# Preprocesamiento de tweets

## *Machine learning*

## Camacho Vázquez Vanessa Alejandra.

El pre-procesamiento de tweets en español implica una serie de tareas comunes para limpiar y estructurar los datos, facilitando su análisis posterior. Aquí hay algunas tareas típicas:

1. **Eliminación de menciones (@usuario):** Las menciones a otros usuarios a menudo no aportan información relevante y pueden eliminarse para simplificar el texto.
2. **Eliminación de enlaces (URLs):** Los enlaces no suelen ser informativos para el análisis de texto y pueden ser eliminados.
3. **Eliminación de caracteres especiales y puntuación:** La limpieza de caracteres no alfabéticos, como signos de puntuación y emojis, ayuda a simplificar el texto y facilita el análisis. 🡪 algunas ocasiones se sustituyen los emojis por palabras o descripciones equivalentes.
4. **Eliminación de stopwords o palabras vacías:** Palabras comunes (como "y", "el", "la") que no aportan mucho significado suelen eliminarse para reducir el ruido en el análisis 🡪 se eliminan las preposiciones y conjunciones.
5. **asdasd**
6. **asdasdad**
7. **Identificación de entidades mencionadas:** Reconocer y etiquetar entidades como nombres de lugares, personas o marcas.
8. **asdasdasdasd**
9. **Normalización de texto:** Convertir todo el texto a minúsculas para garantizar la consistencia y evitar redundancias.

Estas tareas son esenciales para preparar los datos antes de realizar análisis de sentimientos, clasificación de temas o cualquier tarea de procesamiento de lenguaje natural con tweets en español. La combinación de estas técnicas ayuda a obtener un conjunto de datos más limpio y estructurado para análisis posteriores.

## Ejercicio:

Ustedes ya realizaron los puntos 4 y 6.

Ahora deberán llevar a cabo los puntos 1, 2, 3, 5, 8 y 10 🡪 ade más tendrán que eliminar para cada tweet tanto su ID como su fecha y hora de creación.

Noten que los puntos 7 y 9 requieren de un mayor esfuerzo 🡪 por lo tanto, no vamos a solucionarlos en este ejercicio.