

2025-03-26

Лабораторная работа №6: Реализация голосового ввода для управления функциями информационных систем

1. Необходимо реализовать модуль, который записывает команду голосом с микрофона по умолчанию, превращает ее в текст, извлекает из текста ключевые слова, если находит в ключевых словах команду, выполняет ее.
2. Для этого нужно создать виртуальную среду `python -m venv ./venv`
3. Установить библиотеку для записи аудио `pyminiaudio` ([GitHub - irmen/pyminiaudio: python interface to the miniaudio audio playback, recording, decoding and conversion library](#)) или `pyaudio` ([PyAudio: Cross-platform audio I/O for Python, with PortAudio](#))
4. Установить платформу для распознавания речи Whisper ([GitHub - openai/whisper: Robust Speech Recognition via Large-Scale Weak Supervision](#)):
`pip install git+https://github.com/openai/whisper.git`
5. Реализовать препроцессинг голосовой команды и извлечение ключевых слов для вариантов ниже:
 1. для генерации ссылки на поиск, например, в Яндексе произвольного текста в браузере по умолчанию
 2. для открытия видеофайла из файловой системы в плеере по умолчанию или поиска видео на каком-либо видеосервисе
 3. для запуска любого приложения для коммуникации (ВК, Телеграм и т.п.) и ввода сообщения в окне мессенджера
6. Для этого каждую команду необходимо предварительно обработать: токенизировать (изучите NLTK и конкретно `nltk.tokenize`) и нормализовать (лемматизировать) (через `py morphology3`).
7. Реализовать через `subprocess` или иные средства выполнение указанных команд
8. Установить `ollama`: [Ollama](#), выбрать подходящие локальные модели (`deepseek-r1`, `qwen`, `Qwen-2.5-coder` и т.д.)
9. Реализовать взаимодействие с моделью через голосовой ввод и Ollama API.
10. Реализовать на языке C++ алгоритмы сортировки пузырьком, выбором, вставками, подсчетом, блочную сортировку и сортировку простым слиянием

посредством ввода голосовых инструкций LLM (возможные ошибки тоже корректировать через инструктирование модели). Процесс записать на видео

Note

Передавать в Whisper можно как .wav-файл, так и NumPy-массив

Но тут есть нюансы, поэтому рекомендую прочесть эти две ссылки:

- [How to send audio to Whisper in a numpy array ? · openai/whisper · Discussion #450 · GitHub](#)
- [python - How to feed a numpy array as audio for whisper model - Stack Overflow](#)

Сам Whisper можно попробовать заменить на WhisperX, он работает быстрее:

[GitHub - m-bain/whisperX: WhisperX: Automatic Speech Recognition with Word-level Timestamps \(& Diarization\)](#)

Tip

Если есть желание продолжить эксперименты с LLM, то как вдохновение можно рассмотреть ролики из серии, например, «самый иммерсивный Morrowind» ([С NPC можно говорить - и они отвечают голосом ч.1 \(Morrowind + Gemini + ElevenLabs\) - YouTube](#))

Также можно рассмотреть взаимодействие с другими приложениями через MCP-сервер ([Blender MCP | AI-Powered 3D Modeling with Claude|download|use case](#))