

УДК 004.4

СИСТЕМА АВТОМАТИЧЕСКОГО СУБТИТРИРОВАНИЯ ВИДЕОФАЙЛОВ¹

Коробкин Дмитрий Михайлович

*Кандидат технических наук, доцент;
Волгоградский государственный технический университет;
Россия, 400005, Волгоград, пр. им. Ленина, 28;
e-mail: dkorobkin80@mail.ru.*

В данной работе содержится описание процесса автоматического субтитрирования (распознавания звуковой составляющей) видеофайлов в процессе телепроизводства ГТРК «Волгоград-ТРВ». Значимость работы заключается в программно реализованных алгоритмах извлечения аудиодорожки из видеофайла, распознавания речи в аудиофайле, коррекции распознанного текста, добавления субтитров в видеофайл. Разработанный программный модуль автоматического субтитрирования видеофайлов использует технологии: модель распознавания русского языка Golos («Sber»), медиасервер с открытым исходным кодом Emby.

Ключевые слова: субтитры, Golos, Emby.

Для цитирования:

Коробкин Д. М. Система автоматического субтитрирования видеофайлов// Системный анализ в науке и образовании: сетевое научное издание. 2022. № 2. С. 23–27. URL : <http://sanse.ru/download/469>.

SYSTEM OF AUTOMATIC SUBTITLING OF VIDEO FILES

Korobkin Dmitry M.

*PhD in Engineering sciences, associate professor;
Volgograd State Technical University;
28 Lenina Av., Volgograd, 400005, Russia;
e-mail: dkorobkin80@mail.ru.*

This paper contains a description of the process of automatic subtitling (recognition of the audio component) of video files in the process of TV production of the State Television and Radio Broadcasting Company "Volgograd-TRV". The theoretical significance of the work lies in the developed algorithms for extracting an audio track from a video file, speech recognition in an audio file, correcting recognized text, adding subtitles to a video file. The developed software module for automatic subtitling of video files uses the following technologies: the Russian language recognition model Golos ("Sber"), the Emby open source media server.

Keywords: subtitling, Golos, Emby.

For citation:

Korobkin D. M. System of automatic subtitling of video files. System Analysis in Science and Education, 2022;(2):23–27(In Russ). Available from: <http://sanse.ru/download/469>.

Введение

¹Выпуск подготовлен в рамках реализации гранта на разработку программ бакалавриата и программ магистратуры по профилю «Искусственный интеллект», а также на повышение квалификации педагогических работников образовательных организаций высшего образования в сфере искусственного интеллекта (конкурс 2021-ИИ-01 от 10.06.2021)

Процесс телепроизводства, принятый в филиале ВГТРК «ГТРК «Волгоград-ТРВ», требует создания так называемых «микрофонных папок», в которых шеф-редактор согласовывает выход в эфир информационных сюжетов и тематических программ. Обычно в процессе телесъемки осуществляется запись как закадрового текста корреспондента (автора), так и синхрон («прямая речь») с участниками сюжета (программы). Далее эта речь должна быть расшифрована корреспондентом (автором), и тексты синхрон вписаны в микрофонную папку. Этот процесс «ручного» распознавания синхрон весьма трудозатратен. Возможность использования сторонних сервисов, типа youtube, затруднена тем, что программа (сюжет) попадает на эти ресурсы только после выхода программы в телеэфир. Соответственно, такое «несвоевременное» распознавание синхрон грозит штрафами как корреспонденту со стороны администрации ГТРК, так и самой ГТРК со стороны контролирующих органов.

Поэтому, актуальной является проблема автоматического субтитрирования (распознавания звуковой составляющей) видеофайлов в процессе телепроизводства ГТРК «Волгоград-ТРВ».

В результате анализа предметной области и имеющихся аналогов, прототипов и методов реализации поставленной цели были сформулированы следующие задачи и требования:

1) Использовать модель распознавания русской речи *Golos* (<https://developers.sber.ru/portal/products/golos>), позволяющей свободное использование в исследовательских и коммерческих целях.

2) Обеспечить взаимодействие с ресурсом обмена видеоматериалами ГТРК «Волгоград-ТРВ» на основе платформы с открытым исходным кодом Emby. Данная технология также используется и головным предприятием – ВГТРК. На платформу заливаются видеоматериалы для отмотра корреспондентами (авторами), трафиками (специалистами формирования плейлистов телеканалов) и другими заинтересованными лицами.

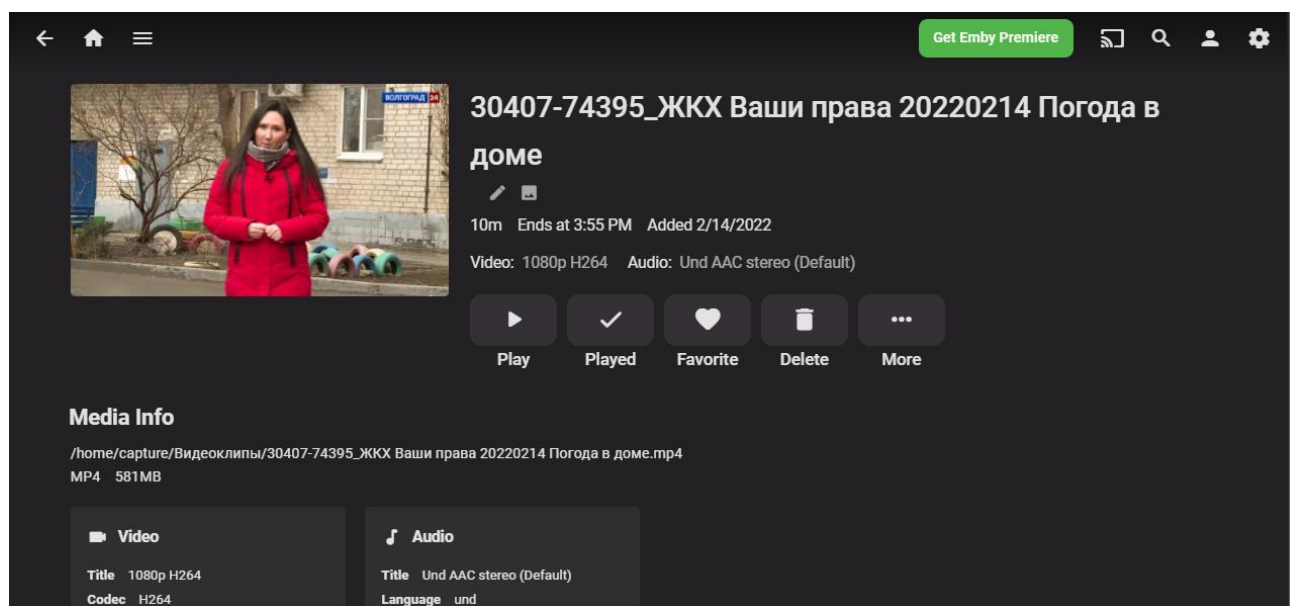


Рис. 1. Медиасервер Emby

3) Обеспечить коррекцию (проверка орфографии и правописания) сформированного текста на основе распознанной речи (*Speech-To-Text*). Данное требование возникает вследствие формирования на основе модели *Golos* текста с большим количеством орфографических ошибок.

1. Проверка функциональности программного модуля

Для разработки программы использовался язык программирования *Python* с дополнительными библиотеками *ffmpeg*, *pysrt*, *pyxdameraulevenshtein*, *nemo_toolkit[asr]*.

Программа состоит из взаимосвязанных скриптов, отвечающих за конвертирование видео в аудио и обратно, а также за запросы к серверу с обученной моделью *Golos*, который генерирует текст из монофонического формата аудиодорожки с расширением *wav*.

Программный модуль выполняет следующие функции:

- извлечение аудиодорожки из *mp4* видеофайла;
- конвертация аудиодорожки в монозвук формата *wav*;
- генерация субтитров;
- коррекция текста (алгоритм Дамерау-Левенштейна);
- преобразование сгенерированного текста из видеофайла в субтитры *srt*;
- преобразование *srt* формата субтитров к формату *ass*;
- создания видеофайла с внедренными субтитрами.



Рис. 2. Алгоритм извлечения аудиодорожки



Рис. 3. Алгоритм распознавания речи в аудиофайле



Рис. 4. Алгоритм коррекции распознанного текста



Рис. 5. Алгоритмы формирования субтитров

Программе передается ссылка на видеофайл в системе *Emby*, для которого необходимо осуществить процесс субтитрирования. Для этого необходимо выбрать заранее загруженное видео и скопировать ссылку на него из адресной строки, например, <http://88.87.79.40:8096/web/index.html#!/item?id=9&serverId=62323a6950f148399a5211ce89603558>.

После запуска приложение выведет на экран настройки, необходимые для дальнейшей работы (см. рис. 6).

```
В данный момент программа использует следующие данные:  
Api_Key - faa94a49edec4a80aade7420e6a5b4a6  
  
Video_Path - /home/gurke/TempEMBY  
  
Emby_Path - /home/gurke/MediaEMBY  
Желаете их изменить? y/n
```

Рис. 6. Настройки программы

Далее требуется указать ссылку на видеоматериал с медиасервера *Emby*, скопированную ранее. А также указать финальное название видеофайла.

В Ростове-на-Дону назовут победителей Всероссийского научно-технического конкурса «Первый шаг»



в россии в третий раз на федеральном уровне проходит молодежный научно технический конкурс первый шаг это возможность оценить лучшие разработки в области кинопроизводства телерадиовещания и телекоммуникаций сейчас юг россии готовится к финальным мероприятиям они пройдут в конце мая в ростове-на-дону конкурс первый шаг ежегодные его учредители российская государственная телевизионная

Рис. 7. Результат работы программы

Дальнейшие этапы будут выполняться автоматически, вплоть до загрузки финального видео в папку медиатеки *Emby*. Стоит отметить, что процесс преобразования видео может занять продолжительное время. После завершения работы программы пользователь получит исходное видео с встроенной дорожкой субтитров, все промежуточные файлы будут автоматически удалены.

Заключение

Теоретическая значимость работы заключается в разработанных алгоритмах извлечения аудиодорожки из видеофайла, распознавания речи в аудиофайле, коррекции распознанного текста, добавления субтитров в видеофайл.

Практическая значимость работы заключается в разработанном программном модуле автоматического субтитрирования видеофайлов. Использовались технологии: модель распознавания русского языка *Golos* («Sber»), медиасервер с открытым исходным кодом *Emby*.

Список источников

1. Денисенко А. Golos – самый большой русскоязычный речевой датасет, размеченный вручную // Хабр : [сайт]. Habr, 2006–2022. Дата публикации: 27.05.2021. URL: <https://habr.com/ru/company/sberdevices/blog/559496/>.
2. Emby - The open media solution : [сайт]. Emby LLC, 2022. URL: <https://emby.media/>.
3. ffmpeg Documentation. URL: <https://ffmpeg.org/ffmpeg.html>.