

Aplicaciones de Análisis

Práctica 1: IBM Watson para NLP

Daniel Pérez Efremova

INTROUCCIÓN

Para el desarrollo de la práctica, con el objetivo de cumplir el requerimiento alto se han utilizado dos servicios: Speech to Text y Natural Language Understanding (NLU) en dos casos de uso sencillos.

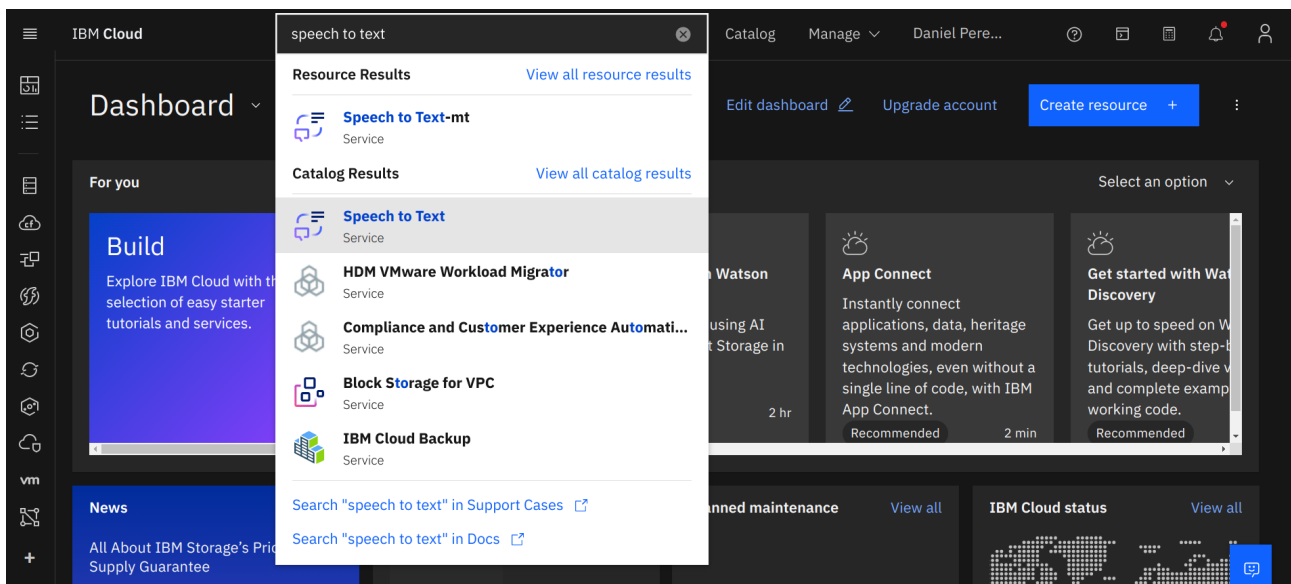
El primer caso de uso es Speech to Text donde se aplica el servicio para resolver Captchas v2 de Google. En el segundo caso, se utiliza NLU para analizar el sentimiento y las palabras clave de comentarios en Amazon sobre un producto a través de su url en la página de Amazon.

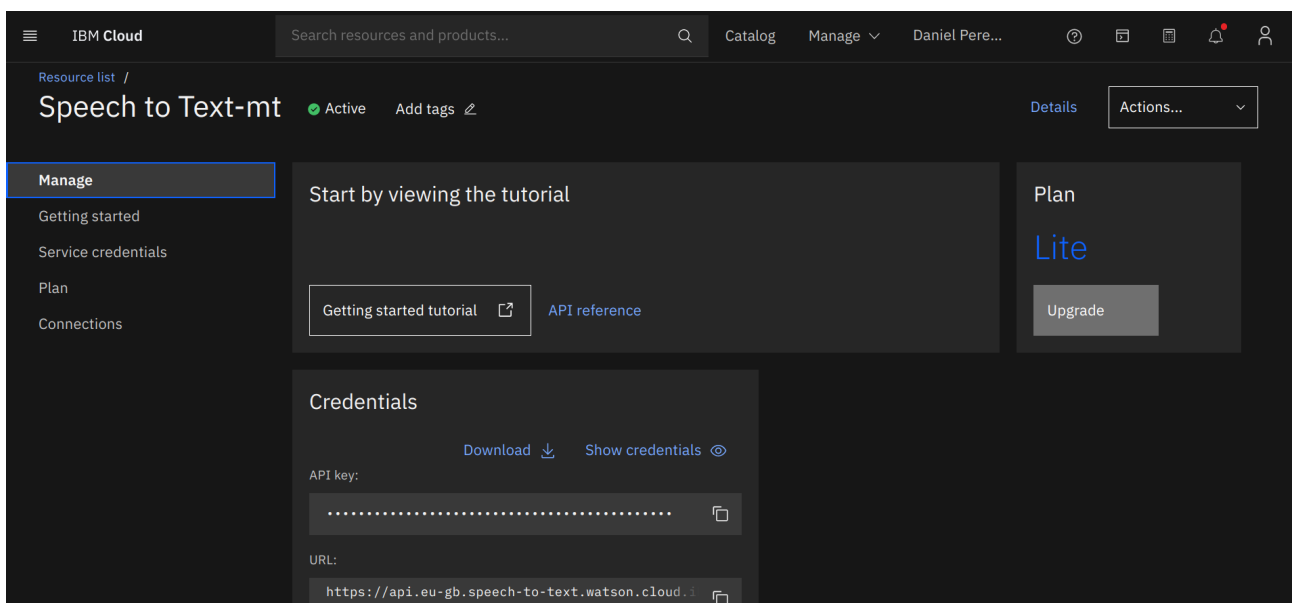
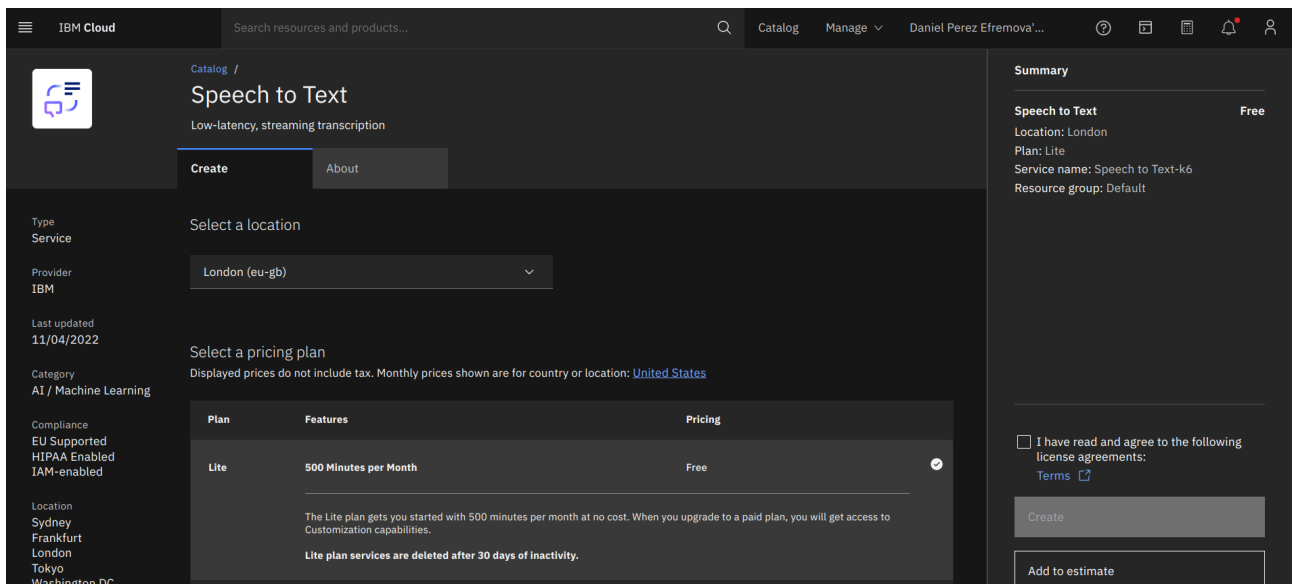
PARTE 1: Speech to Text aplicado a resolver captchas v2 de Google

Se ha creado un repositorio en github que contiene toda la implementación (github.com/danielp2797/text-to-speech-IBM-Watson)

Aquí se mencionan los aspectos más relevantes.

Para este caso de uso se ha creado una instancia Lite de Speech to Text.





Despues se ha escrito un programa en Python para conectar a la API y, mediante el archivo de un audio .mp3, convertir a texto y mostrar el resultao por consola. A continuación se muestran imágenes de las pruebas con recaptcha.

Sample Form with ReCAPTCHA

First Name
Jane

Last Name
Smith

Email
stopallt

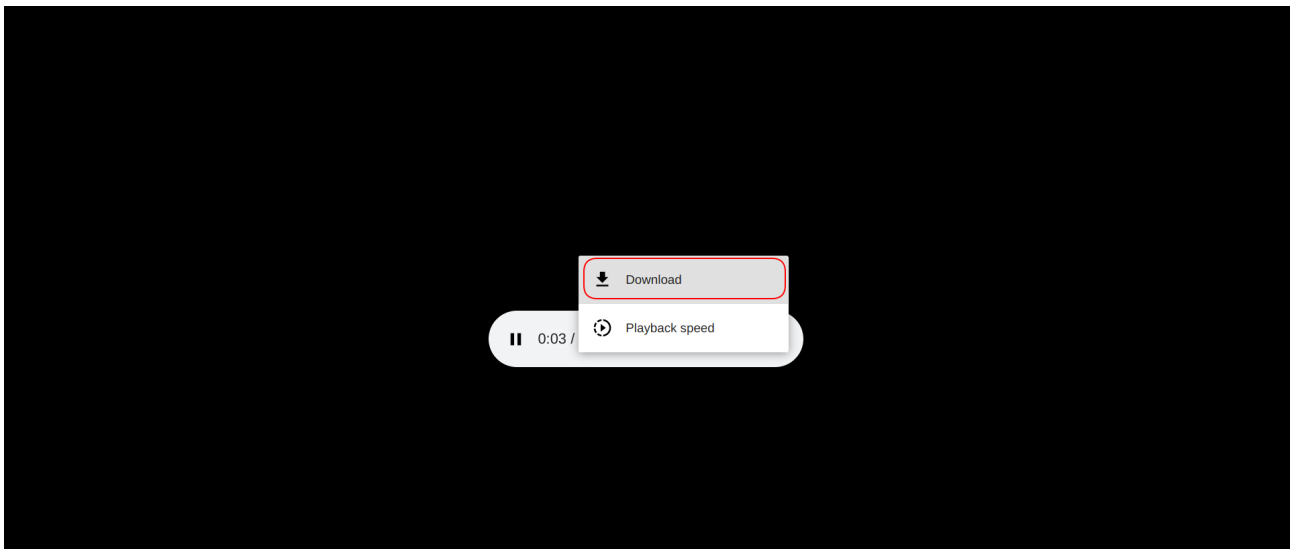
Pick your color
☒ Red
☐ Green

☐

Submit

Select all images with
stairs

⏮️ 🔊 👤 VERIFY



```
(IBMTextToSpeech) daniel@daniel:~/Desktop/IBM speech to text/IBMSpeechToText$ python3 main.py "conversion/audio/audio_test.mp3"
```

arithmetic mean also known

Sample Form with ReCAPTCHA

First Name
Jane

Last Name
Smith

Email
stopallt

Pick your color
☒ Red
☐ Green

☐

Submit

Press PLAY to listen

PLAY

Enter what you hear

arithmetic mean also known

⏮️ 👁️ 👤 VERIFY

Sample Form with ReCAPTCHA

First Name

Jane

Last Name

Smith

Email

stopallbots@gmail.com

Pick your favorite color:

☒ Red

☐ Green



I'm not a robot



Submit

CONCLUSIONES PARTE 1:

- Se ha aprendido cómo implementar mediante API un servicio de Watson
- Se ha comprobado su efectividad probando la transcripción contra un servicio cuya misión es detectar si el usuario no es una máquina y, de integrarse en un bot, el bot sería capaz de resolver captchas v2 con un grado de fiabilidad aceptable de manera gratuita.

PARTE 2: NLU aplicado a comentarios de amazon

Se ha creado un repositorio en github que contiene toda la implementación y un markdown a detalle de la implementación (github.com/danielp2797/Amazon-reviews-scrap-and-sentiment-analysis-with-IBM-Watson)

Se ha desarrollado un programa en python que hace scraping de una página de reseñas de amazon y las guarda en local. Una vez termina, se cargan los datos en un notebook para hacer análisis de sentimiento y hacer un gráfico word cloud.

Se hace una prueba con una página aleatoria de Amazon:

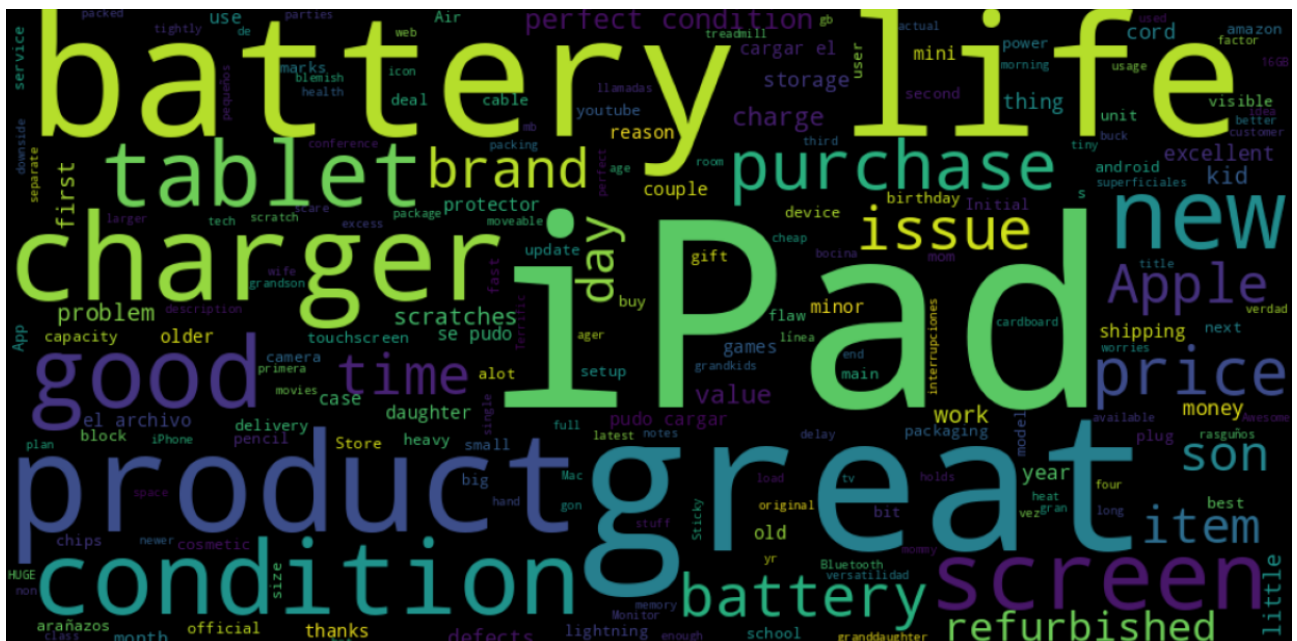
https://www.amazon.com/-/es/Pantalla-Retina-pulgadas-espacial-renovado/product-reviews/B07F93611L/ref=cm_cr_dp_d_show_all_btm?ie=UTF8&reviewerType=all_reviews

Despues se ejecuta el programa para extraer las reseñas.

```
[Amazon_Products_Reviews] daniel@dm101:~/Desktop/Amazon_Products_Reviews$ python3 main.py "https://www.amazon.com/-/es/Pantalla-Retina-pulgadas-especial-renovado/product-reviews/B07F93611l/ref=cm_cr_dp_d_show_all_btm?ie=UTF8&reviewerType=all_reviews"
Downloading https://www.amazon.com/-/es/Pantalla-Retina-pulgadas-especial-renovado/product-reviews/B07F93611l/ref=cm_cr_dp_d_show_all_btm?ie=UTF8&reviewerType=all_reviews
https://www.amazon.com/-/es/Pantalla-Retina-pulgadas-especial-renovado/product-reviews/B07F93611l/ref=cm_cr_dp_d_show_all_btm?ie=UTF8&reviewerType=all_reviews
Downloading https://www.amazon.com/-/es/Pantalla-Retina-pulgadas-especial-renovado/product-reviews/B07F93611l/ref=cm_cr_dp_d_paging_btm_2?ie=UTF8&pageNumber=26&reviewerType=all_review
s
https://www.amazon.com/-/es/Pantalla-Retina-pulgadas-especial-renovado/product-reviews/B07F93611l/ref=cm_cr_dp_d_paging_btm_2?ie=UTF8&pageNumber=26&reviewerType=all_reviews
Downloading https://www.amazon.com/-/es/Pantalla-Retina-pulgadas-especial-renovado/product-reviews/B07F93611l/ref=cm_cr_dp_d_paging_btm_3?ie=UTF8&pageNumber=36&reviewerType=all_review
s
```

D

En la notebook del repositorio se hace el tratamiento de datos y se obtiene la proporción de reseñas positivas, negativas y neutrales, así como el gráfico word cloud.



Se ve que las palabras clave más comunes hacen referencia a:

- La duración de la batería
- El cargador
- La calidad del producto (Ipad reacondicionado)

En caso de querer comprarlo se debe prestar atención a estos factores.

CONCLUSIONES PARTE 2:

- **Se ha visto como implementar un solución sencilla mediante la API del servicio**
- **Se han extraído conclusiones básicas de las reseñas de un producto real**