
Trabajo Práctico Nº 3

Tema: Unidad 2 - Crawler y Scraper

Fecha Inicio: 06/05/2025 **Fecha de Entrega:** 20/05/2025

Actividades:

- 1) Planifique, diseñe y construya un crawler para recolectar al menos 30 enlaces de noticias de la sección <https://www.infobae.com/deportes> . De cada una de estas noticias, extraiga todos los enlaces a otras notas de deporte. Cree las estructuras de datos necesarias para mantener dichas referencias (un enlace de nota apuntando a otros enlaces de notas) y construya una representación tipo grafo de lo recolectado, preste especial atención al tema de las aristas entre los nodos, ya que pueden existir enlaces de doble entrada (entrada y salida) o enlaces “más fuerte” cuando existen más de un enlace entre 2 notas distintas. El Grafo puede ser desarrollado con cualquier herramienta o librería gráfica, pero el crawler debe ser capaz de exportar o proveer la estructura de datos necesaria para dicha representación.

- 2) Realice un web scraping de la siguiente URL:

<https://www.infobae.com/economia/>

De esta URL recolecte las primeras 10 noticias, identificando por cada una el Título, Resumen, Autor de la nota, Listado de imágenes (ubicación del archivo) y el Cuerpo de la misma. A continuación realice un análisis textual sencillo, tokenize dichos documentos, elimine las stop-words y liste los 100 términos más frecuentes. En el mismo sentido realice un stemming y vuelva a listar los 100 términos más frecuentes.