Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение

высшего образования

«Уральский федеральный университет

имени первого Президента России Б.Н. Ельцина»

Институт радиоэлектроники и информационных технологий - РТФ

Школа бакалавриата

ДОПУСТИТЬ К ЗАЩИТЕ ПЕРЕД ГЭК

РОП 09.03.01 Спиричева Н.Р.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись)

«\_\_\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2023 г.

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

Пояснительная записка

12312312312

|  |  |
| --- | --- |
| Руководитель: Д.Б. Шадрин  ст. преподаватель | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Нормоконтролер: В.Ф. К  ст. преподаватель | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Студент группы РИ-490002 И.А. Пономарев | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

Екатеринбург

2023

# **РЕФЕРАТ**

Выпускная квалификационная работа бакалавра 46 с., 4 рис., 25 источников.

БЕКЕНД, РАЗРАБОТКА СЕРВИСА ОТПРАВКИ УВЕДОМЛЕНИЙ, ИНТЕГРАЦИЯ, EXCHANGE, TALK.

Разработка интеграций с Exchange и Talk, модуля отправки писем и внутренних сервисов для продукта «Контур.Встреча».

Цель работы – разработка сервиса для отправки писем и синхронизации календарей пользователей с помощью веб-приложения «Контур.Встреча».

Объект работы – веб-приложение сервиса «Контур.Встреча».

Методы исследования: анализ аналогов, проектирование приложения, разработка приложения.

Результаты работы: приложение для отправки писем и интеграции с календарями для веб-приложения «Контур.Встреча».

Выпускная квалификационная работа выполнена в текстовом редакторе Microsoft Word (формат А4, шрифт Times New Roman, кегль 14) и представлена в электронном формате [1].

**Содержангие**

# **ВВЕДЕНИЕ**

Электронный документооборот позволил упростить работу с документами. Но в то же время, сложно представить себе ценность разрозненной информации и данных.

Во многих государствах существует тренд на открытость и прозрачность работы государственного аппарата, Российская федерация не является исключением, и также предоставляет доступ к государственным реестрам открытых данных.

Ручной сбор и анализ документов, связанных с какой-либо конкретной организацией, может являться довольно трудозатратным. Наиболее подходящим решением данной проблемы является автоматизация этих процессов.

Целью данной дипломной работы является разработка модуля для анализа судебных документов

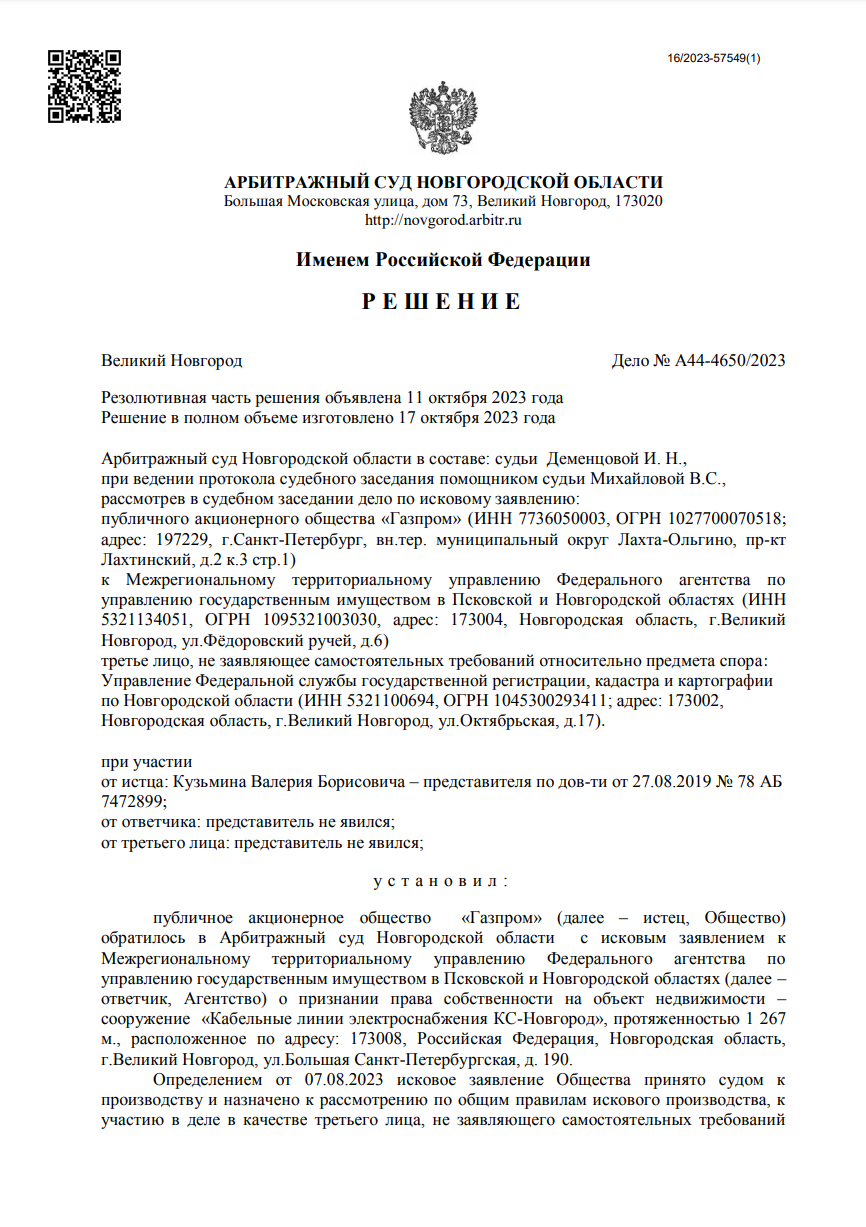
Для достижения цели были поставлены следующие задачи:

1. проанализировать требования к системе и составить частное техническое задание;
2. сравнить существующие решения для анализа текстовых данных;
3. создать датасет;
4. обучить модель;
5. оценить производительность модели;

**Анализ требований**

Сервис проверки партнеров является одной из составляющих экосистемы СБИС. Он позволяет проанализировать различные метрики компании будь то – финансовые показатели, открытые вакансии, надежность.

Проверка компании осуществляется с помощью анализа и разбора открытых данных. Одним показателей надежности компании/контрагента являются связанные с ним судебные дела. Каждое судебное дело может много сказать о текущей ситуации в компании. Например, решением суда у компании могут отсудить имущество или значительные суммы денежных средств.



Одним из способов анализа текстовых данных являются модели машинного обучения для распознавания именованных сущностей (NER)

**Сравнение существующих решений**

# 1)фцвф

2)Сравнение с регулярками?

**Создание датасета**

**Сбор исходных документов**

Для дообучения готовой модели не требуется настолько значительный объем даннных как для ее первоначального обучения. Поэтому для начала соберем и разметим 10 документов.

Сбор документов произведем с помощью сервиса «Картотека арбитражных дел (kad.arbitr.ru)». Существует множество видов документов, таких, как: само исковое заявление, определение, решение и постановления, а также прочие судебные документы. Информацию, нужную для нашей задачи мы можем извлечь из документов вида «Решения и постановления».

Исходные судебные документы представлены в формате PDF. Так как мы не можем напрямую работать с данным форматом, то для начала переведем их в текст.

Перевод документов осуществляем с помощью модуля …

Таким образом получим текст …

Картинка

**Разметка датасета**

Распознавание именованных сущностей в тексте подразумевает под собой извлечение нужных фраз или слов, которые мы можем разделить на множество различимых смысловых групп, будь то названия организаций, брендов, имен людей или дат.

Таким образом, для разметки датасета следует выделить в тексте требуемые поля, а именно:

* реквизиты исца, ответчика
* название суда
* решение суда
* …

Для разметки датасетов существует множество способов и инструментов, например – Толока, разметка по паттернам, или ручная разметка.

Из вышеперечисленных способов я решил выбрать ручную разметку, так как требуется разметить сравнительно небольшое количество документов. Разметку буду производить при помощи инсутремента …

Исходный док картинка

Размеченный док картинка

**Как вариант, сравнение способов разметки – попробовать непрямую и автоматическую(snorkeler)**

**Обучение модели**

**Оценка модели**

# **ЗАКЛЮЧЕНИЕ**

В результате выпускной квалификационной работы создана система для ыуууууууууууууууууууууууууууууууууууууууууууууууууууууууууууууууууууууууууууууууу

# **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. Алферьева Т.И., Васина В.Н., Шадрин Д.Б. Методические рекомендации для подготовки и оформления выпускных квалификационных работ. Методические рекомендации для обучающихся всех форм обучения / составители Т.И. Алферьева, В.Н. Васина, Д.Б. Шадрин. – Екатеринбург: ГОУ ВПО «УрФУ-УПИ», 2019. – 93 с. (дата обращения 01.03.2023)
2. First Round Review How Design Thinking. Transformed Airbnb from a Failing Startup to a Billion Dollar Business: сайт. – URL: https://review.firstround.com/How-design-thinking-transformed-Airbnb-from-failing-startup-to-billion-dollar-business (дата обращения: 23.03.2023).
3. Medium Why is prototyping important: сайт. – URL: https://blog.zipboard.co/why-is-prototyping-important-13150d76abc4 (дата обращения: 23.03.2023).