, в котором смонтирован

этот заземлитель (“впитывание” тока). Согласно ПЭУ номинальное сопротивление заземления должно быть не более 4 Ом.

К основным электрозащитным средствам в электроустановках напряжением до 1000 В относятся:

- изолирующие штанги;
- изолирующие и электроизмерительные клещи;
- диэлектрические перчатки; изолированный инструмент.

Работать со штангой разрешается только специально обученному персоналу в присутствии лица, контролирующего действия работающего. При операциях с изолирующей штангой необходимо пользоваться дополнительными изолирующими защитными средствами — диэлектрическими перчатками и изолирующими основаниями (подставками, ковриками) или диэлектрическими ботами.

Изолирующие клещи применяют в электроустановках до 35 кВ для операций под напряжением с плавкими вставками трубчатых предохранителей, а также для надевания и снятия изолирующих колпаков на ножи однополюсных разъединителей. Изолирующие клещи выполняют из пластмассы.

При пользовании изолирующими клещами оператор должен надевать диэлектрические перчатки и быть изолированным от пола или грунта; при смене патронов трубчатых предохранителей он должен быть в очках. Клещи нужно держать в вытянутых руках.

К дополнительным изолирующим электрозащитным средствам относятся диэлектрические перчатки, боты, резиновые коврики и дорожки, изолирующие подставки на фарфоровых изоляторах и переносные заземления.

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.

Помещения, в которых сила тока не превышает 0.1а, а напряжение меньше 36В, считаются электробезопасными, так как данные величины можно считать безопасными для организма человека. Значение этих показателей могут меняться в зависимости от окружающих условий.

4.10. Защита в чрезвычайных ситуациях

В Томске преобладает континентально-циклонический климат. Природные явления (землетрясения, наводнения, засухи, ураганы и т. д.) отсутствуют.

Возможными ЧС могут быть сильные морозы, диверсия и пожар.

Для Сибири в зимнее время года характерны морозы. Достижение критически низких температур приведет к авариям систем теплоснабжения и жизнеобеспечения, приостановке работы, обморожениям и даже жертвам среди населения. В случае переморозки труб должны быть предусмотрены запасные обогреватели. Их количества и мощности должно хватать для того, чтобы работа на производстве не прекратилась.

Чтобы избежать потерь при авариях на электросетях у каждого компьютера должен быть блок бесперебойного питания, а также должен быть аварийный генератор, который обеспечит дальнейшую бесперебойную работу персонала.

Чрезвычайные ситуации, возникающие в результате диверсий, возникают все чаще.

Зачастую такие угрозы оказываются ложными. Но случаются взрывы и в действительности.

Для предупреждения вероятности осуществления диверсии предприятие необходимо оборудовать системой видеонаблюдения, круглосуточной охраной, пропускной системой, надежной системой связи, а также исключения распространения информации о системе охраны объекта, расположении

помещений и оборудования в помещениях, системах охраны, сигнализаторах, их местах установки и количестве. Должностные лица раз в полгода проводят тренировки по отработке действий на случай экстренной эвакуации.

4.11. Пожарная безопасность

Для обеспечения безопасности людей и сохранения материальных ценностей существует пожарная безопасность, основными системами которой являются системы предотвращения пожара и противопожарной защиты, включая организационно-технические мероприятия.

Пожар представляет большую опасность и наносит огромный ущерб, поскольку грозит уничтожением приборов, компьютеров, инструментов и комплектов документов, представляющих значительную ценность. Кроме того, пожар характеризуется опасностью для жизни человека. Возникновение пожара в комнате может быть обусловлено следующими факторами: в современных ПЭВМ очень высокая плотность размещения электронных схем. При протекании по ним электрического тока выделяется значительное количество тепла, что может привести к повышению температуры отдельных узлов до 100°C. При этом возможно оплавление изоляции соединительных проводов, их оголение, и, как следствие, короткое замыкание, сопровождаемое искрением [16].

Поэтому во избежание пожаров проводится пожарная профилактика – комплекс организационных и технических мероприятий, направленных на обеспечение безопасности людей, на предотвращение пожара, ограничение его распространения, а также на создание условий для успешного тушения пожара. Успех борьбы с пожаром во многом зависит от его своевременного обнаружения и быстрого принятия мер по его ограничению и ликвидации. Основы противопожарной защиты предприятий определены в стандартах ГОСТ 12.1.004-76 и ГОСТ 12.1.010-76.

Огнетушители являются самым эффективным средством первичного пожаротушения. Огнетушитель — переносное или передвижное устройство

для тушения очагов пожара за счет выпуска запасенного огнетушащего вещества.

Наиболее распространены ручные огнетушители:

- химические пенные ОХП-10 – заполнены составом щелочей и кислот;
- воздушно-пенные ОВП-5 и ОВП-10 – заполнены 6-ти% водородным раствором пенообразователя (нельзя использовать для тушения электроустановок);
- углекислотные ОУ-2, ОУ-5, ОУ-8 – заполнены сжиженным диоксидом углерода;
- углекислотные бромэтиловые ОУБ-3, ОУБ-7;
- порошковые ОПС-10 – применяются для предотвращения горения щелочных металлов и электрооборудования.

Помещение, в котором выполнялась работа, относится к категории «Д», с пониженной пожароопасностью.

4.12. Экологическая безопасность

В современном мире работа почти любого производства (промышленного, сельскохозяйственного и т. п.) сопровождается образованием отходов. Они поступают в окружающую среду в виде выбросов в атмосферу, сбросов в водоемы, твердых промышленных и бытовых отходов и мусора на поверхность и в недра Земли. Поэтому возникает такая проблема, как охрана окружающей среды. Наиболее активная форма её решения – это сокращение вредных выбросов промышленных предприятий через полный переход к безотходным или малоотходным технологиям производства. Охрану природы можно представить, как комплекс государственных, международных и общественных мероприятий, направленных на рациональное использование природы, восстановление, улучшение и охрану природных ресурсов.

В ходе данной работы были использованы следующие ресурсы:

- электроэнергия для работы компьютера;

- бумага;
- люминесцентные лампы.

С точки зрения потребления ресурсов компьютер потребляет сравнительно небольшое количество электроэнергии, что положительным образом сказывается на общей экономии потребления электроэнергии в целом.

При написании ВКР вредных выбросов в атмосферу, почву и водные источники не производилось, радиационного заражения не произошло, чрезвычайные ситуации не наблюдались, поэтому не оказывались существенные воздействия на окружающую среду, и никакого ущерба окружающей среде не было нанесено.

В связи с тем, что огромная масса информации содержится на бумажных носителях, уничтожение бумаги играет очень важную роль. Среди основных методов уничтожения, которые применяются на сегодняшний день для бумажных документов, следует отметить следующие:

- Сжигание документов.
- Шредирование.
- Закапывание.
- Химическая обработка.

Переработка оргтехники включает в себя несколько этапов:

Первый этап - удаление всех опасных компонентов.

Второй этап - удаление всех крупных пластиковых частей. В большинстве случаев эта операция также осуществляется вручную. Оставшиеся после разборки части отправляют в большой измельчитель, и все дальнейшие операции автоматизированы.

Третий этап - измельченные в гранулы остатки компьютеров подвергаются сортировке. Сначала с помощью магнитов извлекаются все железные части. Затем приступают к выделению цветных металлов, которых в ПК значительно больше.

Перегоревшие люминесцентные лампы можно отнести в свой районный ДЕЗ или РЭУ, где установлены специальные контейнеры. Там их должны бесплатно принять.

4.13. Научно техническая документация

1. ГОСТ 54 30013-83 Электромагнитные излучения СВЧ. Предельно допустимые уровни облучения. Требования безопасности
2. ГОСТ 12.4.154-85 "ССБТ. Устройства, экранирующие для защиты от электрических полей промышленной частоты"
3. ГН 2.2.5.1313-03 Предельно допустимые концентрации (ПДК) вредных веществ в воздухе рабочей зоны
4. СанПиН 2.2.4/2.1.8.055-96 "Электромагнитные излучения радиочастотного диапазона (ЭМИ РЧ)".
5. СанПиН 2.2.4.548-96. Гигиенические требования к микроклимату производственных помещений.
6. СанПиН 2.2.1/2.1.1.1278-03. Гигиенические требования к естественному, искусственному и совмещенному освещению жилых и общественных зданий.
7. СН 2.2.4/2.1.8.562-96. Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки.
8. ГОСТ 12.4.123-83. Средства коллективной защиты от инфракрасных излучений. Общие технические требования.
9. ГОСТ Р 12.1.019-2009. Электробезопасность. Общие требования и номенклатура видов защиты.
10. ГОСТ 12.1.030-81. Электробезопасность. Защитное заземление. Зануление.
11. ГОСТ 12.1.004-91. Пожарная безопасность. Общие требования. ГОСТ 12.2.037-78. Техника пожарная. Требования безопасности
12. СанПиН 2.1.6.1032-01. Гигиенические требования к качеству атмосферного воздуха
13. ГОСТ 30775-2001 Ресурсосбережение. Обращение с отходами. Классификация, идентификация и кодирование отходов.
14. СНиП 21-01-97. Противопожарные нормы.
15. ГОСТ 12.4.154. Система стандартов безопасности труда. Устройства, экранирующие для защиты от электрических полей промышленной частоты. Общие технические требования, основные параметры и размеры
16. СНиП 23-05-95 "Естественное и искусственное освещение"

4.14. Выводы и рекомендации

Проанализировав условия труда на рабочем месте, где была разработана магистерская работа, можно сделать вывод, что помещение удовлетворяет необходимым нормам и в случае соблюдения техники безопасности и правил пользования компьютером работа в данном помещении не приведет к ухудшению здоровья работника.

Само помещение и рабочее место в нем удовлетворяет всем нормативным требованиям. Кроме того, действие вредных и опасных факторов сведено к минимуму, т.е. микроклимат, освещение и электробезопасность соответствуют требованиям, предъявленным в соответствующих нормативных документах.

Относительно рассмотренного вопроса об экологической безопасности можно сказать, что деятельность помещения не представляет опасности окружающей среде.

Важно добавить, что монитор компьютера служит источником ЭПМ – 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.

Заключение: Разработан интерфейс для работы со Scopus API. Получена информация о 78354 статьях из области “Экономика, Эконометрика и Финансы”, информация о 22712 авторах и их принадлежности к университетам или научным организациям. Произведен первичный анализ полученных данных. Все авторы были разбиты на кластеры, которые основываются на взаимосвязи авторов через совместное написание работ. Выявлена тенденция к неоднократному сотрудничеству авторов, занимающихся одной или смежными областями науки. Интерфейс реализован на языке PHP, данные были выгружены в реляционную базу данных PostgreSQL, анализ произведён при помощи языка R. Максимальное количество соавторов для одной работы из статей, попавших в выборку, равняется 39.

Список литературы

- 1) T.A. Brooks , Private acts and public objects: an investigation of citer motivations, Journal of the American Society for Information Science 36(4) (1985) 223-229.
- 2) Bárbara S. Lancho-Barrantes, Vicente P. Guerrero-Bote, Félix Moya-Aneón. What lies behind the averages and significance of citation indicators in different disciplines? Article first published online: April 13, 2010. Volume: 36 issue: 3, page(s): 371-382.
- 3) Cristopher J. Date, Introduction to Database Systems, 2003.
- 4) Иванова Г.С. Основы программирования: Учебник для вузов. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2002. - 416 с.
- 5) Welling L., Thomson L. PHP and MySQL Web Development (4th Edition). Addison-Wesley Professional, 2008 – 1008 p.
- 6) Noorden, R. V., 5 2014. Global scientific output doubles every nine years. Mathematics & Economics Nature News Blog, <http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html>.
- 7) Asadi, S., Dahlan, H. M., et al., 2017. Organizational research in the field of green it: A systematic literature review from 2007 to 2016. Telematics and Informatics.
- 8) Elsevier, 2017. Scopus database API interface. <https://dev.elsevier.com/> .
- 9) Uddin S., Hossain L., Rasmussen K. 2013. Network effect on scientific collaborations. PloS One 8 (2), e57546
- 10) Schlattmann S., 2017. Capturing the collaboration intensity of research institutions using social network analysis. Procedia Computer Science 106, 25–31.
- 11) Baghdadi, H. S., Ranaivo-Malan, B., 2011. An automatic topic identification algorithm. Journal of Computer Science 7 (9), 1363.

12) Can˜as-Guerrero, I., Mazarron, F., Pou-Merina, A., Calleja-Perucho, C., Diaz Rubio, G., 2013. Bibliometric analysis of research activity in the “agronomy” category from the web of science, 1997-2011. *European Journal of Agronomy* 50 19–28.

13) Szymanski, B., De La Rosa, J., Krishnamoorthy, M., 2012. An internet measure of the value of citations. *Information Sciences* 185 (1), 18–31.

Приложение А. Пример данных о статье (XML формат)

```
<entry>
  <prism:url>
    http://api.elsevier.com/content/abstract/scopus_id/84888033533
  </prism:url>
  <dc:identifier>SCOPUS_ID:84888033533</dc:identifier>
  <dc:title>
    Characteristics and oxidative stability of soy protein-stabilized oil-in-water emulsions: Influence
    of ionic strength and heat pretreatment
  </dc:title>
  <dc:creator>Shao, Y.</dc:creator>
  <prism:publicationName>Food Hydrocolloids</prism:publicationName>
  <prism:issn>0268005X</prism:issn>
  <prism:volume>37</prism:volume>
  <prism:pageRange>149-158</prism:pageRange>
  <prism:coverDate>2014-06-01</prism:coverDate>
  <prism:coverDisplayDate>June 2014</prism:coverDisplayDate>
  <prism:doi>10.1016/j.foodhyd.2013.10.030</prism:doi>
  <dc:description>
    Some physicochemical characteristics, microstructure and stability of native and preheated (95°C,
    15min) soy protein isolate (SPI)-stabilized emulsions, formed at varying protein concentrations (c;
    0.5-4.0%, w/v) in the absence or presence of 300mM NaCl, were characterized.
  </dc:description>
  <citedby-count>0</citedby-count>
  <affiliation>
    <affiliation-url>
      http://api.elsevier.com/content/affiliation/affiliation_id/60024542
    </affiliation-url>
    <afid>60024542</afid>
    <affilname>South China University of Technology</affilname>
    <affiliation-city>Guangzhou</affiliation-city>
    <affiliation-country>China</affiliation-country>
    <name-variant>South China University of Technology</name-variant>
    <name-variant>South China Univ. of Technol.</name-variant>
  </affiliation>
  <prism:aggregationType>Journal</prism:aggregationType>
  <subtype>ar</subtype>
  <subtypeDescription>Article</subtypeDescription>
  <author>
    <author-url>
      http://api.elsevier.com/content/author/author_id/55673455300
    </author-url>
    <authid>55673455300</authid>
    <authname>Shao, Y.</authname>
    <given-name>Yun</given-name>
    <surname>Shao</surname>
    <initials>Y.</initials>
    <afid>60024542</afid>
  </author>
  <authkeywords>
    Emulsifying property | Emulsion stability | Oxidative stability | Protein
  </authkeywords>
</entry>
```

Приложение Б. Метод класса отправки запросов

```
function get_json_request_data($get_request) {  
    $curl = curl_init();  
    curl_setopt($curl, CURLOPT_URL, $get_request);  
    curl_setopt($curl, CURLOPT_RETURNTRANSFER, true);  
    $out = curl_exec($curl);  
    curl_close($curl);  
    return json_decode($out,true);  
}
```

Приложение В. Метод получения данных по номеру журнала

```
include("class_db.php");
include("functions.php");
header ("Content-Type: text/html; charset=utf-8");
$params = [];
$sql = "SELECT issn FROM articles.sources WHERE status_id = 1";
$records = get_allrecords($dbconn, $sql, $params);
foreach ($records as $issn_id) {
    $issn = $issn_id['issn'];
    $curl = curl_init();
    $query =
"https://api.elsevier.com/content/search/scopus?start=0&query='issn={$issn}'&
count=1&subj=ECON&apiKey=c4bc231feb39d5388147831c6ab0b7dd&xml-
decode=true&field=title";
    curl_setopt($curl, CURLOPT_URL, $query);
    curl_setopt($curl, CURLOPT_RETURNTRANSFER, true);
    $out = curl_exec($curl);
    curl_close($curl);
    $output = json_decode($out, true);
    $articles_count = $output['search-results']['opensearch:totalResults'];
    if ($articles_count != 0) {
        $query_1 =
"https://api.elsevier.com/content/search/scopus?start=0&query=issn={$issn}&co
unt=200&subj=ECON&apiKey=c4bc231feb39d5388147831c6ab0b7dd&xml-
decode=true&field=title,citedby-count,publicationName,affilname,affiliation-
city,affiliation-country,afid,authid,authname,authkeywords,dc:creator";
        $curl = curl_init();
        curl_setopt($curl, CURLOPT_URL, $query_1);
        curl_setopt($curl, CURLOPT_RETURNTRANSFER, true);
        $out = curl_exec($curl);
        curl_close($curl);
        $output = json_decode($out, true);
        foreach ($output['search-results']['entry'] as $entry) {
            $affiliations = [];
            $authors = [];
            $url = $entry['prism:url'];
            $h_count = $entry['citedby-count'];
            $publicationName = $entry['prism:publicationName'];
            $title = $entry['dc:title'];
            $coverDate = $entry['prism:coverDate'];
            $main_author = $entry['dc:creator'];
            $main_author_id = 0;
            if (!empty($entry['authkeywords'])) {
                $keywords_array = explode("|", $entry['authkeywords']);
                $author_keywords_json = json_encode($keywords_array);
            } else {
                $author_keywords_json = json_encode([]);
            }
            if (!empty($entry['affiliation'])) {
                foreach ($entry['affiliation'] as $affil) {
```

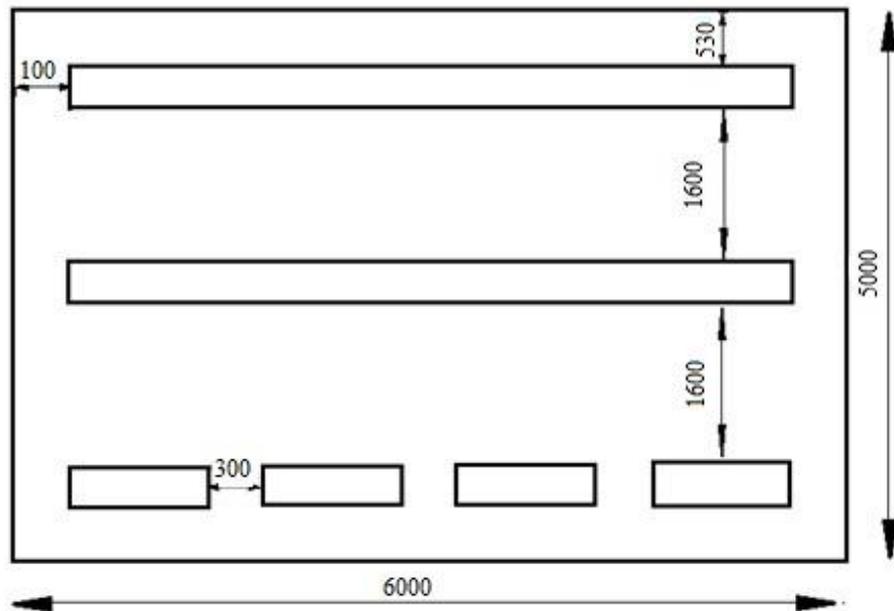
```

        $id = $affil['afid'];
        $name = $affil['affilname'];
        $country = $affil['affiliation-country'];
        $city = $affil['affiliation-city'];
        if (!empty($id)) {
            pg_query_params($dbconn, "INSERT INTO
articles.affiliations(id,name,country,city) VALUES($1,$2,$3,$4) ON CONFLICT DO
NOTHING", array($id, $name, $country, $city));
            $affiliations[] = $id;
        }
    }
}
if (!empty($entry['author'])) {
    foreach ($entry['author'] as $author) {
        $author_af = [];
        $name = $author['authname'];
        $id = $author['authid'];
        if ($name == $main_author) {
            $main_author_id = $id;
        }
        if (!empty($author['afid'])) {
            foreach ($author['afid'] as $author_afid) {
                $author_af[] = $author_afid['$'];
            }
        }
        $author_afid_json = json_encode($author_af);
        pg_query_params($dbconn, "INSERT INTO
articles.authors(id,name,affiliations) VALUES($1,$2,$3) ON CONFLICT DO
NOTHING", array($id, $name, $author_afid_json));
        if (!empty($author['authid']) && $name!=$main_author)
    {
        $authors[] = $author['authid'];
    }
}
}
$affiliations_json = json_encode($affiliations);
$authors_json = json_encode($authors);
pg_query_params($dbconn, "INSERT INTO
articles.articles(issn,author_id,subauthors,
author_key_words,affiliations,title, cetings_count) VALUES($1,$2,$3,$4,$5,$6,
$7) ON CONFLICT DO NOTHING", array($issn, $main_author_id, $authors_json,
$author_keywords_json, $affiliations_json, $title, $h_count));
}
pg_query_params("UPDATE articles.sources SET status_id = 2 WHERE
issn = $1", [$issn]);

```

Приложение Г (обязательное)

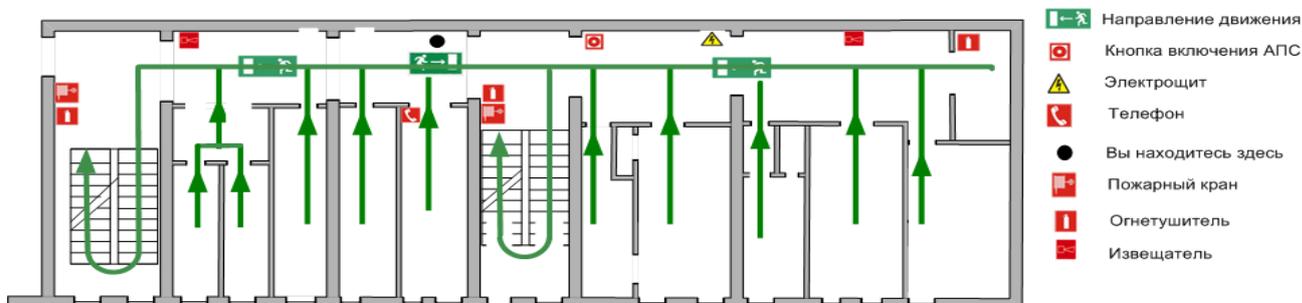
План помещения и размещения светильников с люминесцентными лампами



Приложение Д (обязательное)

План эвакуации в случае пожара

ПЛАН ЭВАКУАЦИИ 2-го этажа



- Направление движения
- Кнопка включения АПС
- Электрощит
- Телефон
- Вы находитесь здесь
- Пожарный кран
- Огнетушитель
- Извещатель

Действия при пожаре Сохранять спокойствие

1	Сообщить по телефону		<ul style="list-style-type: none"> • Адрес объекта • Место возникновения пожара • Свою фамилию
2	Эвакуировать людей		<ul style="list-style-type: none"> • Ориентироваться по знакам направления движения • Взять с собой пострадавших
3	По возможности принять меры по тушению пожара		<ul style="list-style-type: none"> • Использовать средства противопожарной защиты • При необходимости обесточить помещение

Ответственный за эвакуацию и включение системы оповещения

Приложение Е

Магистерской диссертации на английском языке

Formation of scientific groups in the subject field "Economics, Econometrics and Finance" after analyzing the data extracted from the science-based Scopus database

Студент:

Группа	ФИО	Подпись	Дата
0ВМ61	Неретин Артём Александрович		

Консультант школы отделения экспериментальной физики ИШЯТ:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Крицкий Олег Леонидович	кандидат ф-м наук		

Консультант – лингвист отделения иностранных языков ШБИП:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Смирнова Ульяна Александровна			

Introduction

Now, as never before, the scientific community has the material and human resources to carry out relevant research activities in multiple disciplines and research fields. According to some studies, world scientific production doubles every nine years [1]. It is now possible to access most of the scientific literature through the Internet [2], which allows transferring knowledge effectively. However, this process also leads to increasing competition between researchers and higher quality standards that increase the demands on research productivity. These changes encourage researchers, especially those in the early stages of their careers, to establish collaborative relationships with other researchers from their own or another institution [3]. The dynamism of these collaborations makes the analysis of the scientific collaboration network a topic of great interest, both for researchers and for funding agencies.

However, evaluating scientific quality is a notoriously difficult problem that has no standard solution. In fact, the evaluation criteria used to assign these budgets are usually based on scientometric indicators retrieved through web search engines such as Scopus, Google Scholar, etc., which include the number and quality of publications by researchers based on journal and author metrics. Nevertheless, these evaluation criteria do not explicitly take into account the relationships between researchers, even though the globalization of science in the modern world due to collaborations between researchers of the same or different institutions has become an important issue to analyse. To respond to this demand, this paper presents a method to automatically determine the structures of collaborations between researchers from different institutions and disciplines and to provide a fast and reliable method for determining the most important specialty areas of a given institution. Collaboration in publications is used as a criterion because, according to different studies, it is more important for knowledge transfer than patents [4]. This method is based on programming scripts to automatically obtain bibliographic data from scientific publications through the use of the Scopus Database API Interface

(Elsevier) according to different criteria (affiliation, number of citations, department, etc.). This information is then adapted for processing using parsing techniques and the refinement of structured text. Subsequently, the collaboration network is represented using specialized graph visualization software [5] and analysed using statistical metrics.

Scientific databases

Currently, most researchers have access to high-quality multidisciplinary scientific databases. Some of them are open access databases (e.g., Google Scholar), while others are accessed thanks to the subscriptions made by public or private organizations for which these researchers work (e.g., Web of Science, Scopus [6]). Likewise, there are databases on specific fields or areas, such as PubMed, which includes citations for biomedical literature from MEDLINE, life science journals and online books. Different investigations have analysed the characteristics of scientific databases. The main sources of information, such as scientific databases (e.g: Scopus), search engines (e.g: Google Scholar) and social networks (e.g: academia.edu or ResearchGate), are analysed by Asadi [2]. An extensive discussion has been devoted to the advantages and disadvantages of these sources of information. For example, an important advantage of Google Scholar is that the information is updated periodically taking advantage of existing information on the Internet. However, Google Scholar only includes profiles of researchers who have voluntarily discharged themselves, with numerous errors in linking documents to authors, so that many profiles of researchers have publications that do not really correspond to that author. This inconvenience is partially solved by using unique and global identifiers such as the Open Researcher and Contributor ID (ORCID), which assigns a unique identifier to each author, although such information must be included or linked manually. These deficiencies are limited in literature databases whose data sources are editorial, professional associations, etc. [2]. This is the case for Web of Science and Scopus, whose information is considered more reliable,

although it includes a lower number of documents than other systems, such as Google Scholar [7]. To verify the reliability and accuracy of these databases, different studies have analysed the typical errors in the information contained therein [7]. For example, among the most common errors of Scopus is the existence of several authorIDs for the same individual due to their presence in different research institutions or to the fact that they used different signature formats in different publications. Thus, the reliability of any bibliometric study based on these data is limited. In the present study, these inaccuracies have been verified and have been minimized as much as possible thanks to the use of text refinement and data reconciliation tools. It is to be expected that the reliability of the information contained in Scopus will increase in the future, and as a result, the accuracy of the results compiled by the method presented here will be improved. All these databases are often oriented to allow a search for individual researchers and documents on a specific research topic. However, its use for the extraction and analysis of aggregate data has not been of interest until recent years. Fortunately, some of these databases have recently incorporated tools that allow to perform search and to collect data to develop studies of a different nature. This is the case of the Scopus Database API Interface, which is available to the public and allows obtaining raw data from the Scopus database based on different criteria. More specifically, documents in Scopus are classified under four broad subject areas: Economics, Econometrics and Finance, Life Sciences, Physical Sciences, Health Sciences, and Social Sciences & Humanities. These general subjects are further divided into 27 major thematic categories as well as into more than 300 specific subject categories, although a given document can be simultaneously included in different categories. For analyze we choose only one subject area: Economics, Econometrics and Finance.

A framework to analyse research network collaborations

This section presents a method for analysing a large volume of data from any academic or research institution in order to determine different bibliometric indicators of individual or group researchers, as well as the collaboration networks of these researchers in terms of publications. This method, which can be of great interest for researchers, research institutions, and funding agencies, takes advantage of the features provided by the Scopus Database API Interface to automate the search of manuscripts published by authors and institutions, such that information is treated and analysed in a later stage for different specific purposes. Elsevier's Scopus is also the largest database of peer-reviewed literature in different scientific fields.

We created the automatic information extraction script of the Scopus Database. The operation of this script, which allows the collection of all the data required for different analyses, can be divided into three subsequent phases:

1. **Get articles data:** In this phase, the script retrieves all of the information from the publications of authors who have ever published a manuscript using a given affiliation identifier (afid). To perform this task, the information in each of the papers is extracted, and the "authorid" (unique identifier of an author in Scopus) will be stored in the database of the script if it has not previously been registered.
2. **Get authors data:** From the list of researchers whose affiliation coincides with the desired one, script retrieves and stores in its database the information that Scopus has for each one of these authors, including the registered publications, dates of these publications, history of affiliations, h-index, number of citations, etc.
3. **Get the collaboration networks:** The individual information about each researcher is used to establish collaborative relationships based on the co-authoring of papers. More specifically, in this phase, the script applies an iterative process for each of the authors of the institution, then obtains

information such as the name of the institution, city, country, number of co-authors, current affiliation of these collaborators, etc.

Tag Cloud

This is a visual representation of a list of keywords (or tags, also called labels, labels, etc.). Typically used to represent unformatted text. Keywords are often single words or small phrases, and the importance of each keyword is indicated by the font size or color. This view is convenient for quickly perceiving frequently used terms and for distributing keywords by popularity relative to each other.

To build tag clouds, we will use clustering, in particular using the k-means method. To display the tag cloud, you need to break all keywords into groups (clusters) by the frequency of their appearance and change the word size for each group separately. After collection of the required data we did a primary analysis of the author key words, that's give us next Tag Cloud (Image 2).

The construction algorithm is shown in Image 1.

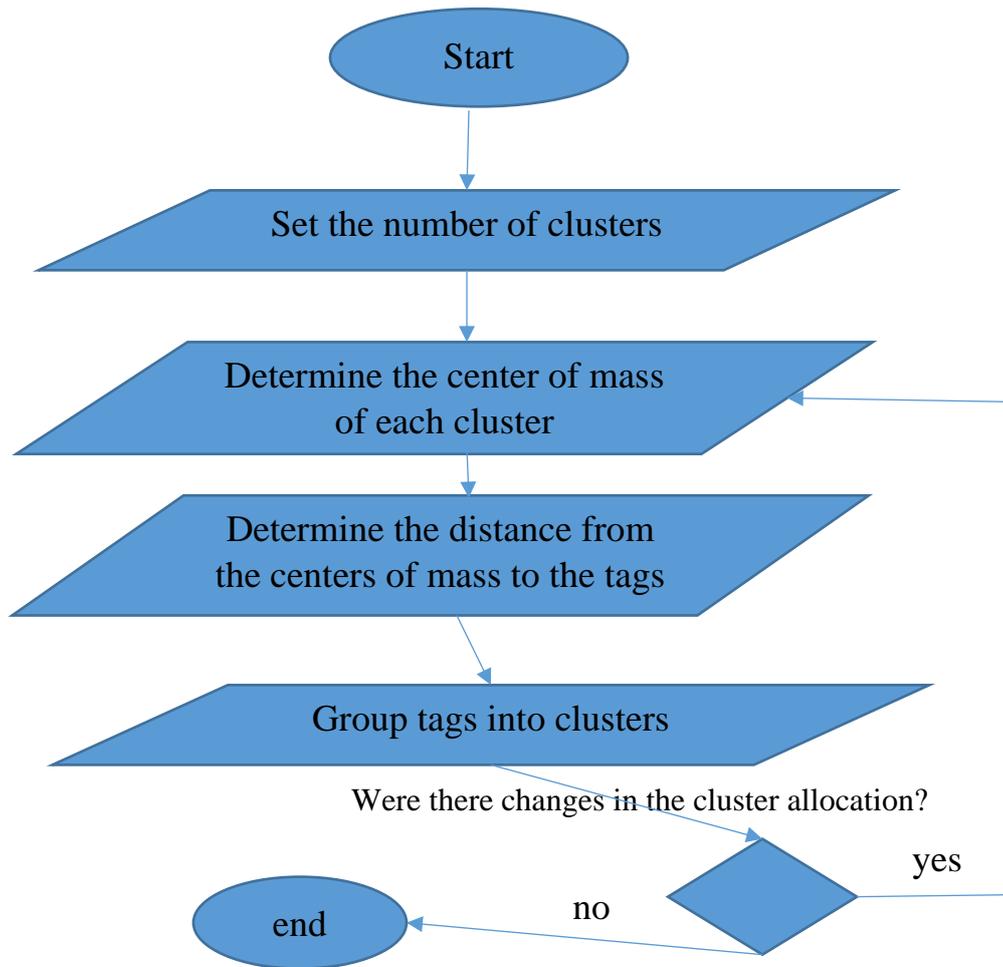


Image 1. Construction algorithm of the Tag Cloud



Image 2. Tag Cloud for "Economics, Econometrics and Finance" subject area.

Cluster analysis

Cluster analysis is an analysis that allows you to break up a large amount of data into classes or groups according to some criterion or a combination of them. It is considered that there is no additional information on the nature of the initial data.

Cluster analysis allows:

1. to classify the objects taking into account the characteristics that reflect their essence;
2. check the assumptions about the presence of some structure taken together of these objects;
3. to construct new classifications for weakly studied phenomena, when it is necessary to establish the existence of connections within the population and to structure them.

The methods of cluster analysis are divided into:

1. agglomerative (combining);
2. divisive (dividing);
3. iterative.

The first consistently combine individual objects into clusters, the latter, on the contrary, dismember clusters into objects. Still others unite the first two. Their peculiarity is the formation of clusters, based on the partitioning conditions (the so-called parameters), which can be changed during the operation of the algorithm to improve the quality of the partition. Iterative methods are usually used to classify large amounts of information.

To carry out the classification of data use the notion of metric or distance.

Definition: A metric is a function that maps some metric space into a space of real numbers and has the following properties (the axioms of a metric):

References

1. Noorden, R. V., 5 2014. Global scientific output doubles every nine years. Mathematics & Economics Nature News Blog, <http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html>.
2. Asadi, S., Dahlan, H. M., et al., 2017. Organizational research in the field of green it: A systematic literature review from 2007 to 2016. Telematics and Informatics.
3. Lewis, J. M., Ross, S., Holden, T., 2012. The how and why of academic collaboration: Disciplinary differences and policy implications. Higher Education 64 (5), 693–708.
4. Ajay, A., Henderson, R. M., 2002. Putting patents in context: Exploring knowledge transfer from mit. Management Science 48 (1), 44–60.
5. Beck, F., Burch, M., Diehl, S., Weiskopf, D., 2016. A taxonomy and survey of dynamic graph visualization. In: Computer Graphics Forum. Wiley Online Library
6. Ma, T., McGroarty, F., 2017. Social machines: how recent technological advances have aided financialisation. Journal of Information Technology
7. Leslie, S. A., Rensleigh, C., 2013. Web of science, scopus and google scholar: A content comprehensiveness comparison. The Electronic Library 31 (6), 727–744