

Transcrição de Áudio utilizando a API Google Cloud Speech-To-Text

Aluno: Gustavo Nunes de Oliveira

Matrícula: 11621BCC032

1. Introdução

O objetivo do trabalho é fazer a transcrição automática de um áudio para texto via streaming(simulando a legenda automática de um vídeo).

Para realizar essa tarefa será utilizada a API Speech-To-Text provida pela Google na plataforma da Google Cloud.

A linguagem escolhida para implementação do projeto foi a linguagem Python, devido à sua simplicidade e aos recursos oferecidos pela linguagem. Dentre os recursos oferecidos pela linguagem utilizaremos para o objetivo principal do projeto os módulos:

- PyAudio: utilizado para fazer o streaming de áudio do microfone para a entrada da aplicação
- Google Cloud Speech-To-Text API: API da google para processamento de linguagem natural.

2. Preparação do ambiente

A primeira coisa a se fazer é criar uma conta na Google Cloud(<https://cloud.google.com>). Após a criação da conta, siga os passos no tutorial oferecido pela plataforma para preparar o ambiente no seu computador(link = https://cloud.google.com/speech-to-text/docs/quickstart-protocol?_ga=2.44512803.-649697660.1561550911&_gac=1.47585365.1562171018.Cj0KCQjwpPHoBRC3ARIsALfx-LA3apTvOpiHTbzfXx-l1nIh0FWuFub7iuK6LvIZI5fEaTSIENHMRlaAskZEALw_wcB)

3. Testando a aplicação

A aplicação foi feita para receber um arquivo de áudio com as seguintes características:

- Idioma suportado pela API da Google(en-US, pt-BR, etc)
- Frequência do arquivo igual a 16000 Hz
- Apenas 1 canal de audio(mono)
- Áudio de 16 ou 24 bits
- Codificação FLAC
- Duração do áudio menor que 1 minuto

Além disso, para fazer uso da API da Google é necessário gerar uma chave para seu projeto na Google Cloud(vá até a página da API desejada no console do seu projeto, credenciais, gerenciar contas de serviço, actions, criar chave, criar chave .json).

Salve a chave criada na pasta do programa com o nome "key.json".

Para instalar os módulos necessários execute os seguintes comandos(no linux):

```
$ virtualenv env
$ source env/bin/activate
(env) $ pip install --upgrade google-cloud-speech
(env) $ pip install --upgrade pyaudio
(env) $ export GOOGLE_APPLICATION_CREDENTIALS="key.json"
```

Feito isso o seu programa irá executar em um ambiente virtual do python (virtualenv).

Para realizar a conversão de um áudio vindo de um vídeo, utilize o programa ffmpeg da seguinte maneira:

```
$ ffmpeg -i video.mp4 -ac 1 -ar 16000 -vn -acodec flac output_audio.flac
```

Dessa forma já é possível fazer a transcrição do arquivo "output_audio.flac" utilizando o programa "transcricao_audio.py" do próximo capítulo.

Para realizar a transcrição em tempo real apenas execute o programa "mic_streaming.py" dos capítulos seguintes.

4. Códigos

4.1 transcricao_audio.py

Código fonte disponível no link:

https://github.com/nunesg/IA/blob/master/Trabalho/transcricao_audio.py

4.2 mic_streaming.py

Código fonte disponível no link:

https://github.com/nunesg/IA/blob/master/Trabalho/mic_streaming.py