Pre-requisitos

1. **Nota:** este taller se debe realizar en **grupos de tres personas**.

2. **Nota 2:** la entrega de este taller consiste en un **cuaderno de jupyter** ejecutable en Google Colab. El cuaderno debe contener todo el código y respuestas a las preguntas propuestas.

Parte 1: Tokenizers

Tome como base el cuaderno mml_taller2_tokenizers_starter que encontrará en Bloque Neón.

1. Ejecute paso a paso el cuaderno siguiendo las instrucciones.

2. Ahora cree otro cuaderno (su cuaderno de entrega) en el que, usando el mismo modelo, emplee otros prompts para modificar los tokens de entrada y salida. Use **3 prompts** diferentes que permitan ilustrar comportamientos diferentes del tokenizador. Comente sus resultados.

3. Selecciones **3 tokenizadores** diferentes al de base, puede ser de los incluidos en el cuaderno u otros que puede encontrar en `https://huggingface.co/models?library=transformers`. Para cada tokenizador emplee los 3 prompts anteriores, explore los tokens generados y compárelos. Comente sus resultados.

Parte 2: Embeddings

Tome como base el cuaderno mml_taller2_embeddings_starter que encontrará en Bloque Neón.

1. En su cuaderno de entrega, descargue el embedding glove-wiki-gigaword-50.

2. Realice una exploración en la que emplee entre 10 y 15 palabras con significados similares y distintos para ilustrar el comportamiento del embedding. Comente sus resultados.

3. En su cuaderno de entrega, entrene el modelo Word2Vec con el dataset de playlists con la configuración inicial. Realice una exploración con entre 10 y 15 canciones que le permita ilustrar el comportamiento del embedding generado. Comente sus resultados.

4. Seleccione **uno** de los parámetros de la función, como vector size o window, y reentrene el modelo empleando 3 valores diferentes a los iniciales. Compare y comente sus resultados.