

Métodos de NLP para la transcripción y sumarización de lenguaje

24 de noviembre de 2023

1. Introducción

Vamos a crear una app en streamlit a la cual le podremos subir un archivo de audio y esta transcribirá el audio y además generará los puntos claves de lo que se dijo en este audio. Para este caso utilizaremos el modelo *whisper-1* de openai y *gtp3.5-turbo* para realizar los puntos clave.

2. Código

Es bastante simple nuestro código el cual queda de esta forma:

```
import streamlit as st
import openai
import ssl
import certifi

ssl._create_default_https_context = lambda: ssl.create_default_context(cafile=certifi.where())
print(ssl.get_default_verify_paths())

st.title("Generador de puntos clave a través de audio")

OPENAI_API_KEY = "sk-4XknLFjnwWiFMdQS3BHcT3B1bkFJN3Xu1LGJjXiyonGCX7kL"
client = openai.OpenAI(api_key=OPENAI_API_KEY)

def transcribe_audio(file):
    transcript = client.audio.transcriptions.create(
        model="whisper-1",
        file=file)

    return transcript.text

def custom_chatgpt(user_input):
    messages = [
        {"role": "system", "content": "You are an office administrator, summarize the text in key points. Please give your answer in the language the text is in."},
        {"role": "user", "content": user_input}
    ]
    response = client.chat.completions.create(
        model="gpt-3.5-turbo",
        messages=messages
    )
```

```
chatgpt_reply = response.choices[0].message.content
return chatgpt_reply

uploaded_audio = st.file_uploader("Elige_un_archivo_de_audio", accept_multiple_files=False)

if uploaded_audio:
    st.audio(uploaded_audio, format="audio")

    # Transcribe audio
    st.subheader("Transcripción:")
    transcription = transcribe_audio(uploaded_audio)
    st.write(transcription)

    # Summarize using ChatGPT
    st.subheader("Puntos_clave:")
    summary = custom_chatgpt(transcription)
    st.write(summary)
```

Como vemos primero declaramos un objeto `file_uploader` para poder agregar el archivo de audio que queremos. Luego, añadimos la opción de desplegar el audio para que pueda ser escuchado en la app.

Posteriormente generamos la transcripción utilizando el modelo *whisper-1* el cual tiene soporte para varios lenguajes. Luego mandamos esto a la api de chat gtp en la cual le pedimos que nos genere los puntos clave en el idioma en el que reciba el audio y guardamos la respuesta en memoria.

3. Fotografías de app funcionando



Figura 1: Muestra de app al iniciar.

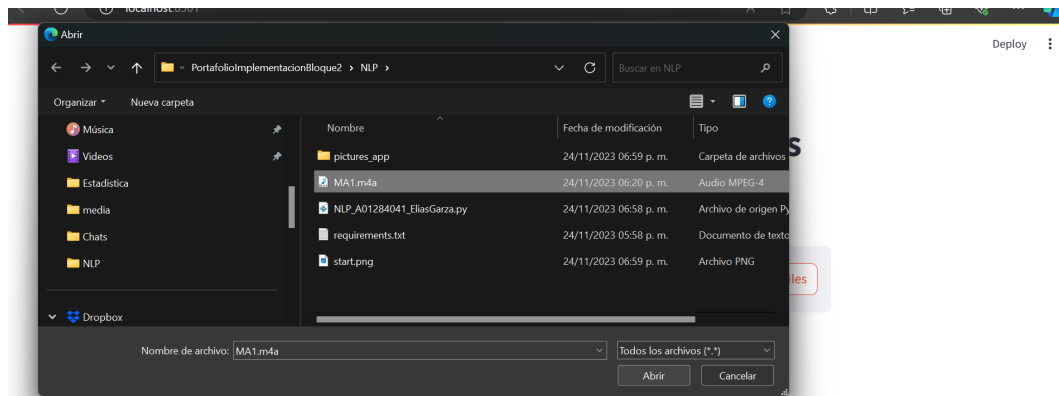
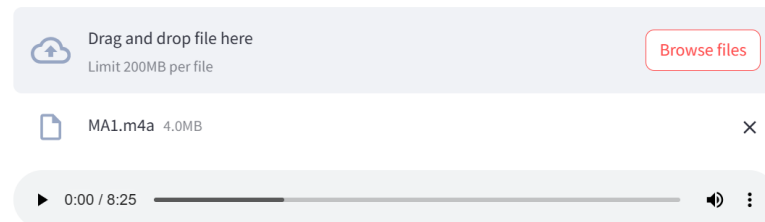


Figura 2: Se selecciona un archivo de nuestra computadora.

Generador de puntos clave a través de audio

Elige un archivo de audio



Transcripción:

Especially, and I would even argue in the last four weeks, you really can't build fitness for the Ironman distance, but you can ruin your fitness for the Ironman distance, especially within the last two to three weeks, by, because, so it takes, it takes your body to adapt to a full load, a full Ironman load, it takes four

Figura 3: Se genera una transcripción del audio.

bumped. I think that's what happened. Yeah, I mean, was it hot? It wasn't really hot. It's just my body, I mean, I don't want to eat. That's a part, right? Yeah, there's a part where your stomach is saturated. Salt, a lot of times, I don't know, maybe that's where Mike was headed with his question, but salt, a lot of times, is one of the things that you need to have more of. If you don't have a proper balance of salt in your stomach, you're not able to absorb the carbohydrates. So even if you keep slamming gels, you end up with this, like, lead weight in your stomach because you aren't able to, the stomach's not able to process it because it doesn't have enough salt in your stomach cavity to absorb the liquid. And you really need to dial in the right amount for you because there's a huge range on what people need for salt. There's really heavy salt sweaters and there's light, and you can do too much salt. And you have this, like, really similar, like, cramping from too much salt, so you really need to dial in in your training.

Puntos clave:

Here are the key points of the text:

- It takes 4 to 6 weeks for the body to fully adapt to an Ironman distance race.
- Training too hard in the last two weeks before the race can ruin fitness gains.
- Short, quick efforts are recommended during the taper phase.
- Weight training should not be done in the last two to four weeks before the race.
- Volume should be gradually reduced in the weeks leading up to the race.
- Tapering strategies may vary for each individual.
- Practice nutrition during training to find what works best.

Figura 4: Se escriben los puntos clave en el idioma del audio.

El código de este proyecto se puede encontrar en

<https://github.com/EliasGarzaV/PortafolioImplementacionBloque2/tree/main/NLP>