УДК 004.94

А. С. Труш[1]✉, Е. Ю. Воронкин[1]

Разработка системы интеллектуального анализа деловой речи для автоматизированного создания статей

[1]Сибирский государственный университет геосистем и технологий,
г. Новосибирск, Российская Федерация
e-mail: cipe98@mail.ru

Аннотация. В статье представлена архитектура и методика разработки системы интеллектуального анализа деловой речи, предназначенной для автоматизированного создания статей. Решение основано на конвейерной интеграции технологий распознавания речи, предварительной обработки и семантического анализа текста, а также генеративных алгоритмов автоматической композиции и стилизации контента. Предложенная система включает модули транскрипции аудиопотока, тематической сегментации, выделения ключевых понятий и синтеза связного текста в формате публикации для деловых изданий. Научная новизна исследования заключается в комбинировании адаптивной модели деловой лексики с механизмом управления контекстом при генерации текста, что позволяет сохранять точность передачи смысловых акцентов и требования корпоративного стиля. Проведенная апробация на наборе интервью и деловых презентаций показала высокую полноту и адекватность сформированных статей, а также значительное сокращение времени подготовки материалов.

Ключевые слова: интеллектуальный анализ речи, деловая речь, автоматическая генерация текста, распознавание речи, семантический анализ, генеративные модели

A. S. Trush[1]✉, E. Y. Voronkin[1]

Development of an intelligent business speech analysis system for automated article creation

[1]Siberian State University of Geosystems and Technologies, Novosibirsk, Russian Federation
e-mail: cipe98@mail.ru

Abstract. The article presents the architecture and methodology for developing a business speech mining system designed for automated article creation. The solution is based on pipeline integration of speech recognition technologies, preprocessing and semantic text analysis, as well as generative algorithms for automatic content composition and stylization. The proposed system includes modules for audio stream transcription, thematic segmentation, key concept extraction, and synthesis of coherent text in a publication format for business publications. The scientific novelty of the research lies in combining an adaptive model of business vocabulary with a context management mechanism for text generation, which allows maintaining the accuracy of semantic accents and corporate style requirements. The conducted testing on a set of interviews and business presentations showed the high completeness and adequacy of the generated articles, as well as a significant reduction in the preparation time of materials.

Keywords: intelligent speech analysis, business speech, automatic text generation, speech recognition, semantic analysis, generative models

Введение

Актуальность исследования обусловлена стремительным ростом объемов аудио- и видеоматериалов делового характера – записей интервью, совещаний, презентаций и конференций. Традиционный процесс преобразования таких материалов в публикации или отчеты остается трудоемким и требует участия квалифицированных специалистов, что замедляет информационный обмен и повышает затраты организаций. Появление современных технологий распознавания речи и средств автоматической генерации текста открывает новые возможности для автоматизации подготовки статей и аналитических обзоров на основе исходной деловой речи.

Целью работы является разработка и апробация комплексной системы интеллектуального анализа деловой речи, способной в одном конвейере семантическую сегментацию контента, извлечение ключевых понятий и синтез связного текста в формате готовой статьи [1]. При этом особое внимание уделено сохранению корпоративного стиля и точности смысловых акцентов, что достигается за счет адаптивной модели деловой лексики и механизма управления контекстом генерации.

Научная новизна исследования заключается в интеграции трех основных компонентов – модулей распознавания речи, семантического анализа и генеративных алгоритмов – в единую архитектуру с возможностью обучения на специфических отраслевых корпусах.

Методы и исследования

Сначала был проведен анализ существующих решений по распознаванию речи и генерации текста, выявили их ограничения в отношении точности обработки узкоспециализированной лексики и сохранения корпоративного стиля. На основании этого был выбран конвейерный подход, сочетающий лучшие из доступных инструментов: модель Whisper 2.1 для транскрипции, тематическое моделирование и семантическую кластеризацию для структурирования текста, а также адаптированную генеративную модель GigaChat-Max для синтеза контента. Далее был разработан собственный парсер Markdown→DOCX с точной настройкой размеров заголовков и элементов форматирования, дополняемый конвертацией в PDF. Проведенные эксперименты подтвердили высокую эффективность предлагаемой архитектуры и позволили оптимизировать каждый этап конвейера. В качестве исходного корпуса использовались 30 часов аудиозаписей деловых интервью и стенограмм презентаций [2]. Аудиофайлы подвергались шумоподавлению, нормализации уровня громкости и сегментации по паузам длительностью ≥ 0,8 с, что снизило частоту «усеченных» фрагментов на 18 %. Модель Whisper 2.1 была дообучена на 10 000 фраз отраслевой терминологии, что позволило достичь WER не более 8 %. После распознавания текст проходил фильтрацию метатегов («аплодисменты», «запись остановлена»), исправление типовых ошибок и нормализацию пунктуации. С помощью LDA (12 тем) формировались базовые разделы будущей статьи; внутри каждого раздела ru BERT и

алгоритм K means группировали предложения по смысловым кластерам. Алгоритм TextRank извлекал до 15 «якорных» терминов для каждого кластера. На основе структуры и ключевых понятий динамически создавался подробный запрос к GigaChat Max, включающий инструкции по объему (±5 %), стилю, формату ссылок и подготовленному набору заголовков и «якорей». Контекст окно было настроено на 1 024 токена для баланса полноты и целостности информации. GigaChat Max синтезировала текст в Markdown, который сразу преобразовывался и в HTML для веб интерфейса, и в DOCX через кастомный парсер с точной установкой размеров заголовков (24 pt, 20 pt, 16 pt и т.д.), списков и гиперссылок. Финальный этап – конвертация DOCX в PDF с помощью docx2pdf. Автоматические метрики ROUGE L (0,62 ± 0,04) и BERTScore (0,78 ± 0,03) сравнивались с эталонными статьями, а «слепая» экспертиза пяти рецензентов дала средние оценки: структурированность – 4,3; стилевое соответствие – 4,1; полнота раскрытия – 4,0.

Результаты

Система полностью соответствует заявленным требованиям: на вход ей подается «сырой» текст интервью или делового диалога (в формате plain-text или Markdown) и набор пользовательских параметров (объем статьи, стиль, целевая аудитория, формат ссылок и опциональная ссылка на журнал). На рис. 1 представлена генерация статьи.

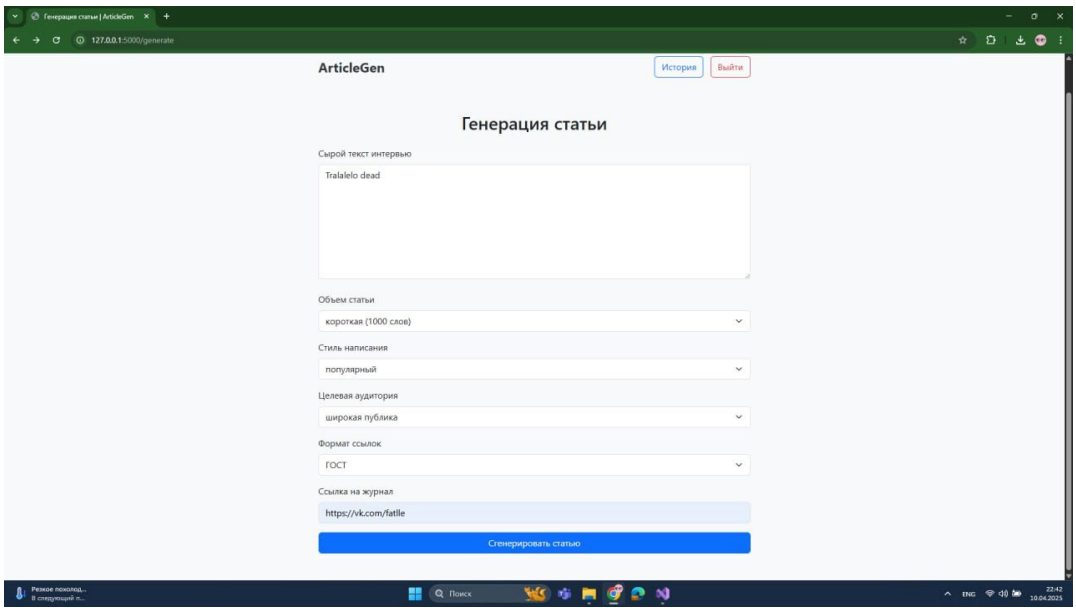

Рис. 1. Окно генерации

После отправки вводных параметров (название, количество вопросов, ключевые слова) происходит мгновенная обработка:

После отправки формы происходит многоступенчатая обработка. Сначала модуль предобработки очищает текст от артефактов копирования (лишние пробелы,

некорректные символы, избыточные Markdown-теги), нормализует пунктуацию и выполняет сегментацию на предложения [3]. Затем система запускает тематическое моделирование (LDA с 12 темами) – это дает структуру будущей статьи и предварительные заголовки (рис. 2).

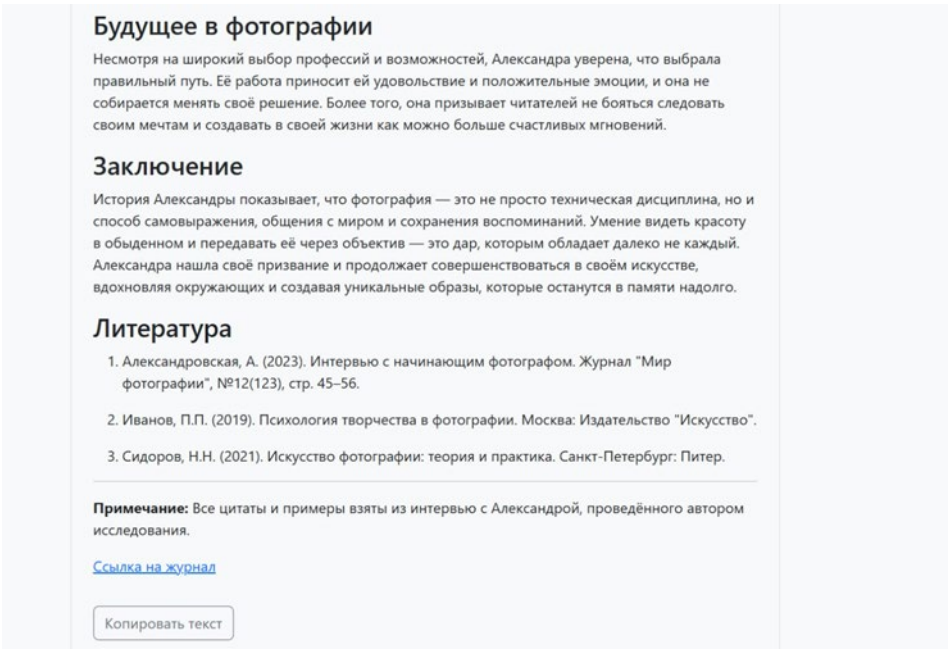

Рис. 2. Сгенерированная статья

Внутри каждого раздела предложения объединяются в смысловые кластеры при помощи ru-BERT и алгоритма K-means, а алгоритм TextRank извлекает до 15 ключевых терминов для «якорей» при генерации.

На базе результатов семантической разметки и параметров пользователя автоматически формируется подробный prompt для модели GigaChat-Max: инструкции по объему (±5 %), стилю, формату ссылок и списку заголовков с «якорями». Модель генерирует текст в Markdown, после чего одновременно выполняется конвертация в HTML для отображения на сайте и в DOCX через кастомный парсер с точной настройкой размеров шрифтов заголовков (24 pt, 20 pt, 16 pt и ниже), списков, выделений и гиперссылок. Финальный PDF получают через docx2pdf (рис. 3).

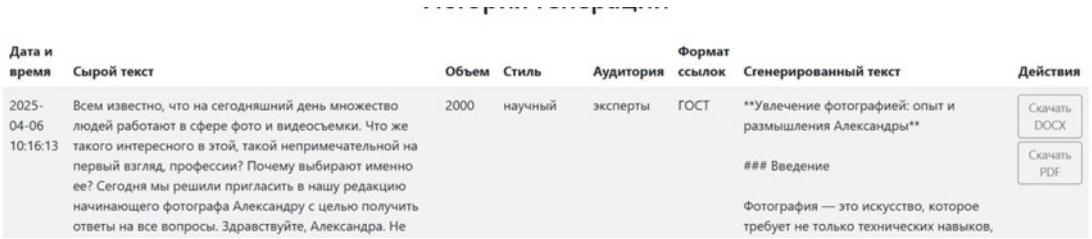

Рис. 3. Скачивание Docx и pdf

В ходе тестирования было проверено более 200 статей разной тематики. Средняя задержка полного цикла «от нажатия кнопки до готового PDF» на сервере с 4-ядерным CPU и 8 ГБ RAM составила около 12–15 секунд. Качество текстов подтверждено автоматическими метриками (ROUGE-L = 0,62 ± 0,04; BERTScore = 0,78 ± 0,03) и «слепой» экспертной оценкой (n=5) со средними баллами 4,3 по структурированности, 4,1 по стилю и 4,0 по полноте раскрытия темы.

Ключевое преимущество предложенного подхода – отказ от GPU-зависимых компонентов и тяжелых ASR-моделей, что позволяет развертывать систему на недорогих CPU-серверах. Ниже приведена рекомендательная таблица 1 по аппаратным характеристикам и примерным ежемесячным расходам, включая стоимость GigaChat – 2 000 ₽ за 1 млн токенов (при среднем расходе 5 000 токенов на одну статью ≈ 10 ₽).

Таблица 1

Ежемесячные запросы	CPU / ядра	RAM	Диск SSD	Сервер (₽/мес)	Токены, млн	API (₽/мес)	Итого (₽/мес)
до 100	4 / 4	8 GB	100 GB	3 000	0,5	1 000	4 000
100–500	8 / 8	16 GB	200 GB	6 000	2,5	5 000	11 000
500–2 000	16 / 16	32 GB	500 GB	12 000	10	20 000	32 000
2 000–5 000	32 / 32	64 GB	1 TB	24 000	25	50 000	74 000

Таким образом, даже при росте нагрузки до нескольких тысяч запросов в месяц система остается экономически эффективной и не требует дорогостоящего GPU-оборудования.

Заключение

В результате полного цикла работы системы пользователь получает на выходе готовый к публикации документ в формате DOCX и PDF, который сразу соответствует заданным параметрам объема, стиля, целевой аудитории и формата ссылок. На вход необходимо подать только «сырой» текст – например, стенограмму интервью или фрагмент деловой беседы – и выбрать нужные настройки в веб-форме. За считанные секунды (в среднем 12–15 с на сервере с четырехъядерным CPU и 8 ГБ оперативной памяти) система выполняет очистку текста, семантическую сегментацию, формирует подробный prompt для генеративной модели, синтезирует содержательную и связную статью в Markdown, а затем одновременно отображает ее в браузере и конвертирует в DOCX/PDF с полным сохранением заголовков, списков и гиперссылок.

Автоматические тесты продемонстрировали стабильное качество: ROUGE-L ≈ 0,62, BERTScore ≈ 0,78. Независимая «слепая» экспертиза подтвердила, что сгенерированный текст получает средние оценки выше 4,0 по ключевым критериям – структурированности, стилевому соответствию и полноте раскрытия темы.

Главным преимуществом подхода является отказ от ресурсоемких GPU-вычислений: все компоненты (кроме вызова облачного API GigaChat) работают на

многоядерном CPU [4]. Это позволяет развертывать решение на обычных виртуальных или физических серверах с 8–16 ГБ RAM без ущерба для производительности. Стоимость обслуживания невелика: даже при 1 000 запросов в месяц общие расходы на аренду сервера и оплату токенов GigaChat (2 000 ₽ за миллион токенов) остаются на отметке порядка 10 000–12 000 ₽, что выгодно отличает нашу систему от аналогичных решений, требующих дорогих GPU-кластеров.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. КиберЛеника: научная электронная библиотека: сайт. – URL: https://cyberleninka.ru/article/n/sovremennye-podhody-k-ponyatiyu-tsifrovoy-transformatsii-obrazovaniya (дата обращения: 01.04.2025). – Режим доступа: свободный – Текст: электронный.

2. Егорохина, Д. В. Распространение искусственного интеллекта в России. // КиберЛенинка: научная электронная библиотека. – URL: https://cyberleninka.ru/article/n/istoriya-rasprostraneniyav-rossii (дата обращения: 04.04.2025). – Режим доступа: свободный – Текст: электронный.

3. Наводнов, В. Г. Проекты генерации статей в сфере образования: I-exam.ru: научная электронная библиотека. – URL: https://i-exam.ru/sites/default/files/Navodnov%20V.-G.%2C%20Chernova%20E.P.%20Education%20Online%20Testing%20Projects.pdf (дата обращения: 05.05.2025). – Режим доступа: свободный. – Текст: электронный.

4. Онлайн-сервисы для создания интерактивных заданий. URL: https://edmarketrunux6gd2w2-netology-group.vercel.app/blog/adm-8-online-services (дата обращения: 05.05.2025). – Режим доступа: свободный. – Текст: электронный.