



## Tarea 3.2.- Preprocesamiento de texto.

- Detallar los pasos necesarios para el preprocesamiento de texto: tokenización, eliminación de stop words, lematización, etc. Entre 300 y 400 palabras.
- Utilizando el corpus de tweets del MundoToday, realizar un ejemplo práctico de preprocesamiento de texto utilizando spaCy. Al menos debéis contemplar los siguientes pasos:
  - **Tokenización:** Dividir el texto en unidades más pequeñas, como palabras o caracteres.
  - **Eliminación de Caracteres Especiales:** Eliminar signos de puntuación, emojis u otros caracteres no deseados.
  - **Eliminación de Stop Words:** Filtrar palabras comunes que no aportan significado al análisis.
  - **Lematización:** Reducir las palabras a su forma base o lema para normalizar el texto.
  - **Eliminación de tokens concretos:** Artículos, pronombres, conjunciones, ...

Este contenido lo prepararás en un Jupyter Notebook de Google Colab titulado **Apellido1\_Nombre\_PLN\_2**

Podéis partir del tutorial de spaCy disponible en:

<https://course.spacy.io/es/chapter1>