

Инструмент анализа голосов на аудиозаписи

Лидер, студент гр. 6304

Пискунов Ярослав Андреевич

Команда, студенты гр. 8303

Быков Андрей Владимирович

Деркач Никита Владимирович

Логинов Евгений Анатольевич

Суть проекта

Цель: создать приложение командной строки, которое будет принимать на вход аудиофайлы (mp3) и обрабатывать их.

Функционал:

- идентификация диктора по набору контрольных образцов,
- идентификация наличия нескольких дикторов на аудио,
- вычисление и вывод фрагментов аудиозаписи для каждого из дикторов в файлы,
- распознавание речи для каждого диктора (с указанием меток времени) с выводом в файл.

Задачи на вторую итерацию

- Разработка функционала идентификации наличия нескольких дикторов на аудио;
- Разработка функционала идентификации диктора по набору контрольных образцов;
- Сбор звуковых файлов для тестирования;
- Написание юнит-тестов для разработанного функционала.

Макет

```
yp@yp-VirtualBox:~/Documents/mse_voice_analysis$ mse_voice_analysis -h
usage: mse_voice_analysis [-h] [-i] [-m] [-d] [-s] file

positional arguments:
  file                  mp3 file

optional arguments:
  -h, --help            show this help message and exit
  -i, --identify        Function which can identify dictor
  -m, --many            Function which can identify many dictors
  -d, --distribution    Function which can distribute all dictors into separate
                        files
  -s, --speech          Function which provide speech recognition for each
                        dictor
```

Рисунок 1 - Справка

```
yp@yp-VirtualBox:~/Documents/mse_voice_analysis$ mse_voice_analysis ../../Downloads/Tony_Igy_-_Astronomia_14893.mp3 --identify
Processing file
Speaker is Ivan Ivanov
```

Рисунок 2 – Идентификация дикторов по набору контрольных образцов

```
yp@yp-VirtualBox:~/Documents/mse_voice_analysis$ mse_voice_analysis ../../Downloads/Tony_Igy_-_Astronomia_14893.mp3 --many
Processing file
There are 5 speakers on this record
```

Рисунок 3 - Идентификация наличия нескольких дикторов на аудио

Методы и технологии

Resemblyzer – библиотека для языка python, которая обрабатывает звуковые файлы, используя глубокое обучение.

Основной метод, используемый для того, чтобы отличить говорящих – анализ частот. Для каждого звукового файла создается вектор из 256 значений, которые описывают характеристики звучащего голоса.

Достигнутый результат

- ✓ Разработан функционал, позволяющий делать выводы о количестве говорящих;
- ✓ Разработан функционал, позволяющий делать выводы о сходстве голоса на звуковом файле с контрольными образцами;
- ✓ Собраны звуковые файлы для тестов;
- ✓ Для разработанного функционала созданы юнит-тесты;

Все поставленные задачи выполнены в срок.

Репозиторий с проектом - https://github.com/moevm/mse_voice_analysis

Достигнутый результат. Юнит-тесты

```
Testing started at 14:35 ...  
C:\Users\ab319ur\PycharmProjects\pythonProject\venv\Scripts\python.exe "D:\PyCharm\PyCharm 2020.2.2\plugins\python\helpers\pydev\pydevconsole.py"  
Launching unittests with arguments python -m unittest C:/Users/ab319ur/PycharmProjects/pythonProject/manyDictorsTest.py  
  
Ran 3 tests in 97.166s  
  
OK
```

Рисунок 4 – Тесты идентификации наличия нескольких дикторов на аудио

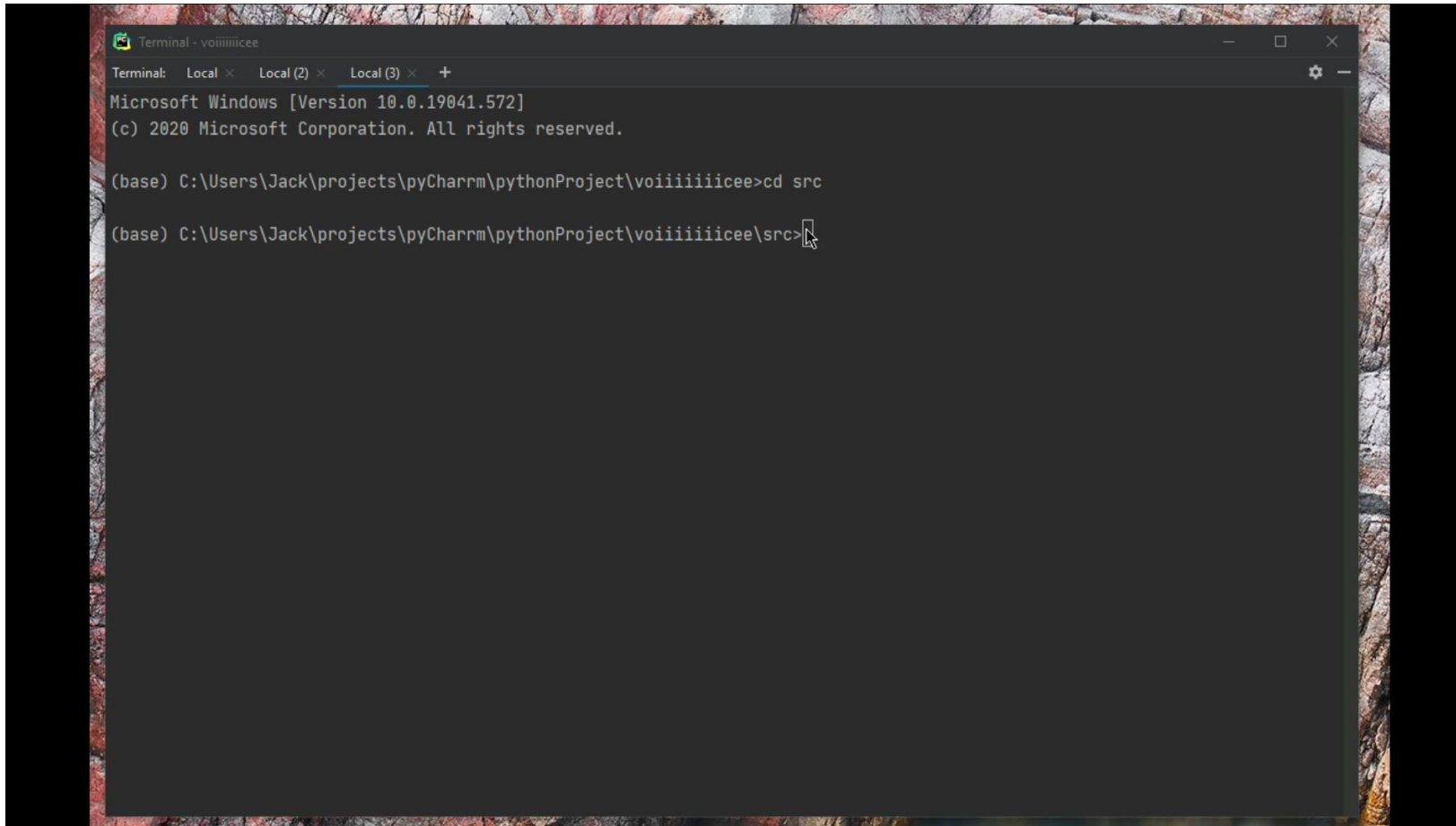
```
Testing started at 14:41 ...  
C:\Users\BigSmoke\PycharmProjects\pythonProject\venv\Scripts\python.exe "D:\pycharm\PyCharm 2020.2.2\plugins\python\helpers\pydev\pydevconsole.py"  
Launching unittests with arguments python -m unittest test_oneDictor.Test.test_one_dictor_identification in C:/Users/BigSmoke/PycharmProjects/pythonProject/test_oneDictor.py  
  
Ran 1 test in 20.024s  
  
OK
```

Рисунок 5 – Тесты идентификации диктора по набору контрольных образцов

Звуковые файлы, использованные в тестах

https://github.com/moevm/mse_voice_analysis/tree/debug/src/audio_data_test

Достигнутый результат. Демонстрация

A screenshot of a Windows terminal window titled "Terminal - voiiiiiiiicee". The terminal shows the following text: "Microsoft Windows [Version 10.0.19041.572] (c) 2020 Microsoft Corporation. All rights reserved. (base) C:\Users\Jack\projects\pyCharm\pythonProject\voiiiiiiiicee>cd src (base) C:\Users\Jack\projects\pyCharm\pythonProject\voiiiiiiiicee\src>". The cursor is at the end of the second command line.

```
Terminal - voiiiiiiiicee
Microsoft Windows [Version 10.0.19041.572]
(c) 2020 Microsoft Corporation. All rights reserved.

(base) C:\Users\Jack\projects\pyCharm\pythonProject\voiiiiiiiicee>cd src
(base) C:\Users\Jack\projects\pyCharm\pythonProject\voiiiiiiiicee\src>
```


Задачи на третью итерацию

- Разработка функционала вычисления и вывода фрагментов аудиозаписи для каждого из дикторов в файлы;
- Разработка функционала распознавания речи для каждого диктора (с указанием меток времени) с выводом в файл;
- Написание юнит-тестов для указанного выше функционала с использованием собранных звуковых файлов;
- Расширение количества юнит-тестов для разработанного на второй итерации функционала за счет собранных звуковых файлов;
- Приведение функционала в соответствие с макетом.

Подробнее о задачах - https://github.com/moevm/mse_voice_analysis/projects

Q & A