



**Alumno:**

Oscar Ivan Valenzuela Diaz

**Doctorado:**

Sistemas Computaciones

**Materia:**

Seminario de Programación de Computadoras

**Actividad de Aprendizaje:**

Actividad 2

**Docente de la materia:**

Dr. Gandhi Samuel Hernández Chan

# Desarrollo

## Introducción

En el marco del seminario de Programación de Computadoras, la actividad de la Semana 2 se centró en la extracción y almacenamiento de datos utilizando técnicas de web scraping y bases de datos. La tarea principal fue seleccionar y justificar el tipo de base de datos más adecuado para almacenar los datos extraídos. Para este propósito, se eligió una base de datos relacional (SQLite) debido a varias razones que subrayan su simplicidad, compatibilidad y escalabilidad.

## Semana 2

### Tipo de Base de Datos

Para el almacenamiento de los datos extraídos, se eligió una base de datos relacional (SQLite). La razón para esta elección incluye:

- Simplicidad: SQLite es fácil de usar y no requiere configuración de servidor.
- Compatibilidad: Es compatible con Python y se puede utilizar directamente con pandas para la manipulación de datos.
- Escalabilidad: A pesar de ser una base de datos ligera, es adecuada para proyectos de pequeño a mediano tamaño.

### Proceso de Almacenamiento en SQLite

A continuación, se muestra el script para almacenar los datos en una base de datos SQLite, incluyendo una columna id auto incremental:

```
import sqlite3
from datetime import datetime

# Leer el CSV
df = pd.read_csv('tendencias_elimparcial.csv')

# Agregar una columna de fecha con la fecha actual
df['fecha'] = datetime.now().strftime('%Y-%m-%d')

# Conectar a la base de datos SQLite (se crea si no existe)
conn = sqlite3.connect('noticias.db')
c = conn.cursor()

# Crear la tabla en la base de datos con una columna ID auto
incremental
c.execute(''
```

```

CREATE TABLE IF NOT EXISTS noticias (
    id INTEGER PRIMARY KEY AUTOINCREMENT,
    fuente TEXT,
    enlace TEXT,
    fecha TEXT,
    noticias TEXT
)

'''

# Insertar los datos del DataFrame en la base de datos
df.to_sql('noticias', conn, if_exists='append', index=False)

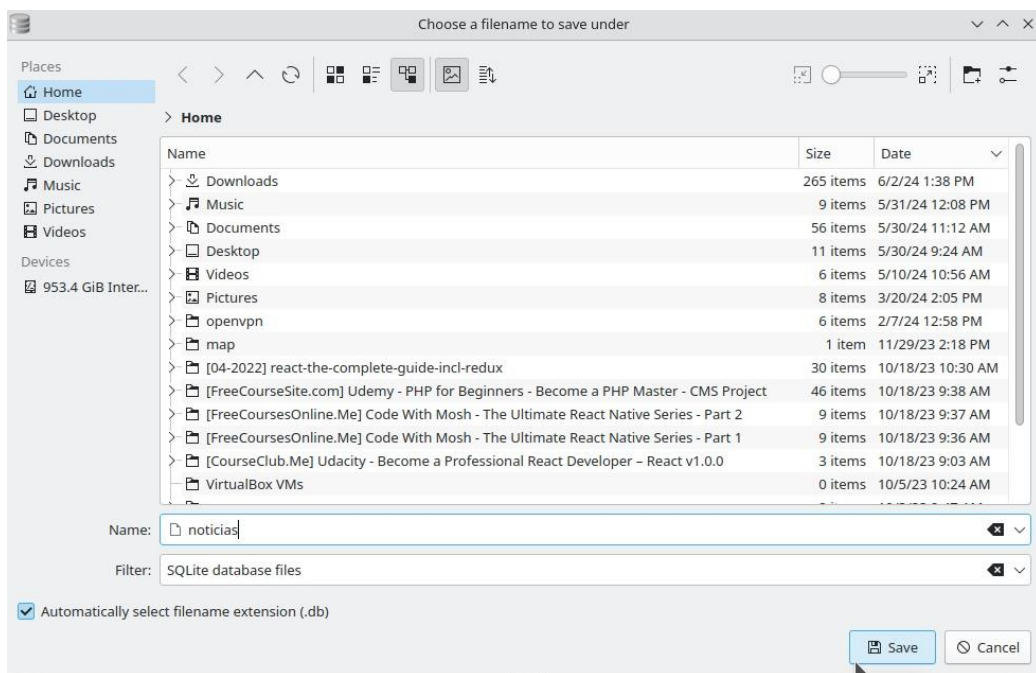
# Confirmar los cambios y cerrar la conexión
conn.commit()
conn.close()

print("Datos insertados correctamente en la base de datos SQLite.")

```

# Imágenes de DB Browser de SQLite

## 1.Creamos la base de datos



## 2. Definimos la tabla

The screenshot shows the 'Edit table definition' window for a table named 'noticias'. The 'Fields' tab is active, showing two fields: 'id' and 'fuente'. Both are of type 'INTEGER'. The 'id' field has 'NN' (Not Null) checked, while 'fuente' does not. Below the fields list, the SQL code is displayed:

```
1 CREATE TABLE "noticias" (  
2   "id" INTEGER NOT NULL,  
3   "fuente" INTEGER  
4 );
```

At the bottom, there are 'OK' and 'Cancel' buttons.

## 3. Podemos visualizar la estructura de la tabla

The screenshot shows the 'DB Browser for SQLite' interface. The 'Database Structure' tab is selected, displaying a tree view of the database contents. The 'noticias' table is highlighted. The 'Edit Database Cell' window is open on the right, showing the 'Text' mode and the 'Type of data currently in cell: Text / Numeric' dropdown.

## 4. Podemos visualizar los datos en esta pestaña

The screenshot shows the 'DB Browser for SQLite' interface with the 'Browse Data' tab selected. The 'noticias' table is displayed, showing columns: 'id', 'fuente', 'enlace', 'fecha', and 'noticias'. The data is listed in a table with 13 rows. The 'Edit Database Cell' window is open on the right, showing the 'Text' mode and the 'Type of data currently in cell: Text / Numeric' dropdown.

	id	fuente	enlace	fecha	noticias
1	1	el imparcial	/ocurioso/2024/06/02/elecciones-mexico-202...	2024-06-02	edicionesel imparcial mexicano elecciones...
2	2	el imparcial	/ocurioso/2024/06/02/elecciones-2024-cuales...	2024-06-02	edicionesel imparcial lo curioso los memes se ...
3	3	el imparcial	/ocurioso/2024/06/02/rosario-lbarra-de-la-...	2024-06-02	edicionesel imparcial lo curioso elecciones s...
4	4	el imparcial	/ocurioso/2024/05/31/senoritas-cantan-cancio...	2024-06-02	edicionesel imparcial lo curioso tiktokseñoras ...
5	5	el imparcial	/ocurioso/2024/06/02/video-juven-graba-...	2024-06-02	edicionesel imparcial lo curioso viral de junio ...
6	6	el imparcial	/ocurioso/2024/06/01/hallan-tesoro-de-...	2024-06-02	edicionesel imparcial lo curioso ...
7	7	el imparcial	/ocurioso/2024/06/01/video-mujer-descubre...	2024-06-02	edicionesel imparcial lo curioso viralvideo ...
8	8	el imparcial	/ocurioso/2024/06/01/video-familia-construye...	2024-06-02	edicionesel imparcial lo curioso viralvideo ...
9	9	el imparcial	/ocurioso/2024/06/02/la-mayoria-de-eda-...	2024-06-02	edicionesel imparcial lo curioso este cambio h...
10	10	el imparcial	/ocurioso/2024/06/02/en-que-año-fueron-las...	2024-06-02	edicionesel imparcial lo curioso eleccionesest...
11	11	el imparcial	/ocurioso/2024/06/02/el-origen-de-la-...	2024-06-02	edicionesel imparcial lo curioso curiosidadesel...
12	12	el imparcial	/ocurioso/2024/06/02/candidatos-...	2024-06-02	edicionesel imparcial lo curioso esta novedad ...
13	13	el imparcial	/ocurioso/2024/06/02/sabias-que-en-el-2006-L...	2024-06-02	edicionesel imparcial lo curioso eleccionesla ...

# Conclusiones

## Actividad 2

El desarrollo de esta herramienta de extracción y almacenamiento de datos permitió aplicar conocimientos de programación, manipulación de datos y técnicas de web scraping. Además, se demostró la viabilidad de utilizar una base de datos SQLite para almacenar y gestionar los datos extraídos. Esta práctica no solo cumple con el objetivo del módulo, sino que también proporciona una comprensión práctica de cómo interactuar con la arquitectura de una computadora y utilizar diversos métodos para establecer comunicación entre la máquina y el ser humano mediante la programación de computadoras.

# Referencias:

## Programación de Computadoras

- Sipser, M. (2012). Introduction to the Theory of Computation (3rd ed.). Cengage Learning.
- Hunt, A., & Thomas, D. (2019). The Pragmatic Programmer: Your Journey to Mastery (20th Anniversary Edition). Addison-Wesley.

## Web Scraping

- Mitchell, R. (2018). Web Scraping with Python: Collecting More Data from the Modern Web (2nd ed.). O'Reilly Media.
- Hajba, G. L. (2018). Web Scraping with Python: A Step-by-Step Guide to Scraping the Web. Apress.

## Manipulación de Datos

- McKinney, W. (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython (2nd ed.). O'Reilly Media.
- Grus, J. (2019). Data Science from Scratch: First Principles with Python (2nd ed.). O'Reilly Media.
- Sweigart, A. (2015). Automate the Boring Stuff with Python: Practical Programming for Total Beginners. No Starch Press.

## Arquitectura de Computadoras

- Patterson, D. A., & Hennessy, J. L. (2017). Computer Organization and Design: The Hardware/Software Interface (5th ed.). Morgan Kaufmann.
- Tanenbaum, A. S., & Bos, H. (2014). Modern Operating Systems (4th ed.). Pearson.