

Lecture 02

Jupyter Notebook - YouTube Transcript API

Marcel Turcotte

2024-09-02

Objectif d'apprentissage

- **Illustrer** le processus d'identification et de résolution des problèmes de bibliothèques manquantes dans Google Colab.

! Important

Cet exemple est destiné à être exécuté dans [Google Colab](#).

API de transcription YouTube

Dans ce notebook, nous visons à utiliser l'[API de transcription YouTube](#) pour télécharger automatiquement la transcription de la vidéo intitulée [Can Machines Think?](#) par Noam Chomsky.

Tout d'abord, importons `YouTubeTranscriptApi` et `TextFormatter` de `youtube_transcript_api`.

```
from youtube_transcript_api import YouTubeTranscriptApi
from youtube_transcript_api.formatters import TextFormatter
```

L'exécution de la cellule de code ci-dessus entraînera une erreur, car la bibliothèque `youtube_transcript_api` n'est pas installée par défaut dans Google Colab.

```
-----
ModuleNotFoundError Traceback (most recent call last)
<ipython-input-1-c8308591d925> in <cell line: 2>()
1 # ! pip install youtube-transcript-api
```

```
----> 2 from youtube_transcript_api import YouTubeTranscriptApi
3 from youtube_transcript_api.formatters import TextFormatter
```

ModuleNotFoundError: No module named 'youtube_transcript_api'

NOTE: If your import is failing due to a missing package, you can manually install dependencies using either !pip or !apt.

To view examples of installing some common dependencies, click the "Open Examples" button below.

Ce problème peut être résolu en ajoutant la ligne de code suivante avant la première instruction import. Essayez-le !

```
! pip install youtube-transcript-api
```

Une fois ce problème résolu, nous pouvons télécharger et imprimer la transcription. Essayez-le !

```
transcript = YouTubeTranscriptApi.get_transcript("Ex9GbzX6tMo")
formatter = TextFormatter()
input_text = formatter.format_transcript(transcript)
print(input_text)
```

Exploration

! permet d'exécuter [des commandes shell Unix/Linux dans IPython](#). Créez une cellule de code et essayez ces commandes.

- ! `uname -a` affiche des informations sur le système.
- ! `ls` affiche le contenu du répertoire actuel.
- ! `ls /` affiche le contenu du répertoire racine.
- ! `pwd` retourne le nom du répertoire de travail.

Ces commandes sont utiles pour déboguer du code, car elles fournissent des informations sur l'environnement informatique, telles que la version du système d'exploitation et le contenu du répertoire local.