

EXPLORANDO LA BIODIVERSIDAD EN CUNDINAMARCA Y BOYACÁ

FASE II. WEBSCRAPING CON SELENIUM

Docente: Emerson Garay

Integrantes:

Fabian Orlando Católico Suárez

David Ricardo Cative Guerra

Jhaireth Jisseth Calleja Medina

Ministerio de Tecnologías de la Información y las Comunicaciones

Universidad Sergio Arboleda

2024

Recopilación de datos:

Para realizar el debido proceso de recolección de datos, primero se indago sobre diferentes páginas las cuales mostraran información de valor, esto en cuanto al tema de biodiversidad en los departamentos de Boyacá y Cundinamarca.

la opción más viable fue la plataforma comunitaria llamada "Naturalista Colombia", la cual incentiva a los usuarios a interactuar con la naturaleza y a contribuir al seguimiento de la biodiversidad en Colombia. Los usuarios de dicha plataforma pueden registrar observaciones, compartirlas con otros naturalistas y colaborar en listas de especies. Además, la NaturalistaCO facilita el aprendizaje sobre la naturaleza e identificar especies. Su objetivo es promover la conservación a través de la ciencia ciudadana.

Enlace de la plataforma: <https://colombia.inaturalist.org/>

Se implementó **Python** como lenguaje de desarrollo y la librería **Selenium** la cual permite automatizar navegadores web y, mediante XPath y selectores CSS, poder navegar y seleccionar elementos HTML de manera consistente.

Procesamiento de datos:

- Creación del archivo "web_scraping.py".
- Importar las librerías de Selenium (webdriver, by, Service,Options, WebDriverWait, EC), la librería time y la librería csv .
- Conectar el driver del navegador Chrome, para ello se debe validar que sea la misma versión del que se tiene instalado.
- Se identifican las clases y XPath necesarias para los obtener datos (la clase del div contenedor, la clase del nombre de la especie, la clase del nombre científico, el XPath de la URL de la imagen, el XPath de la URL del icono y el XPath del enlace de "Siguiente"; este último con el fin de obtener los datos de las páginas siguientes si se tenían).
- Extraer los datos con métodos como find_elements.
- Se valida si la página tiene más resultados en páginas siguientes.
- Guardar los datos en un diccionario por especie "spe_dict".
- Almacenar los diccionarios en una lista "species_list".
- Imprimir los datos obtenidos de cada especie.
- Guardar la lista de diccionarios en un archivo CSV "Biodiversidad en Boyacá y Cundinamarca.csv"

Integración con la interfaz web

- Importar csv.
- Crear la función “leerDatos()” para leer el archivo “Biodiversidad en Boyacá y Cundinamarca.csv” y retornar su contenido por filas.

```
1 def leerDatos():
2     datos = []
3     with open('Biodiversidad en Boyacá y Cundinamarca.csv', 'r',
4               newline='', encoding='utf-8') as csv_file:
5         reader = csv.DictReader(csv_file, delimiter=';')
6         for row in reader:
7             datos.append(row)
8     return datos
```

- Enviar la información del archivo CSV al archivo “biodiversidad.html” para poder manipularla desde allí.

```
1 # Configurar la ruta de biodiversidad
2 @app_flask.route('/biodiversidad')
3 def ir_a_biodiversidad():
4     data = leerDatos()
5     departamento = request.args.get('departamento')
6     clase = request.args.get('clase')
7     tipo = request.args.get('tipo')
8
9     if departamento:
10         data = [d for d in data if d['DEPARTAMENTO'] == departamento]
11     if clase:
12         data = [d for d in data if d['CLASE'] == clase]
13     if tipo:
14         data = [d for d in data if d['TIPO'] == tipo]
15
16     return render_template('biodiversidad.html', datos=data)
```

- Manipulación de los datos e incorporación en la página

```

1 <div class="container">
2   {% for dato in datos %}
3   <div class="card">
4     {% set imagen = dato['ICONO'] if dato['IMAGEN'] == 'No disponible' else dato['IMAGEN'] %}
5     
6     {% if dato['NOMBRE'] == 'No disponible' %}
7     <h2>{{ dato['NOMBRE CIENTÍFICO'] }}</h2>
8     <p><strong>Departamento:</strong> {{ dato['DEPARTAMENTO'] }}</p>
9     <p><strong>Tipo de especie:</strong> {{ dato['TIPO'] }}</p>
10    {% else %}
11    <h2>{{ dato['NOMBRE CIENTÍFICO'] }}</h2>
12    <p><strong>Nombre Científico:</strong> {{ dato['NOMBRE CIENTÍFICO'] }}</p>
13    <p><strong>Departamento:</strong> {{ dato['DEPARTAMENTO'] }}</p>
14    <p><strong>Tipo de especie:</strong> {{ dato['TIPO'] }}</p>
15    {% endif %}
16  </div>
17  {% endfor %}
18 </div>

```

- Resultado de los datos importados en la página web.


BioTechCol

Inicio Biodiversidad Explorar Contacto

Biodiversidad

Departamento: Cundinamarca
Clase: Animales
Tipo: Nativa
Filtrar

Flora y Fauna




Turdus fuscater

Nombre Científico: Turdus fuscater

Departamento: Cundinamarca

Tipo de especie: Nativa




Zonotrichia capensis

Nombre Científico: Zonotrichia capensis

Departamento: Cundinamarca

Tipo de especie: Nativa




Zenaida auriculata

Nombre Científico: Zenaida auriculata

Departamento: Cundinamarca

Tipo de especie: Nativa



Tyrannus melancholicus

Nombre Científico: Tyrannus melancholicus

Departamento: Cundinamarca

Tipo de especie: Nativa