

Práctica tokenización

Esta práctica es individual.

Objetivos

1. Entender qué es un tokenizador
2. Comparar un tokenizador en diferentes idiomas

Tareas a Realizar

- **Entre en el tokenizer de openai (o cualquier tokenizer):**
<https://platform.openai.com/tokenizer>
- **Calcule los tokens que generan las siguientes frases en inglés y español**
 - The capital of Spain is Madrid
 - When I was young, I liked playing basketball with my Friends

Entrega de Resultados

Suba los resultados al Moodle en un fichero llamado solucion.txt en el que incluya los tokens, el número de tokens en cada caso, el tokenizador elegido, el número medio de tokens por palabra y el número medio de tokens por carácter.