

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

Инженерная школа информационных технологий и робототехники
 Направление подготовки 09.04.04 Программная инженерия
 ООП/ОПОП Технологии больших данных
 Отделение школы (НОЦ) Информационных технологий

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРАНТА

Тема работы
Методы, алгоритмы и программное обеспечение обнаружения сленга в онлайн-сообществах социальных медиа

УДК 004.415.2:81`276.16:004.773

Обучающийся

Группа	ФИО	Подпись	Дата
8ПМ11	Балаев Даниил Бахтиярович		

Руководитель ВКР

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

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОСГН ШБИП	Жиронкин С.А.	д.э.н		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ООД ШБИП	Федорчук Ю.М.	д.т.н.		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП, должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Губин Е.И.	к.ф.-м.н.		

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ООП

по направлению 09.04.04 «Программная инженерия»

Код компетенции	Наименование компетенции
Универсальные компетенции	
УК(У)-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий
УК(У)-2	Способен управлять проектом на всех этапах его жизненного цикла
УК(У)-3	Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели
УК(У)-4	Способен применять современные коммуникативные технологии, в том числе на иностранном (-ых) языке (-ах), для академического и профессионального взаимодействия
УК(У)-5	Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия
УК(У)-6	Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки
Общепрофессиональные компетенции	
ОПК(У)-1	Способен самостоятельно приобретать, развивать и применять математические, естественно-научные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте
ОПК(У)-2	Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач
ОПК(У)-3	Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями
ОПК(У)-4	Способен применять на практике новые научные принципы и методы исследований
ОПК(У)-5	Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем
ОПК(У)-6	Способен самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе в новых областях знаний, непосредственно не

	связанных со сферой деятельности
ОПК(У)-7	Способен применять при решении профессиональных задач методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе, в глобальных компьютерных сетях
ОПК(У)-8	Способен осуществлять эффективное управление разработкой программных средств и проектов
Профессиональные компетенции	
ПК(У)-1	Способен к созданию вариантов архитектуры программного средства
ПК(У)-2	Способен разрабатывать и администрировать системы управления базами данных
ПК(У)-3	Способен управлять процессами и проектами по созданию (модификации) информационных ресурсов
ПК(У)-4	Способен проектировать и организовывать учебный процесс по образовательным программам с использованием современных образовательных технологий
ПК(У)-5	Способен осуществлять руководство разработкой комплексных проектов на всех стадиях и этапах выполнения работ

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

Инженерная школа информационных технологий и робототехники
 Направление подготовки (специальность) 09.04.04 Программная инженерия
 ООП/ОПОП Технологии больших данных
 Отделение школы (НОЦ) Информационных технологий

УТВЕРЖДАЮ:
 Руководитель ООП

_____ _____ Губин Е. И.
 (Подпись) (Дата) (Ф.И.О)

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

Обучающийся:

Группа	ФИО
8ПМ11	Балаев Даниил Бахтиярович

Тема работы:

Методы, алгоритмы и программное обеспечение обнаружения сленга в онлайн-сообществах социальных медиа	
Утверждена приказом директора (дата, номер)	№ 37-58/с от 06.02.2023

Срок сдачи студентом выполненной работы:

	10.06.2023
--	------------

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<p>Исходные данные к работе <i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i></p>	<p>Объектом исследования и разработки являются алгоритмы для обнаружения сленга в онлайн-сообществах социальных медиа</p>
<p>Перечень подлежащих исследованию, проектированию и разработке вопросов <i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p>	<ol style="list-style-type: none"> 1. Обзор и анализ существующих решений. 2. Разработка алгоритмы для обнаружения сленга в онлайн-сообществах социальных медиа 3. Оценка точности работы алгоритма

	<p>4. Работа над разделом по финансовому менеджменту, ресурсоэффективности и ресурсосбережению.</p> <p>5. Работа над разделом по социальной ответственности.</p>
<p>Перечень графического материала (с точным указанием обязательных чертежей)</p>	<p>1. Алгоритм для создания таблицы значений векторов</p> <p>2. Алгоритм предобработки данных</p> <p>3. Алгоритм обнаружения сленга</p>
<p>Консультанты по разделам выпускной квалификационной работы (с указанием разделов)</p>	
Раздел	Консультант
Основная часть	Доцент ОИТ ИШИТР к.т.н., доцент Савельев А.О.
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Профессор ОСГН ШБИП, д.э.н. Жиронкин С.А.
Социальная ответственность	Профессор ООД ШБИП, д.т.н Федорчук Ю.М.
Раздел на английском языке	Старший преподаватель ОИЯ к.фил.н. Куркан Н.В.
<p>Названия разделов, которые должны быть написаны на иностранном языке:</p>	
<p>1 Аналитический обзор научной, нормативной и технической документации</p> <p>1.1 Актуальность</p> <p>1.2 Обзор существующих методов и подходов</p> <p>1.3 Постановка задачи исследования</p> <p>1.4 Спецификация требований</p> <p>1.5 Выводы по аналитическому обзору</p>	

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	01.03.2023
---	------------

Задание выдал руководитель ВКР:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Савельев А.О.	к.т.н., доцент		01.03.2023

Задание принял к исполнению обучающийся:

Группа	ФИО	Подпись	Дата
8ПМ11	Балаев Даниил Бахтиярович		01.03.2023

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

Инженерная школа Информационных технологий и робототехники
 Направление подготовки (ООП / ОПОП) 09.04.04 Программная инженерия
 Уровень образования магистратура
 Отделение школы (НОЦ) Информационных технологий
 Период выполнения весенний семестр 2022 /2023 учебного года

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Обучающийся:

Группа	ФИО
8ПМ11	Балаев Даниил Бахтиярович

Тема работы:

Методы, алгоритмы и программное обеспечение обнаружения сленга в онлайн-сообществах социальных медиа
--

Срок сдачи студентом выполненной работы:	10.06.2023
--	------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
10.06.2023	Основная часть	70
10.06.2023	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	10
10.06.2023	Социальная ответственность	10
10.06.2023	Английский язык	10

СОСТАВИЛ:

Руководитель ВКР

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

СОГЛАСОВАНО:

Руководитель ООП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Губин Е.И.	к.ф.-м.н.		

Задание принял к исполнению обучающийся:

Группа	ФИО	Подпись	Дата
8ПМ11	Балаев Даниил Бахтиярович		01.03.2023

РЕФЕРАТ

Выпускная квалификационная работа 85 страниц, 15 рисунков, 28 таблиц, 25 источников, 5 формул, 2 приложения.

Ключевые слова: машинное обучение, нейронная сеть, обработка естественного языка, LSTM, GRU.

Объектом исследования является алгоритм для обнаружения сленга в онлайн-сообществах социальных медиа.

Цель работы – исследование и разработка алгоритмов для обнаружения сленга в онлайн-сообществах социальных медиа.

В процессе исследования был проведен: обзор и анализ существующих решений в сфере обнаружения сленга.

В результате исследования был разработан алгоритм для автоматизированного обнаружения сленга онлайн-сообществ русскоязычных социальных медиа, определена точность классификация текста на текст из IT-сообщества, из маносферы, из неоязыческих сообществ и из популярных групп в социальной сети «ВКонтакте» для нейронной сети с плотно связанным слоем, с LSTM слоем, с Bi-LSTM слоем и с GRU слоем, определена точность работы алгоритма по определению сленговых слов.

Степень внедрения: на текущем этапе внедрение не планируется.

Область применения: социологические и маркетинговые исследования.

Экономическая эффективность/значимость работы: данная разработка позволяет автоматизировать обнаружение сленга онлайн сообществ. Это приводит к увеличению эффективности проведения социологических и маркетинговых исследований.

В будущем планируется проведение оптимизации работы разработанного алгоритма.

СОДЕРЖАНИЕ

РЕФЕРАТ.....	7
ВВЕДЕНИЕ	11
ТЕОРЕТИЧЕСКАЯ ЧАСТЬ.....	12
1 Аналитический обзор научной, нормативной и технической документации .	12
1.1 Актуальность.....	12
1.2 Обзор существующих методов и подходов	14
1.3 Постановка задачи исследования.....	16
1.4 Спецификация требований	17
1.5 Выводы по аналитическому обзору.....	17
МЕТОДЫ, АЛГОРИТМЫ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ОБНАРУЖЕНИЯ СЛЕНГА В ОНЛАЙН-СООБЩЕСТВАХ СОЦИАЛЬНЫХ МЕДИА.....	18
2.1 Анализ набора данных	18
2.2 Векторизация текста.....	19
2.3 Проектирование классификационных моделей.....	22
2.4 Алгоритм обнаружения сленга.....	26
РЕЗУЛЬТАТЫ РАБОТЫ И ИХ ОБСУЖДЕНИЕ.....	28
3.1 Результаты расчета точность классификационных моделей.....	28
3.2 Результаты расчета точности и скорости извлечения сленга.....	30
3.3 Обсуждение результатов и апробация.....	31
ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ	33
4.1 Предпроектный анализ.....	33
4.1.1 Потенциальные потребители результатов исследования	33

4.1.2 SWOT – анализ.....	34
4.1.3 Оценка готовности разработки к коммерциализации	37
4.1.4 Цели и результаты проекта	39
4.2 Планирование научно-исследовательских работ	40
4.2.1 Структура научно-исследовательских работ	40
4.2.2 План проекта	42
4.2.3 Бюджет научно-технического исследования	44
4.2.3.1 Расчет материальных затрат на научное исследование	44
4.2.3.2 Специальное оборудование для научных работ	45
4.2.3.3 Основная заработная плата	45
4.2.3.4 Дополнительная заработная плата научно-производственного персонала	47
4.2.3.5 Отчисления на социальные нужды	47
4.2.3.6 Накладные расходы	48
4.2.3.7 Формирование бюджета затрат научно-исследовательского проекта	48
4.3 Оценка сравнительной эффективности исследования	49
4.3.1 Интегральный показатель финансовой эффективности	49
4.3.2 Интегральный показатель ресурсоэффективности.....	50
4.3.3 Интегральный показатель эффективности	51
4.4 Вывод	53
СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ.....	56
Введение:	56
5.1 Производственная безопасность	56
5.1.1 Вредные факторы.....	56
5.1.1.1 Отклонение показателей микроклимата в помещении	56

5.1.1.2 Превышение уровней шума	58
5.1.1.3 Повышенный уровень электромагнитных излучений	59
5.1.1.4 Недостаточная освещенность	61
5.1.2 Опасные факторы	65
5.1.2.1 Электроопасность; класс электроопасности помещения, безопасные номиналы I, U, Rзаземления, СКЗ, СИЗ	65
5.1.2.2 Пожароопасность, категория пожароопасности помещения, марки огнетушителей, их назначение и ограничение применения.....	66
5.2 Экологическая безопасность	68
5.3 Безопасность в чрезвычайных ситуациях	69
Список источников	71
СПИСОК ЛИТЕРАТУРЫ	72
Приложение А	75
Приложение Б	84

ВВЕДЕНИЕ

На сегодняшний день сленг является неотъемлемой частью языка интернет-пользователей в социальных сетях. По мере того, как они распространяются на другие средства массовой информации, сленг начинает использоваться в статьях, комментариях и разговорах. Значения сленговых слов часто определяют ключевую семантику в предложениях и могут играть решающую роль в приложениях обработки естественного языка (NLP). Таким образом, учет сленга позволяет повысить эффективность семантического анализа текстов.

Целью данной работы является разработка методов и программного обеспечения для автоматизированного обнаружения характерного сленга онлайн-сообществ русскоязычных социальных медиа. Классификация текста на текст содержащий сленг и не содержащий сленг будет проводиться с помощью нейронной сети с плотно связанным слоем, Long Short Term Memory (LSTM) сети, Bidirectional Long Short Term Memory (Bi-LSTM) сети и Gated Recurrent Unit (GRU) сети.

Для классификации в качестве текста, был использован набор данных с текстом комментариев из различных сообществ в социальной сети «ВКонтакте».

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

1 Аналитический обзор научной, нормативной и технической документации

1.1 Актуальность

Сленг – тип неформального языка, состоящий из слов и выражений, используемых в определенных группах. Отличительной чертой сленга является его экспрессивность, воплощенная в гибком использовании слов. Выразительный характер сленга иллюстрирует его социальную функцию, поскольку он обеспечивает эффективный способ общения и обмена знаниями в группах с различными социальными идентичностями [1].

В работе [2] авторы выделяют два типа сленга:

- слова с новым значением: существующие слова в лексиконе, которые приобретают новые сленговые значения, отличные от их устоявшихся значений.

- новые слова: слова, которых нет в стандартном лексиконе.

На сегодняшний день сленг является неотъемлемой частью языка интернет-пользователей в социальных сетях. По мере того, как они распространяются на другие средства массовой информации, сленг начинает использоваться в статьях, комментариях и разговорной речи. Значения сленговых слов часто определяют ключевую семантику в предложениях и могут играть решающую роль в приложениях обработки естественного языка (NLP). Таким образом, учет сленга может помочь в повышении эффективности семантического анализа текстов.

Одна из наиболее распространенных задач семантического анализа текста – это анализ мнений и анализ тональности текста. Основная цель анализа мнений – автоматизация извлечения мнений пользователей из неструктурированных текстов. Он используется для анализа мнений,

отзывов, отношений людей к организациям, людям, проблемам, событиям, и различным темам.

Анализ мнений и анализ тональности текста используется в разных областях человеческой жизнедеятельности. Например, производителям коммерческой продукции он может быть полезным для улучшения продукции, маркетинговым компаниям для получения представления о рыночном спросе, покупателям для получения краткого изложения мнений других клиентов о продукте. Анализ мнений также может быть использован в политике при проведении политических выборов и при принятии решений по улучшению социальной жизни сообщества, а также в финансовой сфере для прогноза стоимости акций или в различных социологических исследованиях [3].

На сегодняшний день значимым источником данных для решения указанных задач являются социальные сети, так как люди используют их в том числе для выражения своего мнения, при этом в них можно найти сообщества посвященный практически любой теме, а совокупный процент активных пользователей социальных сетей составляет 59% населения земли или 4,8 млрд человек [4], при этом 6 наиболее популярных социальных сетей имеют более 1 млрд активных пользователей. Список наиболее популярных социальных сетей представлен на рисунке 1.

Но при этом использование данных социальных сетей сопряжено с такими сложностями, как: большой объем информации, отсутствие структуры, проблема конфиденциальности, низкое качество данных (проблема спама), динамичность информации.

Таким образом единственным потенциальным решением является автоматизация процессов обработки естественного языка.

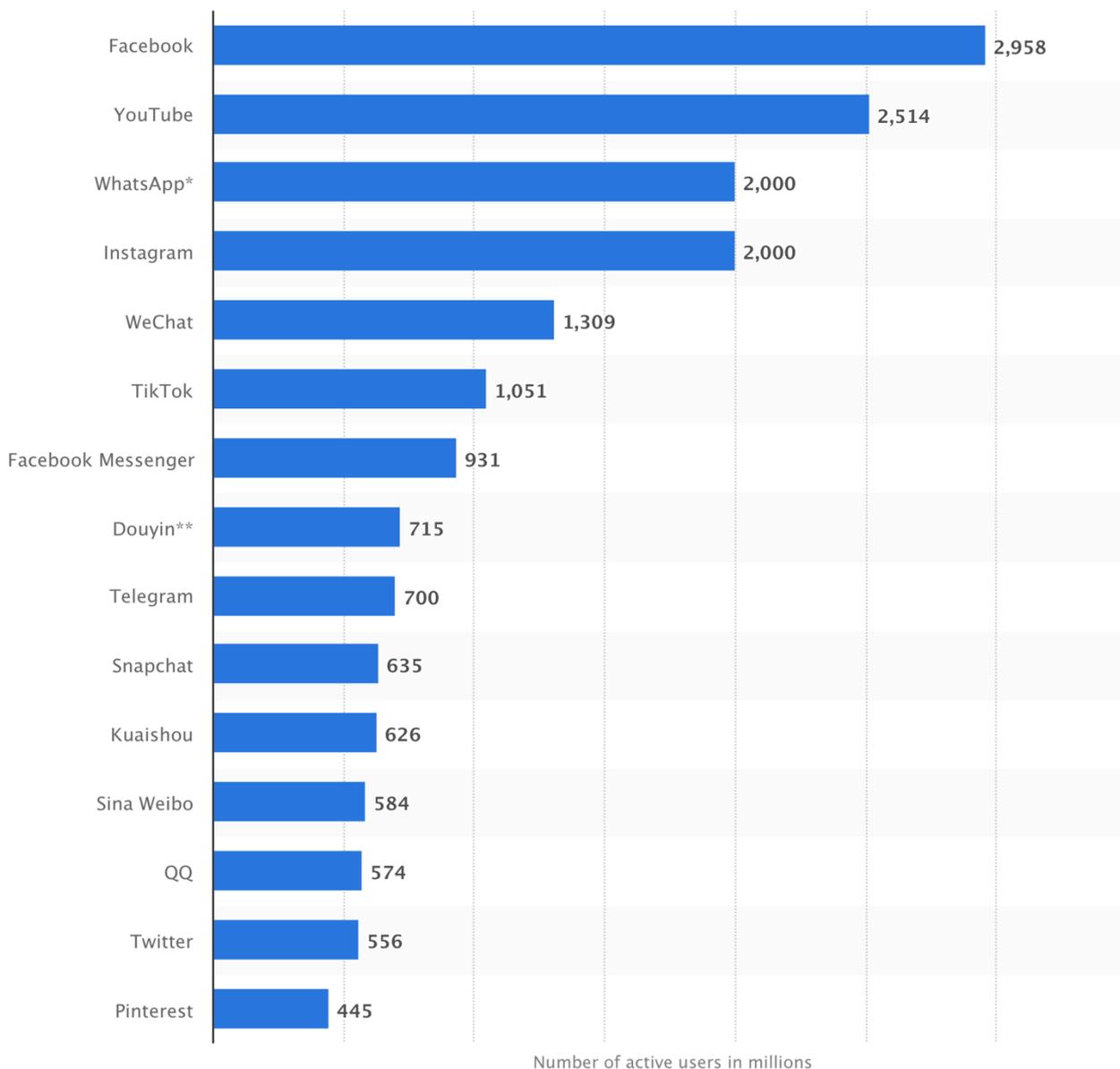


Рисунок 1.1 – Самые популярные социальные сети по количеству активных пользователей в месяц [5]

1.2 Обзор существующих методов и подходов

Существующие на сегодняшний день методы определения сленга можно свести к следующим типам:

1) Лексикографический метод – это метод определения новых слов, основанный на частоте их употребления, например, путем определения наиболее распространенных в корпусе социальных сетей слов, не

являющихся словарными [2, 6]. Этот метод работает для неологизмов, но не будет работать для существующих слов, получивших новое сленговое значение.

2) Метод использующий орфографический плеоназм. Данный метод предложен в статье [7], он основан на предположении о том, что добавление повторяющихся букв подчеркивает (усиливает) значение слова. Таким образом сленгом считаются слова, имеющие наибольшее количество вариантов написания.

3) Тематическое моделирование. Данный метод можно использовать для проверки того, изменилось ли значение слова, возможно, с новым сленговым значением, относительно того, как изменилось его использование в различных темах обсуждения. Однако этот метод не позволяет идентифицировать новые слова и не специфичен для тональности [8, 9].

4) Глубокое обучение. Пример использования этого метода приведен в статьях [10, 11, 12], при его использовании, авторы обнаружили, что отличительной чертой сленга является неожиданное использование слов в разных синтаксических категориях или синтаксический сдвиг. Недостатком этого метода является то, что он может плохо работать на неформально структурированном языке и нуждается в обучающих данных.

5) Частотный метод. Данный способ подходит для определения поляризованного политического сленга. Он предполагает, что такой сленг можно обнаружить, сравнивая, после удаления словарных слов, имен собственных и известного сленга, частоты употребления слов между наборами комментариев, выражающих разную политическую позицию. Таким образом слова, которые чаще употребляются людьми с определенной политической позиций, с большей вероятностью будут являться сленговыми для этой группы [13].

При этом данные методы можно использовать совместно, так, например, в работе [14] предложен способ определения сленга, который предполагает

совместное использование лексикографического метода и глубокого обучения.

Все рассмотренные литературные источники предполагают определение сленга в английском языке и в некоторых других иностранных языках, но в ходе литературного обзора не было обнаружено работ, посвященных обнаружению сленга в русскоязычных текстах. При этом, из-за морфологических и синтаксических особенностей русского языка, методы, хорошо работающие для иностранных языков, могут показать худший результат для русского языка.

1.3 Постановка задачи исследования

Основной целью настоящей работы является разработка комплексного алгоритмического обеспечения для выявления из текстов онлайн-сообществ сленг-слов и словосочетаний. Для достижения поставленной цели в начале необходимо подготовить обучающий, тестовый и валидационный наборы данных, при этом «сырые» данные нужно предварительно предобработать.

Далее должны быть сформированы и обучены несколько различных ML-моделей для автоматизированного выявления сленга и осуществлена реализация программного обеспечения для автоматизированного выявления сленга. На вход программы должна принимать набор текстов из онлайн-сообществ, а на выходе присваивать каждому тексту категорию, указывающую на наличие или отсутствие сленговых слов в рассмотренном тексте, и если в тексте определено наличие сленговых слов, то они должны быть выделены.

На следующем этапе необходимо осуществить апробацию реализованного программного обеспечения на тестовых данных и сделать сравнительную оценку качества работы ML-моделей, а также провести интерпретацию полученных результатов и, если необходимо, провести доработку некоторых моделей.

1.4 Спецификация требований

Разрабатываемое в ходе данной работы алгоритмическое обеспечение для выявления из текстов онлайн-сообществ сленг-слов и словосочетаний должно принимать на вход набор текстов из онлайн-сообществ, а на выходе предсказывать для каждого текста категорию, которая указывает на отсутствие в проанализированном тексте сленговых слов или наличие сленгового слова определенной группы (сленгу участников разных тематических сообществ присвоен разный номер), и если в тексте определено наличие сленговых слов, то эти слова должны быть выявлены как сленговые. При этом точность классификации (accuracy) должна быть выше 50%, а скорость обработки одного текста не должна превышать нескольких секунд.

1.5 Выводы по аналитическому обзору

В итоге обзора литературных источников можно сделать вывод о том, что наиболее универсальным методом определения сленга является метод, основанный на глубоком обучении, так как он в отличие от тематического и лексикографического методов, позволяет обнаруживать как новые сленговые слова, так и слова, получившие новое сленговое значение, при этом, в отличие от частотного метода, глубокое обучение позволяет выявлять сленг в текстах любой тематики.

МЕТОДЫ, АЛГОРИТМЫ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ОБНАРУЖЕНИЯ СЛЕНГА В ОНЛАЙН-СООБЩЕСТВАХ СОЦИАЛЬНЫХ МЕДИА

2.1 Анализ набора данных

В качестве набора данных был использован набор текстов 7029 комментариев из шестнадцати различных сообществ в социальной сети «ВКонтакте». Все эти сообщества можно разделить на 4 группы: IT сообщества («Программирование / itProger», «Люди и код», «Типичный программист»), популярные сообщества («Палата №6», «Мэш»), а также 2 типа сообществ, имеющих признаки радикализации, а именно сообщества маносферы, т.е. сообщества пропагандирующие маскулинность, женоненавистничество и анти-феминизм [15], и неоязыческие сообщества, т.е. сообщества приверженцев новых религиозных движений, для которых характерны синкретизм, активное использование средств массовой информации, коммуникаций, апокалипсизм и миссионерство [16].

Количество комментариев в каждом классе представлено на рисунке 1.

Далее данный набор был проверен на отсутствующие значения и оказалось, что он содержит 926 пустых строк, которые были удалены из него для дальнейшей работы.

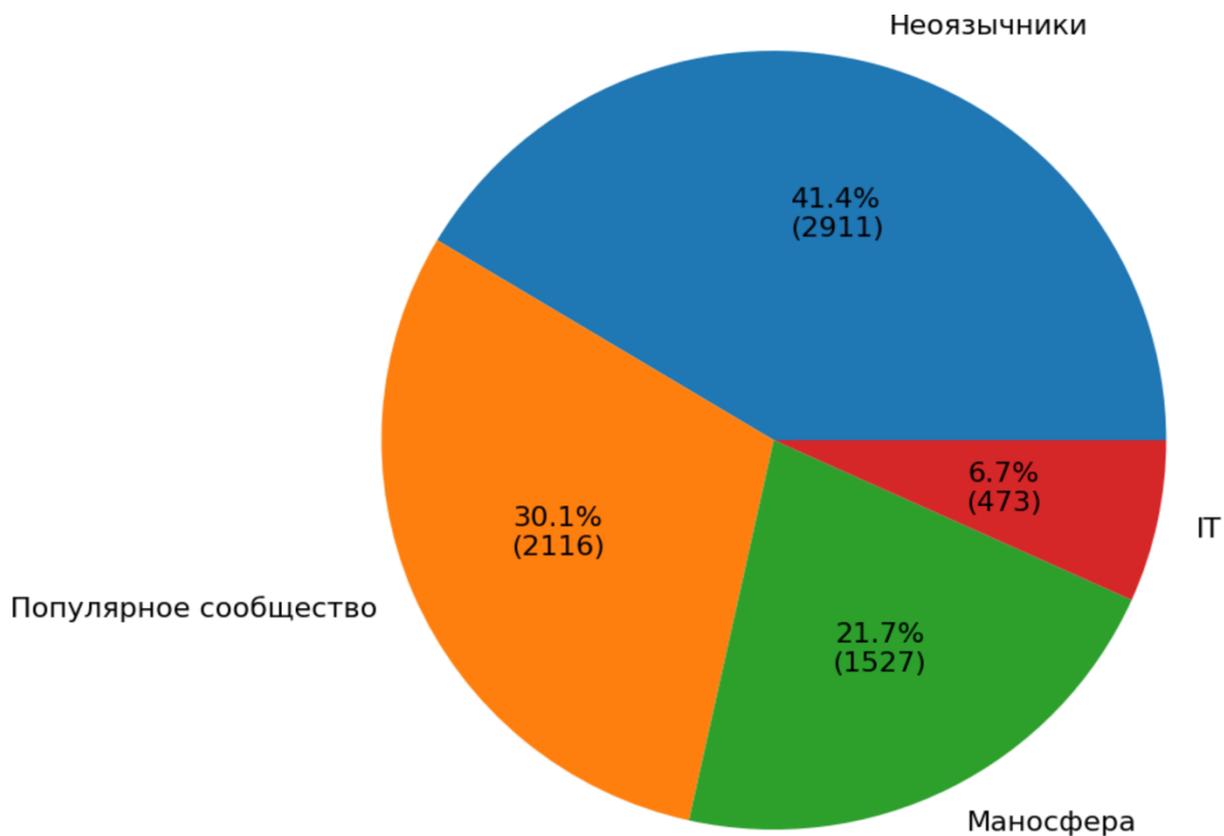


Рисунок 2.1 – Количество комментариев в каждом классе

2.2 Векторизация текста

Далее была проведена векторизация данных, для этого этот набор данных был пропущен через алгоритм, представленный на рисунке 2. Результат работы алгоритма представлен на рисунке 3, а результат работы подпрограммы этого алгоритма для обработки текста представлен на рисунке 4.

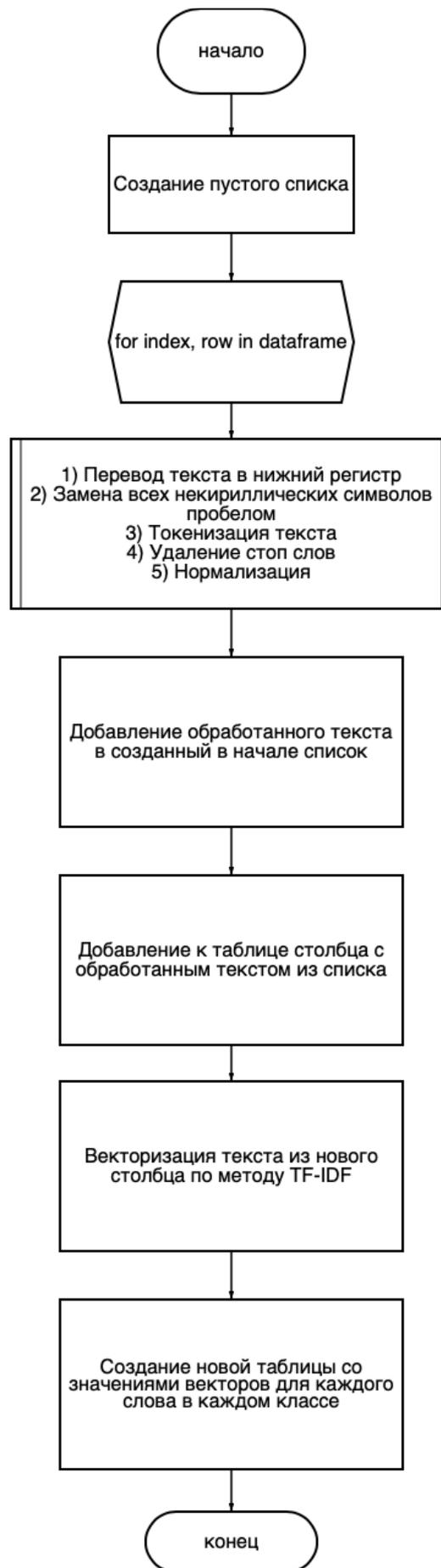


Рисунок 2.2 – Алгоритм для создания таблицы значений векторов

	Doc1	Doc2	Doc3	Doc4
который	0.271561	0.212889	0.263241	0.229221
год	0.213369	0.108955	0.132075	0.236297
это	0.199145	0.333393	0.327735	0.305525
работа	0.173282	0.036151	0.037262	0.110149
разработчик	0.166816	0.000251	0.000318	0.000615

Рисунок 2.3 – Таблица со значениями векторов

	class	text	text_clean
0	IT	nan nan nan Вы старше или моложе JavaScript? 😊...	[старший, молодой, представлять, подборка, нес...
1	Маносфера	...При этом, согласно данным Росстата, сильнее...	[согласно, данные, росстат, сильный, год, сокр...
2	Неоязычники	nan КАК КАЛЕЧИЛИ РУССКИЙ ЯЗЫК В XX ВЕКЕ.\n\nПр...	[калечить, русский, язык, xx, век, процесс, пр...
3	Популярное сообщество	Кошка у меня невероятно старая, в последние ме...	[кошка, невероятно, старый, последний, месяц, ...

Рисунок 2.4 – Результат работы подпрограммы по отчистке, токенизации и нормализации текста

2.3 Проектирование классификационных моделей

Для классификации текста было создано несколько моделей нейронных сетей: с плотно связанным слоем, с LSTM (long short-term memory) слоем, с Bi-LSTM (Bi-directional long short term memory) слоем и GRU (Gated Recurrent Unit) слоем [17].

Для создания этих моделей использовали открытую программную библиотеку для машинного обучения TensorFlow [18]. Полученные модели представлены на рисунках 5 – 8.

```
Model: "sequential"
-----
Layer (type)                Output Shape                Param #
-----
embedding (Embedding)       (None, None, 16)          80000
global_average_pooling1d (G (None, 16)                  0
lobalAveragePooling1D)
dense (Dense)                (None, 24)                 408
dropout (Dropout)           (None, 24)                 0
dense_1 (Dense)              (None, 4)                  100
-----
Total params: 80,508
Trainable params: 80,508
Non-trainable params: 0
```

Рисунок 2.5 – Модель нейронной сети с плотно связанным слоем

```

Model: "sequential_1"

```

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 350, 16)	80000
spatial_dropout1d (SpatialDropout1D)	(None, 350, 16)	0
lstm (LSTM)	(None, 128)	74240
dropout_1 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 4)	516

```

=====
Total params: 154,756
Trainable params: 154,756
Non-trainable params: 0

```

Рисунок 2.6 – Модель нейронной сети с LSTM слоем

```

Model: "sequential_5"

```

Layer (type)	Output Shape	Param #
embedding_5 (Embedding)	(None, 350, 16)	80000
bidirectional_2 (Bidirectional)	(None, 256)	148480
dropout_5 (Dropout)	(None, 256)	0
dense_6 (Dense)	(None, 4)	1028

```

=====
Total params: 229,508
Trainable params: 229,508
Non-trainable params: 0

```

Рисунок 2.7 – Модель нейронной сети с Bi-LSTM слоем

```

Model: "sequential_6"

```

Layer (type)	Output Shape	Param #
embedding_6 (Embedding)	(None, 350, 16)	80000
spatial_dropout1d_2 (SpatialDropout1D)	(None, 350, 16)	0
gru_1 (GRU)	(None, 128)	56064
dropout_6 (Dropout)	(None, 128)	0
dense_7 (Dense)	(None, 4)	516

```

Total params: 136,580
Trainable params: 136,580
Non-trainable params: 0

```

Рисунок 2.8 – Модель нейронной сети с GRU слоем

Для классификации текста с помощью нейронных сетей набор данных был преобработан с помощью алгоритма, представленного на рисунке 9.

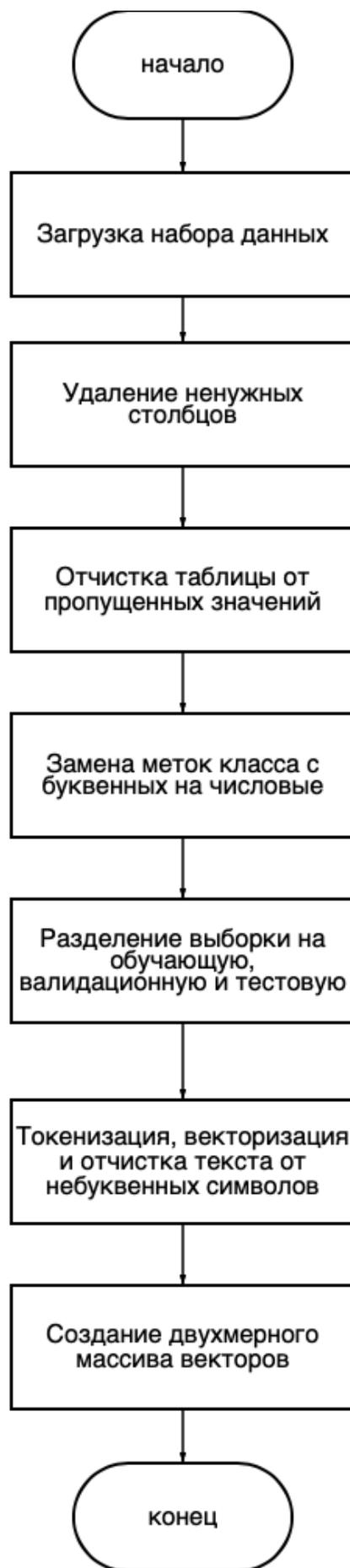


Рисунок 2.9 – Алгоритм предобработки данных

2.4 Алгоритм обнаружения сленга

Для обнаружения сленга был разработан алгоритм, представленный на рисунке 10, результат работы данного алгоритма представлен на рисунке 11.

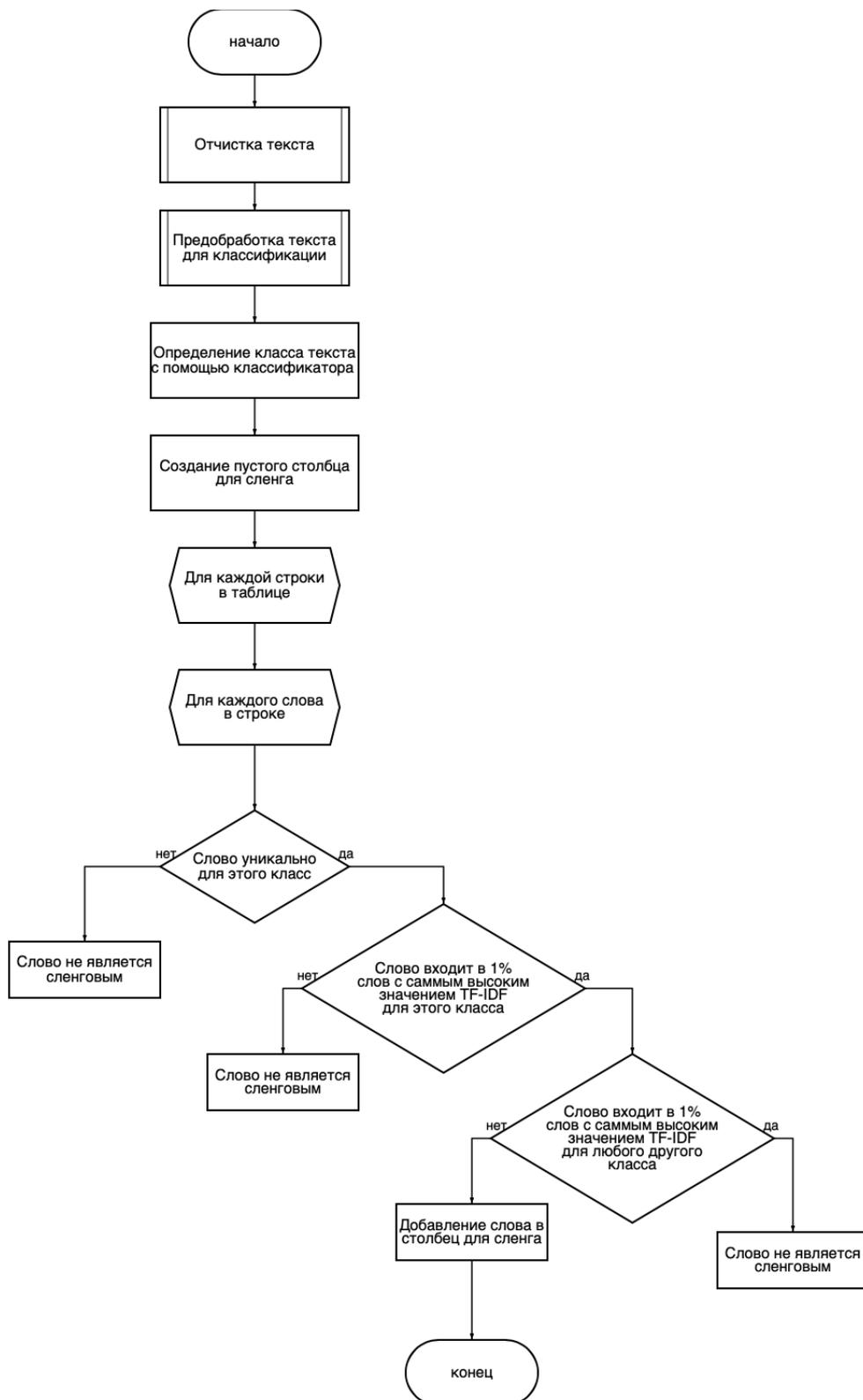


Рисунок 2.10 – Алгоритм обнаружения сленга

	<code>text</code>	<code>slang</code>
990	Женщина в отношениях со мной брезгует целовать...	
12	Очередной поединок бахини и гравитации. Гравит...	[бахинь]
1400	лечить исцелять врачевать ...	□

Рисунок 2.11 – Результат работы алгоритма по обнаружению сленга

Листинг фрагментов кода, содержащих функцию для предобработки текста и функцию для обнаружения сленга представлены в приложении Б.

РЕЗУЛЬТАТЫ РАБОТЫ И ИХ ОБСУЖДЕНИЕ

В данной главе приведены результаты расчетов точности работы классификационных моделей и итогового алгоритма по выявлению сленга.

3.1 Результаты расчета точность классификационных моделей

В таблицах 3.1 – 3.4 приведены достигнутые каждой классификационной моделью уровни точности. Точность проверялась на отложенной выборке, которая составляет 20% от изначального набора данных. Сравнение точности разных классификационных моделей представлено в сводной таблице 3.5.

Таблица 3.1 – Точность нейронной сети с плотно связанным слоем

	precision	recall	f1-score	accuracy
IT	0,00	0,00	0,00	0,82
Маносфера	0,72	0,54	0,61	
Неоязыческие сообщества	0,86	0,97	0,91	
Популярные сообщества	0,81	0,96	0,88	

Таблица 3.2 – Точность нейронной сети с LSTM слоем

	precision	recall	f1-score	accuracy
IT	0,00	0,00	0,00	0,51
Маносфера	0,00	0,00	0,00	
Неоязыческие сообщества	0,92	0,41	0,56	
Популярные сообщества	0,43	1,00	0,6	

Таблица 3.3 – Точность нейронной сети с Bi-LSTM слоем

	precision	recall	f1-score	accuracy
IT	0,77	0,75	0,76	0,86
Маносфера	0,84	0,60	0,70	
Неоязыческие сообщества	0,84	0,93	0,88	
Популярные сообщества	0,90	0,93	0,92	

Таблица 3.4 – Точность нейронной сети с GRU слоем

	precision	recall	f1-score	accuracy
IT	0,00	0,00	0,00	0,72
Маносфера	0,47	0,38	0,42	
Неоязыческие сообщества	0,88	0,90	0,89	
Популярные сообщества	0,66	0,84	0,74	

Таблица 3.5 – Сравнение точности разных нейронных сетей

Модель нейронной сети	precision	recall	f1-score	accuracy
Dense	60%	62%	60%	82%
LSTM	34%	35%	29%	51%
Bi-LSTM	84%	80%	82%	86%
GRU	50%	53%	51%	72%

3.2 Результаты расчета точности и скорости извлечения сленга

Для расчета точности извлечения сленга была вручную проведена экспертная оценка 30 случайных предложений для каждой из трех категорий и для каждой из четырех классификационных моделей. Полученные результаты представлены в таблице 3.6 и 3.7.

Таблица 3.6 – Точность определения сленга

Модель нейронной сети	IT сообщества	Маносфера	Неоязычники	Средняя точность
Dense	73%	57%	68%	66%
LSTM	80%	60%	68%	69%
Bi-LSTM	73%	60%	68%	67%
GRU	77%	57%	68%	67%

Таблица 3.7 – Скорость определения сленга, мс

Модель нейронной сети	IT сообщества	Маносфера	Неоязычники	Средняя скорость
Dense	28	45	264	112
LSTM	177	531	489	399
Bi-LSTM	279	670	634	528
GRU	155	323	589	356

3.3 Обсуждение результатов и апробация

В ходе данной работы была достигнута цель и выполнены следующие задачи:

- проведен обзор и анализ существующих решений в сфере обнаружения сленга.

- разработан алгоритм для автоматизированного обнаружения сленга онлайн-сообществ русскоязычных социальных медиа.

- определена точность классификация текста на текст из IT-сообщества, из маносферы, из неоязыческих сообществ и из популярных групп в социальной сети «ВКонтакте» для нейронной сети с плотно связанным слоем, с LSTM слоем, с Bi-LSTM слоем и с GRU слоем.

- показано, что лучшую точность классификации показала сеть с Bi-LSTM слоем ее значение метрики f1 выше на 182,8%, чем у сети с LSTM слоем, которая показала худшую точность и на 60,8%, чем у сети с GRU слоем, которая показала 2-ю лучшую точность. При этом точность для популярных групп и неоязыческих сообществ выше, чем для IT-сообществ и маносферы, что объясняется тем, что количество текстов для маносферы в 2 раза меньше, чем для неоязыческих сообществ и популярных групп, а у IT-сообщества меньше в 5 раз.

- определена точность работы алгоритма по определению сленговых слов

- показано, что точность определения сленговых слов для для IT-сообществ в среднем выше, чем для маносферы на 28,8% и на 11,8% – чем для неоязыческих сообществ. При этом лучший результат показала сеть с LSTM слоем (на 4,5% выше, чем сеть с Dense слоем, которая показала худший результат). При этом сеть с Dense слоем показала самую высокую скорость работы (в среднем 112 мс на один текст), а сеть с Bi-LSTM – самую низкую (в среднем 528 мс на один текст).

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

Студенту:

Группа	ФИО
8ПМ11	Балаев Даниил Бахтиярович

Школа	ИШИТР	Отделение школы (НОЦ)	ОИТ
Уровень образования	Магистратура	Направление/специальность	09.04.04 «Программная инженерия»

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

1. <i>Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i>	<i>Бюджет проекта – не более 1494683,38 руб., в т.ч. затраты по оплате труда – не более 452506,56 руб.</i>
2. <i>Нормы и нормативы расходования ресурсов</i>	<i>Значение показателя интегральной ресурсоэффективности – не менее 4,45 баллов из 5,00</i>
3. <i>Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	<i>Коэффициент отчислений во внебюджетные фонды – 30,2%</i>

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. <i>Оценка коммерческого и инновационного потенциала НТИ</i>	<i>Анализ потенциальных потребителей, SWOT-анализ.</i>
2. <i>Разработка устава научно-технического проекта</i>	<i>Инициация проекта: определение заинтересованных сторон проекта, целей и результатов проекта</i>
3. <i>Планирование процесса управления НТИ: структура и график проведения, бюджет, риски и организация закупок</i>	<i>Составление календарного плана проекта. Определение бюджета НТИ</i>
4. <i>Определение ресурсной, финансовой, экономической эффективности</i>	<i>Определение ресурсной и финансовой эффективности проекта</i>

Перечень графического материала (с точным указанием обязательных чертежей):

1. <i>Сегментирование рынка</i>
2. <i>Матрица SWOT</i>
3. <i>Цели и результаты проекта</i>
4. <i>Перечень работ и исполнителей</i>
5. <i>График проведения НТИ</i>
6. <i>Бюджет НТИ</i>
7. <i>Оценка ресурсной, финансовой и экономической эффективности НТИ</i>

Дата выдачи задания для раздела по линейному графику	01.03.2023 г.
---	---------------

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОСГН ШБИП ТПУ	Жиронкин Сергей Александрович	д-р. экон. наук		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ПМ11	Балаев Даниил Бахтиярович		

ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ

4.1 Предпроектный анализ

4.1.1 Потенциальные потребители результатов исследования

Разработанный алгоритм обнаружения сленга планируется использовать для улучшения работы приложений по определению тональности текста.

Заинтересованными в продукте могут быть различные компании предоставляющие услуги по обработке естественного языка и маркетинговые компании. Также данный алгоритм можно использовать в социологических исследованиях.

Результаты сегментирования представлены на рисунке 4.1.

		Сфера применения разработки		
		Приложения для обработки естественного языка	Научная отрасль	Маркетинговые предприятия
Размер компании	Крупные			
	Средние			
	Мелкие			

Рисунок 4.1 – Карта сегментирования рынка услуг по обработке естественного языка

	Фирма А		Фирма Б		Фирма В
--	---------	--	---------	--	---------

Таким образом, основными заказчиками, на которые ориентирован программный продукт, являются крупные и средние компании, предоставляющие услуги по обработке естественного языка и маркетинговые компании.

4.1.2 SWOT – анализ

SWOT – Strengths (сильные стороны), Weaknesses (слабые стороны), Opportunities (возможности) и Threats (угрозы) – представляет собой комплексный анализ научно-исследовательского проекта. SWOT–анализ применяют для исследования внешней и внутренней среды проекта [19].

SWOT – анализ состоит из трех этапов. В первом этапе мы анализируем сильные и слабые стороны исследовательской работы (внутренняя среда), а также возможности и угрозы (внешняя среда). Описание выполняется с помощью факторов, не имеющих количественной оценки (таблица 4.1).

Таблица 4.1 – Матрица SWOT

	Сильные стороны научно-исследовательского проекта:	Слабые стороны научно-исследовательского проекта:
	<p>С1. Быстрая скорость выделения сленга.</p> <p>С2. Не требует предварительной обработки текста</p> <p>С3. Отличная от конкурентов технология.</p>	<p>Сл1. Требуется большое количество данных для обучения.</p> <p>Сл2. Относительно невысокая точность выделения сленга.</p>
Возможности:		
<p>В1. Создание более точного алгоритма.</p> <p>В2. Появление дополнительного спроса на технологию.</p>		

Угрозы:		
У1. Появление более совершенного аналога.		
У2. Падение спроса на технологии обработки естественного языка.		

Второй этап состоит в выявлении соответствия сильных и слабых сторон научно-исследовательского проекта внешним условиям окружающей среды. Это соответствие или несоответствие должны помочь выявить степень необходимости проведения стратегических изменений.

В рамках данного этапа построили интерактивную матрицу проекта (таблица 4.2). Ее использование помогает разобраться с различными комбинациями взаимосвязей областей матрицы SWOT. Каждый фактор помечается либо знаком «+» (означает сильное соответствие сильных сторон возможностям), либо знаком «-» (что означает слабое соответствие); «0» – если есть сомнения в том, что поставить «+» или «-».

Таблица 4.2 – Интерактивная матрица проекта

Сильные стороны проекта				
Возможности проекта.		С1.	С2.	С3.
	В1.	-	-	+
	В2.	0	+	+
Угрозы проекта.	У1.	+	-	+
	У2.	0	0	0
Слабые стороны проекта				
Возможности проекта.		Сл1.	Сл2.	
	В1	+	+	
	В2.	-	+	
Угрозы проекта.	У1.	+	+	
	У2.	-	-	

Третий этап заключается в составлении итоговой матрицы SWOT – анализа (таблица 4.3). Она заполняется, исходя из анализа таблицы 4.2.

Таблица 4.3 – Итоговая матрица SWOT-анализ

	Сильные стороны научно-исследовательского проекта:	Слабые стороны научно-исследовательского проекта:
	<p>С1. Быстрая скорость выделения сленга.</p> <p>С2. Не требует предварительной обработки текста</p> <p>С3. Отличная от конкурентов технология.</p>	<p>Сл1. Требуется большое количество данных для обучения.</p> <p>Сл2. Относительно невысокая точность выделения сленга.</p>
<p>Возможности:</p> <p>В1. Создание более точного алгоритма.</p> <p>В2. Появление дополнительного спроса на технологию.</p>	<p>В1С3, В2С2С3. Из данной комбинации видим, что вышеперечисленные сильные стороны благотворно влияют на возможность создания более точного алгоритма и появления дополнительного спроса на технологию.</p>	<p>В1Сл1Сл2, В2Сл2. Анализ показывает, что такие возможности как В1 и В2 могут перекрыть слабые стороны Сл1 и Сл2.</p>
<p>Угрозы:</p> <p>У1. Появление более совершенного аналога.</p> <p>У2. Падение спроса на технологии обработки естественного языка.</p>	<p>У1С1С3. Проведя анализ сильных сторон и угроз, можно сказать, что вышеуказанные сильные стороны проекта могут существенно понизить угрозу У1.</p>	<p>У1Сл1Сл2. Анализ показывает, что такие слабые стороны как Сл1 и Сл2 подвержены угрозе У1, поэтому следует обратить на них внимание при совершенствовании работы алгоритма.</p>

Самым большим преимуществом данной разработки является её отличие от конкурентов, а недостатком – относительно не высокая точность.

4.1.3 Оценка готовности разработки к коммерциализации

Одной из важных задач в ходе выполнения данного раздела является оценка готовности разработки к коммерциализации. Оцениваемыми параметрами являются как научная, так и коммерческая составляющая. Таблица 4.4 представляет собой бланк оценки степени готовности разработки к коммерциализации.

Таблица 4.4 – Бланк оценки степени готовности разработки к коммерциализации

№ п/п	Наименование	Степень проработанности научного проекта	Уровень имеющихся знаний у разработчика
1	Определен имеющийся научно-технический задел	4	4
2	Определены перспективные направления коммерциализации научно-технического задела	3	5
3	Определены отрасли и технологии (товары, услуги) для предложения на рынке	2	3
4	Определена товарная форма научно-технического задела для представления на рынок	2	2
5	Определены авторы и осуществлена охрана их прав	3	3

№ п/п	Наименование	Степень проработанности научного проекта	Уровень имеющихся знаний у разработчика
6	Проведена оценка стоимости интеллектуальной собственности	3	3
7	Проведены маркетинговые исследования рынков сбыта	1	3
8	Разработан бизнесплан коммерциализации научной разработки	1	2
9	Определены пути продвижения научной разработки на рынок	3	3
10	Разработана стратегия (форма) реализации научной разработки	4	4
11	Проработаны вопросы международного сотрудничества и выхода на зарубежный рынок	2	2
12	Проработаны вопросы использования услуг инфраструктуры поддержки, получения льгот	2	3
13	Проработаны вопросы финансирования коммерциализации научной разработки	2	3

№ п/п	Наименование	Степень проработанности научного проекта	Уровень имеющихся знаний у разработчика
14	Имеется команда для коммерциализации научной разработки	1	2
15	Проработан механизм реализации научного проекта	4	4
ИТОГО БАЛЛОВ		<i>37/75</i>	<i>46/75</i>

Таким образом, перспективность научной разработки и ее разработчика к коммерциализации средняя.

4.1.4 Цели и результаты проекта

Информация о заинтересованных сторонах проекта представлена в таблице 4.5.

Таблица 4.5 – Заинтересованные стороны проекта

Заинтересованные стороны	Ожидания
Потенциальные пользователи	Увеличение точности анализа тональности текста
Научный руководитель, студент	Выполненная выпускная квалификационная работа

Цели и результат проекта представлены в таблице 4.6.

Таблица 4.6 – Заинтересованные стороны проекта

Цели проекта:	<ul style="list-style-type: none"> • Изучить предметную область. • Разработать алгоритмы • Провести тестирование работы алгоритма.
Ожидаемые результаты проекта:	Сдана выпускная квалификационная работа.
Критерии приемки результата проекта:	Успешное тестирование функционала в соответствии с функциональным требованием.
Требования к результату проекта	Выполнены все пункты функционального требования

4.2 Планирование научно-исследовательских работ

4.2.1 Структура научно-исследовательских работ

Для выполнения научных исследований сформирована рабочая группа, в состав которой входят научный руководитель и исполнитель.

Для оптимизации работ удобно использовать классический метод линейного планирования и управления. Результатом такого планирования является составление линейного графика выполнения всех работ. Расчет параметров линейного графика требует определения продолжительности работ. При отсутствии нормативов времени на проведение отдельных видов работ используются вероятностные оценки. Порядок составления этапов и работ приведен в таблице 4.7.

Таблица 4.7 – Перечень работ и исполнителей при разработке алгоритма

№	Содержание работ	Исполнитель
1	Выбор научного руководителя работы	Балаев Даниил

		Бахтиярович
2	Составление и утверждение темы работы	Савельев Алексей Олегович, Балаев Даниил Бахтиярович
3	Составление календарного плана-графика выполнения работы	Савельев Алексей Олегович, Балаев Даниил Бахтиярович
4	Подбор и изучение литературы по теме работы	Балаев Даниил Бахтиярович
5	Анализ предметной области	Балаев Даниил Бахтиярович
6	Разработка программных средств	Балаев Даниил Бахтиярович
7	Тестирование работы алгоритма	Балаев Даниил Бахтиярович
8	Согласование выполненной работы с научным руководителем	Савельев Алексей Олегович
9	Выполнение других частей работы (финансовый менеджмент, социальная	Балаев Даниил Бахтиярович
10	Подведение итогов, оформление работы	Савельев Алексей Олегович, Балаев Даниил Бахтиярович

4.2.2 План проекта

Календарный план–график (диаграмма Ганта) проекта представлен в таблице 4.8.

Из таблицы видно, что руководитель потратит на проект около 45 дней, а студент – 120 дней.

Таблица 4.8 – Календарный план–график проведения НИОКР по теме

Вид работ	Исполнители	Тк, кал. дн.	Продолжительности выполнения работ, неделя											
			Февраль			март			апрель			май		
			1	2	3	1	2	3	1	2	3	1	2	
1. Выбор научного руководителя работы	И.	2,93	■											
2. Составление и утверждение темы работы	Н.р. + И.	8,78	■	■										
3. Составление календарного плана-графика выполнения работы	Н.р. + И.	10,00			■									
4. Подбор и изучение литературы по теме работы	И.	4,64				■								
5. Анализ предметной области	И.	8,05				■	■							
6. Разработка программных средств	И.	9,52					■	■	■					
7. Тестирование работы алгоритма	И.	8,78								■				
8. Согласование выполненной работы с научным руководителем	Н.р. + И.	5,00									■			
9. Выполнение других частей работы	И.	14,88										■		
10. Подведение итогов, оформление работы	Н.р. + И.	11,22											■	■

Научный руководитель (Н.р.) –



Исполнитель (И.) –



4.2.3 Бюджет научно-технического исследования

В процессе формирования бюджета научно-технического исследования используется следующая группировка затрат по статьям:

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

4.2.3.1 Расчет материальных затрат на научное исследование

В материальные затраты научного исследования вошли затраты на канцелярские принадлежности в сумме 1000 рублей (таблица 4.9).

Таблица 4.9 – Расчет затрат на по статье “Материальные затраты на научное исследование”

Наименование	Марка, размер	Кол-во	Цена за единицу, руб	Сумма, руб
Канцелярский набор	ErichKrause	1	1000	1000
Всего за материалы				1000
Транспортно-заготовительные расходы (3-5%)				50
Итого по статье С _м				1050

Таким образом, общая сумма материальных затрат составляет 1050 рублей.

4.2.3.2 Специальное оборудование для научных работ

Поскольку для данной работы было использовано бесплатное программное обеспечение, затраты на оборудование включают в себя только затраты на оборудование студента.

Во время проведения научного исследования использовался ПК стоимостью 132000 рублей. Расчёт затрат на амортизацию представлен в таблице 4.10.

Таблица 4.10 – Расчет затрат на 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	Персональный компьютер	1	132000	132000
Итого				132000

Таким образом, общая сумма затрат на специальное оборудование для научных работ составляет 132000 рублей.

4.2.3.3 Основная заработная плата

Статья включает основную заработную плату научных и инженерно-технических работников, участвующих в выполнении работ по данной теме.

В таблице 4.11 приведён баланс рабочего времени для участников разработки.

Таблица 4.11 – Баланс рабочего времени участников разработки

Показатели рабочего времени	Руководитель	Инженер
Календарное число дней	365	365
Количество нерабочих дней	104	104
– выходные дни	14	14
– праздничные дни		
Потери рабочего времени	24	24
– отпуск	–	–
– невыходы по болезни		
Действительный годовой фонд рабочего времени	219	219

Расчёт основной заработной платы исполнителей представлен в таблице 4.12.

Таблица 4.12 – Расчет основной заработной платы

Исполнители	Z_b , руб.	$k_{пр}$	k_d	k_p	Z_m , руб.	$Z_{дн}$, руб.	T_p раб. дн.	$Z_{осн}$, руб.
Руководитель	30000	0,3	0,35	1,3	64350	3290,96	45	148093,2
Исполнитель	20000	0,3	0,35	1,3	42900	2193,97	120	263276,4
Итого по статье								411369,6

4.2.3.4 Дополнительная заработная плата научно-производственного персонала

В данную статью включается сумма выплат, предусмотренных законодательством о труде. Дополнительная заработная плата рассчитывается исходя из 10-15% от основной заработной платы. Заработная плата исполнителей НТИ представлена в таблице 4.13.

Таблица 4.13 –Заработная плата исполнителей НТИ, руб.

Заработная плата	Руководитель	Исполнитель
Основная зарплата	263276,4	148093,2
Дополнительная зарплата	26327,64	14809,32
Зарплата исполнителя	289604,04	162902,52
Итого по статье С _{зп}	452506,56	

4.2.3.5 Отчисления на социальные нужды

Статья включает в себя отчисления во внебюджетные фонды. Коэффициент отчислений на уплату во внебюджетные фонды составил 0,3. Размер отчислений на социальные нужды представлены в таблице 4.14.

Таблица 4.14 – Отчисления на социальные нужды, руб.

Заработная плата	Руководитель	Исполнитель
Зарплата исполнителя	289604,04	162902,52
Отчисления на социальные нужды	86881,21	48870,76
Итого	135751,97	

4.2.3.6 Накладные расходы

В эту статью включаются затраты на управление и хозяйственное обслуживание. Накладные расходы составляют 80-100% от суммы основной и дополнительной заработной платы. Расчёт накладных расходов представлен в таблице 4.15.

Таблица 4.15 – Накладные расходы, руб.

Заработная плата	Руководитель	Исполнитель
Зарплата исполнителя	289604,04	162902,52
Накладные расходы	231683,23	130322,02
Итого	362005,25	

4.2.3.7 Формирование бюджета затрат научно-исследовательского проекта

Рассчитанная величина затрат научно-исследовательской работы является основой для формирования бюджета затрат проекта, который при формировании договора с заказчиком защищается научной организацией в качестве нижнего предела затрат на разработку научно-технической продукции. Все категории трат сведены в таблице 4.16.

Таблица 4.16 – Бюджет затрат НТИ

Наименование статьи	Сумма, руб.
1. Материальные затраты	1050
2. Специальное оборудование	132000
3. Основная заработная плата	411369,6
4. Дополнительная заработная плата	452506,56
5. Отчисления на социальные нужды	135751,97
6. Накладные расходы	362005,25

Наименование статьи	Сумма, руб.
7. Общий бюджет	1494683,38

В результате было вычислено, что бюджет на разработку НТИ составит 1494683,38 рублей.

4.3 Оценка сравнительной эффективности исследования

Определение эффективности происходит на основе расчета интегрального показателя эффективности научного исследования. Его нахождение связано с определением двух средневзвешенных величин: финансовой эффективности и ресурсоэффективности.

4.3.1 Интегральный показатель финансовой эффективности

Интегральный показатель финансовой эффективности научного исследования получают в ходе оценки бюджета затрат трёх (или более) вариантов исполнения исследования. Для этого наибольший интегральный показатель принимается за базу расчёта, с которым соотносятся финансовые значения по всем вариантам исполнения.

Расчёт интегрального показателя финансовой эффективности приведен ниже.

Интегральный финансовый показатель разработки определяется как:

$$I_{\phi}^p = \frac{\Phi_{pi}}{\Phi_{max}} \quad (4.1)$$

где I_{ϕ}^p – интегральный финансовый показатель разработки;

Φ_{pi} – стоимость i -го варианта исполнения;

Φ_{max} – максимальная стоимость исполнения научно-исследовательского проекта.

$$I_{\phi}^p = \frac{1494683,38}{1494683,38} = 1$$

$$I_{\phi}^a = \frac{1394683,38}{1494683,38} = 0,93$$

$$I_{\phi}^b = \frac{1480683,38}{1494683,38} = 0,99$$

$$I_{\phi}^c = \frac{1434683,38}{1494683,38} = 0,96$$

Полученные величины интегрального финансового показателя эффективности разработки говорят о том, что альтернативные варианты исполнения исследования позволят сократить бюджет.

4.3.2 Интегральный показатель ресурсоэффективности

Расчёт интегрального показателя ресурсоэффективности приведен в таблице 4.17.

Таблица 4.17 – Сравнение характеристик вариантов исполнения проекта

Критерии \ ПО	Вес	Текущий проект	Аналог 1	Аналог 1	Аналог 1
Удобство использования	0,25	5	4	5	4
Функциональность	0,25	4	4	5	3
Срок разработки	0,2	5	3	2	4
Точность	0,3	4	4	5	3
ИТОГО	1	4,45	3,80	4,40	3,45

$$I_p = 0,25 \cdot 5 + 0,25 \cdot 4 + 0,2 \cdot 5 + 0,3 \cdot 4 = 4,45$$

$$I_{p2} = 0,25 \cdot 4 + 0,25 \cdot 4 + 0,2 \cdot 3 + 0,3 \cdot 4 = 3,80$$

$$I_{p3} = 0,25 \cdot 5 + 0,25 \cdot 5 + 0,2 \cdot 2 + 0,3 \cdot 5 = 4,40$$

$$I_{p4} = 0,25 \cdot 4 + 0,25 \cdot 3 + 0,2 \cdot 4 + 0,3 \cdot 3 = 3,45$$

4.3.3 Интегральный показатель эффективности

Интегральный показатель эффективности вариантов исполнения разработки:

$$I_{\text{исп.}i} = \frac{I_{p1}}{I_{\text{финр}}^i} \quad (4.2)$$

$$I_{\text{исп.}1} = \frac{4,45}{1} = 4,5$$

$$I_{\text{исп.}2} = \frac{4,09}{0,93} = 45,5$$

$$I_{\text{исп.}3} = \frac{4,40}{0,99} = 4,5$$

$$I_{\text{исп.}4} = \frac{3,45}{0,96} = 45,5$$

С целью определения наиболее целесообразного варианта из предложенных сравним интегральные показатели эффективности вариантов исполнения разработки и определим сравнительную эффективность проекта:

$$\mathcal{E}_{\text{ср}} = \frac{I_{\text{исп.}2}}{I_{\text{исп.}1}} \quad (4.3)$$

$$\mathcal{E}_{\text{ср}1} = \frac{I_{\text{исп.}2}}{I_{\text{исп.}1}} = \frac{4,09}{4,45} = 0,92$$

$$\mathcal{E}_{\text{ср}2} = \frac{I_{\text{исп.}3}}{I_{\text{исп.}1}} = \frac{4,44}{4,45} = 0,99$$

$$\mathcal{E}_{\text{ср}3} = \frac{I_{\text{исп.}4}}{I_{\text{исп.}1}} = \frac{3,59}{4,45} = 0,81$$

Сравнение интегрального показателя эффективности текущего проекта и аналогов представлено в таблице 4.18.

Таблица 4.18 – Сравнительная эффективность разработки

№	Показатели	Разработка	Аналог 1	Аналог 2	Аналог 3
1	Интегральный финансовый показатель разработки	1	0,93	0,99	0,96
2	Интегральный показатель ресурсоэффективности и разработки	4,45	3,80	4,40	3,45
3	Интегральный показатель эффективности	4,45	4,09	4,44	3,59
4	Сравнительная эффективность вариантов исполнения	1	0,92	0,99	0,81

Таким образом, исходя из результатов сравнения значений интегральных показателей эффективности, можно сказать, что выбранный вариант исполнения технической задачи является наиболее приемлемым с позиции финансовой и ресурсной эффективности.

4.4 Вывод

В ходе выполнения данной дипломной работы был составлен календарный план по НИР, рассчитана смета затрат на НИР, общая сумма затрат на НИР равна 1494683,38, время, необходимое для его выполнения 120 календарных дня. Произведена сравнительная характеристика эффективности разработки на основании интегрального показателя эффективности. Полученные данные приведены в таблице 4.18. Сравнив значения интегральных показателей эффективности, можно сделать вывод, что реализация НИР в первом исполнении является более эффективным вариантом решения задачи, поставленной в данной работе с позиции финансовой и ресурсной эффективности.

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»**

Студенту:

Группа	ФИО
8ПМ11	Балаев Даниил Бахтиярович

Школа	ИШИТР	Отделение (НОЦ)	ОИТ
Уровень образования	Магистратура	Направление/специальность	09.04.04 «Программная инженерия»

Тема магистерской диссертации:

Методы, алгоритмы и программное обеспечение обнаружения сленга в онлайн-сообществах социальных медиа

Исходные данные к разделу «Социальная ответственность»:

1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Исследование и разработка алгоритмов для обнаружения сленга в онлайн-сообществах социальных медиа
--	---

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

<p>5.1. Производственная безопасность</p> <p>5.1.1. Анализ выявленных вредных факторов:</p> <ul style="list-style-type: none"> • Природа воздействия • Действие на организм человека • Нормы воздействия и нормативные документы (для вредных факторов) • СИЗ коллективные и индивидуальные <p>5.1.2. Анализ выявленных опасных факторов:</p> <ul style="list-style-type: none"> • Термические источники опасности • Электробезопасность • Пожаробезопасности 	<p>1. Вредные факторы:</p> <p>1.1. Нарушения микроклимата, оптимальные и допустимые параметры;</p> <p>1.2. Шум, ПДУ, СКЗ, СИЗ;</p> <p>1.3. Повышенный уровень электромагнитного излучения, ПДУ, СКЗ, СИЗ;</p> <p>1.4. Недостаточная освещенность. Проведен расчет освещения рабочего места;</p> <p>2. Опасные факторы:</p> <p>2.1. Электроопасность; класс электроопасности помещения, безопасные номиналы I, U, R_{заземления}, СКЗ, СИЗ;</p> <p>2.2. Пожароопасность, категория пожароопасности помещения, марки огнетушителей, их назначение и ограничение применения; Приведена схема эвакуации.</p>
<p>5.2. Экологическая безопасность:</p> <ul style="list-style-type: none"> • Выбросы в окружающую среду • Решения по обеспечению экологической безопасности 	<p>Наличие промышленных отходов (вторцвет- и чермет, пластмасса, перегоревшие люминесцентные лампы, оргтехника) и способы их утилизации;</p>

<p>5.3. Безопасность в чрезвычайных ситуациях:</p> <ol style="list-style-type: none"> 1. перечень возможных ЧС при разработке и эксплуатации проектируемого решения; 2. разработка превентивных мер по предупреждению ЧС; 3. разработка действий в результате возникшей ЧС и мер по ликвидации её последствий. 	<p>Рассмотрены 2 ситуации ЧС:</p> <ol style="list-style-type: none"> 1) природная – сильные морозы зимой, (аварии на электро-, тепло-коммуникациях, водоканале, транспорте); 2) техногенная – несанкционированное проникновение посторонних на рабочее место (возможны проявления вандализма, диверсии, промышленного шпионажа), представлены мероприятия по обеспечению устойчивой работы производства в том и другом случае.
<p>4. Перечень нормативно-технической документации.</p>	<p>– ГОСТы, СанПиНы, СНИПы</p>

Дата выдачи задания для раздела по линейному графику	16.02.2023
--	------------

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ТПУ	Федорчук Ю.М.	Д.Т.Н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ПМ11	Балаев Даниил Бахтиярович		

СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ

Введение:

Социальная ответственность – ответственность отдельного ученого и научного сообщества перед обществом. Первостепенное значение при этом имеет безопасность применения технологий, которые создаются на основе достижений науки, предотвращение или минимизация возможных негативных последствий их применения, обеспечение безопасного как для испытуемых, как и для окружающей среды проведения исследований.

В ходе данной работы разработка и исследование высокоэффективного источника питания для телекоммуникационного оборудования. Работа выполнялась в лаборатории КЦ ТПУ. Все работы выполнялись с использованием компьютера. Раздел также включает в себя оценку условий труда на рабочем месте, анализ вредных и опасных факторов труда, разработку мер защиты от них.

5.1 Производственная безопасность

5.1.1 Вредные факторы

5.1.1.1 Отклонение показателей микроклимата в помещении

Проанализируем микроклимат в помещении, где находится рабочее место. Микроклимат производственных помещений определяют следующие параметры: температура, относительная влажность, скорость движения воздуха. Эти факторы влияют на организм человека, определяя его самочувствие.

Оптимальные и допустимые значения параметров микроклимата приведены в таблице 5.1 и 5.2.

Таблица 5.1 – Оптимальные нормы микроклимата

Период года	Температура воздуха, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	23–25	40	0,1
Тёплый	23–25		0,1

Таблица 5.2 – Допустимые нормы микроклимата

Период года	Температура воздуха, °С		Относительная влажность воздуха, %	Скорость движения воздуха, м/с
	Нижняя допустимая граница	Верхняя допустимая граница		
Холодный	15	24	20–80	<0,5
Тёплый	22	28	20–80	<0,5

Общая площадь рабочего помещения составляет 42 м², объем составляет 147 м³. По СанПиН 2.2.2/2.4.1340-03 санитарные нормы составляют 6,5 м² и 20 м³ объема на одного человека. Исходя из приведенных выше данных, можно сказать, что количество рабочих мест соответствует размерам помещения по санитарным нормам.

После анализа габаритных размеров рассмотрим микроклимат в этой комнате. В качестве параметров микроклимата рассмотрим температуру, влажность воздуха, скорость ветра.

В помещении осуществляется естественная вентиляция посредством наличия легко открываемого оконного проема (форточки), а также дверного

проема. По зоне действия такая вентиляция является общеобменной. Основным недостаток – приточный воздух поступает в помещение без предварительной очистки и нагревания. Согласно нормам СанПиН 2.2.2/2.4.1340-03 объем воздуха необходимый на одного человека в помещении без дополнительной вентиляции должен быть более 40м³ [20]. В нашем случае объем воздуха на одного человека составляет 42 м³, из этого следует, что дополнительная вентиляция не требуется. Параметры микроклимата поддерживаются в холодное время года за счет систем водяного отопления с нагревом воды до 100°С, а в теплое время года – за счет кондиционирования, с параметрами согласно [21]. Нормируемые параметры микроклимата, ионного состава воздуха, содержания вредных веществ должны соответствовать требованиям [22].

5.1.1.2 Превышение уровней шума

Одним из наиболее распространенных в производстве вредных факторов является шум. Он создается вентиляционным и рабочим оборудованием, преобразователями напряжения, рабочими лампами дневного света, а также проникает снаружи. Шум вызывает головную боль, усталость, бессонницу или сонливость, ослабляет внимание, память ухудшается, реакция уменьшается.

Основным источником шума в комнате являются компьютерные охлаждающие вентиляторы. Уровень шума варьируется от 35 до 42 дБА. Согласно СанПиН 2.2.2 / 2.4.1340-03, при выполнении основных работ на ПЭВМ уровень шума на рабочем месте не должен превышать 80 дБА [23].

При значениях выше допустимого уровня необходимо предусмотреть средства индивидуальной защиты (СИЗ) и средства коллективной защиты (СКЗ) от шума.

Средства коллективной защиты:

1. устранение причин шума или существенное его ослабление в

источнике образования;

2. изоляция источников шума от окружающей среды (применение глушителей, экранов, звукопоглощающих строительных материалов, например любой пористый материал – шамотный кирпич, микропористая резина, поролон и др.);

3. применение средств, снижающих шум и вибрацию на пути их распространения;

Средства индивидуальной защиты;

1. применение спецодежды и защитных средств органов слуха: наушники, беруши, антифоны.

5.1.1.3 Повышенный уровень электромагнитных излучений

Источником электромагнитных излучений в нашем случае являются дисплеи ПЭВМ. Монитор компьютера включает в себя излучения рентгеновской, ультрафиолетовой и инфракрасной области, а также широкий диапазон электромагнитных волн других частот. Согласно СанПиН 2.2.2/2.4.1340-03 напряженность электромагнитного поля по электрической составляющей на расстоянии 50 см вокруг ВДТ не должна превышать 25 В/м в диапазоне от 5 Гц до 2 кГц, 2,5 В/м в диапазоне от 2 до 400 кГц [20]. Плотность магнитного потока не должна превышать в диапазоне от 5 Гц до 2 кГц 250 нТл, и 25 нТл в диапазоне от 2 до 400 кГц. Поверхностный электростатический потенциал не должен превышать 500В [20]. В ходе работы использовалась ПЭВМ типа Acer VN7-791 со следующими характеристиками: напряженность электромагнитного поля 2,5 В/м; поверхностный потенциал составляет 450 В (основы противопожарной защиты предприятий ГОСТ 12.1.004 и ГОСТ 12.1.010 – 76.) [24].

При длительном постоянном воздействии электромагнитного поля (ЭМП) радиочастотного диапазона при работе на ПЭВМ у человеческого организма сердечно-сосудистые, респираторные и нервные расстройства,

головные боли, усталость, ухудшение состояния здоровья, гипотония, изменения сердечной мышцы проводимости. Тепловой эффект ЭМП характеризуется увеличением температуры тела, локальным селективным нагревом тканей, органов, клеток за счет перехода ЭМП на теплую энергию.

Предельно допустимые уровни (ПДУ) облучения (по ОСТ 54 30013-83):

- а) до 10 мкВт/см², время работы (8 часов);
- б) от 10 до 100 мкВт/см², время работы не более 2 часов;
- в) от 100 до 1000 мкВт/см², время работы не более 20 мин. при условии пользования защитными очками;
- г) для населения в целом ППМ не должен превышать 1 мкВт/см².

Защита человека от опасного воздействия электромагнитного излучения осуществляется следующими способами:

СКЗ

1. защита временем;
2. защита расстоянием;
3. снижение интенсивности излучения непосредственно в самом источнике излучения;
4. заземление экрана вокруг источника;
5. защита рабочего места от излучения;

СИЗ

1. Очки и специальная одежда, выполненная из металлизированной ткани (кольчуга). При этом следует отметить, что использование СИЗ возможно при кратковременных работах и является мерой аварийного характера. Ежедневная защита обслуживающего персонала должна обеспечиваться другими средствами.
2. Вместо обычных стекол используют стекла, покрытые тонким слоем золота или диоксида олова (SnO₂).

5.1.1.4 Недостаточная освещенность

Для обеспечения требуемой освещенности необходимо использовать совмещенное освещение, создаваемое сочетанием естественного и искусственного освещения. При данном этапе развития осветительной техники целесообразно использовать люминесцентные лампы, которые по сравнению с лампами накаливания имеют большую светоотдачу на ватт потребляемой мощности и более естественный спектр.

Минимальный уровень средней освещенности на рабочих местах с постоянным пребыванием людей должен быть не менее 200 лк.

В расчётном задании должны быть решены следующие вопросы:

- выбор системы освещения;
- выбор источников света;
- выбор светильников и их размещение;
- выбор нормируемой освещённости;
- расчёт освещения методом светового потока.

В данном расчётном задании для всех помещений рассчитывается общее равномерное освещение.

Таблица 5.3 – Параметры помещения

Параметр	Обозначение	Значение, м
Длина	A	12
Ширина	B	10
Высота помещения	H	3,5

Расчёт общего равномерного искусственного освещения горизонтальной рабочей поверхности выполняется методом коэффициента светового потока, учитывающим световой поток, отражённый от потолка и стен.

Световой поток лампы определяется по формуле:

$$\Phi_{\text{рас}} = E_{\text{н}} * S * K_{\text{з}} * Z / N * \eta \quad (5.1)$$

Где $E_{\text{н}}$ – нормируемая минимальная освещённость по СНиП 23-05- 95, лк; S – площадь освещаемого помещения, м²; $K_{\text{з}}$ – 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 $E_{\text{ср}}/E_{\text{min}}$. Для люминесцентных ламп при расчётах берётся равным 1,1; N – число ламп в помещении; η – коэффициент использования светового потока.

Коэффициент использования светового потока показывает, какая часть светового потока ламп попадает на рабочую поверхность. Он зависит от индекса помещения i , типа светильника, высоты светильников над рабочей поверхностью h и коэффициентов отражения стен $\rho_{\text{с}}$ и потолка $\rho_{\text{п}}$.

Индекс помещения определяется по формуле:

$$i = S / h(A + B) \quad (5.2)$$

Проведем расчет индекса помещения:

Площадь помещения:

$$S = A * B = 12 * 10 = 120 \text{ м}^2$$

Индекс:

$$i = \frac{S}{h * (A + B)} = \frac{120}{2,35 * (12 + 10)} = 2,32$$

Согласно этим данным, коэффициент использования светового потока будет равен 56 % или в долях = 0,56.

Коэффициенты отражения оцениваются субъективно (табл. 4.10) [БЖД Практикум 2009-2020].

Согласно указанной методике, выбираем тип источника света.

Наиболее подходящим вариантом является 40 ваттная лампа ЛБ, у которой $\Phi = 2800$ лм. Для выбранного типа лампы подходит светильник ОД-2-40 с размерами: длина = 1230 мм, ширина = 266 мм.

Из уравнения 1.5.1 находим количество ламп для помещения

$$N = E_n * S * K_z * Z / \Phi * \eta = 200 * 120 * 1,3 * 1,1 / 2800 * 0,56 = 21,875;$$

Принимаем $N=24$ лампы или 12 светильников.

Размещаем светильники в 3 ряда по 4 светильника в ряду с соблюдением условий:

L – расстояние между соседними светильниками или рядами (если по длине (А) и ширине (В) помещения расстояния различны, то они обозначаются L_A и L_B),

l – расстояние от крайних светильников или рядов до стены.

Оптимальное расстояние l от крайнего ряда светильников до стены рекомендуется принимать равным $L/3$.

Сначала определим световой поток расчетный.

$$\Phi = E_n * S * K_z * Z / \eta = 200 * 120 * 1,3 * 1,1 / 24 * 0,56 = 2554 \text{ лм};$$

Проведем проверку выполнения условия соответствия:

$$- 10\% \leq ((\Phi_{\text{расч}} - \Phi_{\text{станд}}) / \Phi_{\text{расч}}) * 100\% \leq + 20\%$$

Подставляя численные значения получаем:

$$- 10\% \leq (2800 - 2554) / 2554 * 100\% \leq + 20\%$$

$$- 10\% \leq +9,6\% \leq + 20\%$$

Результат расчета укладывается в допустимые пределы.

Определим мощность осветительной установки:

$$P = N * P_i = 24 * 40 \text{ Вт} = 960 \text{ Вт}.$$

Теперь определим расстояния между светильниками по длине и ширине помещения.

$$12000 = 3 * L_A + 4 * 1230 + 2/3 * L_A; L_A = (12000 - 4920) * 3/11 = 1930 \text{ мм};$$

$$L_A / 3 = 644 \text{ мм};$$

$$10000 = 2 * L_B + 3 * 266 + 2/3 * L_B; L_B = (10000 - 798) * 3/8 = 3450 \text{ мм};$$

$$L_B / 3 = 1150 \text{ мм}.$$

Рисуем схему размещения светильников на потолке для обеспечения общего равномерного освещения.

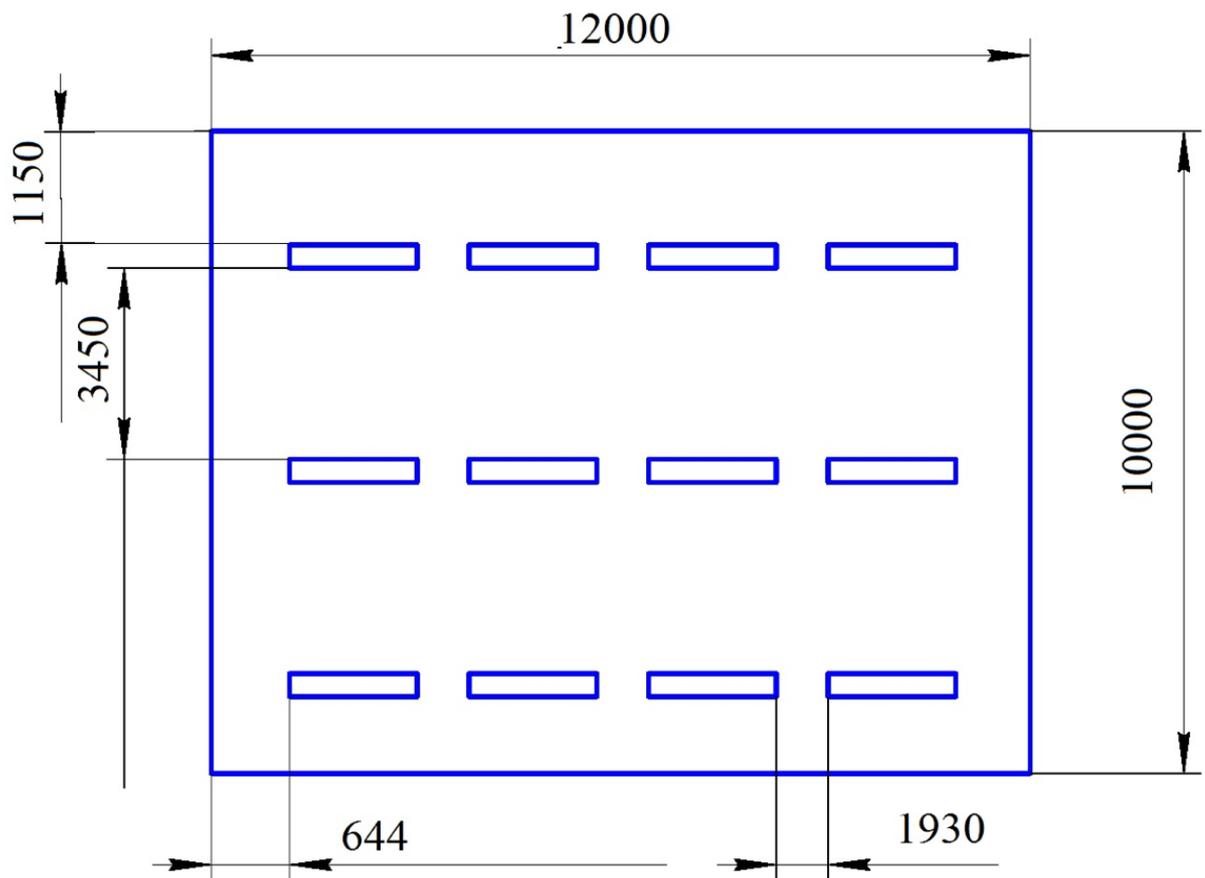


Рисунок 5.1 – План размещения светильников на потолке

Проведем проверку выполнения условия соответствия:

$$-10\% \leq \frac{\Phi_{\text{л.станд}} - \Phi_{\text{л.расч}}}{\Phi_{\text{л.станд}}} * 100\% \leq +20\%$$

Подставляя численные значения получаем:

$$\begin{aligned} -10\% &\leq (2800 - 2554)/2554 * 100\% \leq +20\% \\ -10\% &\leq 9,6\% \leq +20\% \end{aligned}$$

Результат расчета укладывается в поле допуска.

Определим мощность осветительной установки:

$$P = N_i * P_i = 40 * 80 = 3200 \text{ Вт}$$

5.1.2 Опасные факторы

5.1.2.1 Электроопасность; класс электроопасности помещения, безопасные номиналы I, U, R_{зазем}, СКЗ, СИЗ

Поражение электрическим током

К опасным факторам можно отнести наличие в помещении большого количества аппаратуры, использующей однофазный электрический ток напряжением 220 В и частотой 50Гц. По опасности электропоражения комната относится к помещениям без повышенной опасности, так как отсутствует повышенная влажность, высокая температура, токопроводящая пыль и возможность одновременного сприкосновения токоведущих элементов с заземленными металлическими корпусами оборудования [25].

Лаборатория относится к помещению без повышенной опасности поражения электрическим током. Безопасными номиналами являются: $I < 0,1$ А; $U < (2-36)$ В; $R_{\text{зазем}} < 4$ Ом.

Для защиты от поражения электрическим током используют СИЗ и СКЗ.

Средства коллективной защиты:

1. – защитное заземление, зануление;
2. – малое напряжение;
3. – электрическое разделение сетей;
4. – защитное отключение;
5. – изоляция токоведущих частей;
6. – оградительные устройства.
7. Использование щитов, барьеров, клеток, ширм, а также заземляющих и шунтирующих штанг, специальных знаков и плакатов.

Средства индивидуальной защиты:

1. Использование диэлектрических перчаток, изолирующих клещей и штанг, слесарных инструментов с изолированными рукоятками, указатели величины напряжения, калоши, боты, подставки и коврики.

5.1.2.2 Пожароопасность, категория пожароопасности помещения, марки огнетушителей, их назначение и ограничение применения

По взрывопожарной и пожарной опасности помещения подразделяются на категории А, Б, В1-В4, Г и Д.

Согласно НПБ 105-03 лаборатория относится к категории В– горючие и трудно горючие жидкости, твердые горючие и трудно горючие вещества и материалы, вещества и материалы, способные при взаимодействии с водой, кислородом воздуха или друг с другом только гореть, при условии, что помещения, в которых находится, не относятся к категории наиболее опасных А или Б.

По степени огнестойкости данное помещение относится к 1-й степени огнестойкости по СНИП 2.01.02-85 (выполнено из кирпича, которое относится к трудностгораемым материалам).

Возникновение пожара при работе с электронной аппаратурой может быть по причинам как электрического, так и неэлектрического характера.

Причины возникновения пожара неэлектрического характера:

а) халатное неосторожное обращение с огнем (курение, оставленные без присмотра нагревательные приборы, использование открытого огня);

Причины возникновения пожара электрического характера: короткое замыкание, перегрузки по току, искрение и электрические дуги, статическое электричество и т. п.

Для локализации или ликвидации загорания на начальной стадии используются первичные средства пожаротушения. Первичные средства пожаротушения обычно применяют до прибытия пожарной команды.

Огнетушители водо-пенные (ОХВП-10) используют для тушения очагов пожара без наличия электроэнергии. Углекислотные (ОУ-2) и порошковые огнетушители предназначены для тушения электроустановок, находящихся под напряжением до 1000 В. Для тушения токоведущих частей и электроустановок применяется переносной порошковый огнетушитель, например ОП-5.

В общественных зданиях и сооружениях на каждом этаже должно размещаться не менее двух переносных огнетушителей. Огнетушители следует располагать на видных местах вблизи от выходов из помещений на высоте не более 1,35 м. Размещение первичных средств пожаротушения в коридорах, переходах не должно препятствовать безопасной эвакуации людей.

Для предупреждения пожара и взрыва необходимо предусмотреть:

1. специальные изолированные помещения для хранения и разлива легковоспламеняющихся жидкостей (ЛВЖ), оборудованные приточно-вытяжной вентиляцией во взрывобезопасном исполнении - соответствии с ГОСТ 12.4.021-75 и СНиП 2.04.05-86;

2. специальные помещения (для хранения в таре пылеобразной канифоли), изолированные от нагревательных приборов и нагретых частей оборудования;

3. первичные средства пожаротушения на производственных участках (передвижные углекислые огнетушители ГОСТ 9230-77, пенные огнетушители ТУ 22-4720-80, ящики с песком, войлок, кошма или асбестовое полотно);

4. автоматические сигнализаторы (типа СВК-3 М 1) для сигнализации о присутствии в воздухе помещений предвзрывных концентраций горючих паров растворителей и их смесей.

Лаборатория полностью соответствует требованиям пожарной безопасности, а именно, наличие охранно-пожарной сигнализации, плана эвакуации, изображенного на рисунке 5.2, порошковых огнетушителей с

поверенным клеймом, табличек с указанием направления к запасному (эвакуационному) выходу.



Рисунок 5.2 – План эвакуации

5.2 Экологическая безопасность

В компьютерах огромное количество компонентов, которые содержат токсичные вещества и представляют угрозу, как для человека, так и для окружающей среды.

К таким веществам относятся:

- свинец (накапливается в организме, поражая почки, нервную систему);
- ртуть (поражает мозг и нервную систему);
- никель и цинк (могут вызывать дерматит);
- щелочи (прожигают слизистые оболочки и кожу);

Поэтому компьютер требует специальных комплексных методов утилизации.

Таким образом утилизацию компьютера можно провести следующим образом:

- отделить металлические детали от неметаллов;
- разделить углеродистые металлы от цветмета;
- пластмассовые изделия (крупногабаритные) измельчить для уменьшения объема;
- копир-порошок упаковать в отдельную упаковку, точно также, как и все проклассифицированные и измельченные компоненты оргтехники, и после накопления на складе транспортных количеств отправить предприятиям и фирмам, специализирующимся по переработке отдельных видов материалов.

Люминесцентные лампы утилизируют следующим образом. Не работающие лампы немедленно после удаления из светильника должны быть упакованы в картонную коробку, бумагу или тонкий мягкий картон, предохраняющий лампы от взаимного соприкосновения и случайного механического повреждения. После накопления ламп объемом в 1 транспортную единицу их сдают на переработку на соответствующее предприятие. Недопустимо выбрасывать отработанные энергосберегающие лампы вместе с обычным мусором, превращая его в ртутьсодержащие отходы, которые загрязняют ртутными парами

5.3 Безопасность в чрезвычайных ситуациях

Природная чрезвычайная ситуация – обстановка на определенной территории или акватории, сложившейся в результате возникновения источника природной чрезвычайной ситуации, который может повлечь или повлек за собой человеческие жертвы, ущерб здоровью людей и (или)

окружающей природной среде, значительные материальные потери и нарушение условий жизнедеятельности людей.

Производство находится в городе Томске с континентально-циклоническим климатом. Природные явления (землетрясения, наводнения, засухи, ураганы и т. д.), в данном городе отсутствуют.

Возможными ЧС на объекте в данном случае, могут быть сильные морозы и диверсия.

Для Сибири в зимнее время года характерны морозы. Достижение критически низких температур приводит к авариям систем тепло- и водоснабжения, сантехнических коммуникаций и электроснабжения, приостановке работы. В этом случае при подготовке к зиме следует предусмотреть а) газобаллонные калориферы (запасные обогреватели), б) дизель или бензоэлектрогенераторы; в) запасы питьевой и технической воды на складе (не менее 30 л на 1 человека); г) теплый транспорт для доставки работников на работу и с работы домой в случае отказа муниципального транспорта. Их количества и мощности должно хватать для того, чтобы работа на производстве не прекратилась.

В лаборатории КЦ ТПУ наиболее вероятно возникновение чрезвычайных ситуаций (ЧС) техногенного характера.

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.

Список источников

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. СН 2.2.4/2.1.8.562-96. Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки.
7. ГОСТ 12.1.030-81. Электробезопасность. Защитное заземление. Зануление.
8. ГОСТ 12.1.004-91. Пожарная безопасность. Общие требования.
9. ГОСТ 12.2.037-78. Техника пожарная. Требования безопасности.
10. СанПиН 2.1.6.1032-01. Гигиенические требования к качеству атмосферного воздуха.
11. ГОСТ 30775-2001 Ресурсосбережение. Обращение с отходами. Классификация, идентификация и кодирование отходов.
12. СНиП 21-01-97. Противопожарные нормы.

СПИСОК ЛИТЕРАТУРЫ

1. Орлова Н. О. Сленг vs жаргон: проблема дефиниции // Ярославский педагогический вестник. – 2004. – №3 (40). – С. 36-39.
2. Shehzaad D., Diptesh K., Pushpak B. SlangNet: A WordNet like resource for English slang // In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC). – Portoroz: 2016. – Pp. 4329-4332.
3. Hemmatian F., Sohrabi M.K. A survey on classification techniques for opinion mining and sentiment analysis // Artificial Intelligence Review. – 2019. – №52. – Pp. 1495-1545.
4. DIGITAL 2023: GLOBAL OVERVIEW REPORT // datareportal URL: <https://datareportal.com/reports/digital-2023-global-overview-report> (Date of application: 27.03.2023).
5. Most popular social networks worldwide as of January 2023, ranked by number of monthly active users // Statista URL: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/> (Date of application: 27.03.2023).
6. Asghar M.Z. Detection and Scoring of Internet Slangs for Sentiment Analysis Using SentiWordNet // Life science journal. – 2014. – №11. – Pp. 66-72.
7. Thelwall M., This! Identifying New Sentiment Slang Through Orthographic Pleonasm Online: Yasss Slay Gorg Queen Ilysm // IEEE Intelligent Systems. – 2021. – №4. – Pp. 114-120.
8. Matsumoto K., Yoshida M., Tsuchiya S., Kita K., Ren F. Slang Analysis Based on Variant Information Extraction Focusing on the Time Series Topics // International Journal of Advanced Intelligence. – 2016. – №1. – Pp. 84-98.
9. Matsumoto K. Slang feature extraction by analysing topic change on social media // CAAI Transactions on Intelligence Technology. – 2019. – №.1. – Pp. 64-71.

10. Pei Z., Sun Z., Xu Y. Slang detection and identification // Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL). – Hong Kong: 2019. – Pp. 881–889.
11. Seki Y., Liu Y. Multi-task Learning Model for Detecting Internet Slang Words with Two-Layer Annotation // 2022 International Conference on Asian Language Processing (IALP). – Singapore: 2022. – Pp. 212-218.
12. Hamid M.A., Tumpa E.S., Polin J.A., Al Nahian J., Rahman A., Mim N.A. Bengali Slang detection using state-of-the-art supervised models from a given text // Bulletin of Electrical Engineering and Informatics. – 2023. – №4. – Pp. 2381-2387.
13. Hossain N., Tran T.T.T., Kautz H. Discovering political slang in readers' comments // Proceedings of the International AAAI Conference on Web and Social Media. – 2018. – №1. – Pp. 612-615.
14. Pal A.R., Saha D. Detection of Slang Words in e-Data using semi-Supervised Learning // International Journal of Artificial Intelligence & Applications (IJAIA). – 2013. – №5. – Pp. 49-61.
15. Christa Hodapp Men's Rights, Gender, and Social Media. Lexington Books, 2017.
16. Гурко А. В. Неоязычество в Беларуси: предпосылки и условия возникновения, организационные формы, перспективы // Неоязычество на просторах Евразии / Сост. сб. В. А. Шнирельман. – М.: Библейско-богословский институт святого апостола Андрея. – 2001. – С. 68-79.
17. Шолле Ф. Глубокое обучение на Python. – СПб.: Питер, 2018. – 400 с.
18. Module: tf // tensorflow URL: https://www.tensorflow.org/api_docs/python/tf (дата обращения: 05.06.2023).
19. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение: учебно-методическое пособие / Н.А. Гаврикова, Л.Р. Тухватулина, И.Г. Видяев, Г.Н. Серикова, Н.В. Шаповалова; Томский политехнический университет. – Томск: Изд-во Томского политехнического университета, 2014. – 73 с.

20. ГОСТ 54 30013–83. Электромагнитные излучения СВЧ. Предельно допустимые уровни облучения. Требования безопасности: дата введения 1984-01-01. – URL: http://online.budstandart.com/ru/catalog/doc-page?id_doc=47453 (дата обращения: 24.05.2023). – Текст: электронный.
21. ГОСТ 12.4.154–85. ССБТ. Устройства, экранирующие для защиты от электрических полей промышленной частоты: дата введения 1986-07-01. – URL: <https://docs.cntd.ru/document/1200005824> (дата обращения: 24.05.2023). – Текст: электронный.
22. ГН 2.2.5.1313–03. Предельно допустимые концентрации (ПДК) вредных веществ в воздухе рабочей зоны: дата введения 1998-02-04. – URL: <https://docs.cntd.ru/document/1200000525> (дата обращения: 24.05.2023). – Текст: электронный.
23. СанПиН 2.2.4/2.1.8.055–96. Электромагнитные излучения радиочастотного диапазона (ЭМИ РЧ): дата введения 1996-05-08. – URL: <https://docs.cntd.ru/document/1200001537> (дата обращения: 24.05.2023). – Текст: электронный.
24. СанПиН 2.2.4.548–96. Гигиенические требования к микроклимату производственных помещений): дата введения 1996-10-01. – URL: <https://docs.cntd.ru/document/901704046> (дата обращения: 24.05.2023). – Текст: электронный.
25. СН 2.2.4/2.1.8.562–96. Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки: дата введения 1996-10-31. – URL: https://minstroy.gov-murman.ru/files/4.40_-_sn-2.2.42.1.8.562_96-shum-na-rabochikh-mestakh_-v-pomeshcheniyakh-zhilykh_-obshchestvennykh-zdaniy-i-na..._tekst.pdf (дата обращения: 24.05.2023). – Текст: электронный.

Приложение А

(справочное)

Analytical overview

Студент

Группа	ФИО	Подпись	Дата
8ПМ11	Балаев Даниил Бахтиярович		

Руководитель ВКР

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

Консультант-лингвист отделения иностранных языков ШБИП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Куркан Наталия Владимировна	к.фил.н.		

1 Analytical overview

1.1 Relevance

Slang is a type of informal language consisting of words and expressions used in certain groups. A distinctive feature of slang is its expressiveness, enabled by the lexical flexibility. The expressive nature of slang illustrates its social function, as it provides an effective way to communicate and share knowledge in groups with different social identities [1].

In [2], the authors distinguish two types of slang:

- words with new meanings: existing words in the lexicon that acquire new slang meanings that are different from their established meanings.
- new words: words that are not in the standard lexicon.

Today, slang is a significant part of the language of Internet users in social networks. As they spread to other media, slang begins to be used in articles, comments and colloquial speech. Slang word meanings often define key semantics in sentences and can play a critical role in natural language processing (NLP) applications. Thus, slang studies can help improve the efficiency of semantic analysis of texts.

One of the most common tasks of semantic text analysis is opinion mining and text sentiment analysis. The main goal of opinion mining is to automate the extraction of user opinions from unstructured texts. It is used to analyze opinions, feedback, people's attitudes towards organizations, people, issues, events, and various topics.

Opinion mining and sentiment analysis of the text are used in various areas of human life. E.g., it can be useful for commercial manufacturers to improve the quality of their products, for marketing companies to analyze market demand, for customers to get a summary of other customer's opinions about a product. Opinion mining can also be used in politics for political elections and in making decisions that should improve the social life of the community as well as

in the financial sector to predict the stock's value or in various sociological studies. [3].

Today, social networks are a significant source of data for solving these tasks, as people use them to express their opinion, while they can find communities to almost any topic. The total percentage of active users of social networks is 59% of the world's population or 4.8 billion people [4], while the 6 most popular social networks have more than 1 billion active users. The list of the most popular social networks is shown in Figure 1.

At the same time, the use of these networks involves great difficulties such as the amount of data, the lack of structure, the problem of confidentiality, the poor quality of data (the problem of spam) and the dynamism of information.

Thus, the only potential solution is to automate natural language processing.

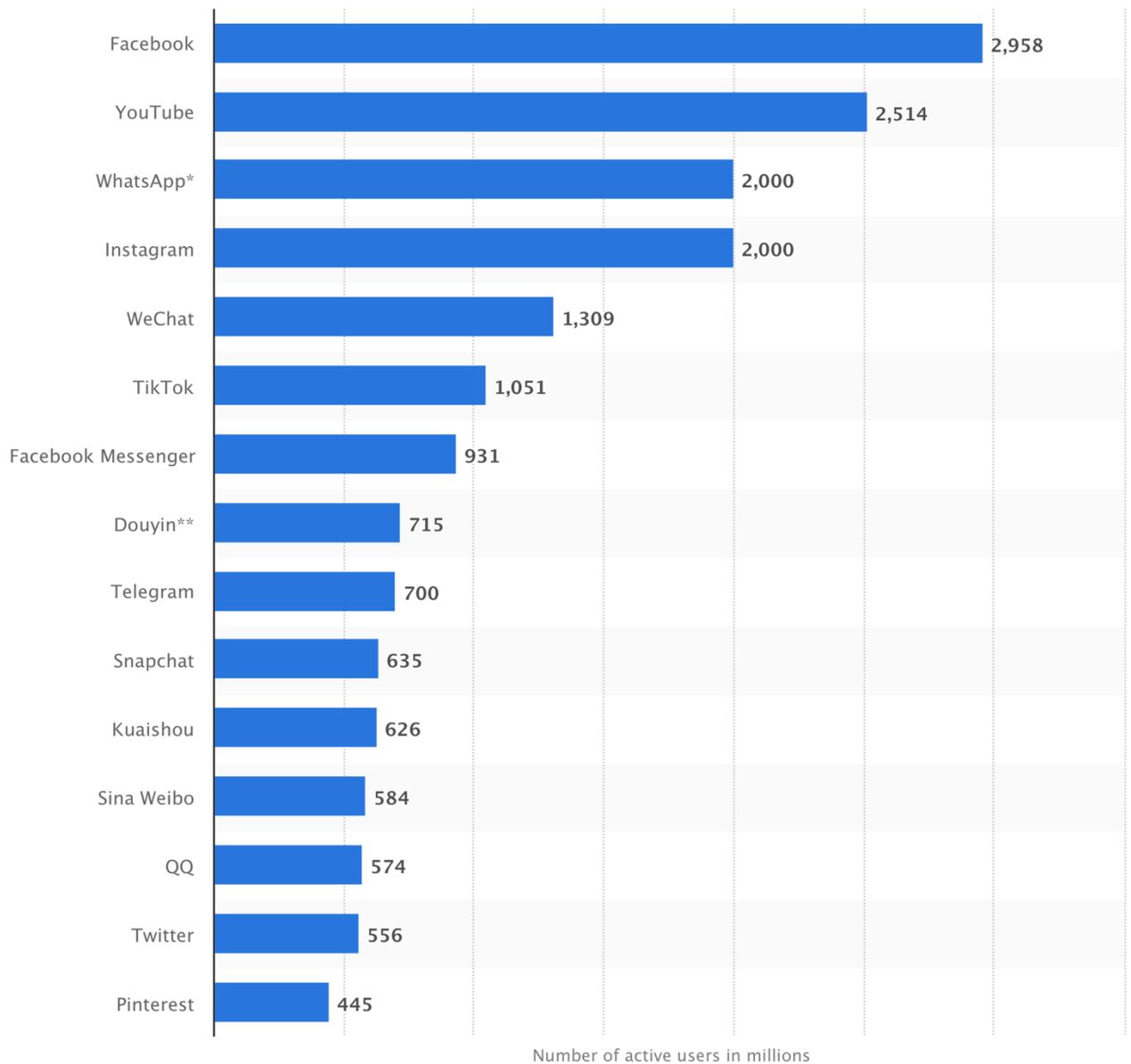


Figure 1 – The top 15 social networks by number of monthly active users [5]

1.2 Overview of existing methods and approaches

The current methods for determining slang can be reduced to the following types:

1) The lexicographic method is a method of identifying new words based on the frequency of their use, e.g., by determining the most common non-dictionary words in the corpus of social networks [2, 6]. This method works for

neologisms, but doesn't work for existing words that have been given a new slang meaning.

2) The method based on orthographic pleonasm. This method proposed in the article [7], based on the assumption that the addition of repeated letters emphasizes (strengthens) the meaning of the word. So the words with the largest number of spellings are considered as slang.

3) Topic modeling. This method can be used to check if the meaning of a word has changed relative to its usage in various discussions. However, this method does not allow identifying new words and is not specific to sentiment [8, 9].

4) Deep learning. The example of using this method is given in articles [10, 11, 12], when using it, the authors found that the indication of slang is the unexpected use of words in different syntactic categories or syntactic shift. The disadvantage of this method is that it can be not effective for a informally structured language and needs training data.

5) Frequency method. This method is suitable for identifying polarized political slang. This method assumes that this slang can be detected by comparing the word frequency between sets of comments with different political positions after removing dictionary words, proper names, and known slang. Thus, the words that are often used in a certain political discourse are likely to be slang for this group [13].

At the same time, these methods can be used together, e.g., in [14], a method for determining slang is proposed, which involves the joint use of the lexicographic method and deep learning.

All considered literary sources suggest the definition of slang in English and in some other languages, however, when reviewing literary, no works devoted to slang identification were found in Russian-language texts. At the same time, the methods that are effective for another's languages can show the worst result for the Russian language due to the morphological and syntactic features of the Russian language.

1.3 A research problem statement

The main goal of this work is to develop a comprehensive algorithmic support for identifying slang words and phrases from the texts of online communities. To achieve this goal, it is necessary to prepare training, test and validation data sets, while the raw data must be pre-processed.

At the next stage different ML models for automated slang detection should be generated and trained, and the software for automated slang detection should be implemented. The program should take a set of texts from online communities as input, and assign each text a category indicating the presence or absence of slang words at the output. If the text contains slang words, those ones should be highlighted.

Furthermore, it is necessary to test the implemented software on test data and make a comparative assessment of the ML-model quality, as well as to interpret the results obtained and, if necessary, refine some models.

1.4 Requirements specification

The algorithmic software for identifying slang vocabulary from the texts of online communities developed in this research should take a set of texts from online communities as input, and predict a category of each text indicating the absence or presence of slang in it at the output. Different thematic slang groups get different numbers assigned. The classification accuracy should be higher than 50%, and the processing speed of one text should not exceed a few seconds.

Conclusion

As a result of the overview of literary sources, it can be concluded that the most universal method for determining slang is the method based on deep learning. Since this method enables us to detect both new slang vocabulary and the vocabulary with a new slang meaning, unlike the thematic and lexicographic methods. In comparison with frequency method, deep learning enables us to identify slang in texts of any subject.

References

1. Orlova N.O. Slang vs jargon: the problem of definition // Yaroslavl Pedagogical Vestnik. – 2004. – №40. – Pp. 36-39.
2. Shehzaad D., Diptesh K., Pushpak B. SlangNet: A WordNet like resource for English slang // In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC). – Portoroz: 2016. – Pp. 4329-4332.
3. Hemmatian F., Sohrabi M.K. A survey on classification techniques for opinion mining and sentiment analysis // Artificial Intelligence Review. – 2019. – №52. – Pp. 1495-1545.
4. DIGITAL 2023: GLOBAL OVERVIEW REPORT // datareportal URL: <https://datareportal.com/reports/digital-2023-global-overview-report> (Date of application: 27.03.2023).
5. Most popular social networks worldwide as of January 2023, ranked by number of monthly active users // Statista URL: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/> (Date of application: 27.03.2023).
6. Asghar M.Z. Detection and Scoring of Internet Slangs for Sentiment Analysis Using SentiWordNet // Life science journal. – 2014. – №11. – Pp. 66-72.
7. Thelwall M., This! Identifying New Sentiment Slang Through Orthographic Pleonasm Online: Yasss Slay Gorg Queen Ilysm // IEEE Intelligent Systems. – 2021. – №4. – Pp. 114-120.
8. Matsumoto K., Yoshida M., Tsuchiya S., Kita K., Ren F. Slang Analysis Based on Variant Information Extraction Focusing on the Time Series Topics // International Journal of Advanced Intelligence. – 2016. – №1. – Pp. 84-98.
9. Matsumoto K. Slang feature extraction by analysing topic change on social media // CAAI Transactions on Intelligence Technology. – 2019. – №.1. – Pp. 64-71.

10. Pei Z., Sun Z., Xu Y. Slang detection and identification // Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL). – Hong Kong: 2019. – Pp. 881–889.
11. Seki Y., Liu Y. Multi-task Learning Model for Detecting Internet Slang Words with Two-Layer Annotation // 2022 International Conference on Asian Language Processing (IALP). – Singapore: 2022. – Pp. 212-218.
12. Hamid M.A., Tumpa E.S., Polin J.A., Al Nahian J., Rahman A., Mim N.A. Bengali Slang detection using state-of-the-art supervised models from a given text // Bulletin of Electrical Engineering and Informatics. – 2023. – №4. – Pp. 2381-2387.
13. Hossain N., Tran T.T.T., Kautz H. Discovering political slang in readers' comments // Proceedings of the International AAAI Conference on Web and Social Media. – 2018. – №1. – Pp. 612-615.
14. Pal A.R., Saha D. Detection of Slang Words in e-Data using semi-Supervised Learning // International Journal of Artificial Intelligence & Applications (IJAIA). – 2013. – №5. – Pp. 49-61.

Приложение Б

Фрагмент скрипта на python

```
nltk.download('stopwords')
nltk.download('punkt')
stopwords_list = stopwords.words("russian")

def clear_text(text: str):
    morphy = pymorphy2.MorphAnalyzer()
    text = text.lower()
    text = re.sub(r"^[А-Яа-я]+", " ", text)
    tokenized_text = nltk.word_tokenize(text, language="russian")
    clear_list = [word for word in tokenized_text if word not in stopwords_list]
    for i in range(len(clear_list)):
        clear_list[i] = morphy.parse(clear_list[i])[0].normal_form
    return clear_list

def slg(df, Pred):
    newdf = df
    newdf["slang"] = ""
    Y_pred = Pred

    for i in Y_pred:
        dict0_list = [l1_0, l2_0, l3_0, l4_0]
        dict0_list.pop(i)
        dict_list = [l1, l2, l3, l4]
        slang_list = []
        L1 = dict_list[i]
        ndict = dict_list.pop(i)
        L2 = ndict[0]
```

```
L3 = ndict[1]
L4 = ndict[2]
L02 = dict0_list[0]
L03 = dict0_list[1]
L04 = dict0_list[2]

for j in range (0, len(newdf['text_clean'].iloc[i])):
    w = newdf['text_clean'].iloc[i][j]
    if w not in L02 and w not in L03 and w not in L04:
        slang_list.append(w)
    else:
        if w in L1 and w not in L2 and w not in L3 and w not in L4:
            slang_list.append(w)
        else: pass
newdf['slang'].iloc[i] = slang_list
```