



EXTRACCIÓN DE DATOS CON PYTHON



EXTRACCIÓN DE DATOS CON PYTHON



La **extracción de datos** es un **proceso de eliminar datos de una fuente y almacenarlos en una forma diferente**. Se puede hacer de forma manual o automática usando Python.

Se usa comúnmente para la extracción de datos, ya que tiene muchas bibliotecas disponibles para automatizar el proceso, como MySQL, Postgresql, hojas de cálculo de Excel, etc.




LIBRERÍAS PARA EXTRACCIÓN DE DATOS




La **librería más popular** para extraer **datos de bases de datos MySQL** es la librería **mysql**.

También hay **otras librerías que se pueden usar para extraer datos** de otras fuentes, como:



 **Pandas:** para extraer conjuntos de datos y tablas en varios formatos en uno o más de los siguientes formatos: CSV, HTML, JSON o libro de Excel (.xlsx)



 **Openpyxl:** para trabajar con archivos de hoja de cálculo en varios formatos en uno o más de los siguientes formatos: CSV, HTML, XML o libro de Excel (.xlsx) - solicitudes: para obtener datos de una URL



 **Datetime:** para manejar objetos de fecha y hora

LIBRERÍAS PARA EXTRACCIÓN DE DATOS



NumPy: una librería de propósito general para computación científica con módulos diseñados para aplicaciones numéricamente intensivas como álgebra lineal, transformadas de Fourier, y generación de números aleatorios



scipy: una librería para computación científica y visualización con muchos módulos, incluidas funciones especiales, álgebra lineal, transformadas de Fourier, estructuras de datos y algoritmos para el procesamiento de señales.



netCDF: software diseñado para leer, escribir, actualizar y manipular grandes conjuntos de datos en un entorno de red. Utiliza una API orientada a objetos que admite C, Fortran y C++

LIBRERÍAS PARA EXTRACCIÓN DE DATOS



BeautifulSoup

La librería de Python llamada **Beautiful Soup** es una de esas herramientas que se puede usar para **extraer datos de archivos HTML y XML**. Sin embargo, **no puede** extraer datos de otros formatos, como archivos **PDF y videos**.

Para extraer datos de un PDF, utilice la misma técnica que para extraer datos de HTML. Use BeautifulSoup para analizar el archivo y luego utilícelo para obtener el contenido de texto de cada elemento HTML.



PROCESO DE EXTRACCIÓN DE DATOS



El **proceso de extracción de datos** con Python es un paso **importante en el desarrollo de una aplicación**. En este tutorial vamos a ver cómo crear una base de datos con los datos que se han recogido durante la investigación. A continuación, veremos el proceso para extraer los datos y cómo utilizar Python para crear una aplicación web con esta información.

Python aplicará el **método de extracción de datos** y los **resultados** se guardarán en **una tabla**.

P La **instrucción SELECT** se utiliza para **obtener ciertos valores** de una tabla.

P La **instrucción DELETE** es utilizada para **quitar datos** de una tabla

P La **instrucción INSERT** se utiliza para **agregar datos** a una tabla.



PROCESO DE EXTRACCIÓN DE DATOS



Para realizar el **procedimiento de extracción de datos con Python**, vamos a empezar nuestro proceso desde el comienzo, donde tecleamos la siguiente **instrucción: SELECT * FROM TableName**.

La **instrucción SELECT** especifica todos los **valores que se pueden obtener de las filas y columnas** en una tabla. **Python** usará estos valores para generar **nuevas columnas de resultados**.

El **valor lógico** que se puede obtener es **NULL**, pero Python también podría tener un **valor de error lógico** como **NOINDEX** o **NOFOO**.



PROCESO DE EXTRACCIÓN DE DATOS



El **resultado** de la **instrucción SELECT** especifica la **lista de columnas en la tabla**.

El **resultado** de **cada instrucción SELECT** especifica el **resultado para cada uno de los elementos** en esta lista.

Puede usar este método para realizar la **misma tarea DOS veces**. Es posible que se necesite una **instrucción SELECT DISTINCT** para hacer esto.

Los **parámetros SELECT* son despleables**. Esto es particularmente útil si el nombre de la tabla es largo.



pontia

¿Estás interesado en aprender más sobre la extracción de datos con Python?

En Pontia tenemos másters en Data Analytics y en Data Science, donde te enseñamos más sobre cómo extraer datos con este lenguaje tan demandado.

www.pontia.tech

Pontia.Tech

