

---

# Práctica 1.2

## Procesamiento del lenguaje natural

### Exploración de temas en novela histórica

---

#### Enunciado

En esta actividad, exploraremos la aplicación del PLN para analizar y caracterizar temas históricos presentes en una selección de libros. Utilizaremos técnicas de *scraping* para obtener información relevante de la página "Books to Scrape", específicamente de la sección de novelas y obra sobre Historia. Posteriormente, aplicaremos etiquetado morfológico mediante la biblioteca Spacy para identificar los sustantivos en las descripciones de los libros. Estos sustantivos nos proporcionarán una idea sobre los temas o eventos históricos con más presencia en el conjunto de los libros.

#### Pasos:

1. Realiza *scraping* de la página de historia de "Books to Scrape" para obtener información sobre los libros.
2. Aplica etiquetado morfológico con Spacy para identificar y crea una lista con todos los sustantivos que contengan las descripciones de los libros.
3. Muestra por pantalla los 20 con mayor número de ocurrencias en la lista (el objeto "counter" de la librería "collections" es útil a tal fin).

**¿Se puede llegar a alguna conclusión sobre las temáticas con más presencia en la lista de libros?**

4. Crea una variante del código previo que trabaje con **bigramas** (unidades de análisis formadas por dos tokens) en lugar de con unigramas (unidades de análisis formadas por un único token). Dado que el texto se encuentra en inglés, puede ser útil explorar bigramas compuestos por un adjetivo seguido de un sustantivo (por ejemplo "white house").

**¿Aporta información más concluyente utilizar bigramas?**

#### Recursos para consulta

- Sección de Historia de la web "[Books to scrape](#)"
- [The Web Scraping Project Perfect for Beginners — Books to Scrape](#)