

DevWeek 2025: Кейс «Поиск видео без метаданных»

Команда: overfitters

Авторы

- Матвей Непомнящий — [@matnepo](#)
- Максим Писарик — [@pmv101](#)
- Всеволод Мартынов — [@AboHolo](#)
- Александр Прокошев — [@Calcifer02](#)
- Александр Фёдоров — [@alexanderflex](#)

Основная идея

Система обрабатывает видеофайл, извлекает из него визуальные и аудиопризнаки, преобразует их в векторные представления (эмбеддинги) и сохраняет в базе. Пользовательский текстовый запрос также преобразуется в embedding, и поиск осуществляется по близости в векторном пространстве с использованием FAISS.

Архитектура

Извлечение признаков

- Видео → кадры → описание (Video-LLaMA)
- Аудио → распознавание речи (Whisper)
- Генерация ключевых слов/хештегов (Qwen2.5)
- Объединение текста → эмбеддинг (MiniLM)

Индексирование

- Сохранение эмбеддингов видео в базу (FAISS + SQLite)

Поиск

- Запрос → embedding → ближайшие видео по cosine similarity

Используемые технологии

Языки

- Python 3.11+

Основные библиотеки и инструменты

- transformers, sentence-transformers, torch
- openai-whisper, faster-whisper
- faiss, numpy, pandas
- opencv-python, moviepy
- KeyBERT, yake, spacy, nltk

Модели

- Whisper large-v3
- Video-LLaMA 3-7B
- paraphrase-multilingual-MiniLM-L12-v2
- Qwen2.5-1.5B-Instruct

Инфраструктура

- FAISS (индексация и быстрый поиск)
- Google Colab / Jupyter
- Git + GitHub

Ветка загрузки видео

video_to_text.ipynb

Загружает список ссылок на видеофайлы в формате S3 и создаёт JSON-файл с описаниями видео. **Результат:** JSON-файл с базовой информацией о видео. **Формат JSON:**

```
[
  {
    "id": "1",
    "filename": "bde2e7b0-0774-4a72-bb3b-57515512acd9.mp4",
    "description": "Здесь текстовое описание видео.",
    "transcription": "Здесь текстовое описание аудио."
  },
  ...
]
```

merge_jsons.ipynb (опционально)

Объединяет несколько JSON-файлов. Используется при поэтапной обработке данных.

create_hashtags.ipynb

Добавляет к JSON хэштеги на основе описания и транскрипции. **Формат JSON:**

```
{
  "id": 6003,
  "filename": "97032cd0-2b26-4392-810b-835c9daaaeeb.mp4",
  "description": "Здесь текстовое описание видео.",
  "transcription": "Здесь текстовое описание аудио.",
  "tags": ["tag_1", "tag_2", ...]
}
```

generate_embeddings.ipynb

Генерирует эмбединги для полей "description" и "transcription".

Ветка поиска

search.ipynb

Работает с подготовленным JSON и осуществляет поиск по:

- описанию
- транскрипции
- комбинированным методам

Пример вывода JSON:

```
{
  "search_results": [
    {
      "id": "6003",
      "filename": "97032cd0-2b26-4392-810b-835c9daaaeeb.mp4",
      "relevance": 0.95
    },
    ...
  ]
}
```

Дополнительные исследования

Проводились эксперименты с различными моделями генерации хэштегов (см. `experiments/*.ipynb`).

Развёрнутое приложение

<https://github.com/v-martin/devweak-yappy>

Идеи улучшения

- Использование более мощных моделей
- Увеличение набора данных
- Масштабирование и интеграция в Yappy
- Разные веса аудио/видео эмбеддингов
- Дообучение моделей под актуальный сленг и контекст

Уникальность решения

Система полностью автоматически формирует метаданные (описание, транскрипцию, хэштеги) и осуществляет поиск по смыслу. Архитектура модульная и масштабируемая. Решение готово к внедрению в реальные сервисы.