



Subtítulació de vídeos en català amb Whisper

Tipus: [1]

Autor: [Gutiérrez Ferrerías, Fernando](#) [2]

Creació: Publicat per [Fernando Gutiérrez Ferrerías](#) [2] el 16/10/2023 - 15:26 | Última modificació: 30/11/2024 - 14:06

Etiquetes: accessibilitat

Etiquetes: vídeo

Etiquetes: youtube

Etiquetes: subtítols

Etiquetes: WCAG

[1 adjunt](#) [3]

La història de l'accessibilitat del contingut web en les administracions públiques no és malauradament una història d'èxit, com demostra el [timeline normatiu del Portal de Administración Electrónica](#) [4]. Un dels requisits per assolir la conformitat amb els criteris d'accessibilitat és oferir [subtítols sincronitzats per al contingut multimèdia](#) [5]. Avui us comparteixo una solució per afegir **subtítols en català de qualitat** als vídeos de YouTube mitjançant [Whisper](#) [6], una intel·ligència artificial (IA) creada per OpenAI (els de ChatGPT) i alliberada com a programari de codi obert (*open source*).

Introducció

Whisper és una IA enfocada en el reconeixement automàtic de la parla (*ASR, Automatic Speech Recognition*), amb unes característiques que la fan totalment disruptiva davant les opcions preexistents, com són haver estat entrenada amb un dataset de més de 600.000 hores d'audio, el 17% d'elles en idiomes diferents de l'anglès, incloent-hi el català amb 1.883 hores en un lloc rellevant (14è entre 75 idiomes). Poden semblar poques 1.883 hores, però la veritat és que són més que suficients per obtenir resultats gairebé professionals, com veurem més endavant.

La situació anterior a Whisper

El català no està entre les 14 [llengües suportades per a la generació automàtica de subtítols en Youtube](#) [7]. De tota manera, la subtítolació automàtica en sí mateixa pot arribar a provocar errors fints i tot mortals com el [descriu en aquest enllaç de la WAI/W3C](#) [8].

Tot i que existien diverses alternatives externes, el cert és que les solucions anteriors a Whisper prometien molt però la realitat és que donaven més feina de la que estalviaven. A tall d'exemple, veieu aquest vídeo del Govern de les Illes Balears demostratiu de les capacitats de Subtitle Edit: [Subtitle Edit. Generar subtítols en català. IBSTEAM \(caib\)](#) [9]. El vídeo és de l'any 2022, per tant coetani a la publicació de Whisper. L'enllaç porta directament al minut 2:37 on es veuen els resultats generats pel programari i les correccions que ha de fer l'usuari al text generat, gairebé l'ha de reescriure de nou:

- Text generat automàticament: "O dona és que tal volta trobes so miquel nebot assessor dir vestim i en El dia de hui no";
- Text corregit manualment: "Molt bones, què tal com te trobes? Som en Miquel Nebot, assessor d'IBSTEAM i en el dia d'avui".

Funcionament de Whisper

Hi ha diverses opcions per executar Whisper, però la més senzilla i immediata que requereix menys recursos és executar-lo en un entorn remot com [Colaboratory de Google](#) [10]. Només us caldrà un compte de Google i unes poques línies de codi. És recomanable consultar la secció de [Primeres passes amb Colab](#) [11].

Pel que fa el compte de Google podeu fer servir el mateix que va crear per configurar el mòbil corporatiu.



1. Entreu a <https://colab.research.google.com/> i valideu-vos amb les vostres credencials.
2. Un cop dintre de l'eina, Cliqueu a sobre de "+ Nuevo Cuaderno" o "+ New Notebook"
3. En el menú superior cliqueu a "**Entorno de ejecución > Cambiar tipo de entorno de ejecución**" o "**Runtime > Change Runtime type**", verifiqueu que estigui seleccionat **Python 3** com a entorn d'execució i **GPU** com a accelerador de hardware
4. A continuació cliqueu a sobre de "+ Código" o "+ Code" es crearà una nova cel·la per introduir instruccions
5. Copieu i enganxeu aquestes 3 línies que instal·laran en l'entorn virtual de Google Colab el programari necessari, no s'instal·larà res en el vostre maquinari:

```
!pip install git+https://github.com/openai/whisper.git
```

1. Executeu la cel·la clicant a la icona de Play situada a l'esquerra, espereu fins que al costat d'aquesta icona aparegui un símbol de verificat i els segons que ha trigat l'execució.
2. Ara si teniu l'àudio del vostre vídeo el podeu carregar a la zona de fitxers de Colab, o bé podeu importar-lo directament de YouTube, amb aquestes línies de codi. Ho farem en una nova cel·la de codi, o sigui que novament "+ Código" i enganxem i executem aquesta cel·la prement novament el botó play (codi editat el 16 octubre de 2024 per fer servir pytubefix en comptes de pytube, més info als comentaris):

```
!pip install pytubefix  
from pytubefix import YouTube
```

1. En una altra cel·la enganxem i executem (Play)

```
url = input("Introdueix la URL del video de YouTube: ")
```

1. Ens demanarà la URL del vídeo que prèviament, o ara mateix, podem copiar (aneu al vídeo i feu botó dret a sobre i Copiar URL del video), l'enganxem amb Ctrl+V i Intro.
2. En una altra cel·la copiem i executem el següent:

```
yt_url = YouTube(url)  
audio_stream = yt_url.streams.filter(only_audio=True).first()  
output_path="/content"  
filename="audio.mp3"  
audio_stream.download(output_path=output_path, filename=filename)  
print(f"S'ha desat l'audio a la següent carpeta del COLAB {output_path}/{filename}")
```

1. Ja estem acabant, si cliqueu en la icona amb forma de carpeta a l'esquerra de Colab, s'obrirà (o es tancarà, alternativament) la zona d'arxius d'usuari, veureu que s'ha creat un fitxer anomenat "audio.mp3".
2. Ara que ja tenim Whisper instal·lat i l'audio del nostre vídeo podem executar la següent línia de codi:

```
!whisper "/content/audio.mp3" --task transcribe --model large --language=ca --verbose False --output_format srt  
--output_dir audio_transcription
```

1. Depenent de la durada de l'àudio trigarà més o menys l'execució. En un primer moment us apareixerà una barra de progrés que només indica la càrrega del model de la IA, que actualment són 2,87 GB. A continuació apareixerà un nou contador, en aquest cas de frames (per exemple: 100% 633027/633027 [48:10<00:00, 219.00frames/s])
2. El fitxer amb els subtítols en format SRT es desarà en una carpeta anomenada audio_transcription. Descarrega-ho quan abans millor (botó dret > Descargar) perquè tots els arxius de l'entorn de Colab desapareixeran en tancar la sessió. El fitxer SRT es pot carregar directament en Youtube, trobareu més de com fer-ho i de com canviar la configuració d'idioma del vídeo -si cal- en l'ajuda de Youtube.



3. Ara podeu tornar a iniciar el procés, per generar els subtítols d'un altre vídeo a partir de les pases en que carreguem l'audio, mentre estigui activa la sessió, no cal tornar a instal·lar els programes.

A tall d'exemple, he fet la prova amb el primer vídeo del canal Diba que actualment és el de la campanya [Foc off!](#) [12] i que en tenir subtítols integrats, serveix per comparar els resultats amb els generats per Whisper que copio a continuació:

1
00:00:00,000 --> 00:00:05,000
Has vist l'incendi descontrolat i tu llençant burilles per la finestra?

2
00:00:06,000 --> 00:00:07,000
Foc off!

3
00:00:07,000 --> 00:00:10,000
Amb la sequera que patim i faràs barbacoes a tocar del bosc?

4
00:00:10,000 --> 00:00:11,000
Foc off!

5
00:00:11,000 --> 00:00:14,000
I encendràs petards en plena onada de calor?

6
00:00:16,000 --> 00:00:17,000
Foc off!

7
00:00:17,000 --> 00:00:20,000
Saps que si utilitzes maquinària on no toca pots provocar un desastre?

8
00:00:20,000 --> 00:00:21,000
Foc off!

9
00:00:21,000 --> 00:00:23,000
Aquesta es diu més que cap altra.

10
00:00:23,000 --> 00:00:24,000
Per les imprudències...

11
00:00:24,000 --> 00:00:25,000
Foc off!

12
00:00:25,000 --> 00:00:28,000
Si veus foc o una columna de fum, truca al 112.

13
00:00:28,000 --> 00:00:29,000
Generalitat de Catalunya.



14

00:00:29,000 --> 00:00:30,000

Sempre endavant.

Només cal corregir:

- "Has vist l" per "El risc d" al núm. 1
- "Aquesta es diu" per "Aquest estiu" al núm. 9
- "Per" per "A" al núm. 10

Sempre cal revisar els resultats, òbviament, sobretot pel que fa noms propis, gentilicis, topònims, etc. però Whisper estalvia més del 90% de la feina. Gaudiu-ne!

Categories: General

Categories: Punt web

Etiquetes: accessibilitat

Etiquetes: vídeo

Etiquetes: youtube

Etiquetes: subtítols

Etiquetes: WCAG

Adjunt

Mida

[e8joamuogmk.txt](#) [13]

33.56 KB

- [14]

URL d'origen: <https://puntweb.diba.cat/blogs/2023/10/16/subtitulacio-de-videos-catala-amb-whisper>

Enllaços:

[1] <https://puntweb.diba.cat/>

[2] <https://puntweb.diba.cat/members/gutierrezff>

[3] <https://puntweb.diba.cat/blogs/2023/10/16/subtitulacio-de-videos-catala-amb-whisper>

[4] https://administracionelectronica.gob.es/pae_Home/pae_Estrategias/pae_Accesibilidad/pae_normativa/pae_eInclusion_Normas_Accesibilidad.html

[5] <https://www.w3.org/WAI/media/av/captions/>

[6] <https://openai.com/research/whisper>

[7] <https://support.google.com/youtube/answer/6373554?hl=ca#zippy=%2Csubt%C3%ADtols-autom%C3%A0tics-en-v%C3%ADdeos-llargs-i-curts>

[8] <https://www.w3.org/WAI/media/av/captions/#automatic-captions-are-not-sufficient>

[9] https://youtu.be/_iWCosx9_Y8?t=156

[10] <https://colab.research.google.com/>

[11] https://colab.research.google.com/#scrollTo=Primeros_pasos

[12] <https://www.youtube.com/watch?v=13KNmFE6TLo>

[13] <https://puntweb.diba.cat/sites/puntweb.diba.cat/files/e8joamuogmk.txt>

[14] <https://puntweb.diba.cat/node/1397>