УДК 004.352.246

А. Р. Сабиров[1]✉, Т. Ю. Бугакова[1]

Разработка системы автоматической стенографии на базе искусственного интеллекта для использования в учебном процессе университета

[1]Сибирский государственный университет геосистем и технологий,
г. Новосибирск, Российская Федерация
e-mail: arsen.sabirov.2020@mail.ru

Аннотация. В современной деловой среде ежедневно проводятся десятки совещаний, требующих аутсорсинга. Традиционное протоколирование остается трудоемким и затратным по времени процессом, создающим значительную нагрузку на специалистов. Существующие электронные системы протоколирования обладают существенными недостатками: высокая стоимость, ограниченная экономия времени и неудовлетворительное качество распознавания. В статье предлагается применение адаптированной модели Whisper от OpenAI, которая предлагает три ключевых преимущества: доступность благодаря открытой архитектуре, эффективность за счет специализации под протоколы и высокое качество точного распознавания профессиональной лексики. Модель демонстрирует высокую скорость обработки информации, превосходящую ручной ввод в 3–5 раз, точное распознавание терминологии и поддержку различных акцентов. Однако решение имеет некоторые ограничения: требует производительных GPU, качественного аудиовхода и редактирования в сложных случаях. Перспективы развития включают интеграцию с системами видеоконференцсвязи, разработку специализированных шаблонов и оптимизацию для работы на менее мощном оборудовании. Предлагаемое решение существенно упрощает работу стенографистов, сохраняя при этом высокое качество фиксации информации.

Ключевые слова: ИИ-технологии, стенография, Whisper-расшифровка, OpenAI, транскрибация, открытый исходный код, метод развертывания

A. R. Sabirov[1]✉, T. Y. Bugakova[1]

Development of an automatic shorthand system based on artificial intelligence for use in the educational process of the university

[1]Siberian State University of Geosystems and Technologies, Novosibirsk, Russian Federation
e-mail: arsen.sabirov.2020@mail.ru

Abstract. In today's business environment, dozens of outsourcing meetings are held daily. Traditional logging remains a laborious and time-consuming process, creating a significant burden on specialists. Existing electronic logging systems have significant disadvantages: high cost, limited time savings, and poor recognition quality. The article suggests the use of OpenAI's adapted Whisper model, which offers three key advantages: accessibility due to an open architecture, efficiency due to specialization in protocols, and high quality accurate recognition of professional vocabulary. The model demonstrates high information processing speed, exceeding manual input by 3-5 times, accurate terminology recognition and support for various accents. However, the solution has some limitations: it requires a high-performance GPU, high-quality audio input, and editing in difficult cases. Development prospects include integration with video conferencing systems, development of specialized templates, and optimization to work on less powerful hardware. The proposed solution significantly simplifies the work of stenographers, while maintaining high-quality information recording.

Keywords: AI technologies, stenography, Whisper transcription, OpenAI, transcription, open source, deployment method

Введение

В условиях интенсивного развития научной коммуникации проблема эффективного документирования совещаний приобретает особую актуальность. Протоколирование является важнейшим инструментом организационного управления и юридическим подтверждением принятых решений и традиционно осуществляется силами стенографистов. Однако данный процесс характеризуется существенными недостатками: высокой трудоемкостью, значительными временными затратами и субъективностью интерпретации информации. Существующие электронные решения, хотя и предлагают определенную автоматизацию, зачастую остаются экономически неэффективными и не обеспечивают существенного выигрыша во времени.

Современные достижения в области искусственного интеллекта, в частности модели автоматического распознавания речи, открывают новые перспективы для совершенствования процесса протоколирования. Особый интерес представляет модель Whisper от OpenAI, демонстрирующая высокую точность транскрибации благодаря обучению на обширных многоязычных данных. Ее способность эффективно работать в условиях фонового шума и с узкоспециализированной терминологией делает ее перспективным решением для задач автоматического протоколирования.

Целью настоящего исследования является разработка доступной, быстрой и качественной системы поддержки работы стенографистов на базе модели Whisper. Whisper это оптимальное решение для организаций, требующих безопасности, гибкости и высокой точности в документооборота. Его открытый исходный код и возможность локального использования делает его актуальным выбором если сравнивать с коммерческими аналогами.

В работе представлены: адаптация модели для задач деловой коммуникации, оценка ее эффективности в сравнении с традиционными методами протоколирования, а также анализ технических аспектов реализации, включая требования к вычислительным ресурсам.

Особое внимание уделено перспективным направлениям развития технологии, таким как интеграция с системами видеоконференцсвязи, реализация семантического анализа для автоматического выделения ключевых решений и создание интеллектуальных шаблонов протоколов.

Методы и исследования

Архитектура Whisper – это современная система, разработанная для обработки и анализа звуковых данных, в частности, для распознавания речи. Whisper использует глубокие нейронные сети, которые обучены на больших объемах данных, что позволяет достигать высокой точности распознавания. Она была создана с целью улучшения качества и точности распознавания, а также для обеспечения высокой скорости обработки [1–3]. Она состоит из:

– модуля обработки звука, который отвечает за предварительную обработку звуковых сигналов и включает в себя фильтрацию шума, нормализацию громкости и преобразование звука в цифровой формат, который может быть использован для дальнейшего анализа;

– модуля распознавания речи, в котором применяются алгоритмы машинного обучения для преобразования звуковых сигналов в текст.

– модуля постобработки, который отвечает за исправление ошибок и улучшение качества текста. Он может включать в себя грамматические и стилистические проверки, а также адаптацию текста под конкретные нужды пользователя [4–7].

В работе была проведена серия экспериментов по оценке качественного перевода модели Whisper. Для экспериментов использовалась облачная платформа Google Colab, на которой бал развернут тестовый стенд и выполнена оценка качества распознавания речи на различных аудиоматериалах. Выявлено ограничение перевода (максимальная длительность обработки 6–8 минут). Для удаления ограничений была разработана методика переноса системы на локальные вычислительные мощности с развертыванием на Ubuntu.

Применение методики развертывания требует установки программного обеспечения:

– последняя версия Python 3.8 (отметить “Add Python to PATH” при установке). Это нужно для того, чтобы запускать Python из командной строки, т.к. дальнейшие действия будут проходить только в командной строке;

– Git – это специальная программа, которая позволит отслеживать любые изменения в файлах и хранить их версии;

– CUDA (в случае установленной NVIDIA GPU) – данная архитектура нужна для использования графических процессоров и для повышения производительности;

– FFmpeg для обработки аудио.

Для установки Whisper в командной строке нужно прописать “pip install”. Модели загружаются автоматически при первом использовании. Можно выбирать размер модели (tiny, base, small, medium, large) с помощью аргумента “--model”. Транскрипция аудио или видео файла запускается при помощи аргумента “whisper audio.mp3 (название и формат файла) --model medium --language ru”. Готовая транскрипция будет сохраняться в файле “audio.txt”

Методика развертывания представлена алгоритмами, изображенными на рис. 1, 2.

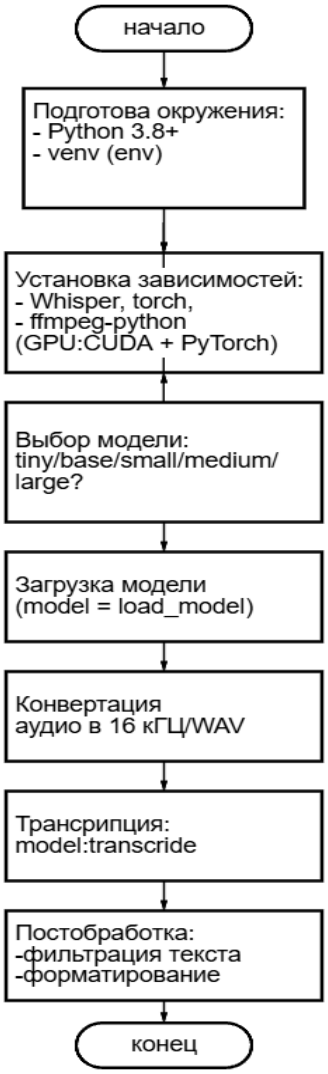


Рис. 1. Алгоритм методики развертывания

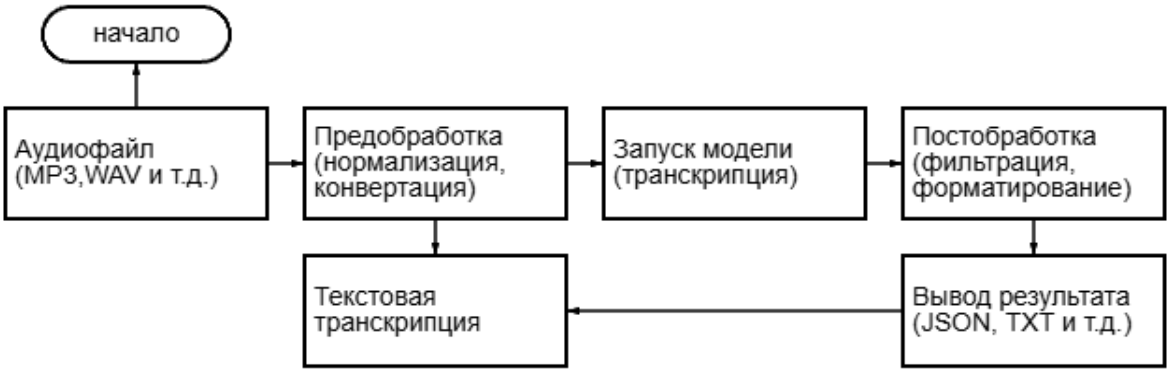


Рис. 2. Алгоритм транскрипции текста

Результаты

После развертывания при помощи Python был получен готовый терминал (рис. 3) на операционной системе Ubuntu. Модель Whisper работает при помощи простых команд например: whisper 123.m3(название файла) model medium.

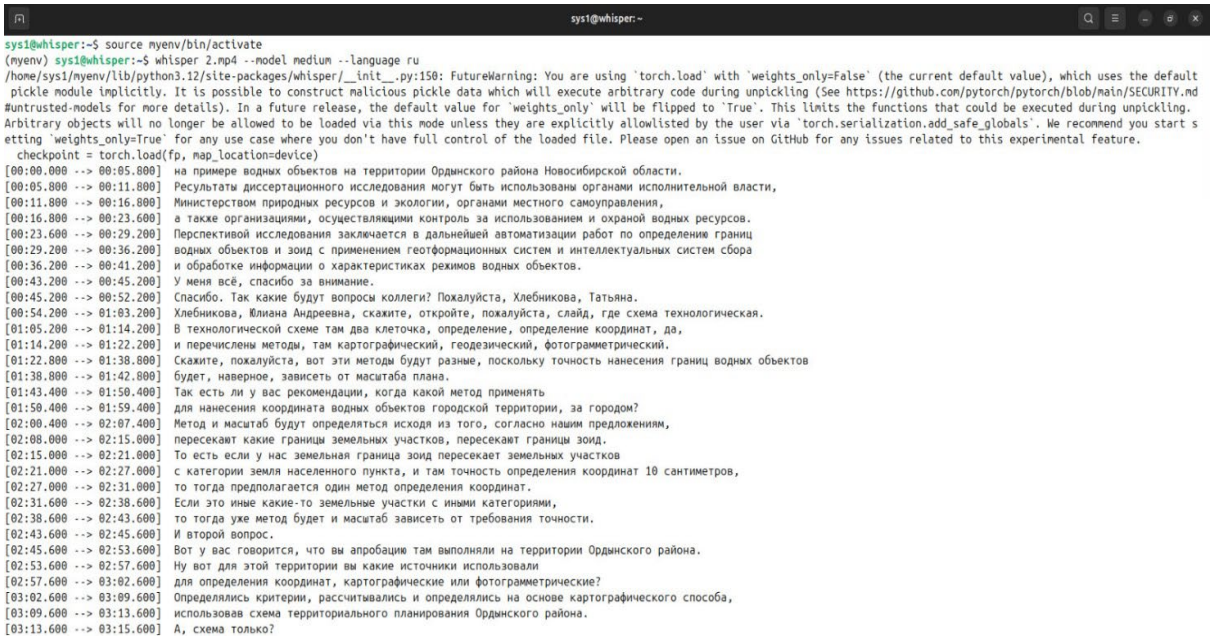
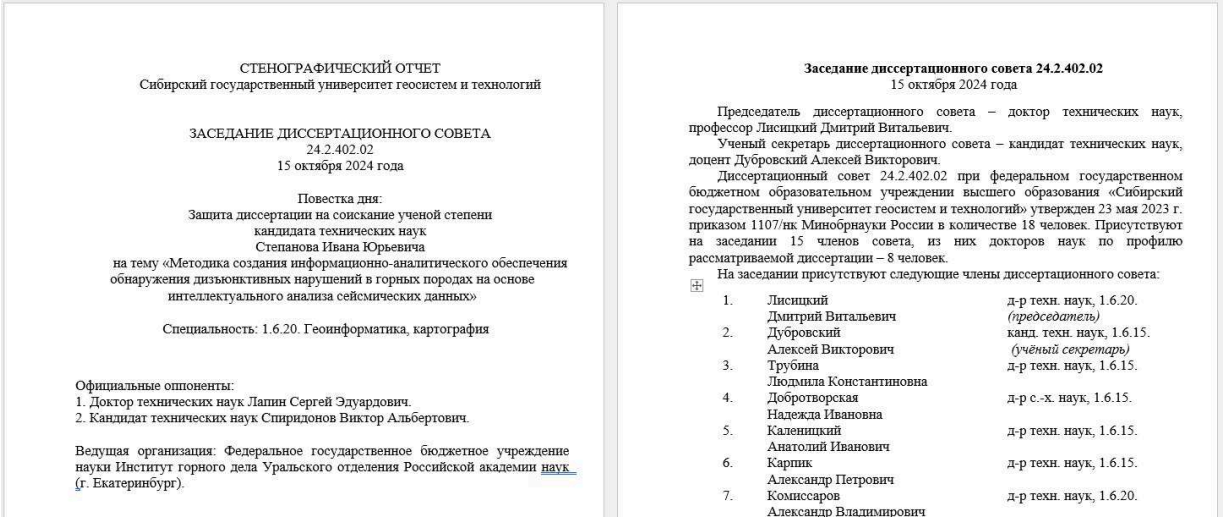


Рис. 3. Терминал Whisper

На основе готового перевода был составлен протокол стенографистом (рис. 4).



СТЕНОГРАФИЧЕСКИЙ ОТЧЕТ
Сибирский государственный университет геосистем и технологий

ЗАСЕДАНИЕ ДИССЕРТАЦИОННОГО СОВЕТА
24.2.402.02
15 октября 2024 года

Повестка дня:
Защита диссертации на соискание ученой степени
кандидата технических наук
Степанова Ивана Юрьевича
на тему «Методика создания информационно-аналитического обеспечения обнаружения дизъюнктивных нарушений в горных породах на основе интеллектуального анализа сейсмических данных»

Специальность: 1.6.20. Геоинформатика, картография

Официальные оппоненты:
1. Доктор технических наук Лапин Сергей Эдуардович.
2. Кандидат технических наук Спиридонов Виктор Альбертович.

Ведущая организация: Федеральное государственное бюджетное учреждение науки Институт горного дела Уральского отделения Российской академии наук (г. Екатеринбург).

Заседание диссертационного совета 24.2.402.02
15 октября 2024 года

Председатель диссертационного совета – доктор технических наук, профессор Лисицкий Дмитрий Витальевич.

Ученый секретарь диссертационного совета – кандидат технических наук, доцент Дубровский Алексей Викторович.

Диссертационный совет 24.2.402.02 при федеральном государственном бюджетном образовательном учреждении высшего образования «Сибирский государственный университет геосистем и технологий» утвержден 23 мая 2023 г. приказом 1107/нк Минобрнауки России в количестве 18 человек. Присутствуют на заседании 15 членов совета, из них докторов наук по профилю рассматриваемой диссертации – 8 человек.

На заседании присутствуют следующие члены диссертационного совета:

1.	Лисицкий Дмитрий Витальевич	д-р техн. наук, 1.6.20. *(председатель)*
2.	Дубровский Алексей Викторович	канд. техн. наук, 1.6.15. *(учёный секретарь)*
3.	Трубина Людмила Константиновна	д-р техн. наук, 1.6.15.
4.	Добротворская Надежда Ивановна	д-р с.-х. наук, 1.6.15.
5.	Калеицкий Анатолий Иванович	д-р техн. наук, 1.6.15.
6.	Карпик Александр Петрович	д-р техн. наук, 1.6.15.
7.	Комиссаров Александр Владимирович	д-р техн. наук, 1.6.20.

Рис. 4. Протокол

Модель Whisper была развернута на Windows 11. В ходе тестирования было зафиксировано время обработки, которое зависело от мощности оборудования: при использовании более производительных графических процессоров (GPU)

файл с большим объемом удалось обработать примерно за 8–9 минут. Результаты показали, что модель Whisper демонстрирует высокую точность перевода. Из огромных нюансов в работе является человеческий фактор, качество исходного файла и четкое произношение, оказали значительное влияние на результат.

Заключение

В результате проведенного исследования была успешно разработан и протестирован небольшой помощник составления протоков на базе модели Whisper, который продемонстрировал высокую точность и эффективность в решениях задач документирования совещаний и рабочих мероприятий, что позволило в 3–5 раз сократить временные затраты по сравнению с традиционным методом. Ключевым преимуществом модели Whisper являются открытый исходный код, возможность глубокой адаптации под специфические требования и локальное развертывание с обеспечением безопасности данных. Дальнейшее развитие включают оптимизацию для маломощного оборудования, разработку интуитивно понятного интерфейса [8]. На данный момент получилось существенно повысить эффективность работы с документацией при сохранении высокого качества протоколирования.

Полученные результаты свидетельствуют, что применение современных ИИ-технологий позволяет не только автоматизировать процесс документирования, но и существенно повысить качество организационного управления за счет точной и оперативной фиксации рабочих дискуссий.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Радфорд А. и др. Robust Speech Recognition via Large-Scale Weak Supervision (Whisper) / А. Радфорд и др. – arXiv, 2022. – URL: https://arxiv.org/abs/2209.12320 (дата обращения: 19.02.2021). – Текст: электронный.

2. OpenAI. Whisper: Robust Speech Recognition via Large-Scale Weak Supervision. – URL: https://cdn.openai.com/papers/whisper.pdf (дата обращения: 17.04.2024). – Текст: электронный.

3. OpenAI Whisper Documentation. – URL: https://github.com/openai/whisper (дата обращения: 09.01.2025). – Текст: электронный.

4. Microsoft Azure Speech-to-Text Documentation. – URL: https://learn.microsoft.com/en-us/azure/cognitive-services/speech-service/ (дата обращения: 11.02.2025). – Текст: электронный.

5. Otter.ai Technical Specifications. – URL: https://otter.ai/technology (дата обращения: 19.02.2024). – Текст: электронный.

6. Whisper: Speech Recognition Model. – URL: https://openai.com/research/whisper (дата обращения: 12.03.2025). – Текст: электронный.

7. Обзор модели Whisper для распознавания речи. – URL: https://habr.com/ru/articles/884992/ (дата обращения: 18.01.2025). – Текст: электронный.

8. docx2pdf 0.1 Documentation / AlJohri и др. – URL: https://pypi.org/project/docx2pdf/ (дата обращения: 19.02.2025). – Текст: электронный.