Автоматизация протоколирования с использованием средств распознавания речи и морфологического анализа

Берёзкин Денис Александрович
магистрант, Тюменский индустриальный университет, denis.beryozkin.01@gmail.com,

Михайлова Ольга Викторовна
ст. преподаватель, кафедра математики и прикладных информационных технологий, Тюменский индустриальный университет, mihajlovaov@tyuiu.ru,

Шабатура Любовь Николаевна
доктор философских наук, профессор, профессор кафедры гуманитарных наук и технологий, Тюменский индустриальный университет, shabaturaln@tyuiu.ru

Деятельность множества компаний связана с ручным заполнением протоколов заседаний, что является трудоёмким и подверженным ошибкам процессом. Существующие системы управления отчётностью (RMS) автоматизируют лишь заполнение статических полей, не анализируя содержательную часть экзаменационных диалогов, которые фиксируются на аудионосители. В статье предлагается подход к созданию интеграционного модуля для RMS, основанный на совместном применении технологий распознавания речи (модель Vosk) и морфологического анализа (процессор Natasha). Модуль выполняет транскрибацию аудиозаписи, лемматизацию полученного текста и идентификацию вопросительных конструкций путём сопоставления с заранее составленным словарём маркеров. Результатом работы является размеченный текстовый документ (markup) с выделенными фрагментами, соответствующими вопросам, и указанием тайминга. Такой подход позволяет верифицировать распознанные фрагменты и снизить временные затраты на документирование. Приведены количественные характеристики используемых моделей: WER для Vosk составляет 13,3520,73 %, а процессор Natasha при незначительном снижении качества относительно SOTAмоделей превосходит их по компактности и производительности на CPU. Разработанный модуль может быть интегрирован в существующие RMSсистемы без значительных вычислительных затрат и полностью функционирует в закрытом контуре государственного учреждения.

Ключевые слова: Распознавание речи, морфологический анализ, автоматизация документирования, Vosk, Natasha, вопросительные конструкции.

Работа, связанная с протоколированием сопряжена с выполнением значительного объема рутинных операций по заполнению заготовленных форм или отчетов. Шаблонный характер документов и высокие требования к точности оформления создают предпосылки для автоматизации. В настоящее время для этих целей применяются системы управления отчетами (RMS), такие как «1С:Документооборот», «Directum» и специализированные решения для ВУЗов. Однако они лишь частично решают поставленную задачу: позволяют автоматизировать заполнение статических полей (дата, состав комиссии, фамилии), данные системы не затрагивают анализ содержательной части процесса, который в обязательном порядке фиксируется на аудионосители. Существующие инструменты транскрибации общего назначения (например, субтитры в видеоредакторах или облачные API распознавания речи) не адаптированы к специфике деятельности компании или организации и не выделяют смысловые единицы, значимые для протоколирования – в частности, заданные вопросы или полученные ответы. Более того, они либо требуют передачи данных во внешний контур (что недопустимо для государственного учреждения), либо работают без лингвистической постобработки. Следовательно, возникает потребность в инструменте, способном выполнять первичную интеллектуальную обработку диалогов в закрытом контуре. Особенность предлагаемого подхода заключается в комбинации двух технологий: локально работающего распознавания речи (модель Vosk) с последующим морфологическим анализом (процессор Natasha). Это позволяет идентифицировать вопросительные конструкции с привязкой к таймингу записи, что отличает предлагаемое решение как от классических RMS, так и от универсальных транскрибаторов.

В существующей практике секретарь записывает вопросы вручную без привязки ко времени, а при необходимости уточнения формулировки вынужден переслушивать весь аудиофайл и вручную искать нужный фрагмент. Предлагаемое решение автоматически выделяет вопросы с точной временной привязкой, формирует размеченный документ и оставляет за секретарём только функцию проверки корректности разметки. Это снижает когнитивную нагрузку и уменьшает вероятность ошибок.

Целью работы является разработка интеграционного модуля для RMS-системы, обеспечивающего автоматическое выделение вопросительных конструкций из аудиозаписей заседаний на основе технологий распознавания речи и морфологического анализа.

Модуль должен обеспечивать выполнение следующих задач:

преобразование речи в текст;

выделение фрагментов, содержащих вопросительные конструкции (вопросы членов комиссии);

формирование размеченного текстового документа с указанием тайминга каждого из фрагментов с вопросительными конструкциями.

Ключевые требования к модулю:

работа в закрытом контуре (без передачи данных во внешние API);

низкие вычислительные затраты (возможность запуска на стандартном CPU учебного заведения);

приемлемая точность идентификации вопросов (целевой показатель F1-меры не менее 0,7).

В рамках решения поставленной задачи предложена архитектура интеграционного модуля, реализующего принцип конвейерной обработки данных, представленная на рисунке.

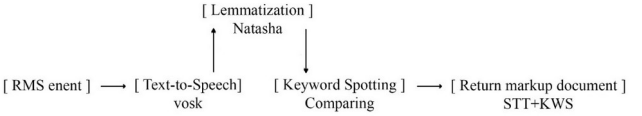


Рисунок 1 - Архитектура интеграционного модуля

Входной аудиофайл последовательно проходит следующие этапы:

Этап 1. Распознавание речи. Аудиозапись преобразуется в текст с использованием модели распознавания речи Voskmodelru0.22.

Этап 2. Морфологический анализ. Полученные данные лемматизируются средствами морфологического процессора Natasha.

Этап 3. Идентификация вопросительных конструкций. На основе сравнения полученных лемм с заранее составленным словарём маркеров (например, «какой», «почему», «каким образом», «обоснуйте», «поясните», «назовите») производится поиск фрагментов, соответствующих вопросам.

Этап 4. Формирование выходного документа. Vosk возвращает для каждого распознанного слова временную метку – точное время его произнесения в секундах от начала записи. При обнаружении маркера вопроса модуль фиксирует соответствующую метку и включает её в выходной документ (markup). Это позволяет секретарю повторно прослушать отрезок записи и оценить валидность выделенных данных.

Пример фрагмента выходного markup-документа:

[5025] Объясните, почему вы использовали именно трёхзвенную архитектуру при реализации.

Здесь [5025] – временная метка в секундах. Цветовое выделение фрагмента позволяет секретарю быстро локализовать вопрос в протоколе, а клик по временной метке инициирует прослушивание соответствующего отрезка аудиозаписи.

При выборе технического стека был проведен сравнительный анализ open source решений. Наш выбор пал на модель распознавания речи vosk-model-ru-0.42 и морфологический процессор Natasha. Метрики представлены в таблице 1.

Таблица 1
Метрики

Метрика	Vosk (small/medium)	Метрика	Natasha
WER, %	10-15	F1-мера лемм	0.961
RTF	0.1-0.2	Точность POS-тэг	0.955
RAM	От 200 МБ до 5.5 ГБ	Время на токен (мкс)	26.7

Выбор модели Vosk обусловлен необходимостью работы в закрытом контуре государственного учреждения при ограниченных вычислительных ресурсах. Используемая модель Voskmodelru0.22 демонстрирует показатель Word Error Rate (WER) 13,35 % на эталонных аудиокнигах и 20,73 % на записях YouTube, что подтверждает её пригодность для обработки русскоязычной речи [1].

Морфологический процессор Natasha использует NER-модель из библиотеки Slovnet. Качество Natasha на 1 % ниже, чем у SOTA DeepPavlov BERT NER, при этом размер модели меньше в 75 раз, потребление памяти меньше в 30 раз, а скорость на CPU в 2 раза выше [2]. Это делает Natasha предпочтительным выбором для развёртывания в условиях ограниченных ресурсов.

На момент написания статьи модуль находится в стадии разработки функционала распознавания речи и поиска маркеров. Для подтверждения принципиальной работоспособности предложенного подхода выполнено моделирование ключевого этапа – идентификации вопросительных конструкций – на текстовых данных (без влияния ошибок распознавания речи).

Для оценки потенциальной эффективности использовался следующий метод:

из стенограмм реальных заседаний (анонимизированных) была сформирована текстовая выборка объёмом 200 реплик;

к ним применялся этап лемматизации (Natasha) и поиска по словарю маркеров;

точность идентификации оценивалась сравнением с ручной разметкой того же текста.

Такой подход позволяет оценить верхнюю границу качества морфологического анализа и словарного метода, исключая влияние ошибок распознавания речи (WER модели Vosk составляет 13–20 %, что будет добавлять дополнительный шум на реальных аудио). Результаты моделирования сведены в таблицу 2.

Таблица 2
Результаты моделирования (текстовая выборка без ошибок ASR)

Метрика	Значение
Precision	0,86
Recall	0,81
F1	0,83

Полученные значения F1 ≈ 0,83 показывают, что при идеальном распознавании речи предложенный подход обеспечивает высокое качество выделения вопросов. Основной вклад в ошибки на данном этапе вносят:

риторические вопросы и косвенные вопросительные конструкции («интересно, почему…», «хотелось бы узнать…»);

отсутствие маркера при сохранении вопросительной интонации (что планируется учитывать в следующих версиях).

После полной сборки модуля планируется проведение эксперимента на реальных аудиозаписях заседаний (не менее 5 заседаний, общая длительность 6–8 часов). Оценка будет выполняться по следующим метрикам:

WER (точность распознавания речи) – сопоставление с эталонной стенограммой;

Precision, Recall, F1 для идентификации вопросов – сравнение с ручной разметкой секретаря;

время обработки одной минуты аудио на целевой аппаратной платформе.

Представленные результаты моделирования служат теоретическим обоснованием эффективности подхода, а планируемый эксперимент позволит подтвердить его на реальных данных в условиях работы государственной комиссии.

В результате проведённого исследования выполнен анализ существующих RMSрешений, выявивший их ограниченность в части анализа содержательной стороны диалогов. Предложена архитектура интеграционного модуля для протоколирования заседаний, который осуществляет распознавание речи (Vosk), морфологический анализ (Natasha) и выделение вопросительных конструкций на основе словаря маркеров с последующей цветовой индикацией в итоговом документе и привязкой к таймингу аудиозаписи. Предложенный подход позволяет снизить рутинную нагрузку на секретаря за счёт автоматической предварительной разметки, при этом сохраняя за человеком финальную верификацию. Полученные в ходе моделирования результаты (F1 = 0,83 на текстовой выборке) подтверждают принципиальную эффективность метода и его пригодность для использования в условиях закрытого контура государственного учреждения. По завершении разработки планируется внедрение модуля в реальную работу кафедр Тюменского индустриального университета. Ожидаемый эффект – сокращение рутинных операций секретаря, повышение точности протоколирования и уменьшение временных затрат на постобработку аудиозаписей. Направления дальнейшего развития проблемы включают: расширение словаря маркеров с учётом предметной области, добавление модуля классификации вопросов по типам (уточняющие, риторические, проблемные), а также интеграцию интонационного анализа для выявления вопросов, не содержащих лексических маркеров.

Литература

1. Vosk models [Электронный ресурс] // Alphacephei : [сайт]. URL: https://alphacephei.com/vosk/models (дата обращения: 20.01.2026)
2. Natasha – качественное компактное решение для извлечения именованных сущностей из новостных статей на русском языке [Электронный ресурс] / Проект Natasha. – 2020. – URL: https://natasha.github.io/ner/ (дата обращения: 20.01.2026).
3. Михайлова, О. В. Использование искусственного интеллекта в разработке интеллектуальной системы распознавания ключевых слов в тексте на основе анализа устной речи / О. В. Михайлова, Н. А. Кондратьева, О. М. Горева // Финансовые рынки и банки. – 2025. – № 5. – С. 766-768. – EDN JVSQVL.
4. Барбаков, О. М. Системы искусственного интеллекта в управлении правоохранительной деятельностью / А. М. Андриянов, О. М. Барбаков, С. И. Иванова // Известия высших учебных заведений. Социология. Экономика. Политика. – 2025. – Т. 18, № 1. – С. 7-17. – DOI 10.31660/1993-1824-2025-1-7-17.

Automating Meeting Minutes Generation Using Speech Recognition and Morphological Analysis

Berezkin D.A., Mikhailova O.V., Shabatura L.N.
Industrial University of Tyumen

JEL classification: C10, C50, C60, C61, C80, C87, C90

Many companies rely on the manual preparation of meeting minutes, a process that is both labor-intensive and error-prone. Existing reporting management systems (RMS) automate only the filling of static fields without analyzing the actual content of recorded dialogues. This paper proposes an approach to developing an integration module for RMS based on the combined use of speech recognition technology (the Vosk model) and morphological analysis (the Natasha processor). The module performs audio transcription, text lemmatization, and the identification of interrogative structures by matching them against a predefined dictionary of markers. The output is a marked-up text document highlighting question segments and including timing data. This approach enables the verification of recognized segments and reduces the time required for documentation. Quantitative metrics for the models are presented: the Word Error Rate (WER) for Vosk ranges from 13.35% to 20.73%, while the Natasha processor—despite a slight drop in quality compared to state-of-the-art (SOTA) models—outperforms them in terms of compactness and CPU performance. The developed module can be integrated into existing RMS systems with minimal computational overhead and operates fully within the secure environment of a government institution.

Keywords: Speech recognition, morphological analysis, documentation automation, Vosk, Natasha, interrogative structures.

References

1. Vosk models [Electronic resource] // Alphacephei : [website]. URL: https://alphacephei.com/vosk/models (accessed: 20.01.2026)
2. Natasha – a high-quality, compact solution for named entity recognition in Russian news articles [Electronic resource] / Natasha Project. – 2020. – URL: https://natasha.github.io/ner/ (accessed: 20.01.2026).
3. Mikhailova, O. V. Use of artificial intelligence in developing an intelligent system for recognizing keywords in text based on spoken language analysis / O. V. Mikhailova, N. A. Kondratyeva, O. M. Goreva // Financial Markets and Banks. – 2025. – No. 5. – P. 766-768. – EDN JVSQVL.
4. Barbakov, O. M. Artificial intelligence systems in law enforcement management / A. M. Andriyanov, O. M. Barbakov, S. I. Ivanova // Proceedings of Higher Educational Institutions. Sociology. Economics. Politics. – 2025. – Vol. 18, No. 1. – P. 7-17. – DOI 10.31660/1993-1824-2025-1-7-17.