Transcription of YouTube Videos with Speech-to-Text Models

Overview

In this exercise, you'll use a Python script and the YouTube API to compile a list of videos along with some metadata. Following this, you'll extract audio from these videos and transcribe the spoken words using a text-to-speech model like Whisper v3. Afterwards, you'll load the transcriptions into a dataframe and analyze the video content using R and the Tidyverse.

This exercise assumes you already completed the previous exercise on getting started with the HPC cluster. If not, go back and complete that exercise first.

Task 1: Test drive the provided Jupyter notebook

In this exercise, you'll need to run code snippets that would go beyond the scope of this course if you had to write them from scratch. Therefore, I prepared a Jupyter notebook with helpful code snippets for you. You can find the notebook in the code repository you cloned in the previous exercise.

Your first task in this exercise is to start a Jupyter Lab server on the HPC cluster and open the notebook `youube_api.ipynb`. The notebook is structured into six sections using markdown cells. In each section, you find the code for a specific task. Your job is to run the code cells in each section and understand what they do. Make sure you run the code from top to bottom, as the code in later cells may depend on the results of earlier cells.

> Note
>
> In the first section of the notebook, you need to set the constant `API_KEY` and assign a valid YouTube Data API key. In the context of the course, I will provide you with a valid key that you can use. Make sure you never share this key with others or commit it to a public repository.