

Автоматическая генерация конспекта по видеокурсу stepik.org

Чаркин Константин

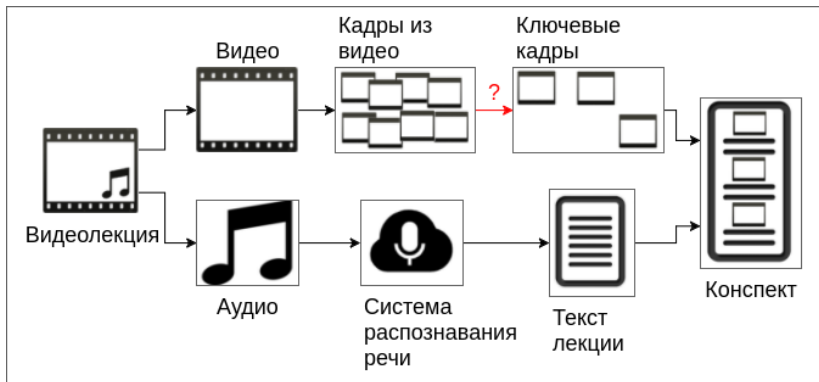
руководитель: Николай Иванович Вяххи

СПб АУ НОЦНТ РАН

2017 г.

- *Stepik.org* – образовательная платформа, на которой есть онлайн-курсы, в том числе видеокурсы.
- У видеокурсов есть особенности, из-за которых их освоение проходит значительно лучше при наличии у студентов конспекта.
- Просить преподавателей предоставить конспект их курса не хотим, т.к. это усложнит им создание курсов
- Решение – генерировать конспекты автоматичеки

Идея автоматической генерации конспекта



Ключевые кадры – минимальный набор кадров, на которых есть вся полезная информация.

Цель:

- Сгенерировать и предоставить пользователям доступ к конспектам видеокурсов

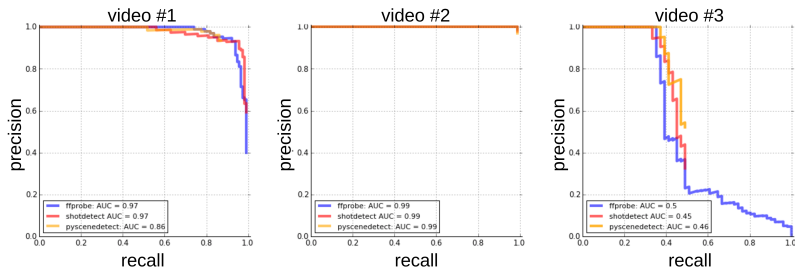
Задачи:

- Реализовать извлечение ключевых кадров
- Реализовать распознавание аудиоряда
- Собрать результаты предыдущих задач в единый документ
- Предоставить пользователям доступ к чтению и совместному редактированию этого документа
- Оформить генератор конспектов в виде инструмента удобного для использования, поддержки и развития.

Существующие решения:

- ffprobe, ffmpeg.org/ffprobe.html
- Shotdetect, johmathe.name/shotdetect.html
- PySceneDetect, pyscenedetect.readthedocs.io

Сравнение:



github.com/albanie/shot-detection-benchmarks

Подходы в исследованиях

- По типу решаемой задачи:
 - В общем случае
 - G.Ciocca et al., An Innovative Algorithm For Key Frame Extraction In Video Summarization
 - В частном случае
 - A.Ekin et al., Automatic Soccer Video Analysis and Summarization
 - S.X.Ju et al., Summarization of Videotaped Presentations: Automatic Analysis of Motion and Gesture
- По способу решения задачи:
 - Видео – это множество кадров – кластеризация
 - P.Mundur et al., Keyframe-based video summarization using Delaunay clustering
 - Видео – это последовательность кадров – сравнение соседних кадров
 - K.Khurana et al., Key Frame Extraction Methodology For Video Annotation
 - H.Zhang et al., Video Parsing and Browsing Using Compressed Data

Извлечение ключевых кадров

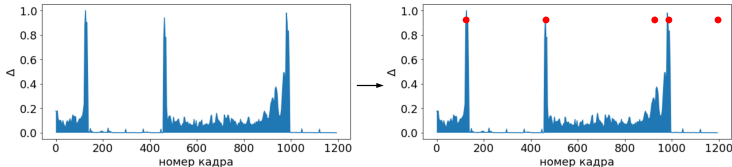
Общая схема извлечения ключевых кадров в случае представления видео в качестве упорядоченной последовательности кадров:

1 Исходное видео – последовательность кадров:

$$F_1 \rightarrow F_2 \rightarrow F_3 \rightarrow \dots \rightarrow F_N$$

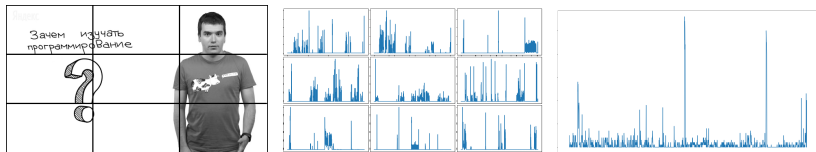
2 $\Delta(i)$ – разница между F_i и F_{i+1}

3



4 На основании графика из пункта 3 делается вывод о том, какие кадры следует добавить во множество ключевых

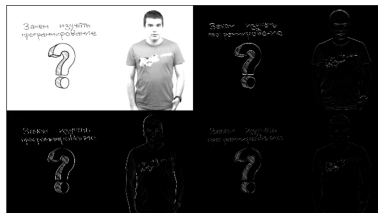
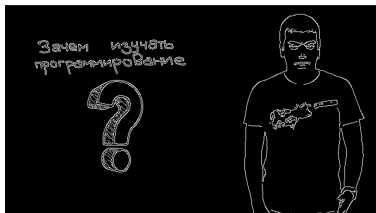
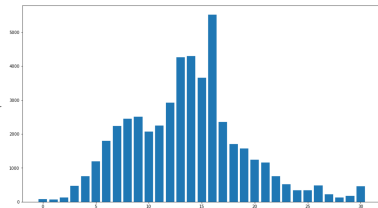
Анализ кадра



- Разбиение кадра на блоки
- Вычисление Δ для каждого блока
- Бинаризация Δ для каждого блока
- Вычисление Δ для всего кадра

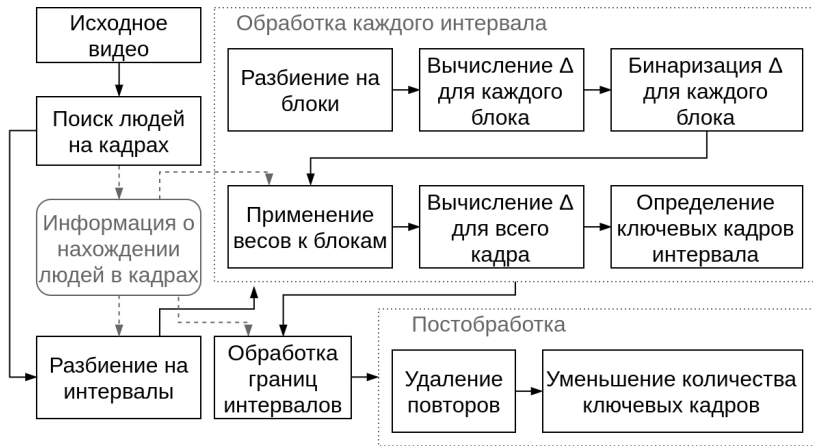
Извлечение ключевых кадров

Вычисление различия кадров.



Извлечение ключевых кадров

Схема алгоритма извлечения ключевых кадров



Извлечение ключевых кадров

Сравнение

Данные: 20 видеолекций, общая продолжительность ~2 часа.

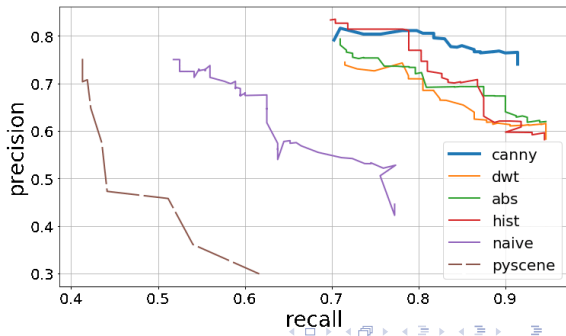
$I = \{[x_1, x'_1], [x_2, x'_2], \dots, [x_n, x'_n]\}$ – разметка: набор интервалов таких, что идеальный конспект – это ровно по 1 кадру из каждого интервала.

F – множество кадров, выданных алгоритмом

T – множество полезных кадров, выданных алгоритмом

$$recall = \frac{|T|}{|I|}$$

$$precision = \frac{|T|}{|F|}$$



Распознавание аудио

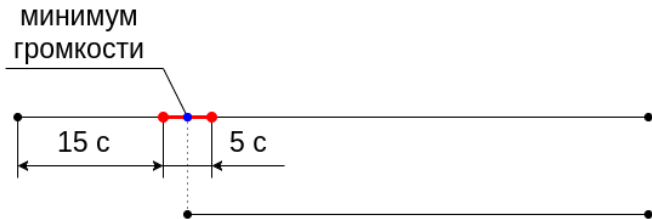
Требования:

- Бесплатность
- Распознавание русского языка

Ограничения:

- 1000 запросов в сутки
- 20 секунд на запрос

*YandexSpeechKit*²



Нарезать аудио по 20 секунд плохо, т.к. часто слова будут разбиваться на 2 части и неправильно распознаваться.

²www.tech.yandex.ru/speechkit/

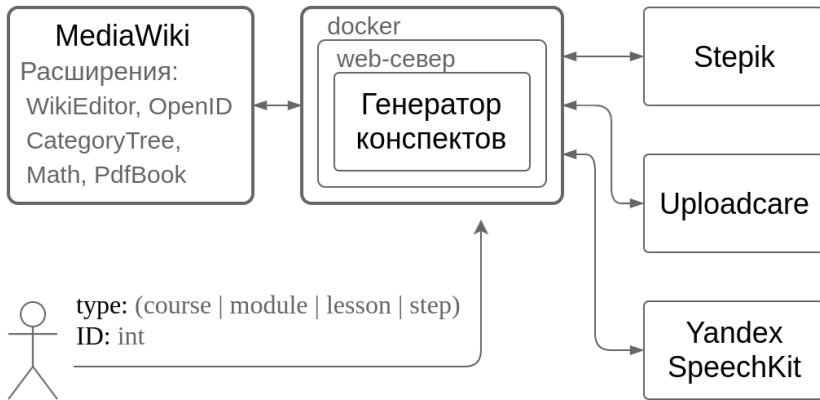
Требования:

- Совместное редактирование
- Разрешения конфликтов
- История изменений
- Система прав доступа
- Поддержка \LaTeX
- Экспорт в pdf
- Поддержка сложной структуры страниц

*MediaWiki*³

³www.mediawiki.org

Сервис по генерации конспекта



Страница курса

The screenshot shows the main page of a course on Stepik. The course title is "Category: Web-технологии (C-154)". Below the title, it says "Page for course 'Web-технологии' with id = 154" and provides the URL "https://stepik.org/course/154/". The page lists various topics under the category, including "Веб-технологии (C-154)", "Статический сайт (M-521)", "Вводный урок курса (L-18632)", "Архитектура Web-приложений (L-12527)", "Step 1 (S-44942)", "Step 2 (S-44678)", "Step 3 (S-44678)", "Step 4 (S-44680)", "Step 5 (S-44681)", "Step 6 (S-44082)", "Step 7 (S-57362)", "Step 8 (S-57386)", "Step 9 (S-57381)", "Step 10 (S-57390)", "Web-клиенты (L-12528)", "Язык разметки HTML (L-14820)", "CSS - язык описания стилей (L-14821)", "Сетевые протоколы (L-14823)", "Протокол HTTP (L-14824)", "Web-сервера (L-14825)", "Архитектура front-end-backend (L-14826)", "Динамический сайт (M-852)", and "Дополнительные темы (M-883)". There is also a "Subcategories" section indicating 3 subcategories and a search bar for "Category: Courses".

Страница шага

The screenshot shows a specific step page titled "Step 3 (S-44678)". The URL is "https://stepik.org/lesson/12527/step/3/". The main content area features a video player with the title "Документы" and a video thumbnail of a man. Below the video, there is a text block starting with "[00:00 - 00:19] давайте подробно остановимся на том какие бывают документы что он из себя представляет а поскольку это слово сетя вот в в 1 очередь нужно сказать что документы бывают разных типов для". Below this is a section titled "Типы документов (MME-типы)" with a list of document types: "application", "audio", "image", "text/css", "text/html", "text/javascript", "text/plain", "text/xml", "video", "video/mp4", and "video/webm". There is another video thumbnail below the list. The text continues with "[00:19 - 00:36] того чтобы указать тип документа используется так называемый mime типа а изначально были придуманы для передачи различной информации через а электронную почту". The next text block is "[00:36 - 00:52] электронная почта текстовый протокол а для того чтобы передать скваем изображение его нужно как то кодировать и указать что эти данные являются изображением была выведена система типов которая потом перешла". The following text block is "[00:52 - 01:08] также в марте эта страница обычно состояла из 2 слов разделенных ссылкой а вот например для документов типа text/css". The final text block is "[01:08 - 01:24] а документ указывает получателю то есть клиенту а способ которым можно интерпретировать данные документы например мы документ можно отображать так вот страничку либо".

<https://wiki.stepik.org>

- Изучено большое количество исследований по теме извлечения ключевых кадров
- Разработан и реализован алгоритм по извлечению ключевых кадров для задачи генерации конспекта по видеокурсу, который, по результатам проведенного сравнения, работает лучше, чем существующие решения.
- В ходе решения поставленных задач была освоена работа со множеством различных инструментов и сервисов, среди которых YandexSpeechKit, Stepik, Uploadcare, MediaWiki, Docker, OpenCV, dlib и др.
- Реализован сервис по генерации конспектов, который инкапсулирует в себе всю логику по созданию конспектов, предоставляя пользователю максимально простой интерфейс.

Q & A