

Установка зависимостей

Для использования Whisper требуется:

1) установить набор ffmpeg:

```
sudo apt install ffmpeg
```

2) установить зависимости:

```
pip install -r requirements.txt
```

Информация о моделях

Существует 5 видов моделей, чем модель тяжелее, тем качественнее получается расшифровываемое сообщение и тем больше она весит, работает и требует оперативной памяти. Модели представлены на рисунке 1.

Size	Parameters	English-only model	Multilingual model	Required VRAM	Relative speed
tiny	39 M	<code>tiny.en</code>	<code>tiny</code>	~1 GB	~32x
base	74 M	<code>base.en</code>	<code>base</code>	~1 GB	~16x
small	244 M	<code>small.en</code>	<code>small</code>	~2 GB	~6x
medium	769 M	<code>medium.en</code>	<code>medium</code>	~5 GB	~2x
large	1550 M	N/A	<code>large</code>	~10 GB	1x

Рисунок 1 – Виды моделей.

Ограничения

Максимальный вес загружаемых файлов должен быть меньше 25 МВ.

Поддерживаемые форматы файлов: mp3, mp4, mpeg, mpga, m4a, wav, webm, ogg.

Запуск моделей

В запусках участвовали модели tiny, base, small, medium. Образцом для перевода служили 9 скачанных из телеграма аудиосообщений, имеющих расширение .ogg. Общая продолжительность аудиосообщений 5:03, первые аудиосообщений 6 – имеют очень хорошее качество записи, слова слышны отчетливо и четко.

Аудиосообщение №7 представляет собой отчетливое произношение исходного текста из рисунка 2.

Вопросы по оценкам:

- Сколько лабораторных работ
- Сколько контрольных работ
- Как получить пятёрку
- Со скольких баллов начинается тройка
- Как составляется оценка
- Сколько баллов можно получить за курс по гиту

Вопросы по организации:

- Можно ли перезащитить лабу
- Что могут спросить на защите
- Какой дедлайн у первой лабы
- Если я не принял приглашение в репозиторий

Общие вопросы

- Что такое курсовая работа
- Как проходит экзамен
- Что, если получил двойку за предмет
- Я заболел и пропустил защиту, что делать
- Можно ли идти на день качества, если не сдана курсовая работа
- Как проходит производственная практика

Рисунок 2 – Исходный текст аудиосообщения №7.

Аудиосообщение №8 представляет собой невнятное произношение с 'соединением' слов, с заиканием текста из рисунка 3.

- Как сделать пул реквест
- Как назвать ветку
- Что делать, если случайно смёрджил пул реквест
- Что делать, если мой пул реквест закрыт
- Как внести изменения в пул реквест
- Как разрешить конфликт между коммитами
- Как и какие метки ставить на пул реквест

Вопросы по оценкам:

- Сколько лабораторных работ
- Сколько контрольных работ
- Как получить пятёрку
- Со скольких баллов начинается тройка
- Как составляется оценка
- Сколько баллов можно получить за курс по гиту

Вопросы по организации:

- Можно ли перезащитить лабу
- Что могут спросить на защите
- Какой дедлайн у первой лабы
- Если я не принял приглашение в репозиторий

Общие вопросы

- Что такое курсовая работа
- Как проходит экзамен
- Что, если получил двойку за предмет
- Я заболел и пропустил защиту, что делать
- Можно ли идти на день качества, если не сдана курсовая работа
- Как проходит производственная практика

Рисунок 3 – Исходный текст аудиосообщения №8.

Аудиосообщение №9 представляет собой среднего качества произношение текста:

“ Как узнать сроки зачетной недели / экзаменационной сессии / дня качества / доп. сессии?

Как понять какая форма аттестации по предмету?

Если у меня перезащита курсовой по программированию на допсессии, допускаюсь ли я до основной сессии по остальным дисциплинам?

Я заболел и пропустил лекцию/практику/контрольную/лабораторную. Что мне делать, чтобы ликвидировать текущие задолженности?

Я заболел во время зачетной недели/сессии/доп. сессии/комиссии. Что мне делать в данном случае?

Я болею дольше месяца. Что делать?”

Результаты запуска (текстовые файлы, представляющие собой переведенные в текст аудиосообщения) находятся в папке *result*. Имя файла отражает запущенную для перевода модель. Образцы (аудиосообщения, скачанные из телеграма) находятся в папке *samples*.

Модель **tiny** отработала за 12 секунд, полученный текст низкого качества, слова разбиты на куски, несуществующие слова. Не воспринимаются слова, касающиеся вопросов, задаваемых на кафедре (pull request, ветку, смержил, и т. д.)

Модель **base** отработала за 16 секунд, получен текст более высокого качества количество разорванных слов значительно уменьшилось, но имена собственные не воспринимаются моделью, также страдают склонения глаголов. Не воспринимаются слова, касающиеся вопросов, задаваемых на кафедре (pull request, ветку, смержил, и т. д.)

Модель **small** отработала за 32 секунды, получен текст приемлемого качества, имена собственные (такие как Черчилль) воспринимаются моделью. Также модель начала распознавать знаки препинания. Плохо воспринимаются слова, касающиеся вопросов, задаваемых на кафедре (pull request, ветку, смержил, и т. д.)

Модель **medium** отработала за 204 секунды, получен текст высокого качества. Достаточно хорошо воспринимаются слова, касающиеся вопросов, задаваемых на кафедре (pull request, ветку, смержил, и т. д.)

Модель **large** не была запущена, так как у меня не хватает видеопамяти.

Параметры программы

`model_size` – определяет тип используемой для расшифровки модели (“tiny”, “base”, “small”, “medium”, “large”)

`model = whisper.load_model(model_size, device="cuda")` – “cuda” – использование GPU для вычислений, “cpu” - CPU.

`fp16=False` – Запрет на использование 16-битного числа с плавающей точкой. Используются 32-битные числа.