

**DEPARTAMENTO:** Ciencias de la Ingeniería **PROFESOR:** Mg. Gustavo Guaigua A.

**CARRERA:** Sistemas de Información

**ESTUDIANTES:** Cando Orlando, Columba Ámbar, Guamba Cristian, Tipán Gissela

**CURSO:** 9 **PARALELO:** "B"

**DESCRIPCIÓN:** Proyecto OpenCV.

**ASIGNATURA:** Tendencias innovadoras de la profesión **SEMESTRE:** 2024 B

## TEMA: EXTRACCIÓN Y VISUALIZACIÓN DE INFORMACIÓN DE PELÍCULAS EN IMDB MEDIANTE WEB SCRAPING

### Marco Teórico sobre Web Scraping

#### Definición de Web Scraping

El web scraping es una técnica utilizada para extraer información de sitios web de manera automática. A través de esta técnica, es posible recopilar datos estructurados, como tablas, listas o contenido específico de una página web, para su posterior análisis o procesamiento. Esta práctica es ampliamente utilizada en ámbitos como el comercio electrónico, el análisis de datos, el marketing y la investigación académica.

#### Cómo se Realiza el Web Scraping

El proceso de web scraping involucra el uso de herramientas o programas que acceden a sitios web, descargan el contenido HTML y analizan su estructura para extraer información relevante. Este procedimiento consta de las siguientes etapas:

1. **Envío de la Petición:** El software realiza una petición HTTP al servidor donde está alojada la página web, obteniendo el código fuente HTML de la misma.
2. **Análisis del HTML:** Una vez obtenido el código fuente, se utiliza un parser para interpretar y analizar la estructura del documento, identificando las etiquetas HTML que contienen los datos deseados.

3. **Extracción de Datos:** El programa selecciona los elementos de interés (por ejemplo, textos dentro de etiquetas, atributos, o valores en tablas) y los organiza en un formato estructurado como JSON, CSV o bases de datos.
4. **Almacenamiento:** Los datos extraídos son almacenados para su posterior procesamiento o análisis.

## Utilidades del Web Scraping

El web scraping tiene múltiples aplicaciones, entre las cuales destacan:

- **Monitorización de Precios:** Empresas de comercio electrónico utilizan esta técnica para rastrear los precios de competidores y ajustar sus estrategias de venta.
- **Generación de Bases de Datos:** Investigadores recopilan grandes volúmenes de datos para análisis estadísticos.
- **Análisis de Tendencias:** Empresas analizan datos de redes sociales o sitios web para identificar patrones de comportamiento del consumidor.
- **Automatización de Tareas:** El web scraping permite automatizar procesos repetitivos como la búsqueda de información.

## Herramientas Comunes para Web Scraping

Existen diversas herramientas y bibliotecas para realizar web scraping, entre las que destacan:

- **Python:** Con bibliotecas como BeautifulSoup, Scrapy y Selenium, Python es una de las opciones más populares para web scraping.
- **R:** Utilizado en entornos académicos y estadísticos, con paquetes como rvest.
- **Herramientas No Code:** Aplicaciones como Octoparse y ParseHub permiten a los usuarios realizar web scraping sin necesidad de programación.

## Aspectos Legales y Éticos

Es importante tener en cuenta las implicaciones legales y éticas del web scraping. Muchas páginas web establecen restricciones en sus términos de uso sobre la

recopilación automática de datos. Además, es crucial respetar la privacidad de los usuarios y no infringir leyes relacionadas con la protección de datos.

## **Tema a Realizar como Trabajo Final**

### **Descripción del Proyecto**

El trabajo consiste en desarrollar una aplicación que permita extraer información relevante de películas desde el sitio web IMDb utilizando técnicas de web scraping. Posteriormente, los datos serán organizados y presentados en un dashboard interactivo creado en Power BI.

### **Objetivos del Proyecto:**

- Implementar una solución en Python para realizar el scraping del sitio IMDb, enfocándose en información como el título de la película, año de lanzamiento, duración y clasificación.
- Almacenar los datos obtenidos en un formato estructurado como CSV para facilitar su análisis.
- Diseñar un dashboard en Power BI que permita visualizar la información extraída de manera intuitiva y dinámica.

### **Justificación**

Este proyecto permite explorar el potencial de las técnicas de web scraping en la extracción de datos masivos y su aplicación en la generación de valor mediante herramientas de visualización. Asimismo, brinda la oportunidad de abordar aspectos prácticos y teóricos del manejo de grandes volúmenes de datos.

### **Resultados Esperados:**

- Una herramienta funcional de scraping que obtenga datos actualizados de IMDb.
- Un dashboard interactivo que permita a los usuarios analizar las tendencias y características de las películas.

## Referencias

- Alvarez, J. (2021). *Web Scraping con Python: Aprende a extraer datos de internet de manera automatizada*. Editorial Independiente.
- Mitchell, R. (2018). *Web Scraping with Python: Collecting More Data from the Modern Web* (2nd ed.). O'Reilly Media.
- Ryan, K. (2020). *Data Wrangling with Python: Tips and Tools to Make Your Life Easier*. Apress.
- Vohra, S. (2019). *Pro Power BI Desktop: Self-Service Analytics and Data Visualization*. Apress.
- Winston, W. (2019). *Microsoft Power BI for Data Analysis and Visualization*. Wiley.