

WEB-ПРИЛОЖЕНИЕ ПРОТОКОЛИРОВАНИЯ МЕРОПРИЯТИЙ С ПОМОЩЬЮ АВТОРЕФЕРИРОВАНИЯ АУДИОЗАПИСЕЙ

Аннотация. В статье представлен анализ проблемы ручного протоколирования мероприятий, предложен вариант решения проблемы — дообучение моделей искусственного интеллекта. Предложен вариант разработки веб-приложения для использования обученных моделей.

Ключевые слова: распознавание речи, автореферирование, протоколирование, мероприятие, информация, веб-приложение.

Введение. В современном мире задача протоколирования мероприятий является одной из значимых, так как протокол содержит в себе последовательную запись обсуждения вопросов и принятия решений в различных отраслях [1]. С приходом цифровых и информационных технологий появилась возможность вести аудиозаписи, которые в дальнейшем можно обрабатывать, извлекая из них текст и реферировать его автоматически [2]. Например, одной из сильных сторон российского электронного арбитражного правосудия является использование аудиопротоколирования при рассмотрении 100% дел [3]. Такой подход помогает сэкономить много времени и ресурсов, которые используются при ручном протоколировании.

Существует не так много инструментов, которые объединяют в себе возможности, как обработки аудиозаписей и их распознавание, так и автореферирования. Одной из них стоит выделить информационную систему «Референт-ПС» [4]. Также стоит отметить, что большинство из данных систем не имеют открытого доступа или не способны работать с русским языком, как «Meeting Decision Tracker» [5].

Проблема исследования. Зачастую при протоколировании мероприятий сталкиваются с такими проблемами, как: затраты большого количества времени, допускаемые ошибки и пропуски, неэффективность при протоколировании, возникающая, когда у человека нет ясного понимания, что нужно документировать. Все это может

привести к неправильной интерпретации текста, ненужной детализации и снижению практической ценности протокола. Кроме того, протоколирование мероприятий может потребовать значительных усилий от сотрудников, что может привести к увеличению расходов на оплату их труда.

Целью данной работы является разработка веб-приложения для автоматического протоколирования с применением моделей искусственного интеллекта для распознавания речи и автореферирования.

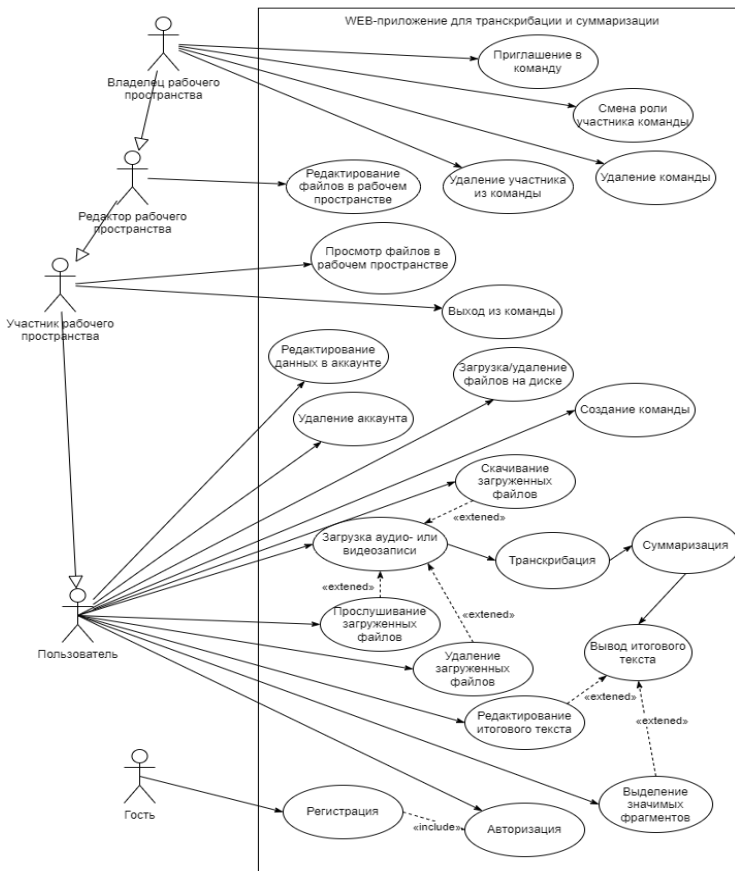


Рис. 1. Диаграмма прецедентов

Материалы и методы. На первом этапе проектирования построена диаграмма прецедентов для описания поведения разрабатываемой системы, представлена на рис. 1. На следующем этапе проектирования построена логическая модель базы данных. Данный процесс необходим для правильного структурирования данных и устранения избыточности и несогласованности данных.

Важным шагом на этапе проектирования является проектирование общей архитектуры приложения. Для разрабатываемого приложения на данном шаге были выделены следующие составляющие архитектуры: серверная и клиентская части веб-приложения, и модуль для работы с моделями искусственного интеллекта. Спроектированная общая архитектура изображена на рис. 2.

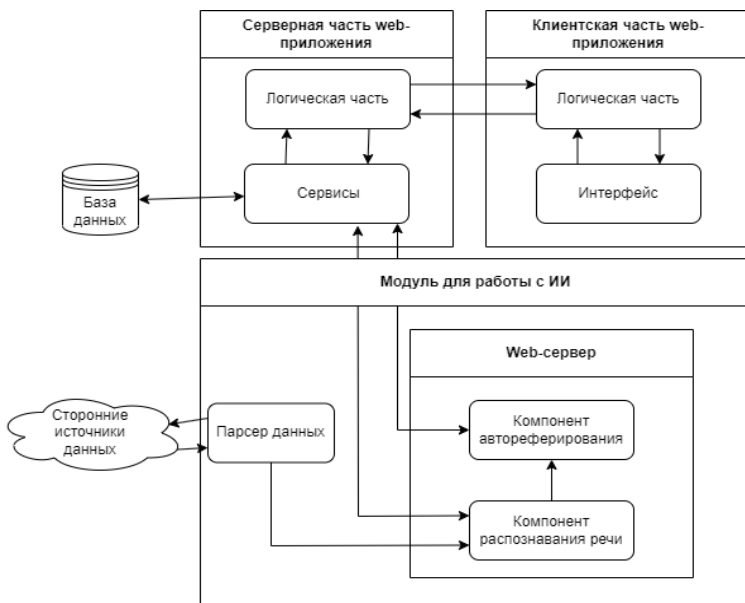


Рис. 2. Архитектура разрабатываемого приложения

Для реализации веб-приложения были выбраны следующие инструменты: React.js для клиентской части, Node.js с фреймворком Express.js для серверной части и PostgreSQL в качестве СУБД.

Для реализации системы распознавания речи было выбрано сочетание акустического и языкового моделирования. В качестве метода реализации акустической модели были выбраны нейронные сети. Языковую модель решено реализовать, как статистическую грамматную модель. Для реализации системы распознавания речи был выбран набор инструментов под названием «NeMo» от компании «NVIDIA».

Для обучения системы были выбраны речевые корпуса на русском языке с разными дикторами «Vox Forge» и «Open STT», которые включают в себя, помимо аудиозаписей, их расшифровки. Для сбора данных с «Vox Forge» был разработан парсер с применением языка программирования «Python» и библиотеки «Beautiful Soup». Объем собранных данных после обработки составил ~150 часов устной речи. При реализации языковой модели была использована часть набора данных «Тайга» размером ~17 ГБ. В случае с разрабатываемой языковой модели была выбрана 3-граммная последовательность.

Для дообучения акустической модели взята модель «STT Ru Quartznet 15x5», которая разработана и обучена на основе архитектуры «QuartzNet» [6]. Она содержит в себе CTC блок, необходимый для выравнивания выходной последовательности и ее дальнейшей обработки [7]. Особенностью модели является то, что она обучается с использованием аугментации данных. В используемый набор инструментов «NeMo» встроен модуль для аугментации SpecAugment, который применяет несколько случайных преобразований к спектрограмме аудиофайла во время обучения модели [8].

При выборе методов реферирования был выбран экстрактивный метод. Для реализации системы автореферирования было выбрано использование нейросетевого подхода с архитектурой трансформера [9]. В качестве инструмента для реализации выбрана библиотека на языке программирования «Python» — «Transformers» с платформы «Hugging Face» [10].

В качестве модели системы автореферирования решено использовать предобученную модель от проекта «DeepPavlov» — RuBERT, которую планируется дообучить под задачу экстрактивного реферирования [11].

В качестве обучающей выборки для системы автореферирования был выбран набор данных «Gazeta», находящийся в открытом доступе.

Результаты. *Модуль распознавания речи.* Проведя процесс дообучения акустической модели, и реализовав языковую n-граммную модель для минимизации получения ошибок решено использовать алгоритм поиска луча [12]. Алгоритм работы реализованной системы распознавания речи приведен на рис. 3. Проведено сравнение метрик качества акустической и языковой моделей. Результаты представлены в табл. 1.

Таблица 1

Результаты сравнения метрик качества используемых моделей

<i>Модель</i>	<i>WER</i>	<i>CER</i>
Акустическая модель до дообучения	47%	12%
Акустическая модель после дообучения	24%	5%
Акустическая модель с языковой моделью	14%	2%

Так как метрика WER у итогового варианта < 25%, то решено считать, что данную систему можно использовать для решения поставленной задачи.

Модуль автореферирования. Результатом обучения модели RuBERT является усовершенствованная дообученная модель для решения задачи экстрактивного реферирования [13].

Веб-приложение. Последним шагом на этапе реализации стала разработка веб-приложения, благодаря которому можно воспользоваться функционалом модулей распознавания речи и автореферирования. Основными выводами приложения являются результат обработки аудиофайла, то есть полученный текст и его сокращенная версия, а также экспорт в удобном формате (рис. 4).

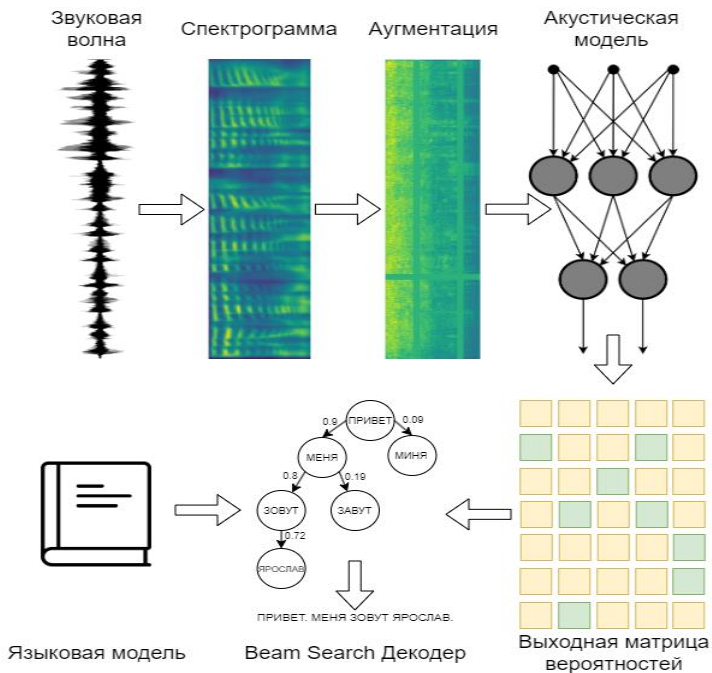


Рис. 3. Алгоритм работы системы распознавания речи

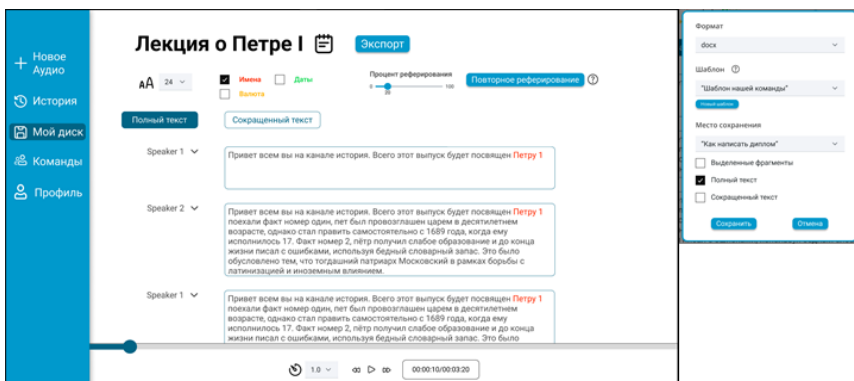


Рис. 4. Пример обработки аудио, окно экспорта

Заключение. В ходе данной работы была выявлена и сформулирована проблема, возникающая в процессе протоколирования мероприятий. Разработано веб-приложение для протоколирования мероприятий с помощью автореферирования аудиозаписей. Выбраны подходящие модели искусственного интеллекта и дообучены под решение поставленных задач. Разработанное веб-приложение позволяет существенно сократить время обработки аудиозаписей различных мероприятий, получить из них итоговый текст и его выжимку с наиболее значимыми предложениями. Помимо этого, разработанное веб-приложение помогает сэкономить человеческие ресурсы.

СПИСОК ЛИТЕРАТУРЫ

1. Ходжаян Е. Г. Проведение и протоколирование коллегиального обсуждения вопросов и принятия решения / Е. Г. Ходжаян. — Текст : непосредственный // Симбирский научный вестник. — 2020. — № 3-4 (41-42). — С. 104-119. — EDN MCTUOL.
2. Vinnarasu A., Deepa V. Jose Speech to text conversion and summarization for effective understanding and documentation / A. Vinnarasu, V. Deepa. — Direct text // International Journal of Electrical and Computer Engineering (IJECE). 2019. Vol. 9, No. 5. Pp. 3642-3648.
3. Информационные технологии в правосудии: состояние и перспективы. Россия и мир. Аналитический доклад / А. В. Кашанин, А. Б. Козырева, Н. А. Курносова, Д. В. Малов; под ред. А. В. Кашанина. — Москва: Центр развития современного права, 2020. — 80 с. // Институт исследований национального и сравнительного права [сайт]. — URL: <https://ilr.hse.ru/data/2020/07/14/B8.pdf> (дата обращения: 29.05.2023). — Текст: электронный.
4. Свидетельство о государственной регистрации программы для ЭВМ № 2015617545 Российская Федерация. «Комплекс протоколирования совещаний Референт-ПС: СПО приема и хранения аудио- и видеоданных, СПО управления протоколированием, СПО АРМ протоколирования совещаний» («Референт-ПС») : № 2015613095 : заявл. 15.04.2015 : опубл. 14.07.2015 / П. В. Конюхов, В. В. Шитьковский; заявитель Закрытое акционерное общество «НТЦ Передовые Системы». — EDN KLSLVU. — Текст : непосредственный.
5. Meeting Decision Tracker: Making Meeting Minutes with De-Contextualized Utterances / Shumpei Inoue, Hy Nguyen, Pham Viet Hoang [and

- etc.]. — <https://doi.org/10.48550/arXiv.2210.11374> // arXiv:2210.11374v1, 2022. (date of the application 26.05.2023). — Text : electronic.
6. Quartznet: Deep automatic speech recognition with 1D time-channel separable convolutions / Samuel Krivan, Stanislav Beliaev, Boris Ginsburg [and etc.] . — Direct text // arXiv:1910.10261v1, 2019. (date of the application 26.05.2023). — Text : electronic.
 7. An Intuitive Explanation of Connectionist Temporal Classification . — URL: <https://towardsdatascience.com/intuitively-understanding-connectionist-temporal-classification-3797e43a86c> (date of the application 22.05.2023). — Text : electronic.
 8. SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition / Daniel S. Park, William Chan, Yu Zhang [and etc.] // arXiv:1904.08779v3, 2019. (date of the application 22.05.2023). — Text : electronic.
 9. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding / Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. // arXiv:1810.04805v2, 2019. (date of the application 26.05.2023). — Text : electronic.
 10. Введение в библиотеку Transformers и платформу Hugging Face: [сайт]. — URL: <https://habr.com/ru/articles/704592/> (дата обращения: 29.05.2023). — Текст: электронный.
 11. Adaptation of Deep Bidirectional Multilingual Transformers for Russian Language / Yuri Kuratov, Mikhail Arkhipov. — Direct text // arXiv:1905.07213v1, 2019.
 12. Word Beam Search: A CTC Decoding Algorithm. — URL: <https://towardsdatascience.com/word-beam-search-a-ctc-decoding-algorithm-b051d28f3d2e> (date of the application 22.05.2023). — Text : electronic.
 13. Yang Liu. Fine-tune BERT for Extractive Summarization / Liu Yang // <https://doi.org/10.48550/arXiv.1903.10318>, 2019. (date of the application 22.05.2023). — Text : electronic.