

How to... transcribe with Whisper ?

Introduction

You're dreaming to make your own subtitles of videos, you have the programs to edit srt file but you're not good to learn a new language. OpenAI Whisper is what you need, it's accuracy with some languages, and it's free !

Now let's start on how you can do it !

1) Upload files to your Google drive

First of all, login to your google account and go to your google drive.

Drag and drop your files or right click on your Google drive. (I choose to upload only the audio of the video, it will take less time).

Now go to :

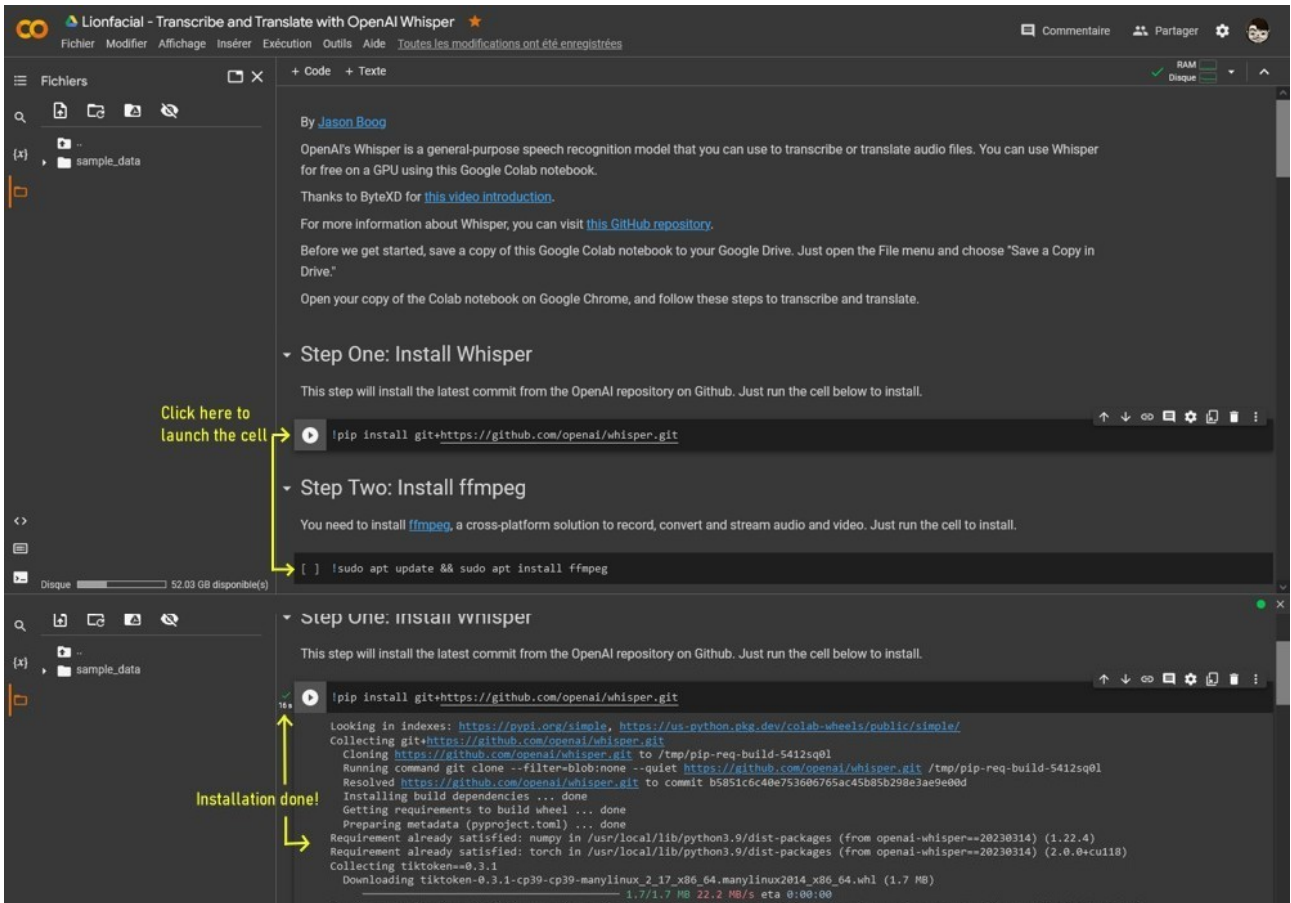
<https://colab.research.google.com/drive/1WLYoBvA3YNKQ0X2lC9udUOmjK7rZgAwr?usp=sharing>

and follow the instructions of the page, it's a tutorial for using whisper but i will add some useful informations next, so only do the introduction before « Step One ».

Open your colab page, you can do it with Google Chrome or Firefox or else, choose your favorite web browser.

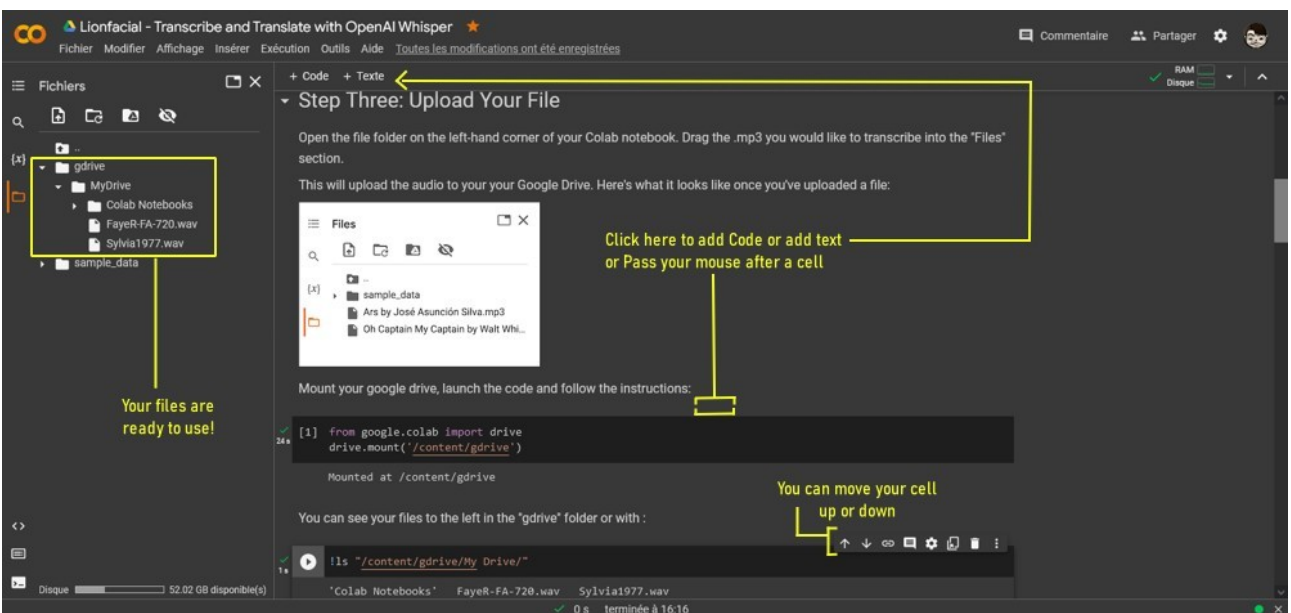
2) Using Colab, follow the steps

To begin use whisper follow the instructions of your page (step one and two) :



Now you need to mount your google drive.

First add the code (at step Three) by clicking [+Code] on the upper page or after a cell and enter :
from google.colab import drive
drive.mount('/content/gdrive')



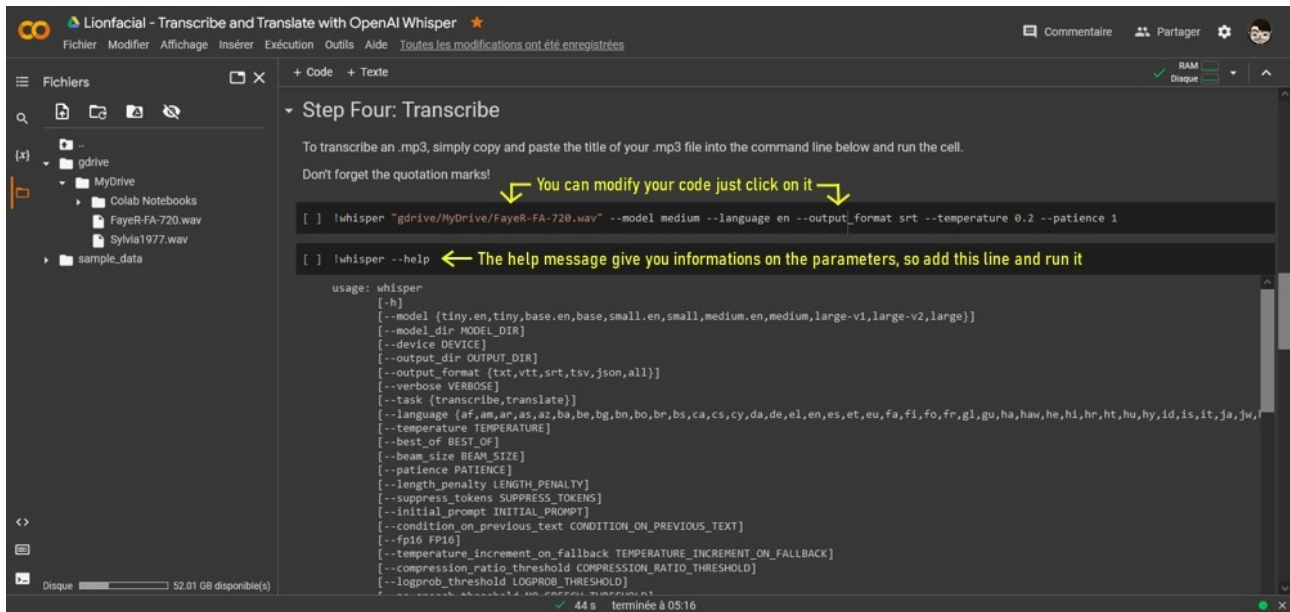
You can see a new folder named « gdrive » to the left (it take a little time), your files are in.
 (You can move the cell with their arrows buttons)

3) Parameters to transcribe with Whisper

You can modify all the cells of your page, just click on it or the button « Modify » of the cell. So change the command line of the step Four, put the path of your file and add parameters if you want.

To see what sort of parameters Whisper have, add the code :

```
!whisper --help
```



And now a description of the most useful parameters :

- The file formats can be mp3, mp4, mpeg, mpga, m4a, wav, or webm. Write with quotes.
- --model

Follow by the name of the AI model. By default is « small », so change by *medium.en* or *medium* for English and *large* for an other language.

- --output_format

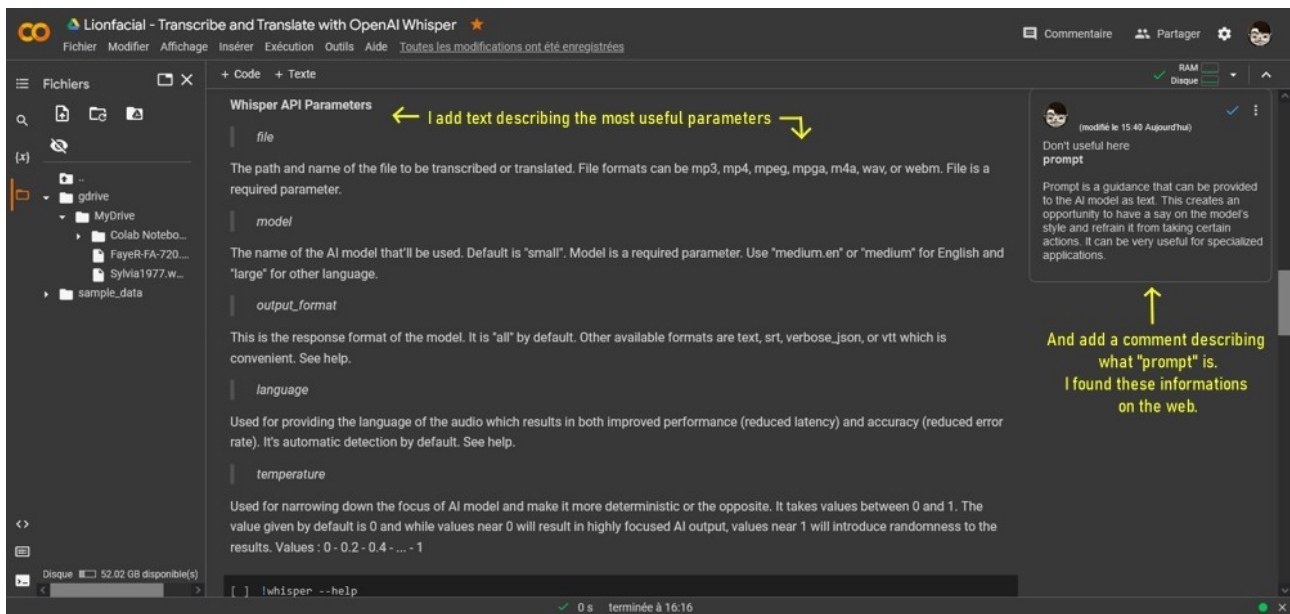
This is the format of the text transcription it create. By default is *all*, so change it to *srt*. See help message for other format.

- --language

Used for providing the language of the audio (improved performance and accuracy). By default it's automatic detection. See help message for the country code.

- --temperature

Used for narrowing down the focus of AI model and make it more deterministic or the opposite. By default is 0 and the max is 1 the step is 0,2. Values near 0 will result in highly focused AI output, near 1 will introduce randomness to the results.



4) Translate with...

Step 5 of the page is useless without a google colab pro, so i never try it to translate a file.

Just download the text file Whisper have created at the left of the page. It's a srt file if you choose only this output format.

Now translate with your favorite translator :

Google translate, DeepL translator, yourself because you're very smart or masochist, or a dictionary because it's still exist and you are really a masochist. Or else...

5) Tips

- Sometimes for transcription, you can have different results with a second run of the command line. But warning, it overwrite the first file.
- If there are missing parts in the transcription of a dialogue, re-try with only this audio part record. Whisper can detect it or not, in this case change parameters like `--temperature`.
- Be careful with google colab, it's using a virtual machine and you can be disconnected if you don't use it after a little period. You will loose all the files created, so download them quickly.

Lionfacial April, 5 2023.