

## ¿QUÉ ES PYQUERY?

PyQuery es un paquete Python de terceros que analiza y extrae datos de páginas "xml" y "html". Está inspirado en la biblioteca jQuery JavaScript y tiene una sintaxis y una API prácticamente igual, lo que le permite leer y cambiar el árbol de documentos utilizando diferentes métodos convenientes y códigos de acceso directo. Esta biblioteca está diseñada para ser simple y fácil de usar, lo que significa que cualquier persona con conocimientos básicos de Python puede utilizarla para extraer información de documentos HTML.

- Si bien PyQuery es una herramienta poderosa, no es la única opción disponible para trabajar con HTML y XML en Python. En última instancia, la elección entre PyQuery y BeautifulSoup depende de sus necesidades y preferencias específicas.

## CÓMO UTILIZAR PYQUERY PARA ANALIZAR DOCUMENTOS HTML

El primer paso para utilizar Pyquery es instalarlo en tu entorno Python utilizando pip.

```
python -m pip install pyquery
```

Después de instalar la biblioteca, importa la clase PyQuery desde el módulo pyquery. Luego, crea un objeto PyQuery utilizando una cadena que contenga el HTML que deseas analizar. Una vez que hayas creado tu objeto PyQuery, puedes utilizar sus métodos para buscar y manipular elementos HTML.

## BÚSQUEDA DE ELEMENTOS HTML CON PYQUERY

Hay varias formas de buscar elementos HTML utilizando Pyquery. Algunas de las formas más comunes son:

- Utilizando un selector CSS
- Buscando elementos por tag
- Buscando elementos por clase

## EXTRACCIÓN DE DATOS DE UNA PÁGINA WEB

Una vez que hayas encontrado los elementos que deseas, puedes extraer información de ellos utilizando los métodos de Pyquery. Algunos de los métodos más comunes son:

- `.text()` - devuelve el texto de un elemento
- `.html()` - devuelve el HTML de un elemento
- `.attr()` - devuelve el valor de un atributo de un elemento

## MANIPULACIÓN DEL DOM

Pyquery también te permite manipular el DOM (Modelo de Objetos del Documento) de la página web. Puedes modificar elementos existentes, añadir nuevos elementos o eliminar elementos existentes.

## WEB SCRAPING CON PYQUERY

Estamos en un momento en el cual tenemos datos en todos los lados donde miremos, tenemos datos abiertos de infinidad de APIS que muchas organizaciones ponen a disposición de los desarrolladores, sin embargo muchas veces no miramos una de las fuentes principales de datos que tenemos a nuestro alcance. La web.

Empresas de ciencias, de marketing, de negocios, traders, y todo mundo en este momento realizan robots o scrappers para obtener información de noticias, comportamiento de mercado, información de ciencia, etc.

Esto lo hacen a travez de diferentes tecnologías que les permiten extraer el HTML de los sitios y posteriormente analizarlos y «escarbar» datos entre tanto html para poder ordenar información, almacenarla en sus servidores propios y posteriormente analizarlos.

Python tiene diferentes librerías para realizar estas tareas, sin embargo, hay uno en particular por la manera sencilla que permite analizar el HTML que obtenemos de los sitios, es algo muy similar a como lo maneja JQUERY mediante selectores. La librería llamada Pyquery

## VENTAJAS

- te permite analizar y manipular documentos HTML utilizando lenguaje Python.
- fácil de usar para trabajar con archivos HTML y XML
- Su API y sintaxis similar a jQuery facilitan el análisis, el recorrido y la manipulación de HTML y XML, así como la extracción de datos.
- es adecuada para diferentes casos de uso. Todo va a depender de lo que quiera el usuario.
- Con respecto a su uso, podemos encontrar mucha información en cuanto a código, métodos y formas de aplicación para distintas oportunidades.

## DESVENTAJAS

- Puede generar una carga adicional de rendimiento en las páginas web, especialmente si se utilizan muchas funciones y animaciones.
- Puede limitar la flexibilidad y personalización en algunos casos, ya que las funciones y herramientas predefinidas pueden no adaptarse exactamente a las necesidades específicas de un proyecto.
- En ocasiones, conlleva problemas de compatibilidad si se utilizan versiones antiguas o incompatibles, o si se combinan con otras bibliotecas o frameworks que tienen conflictos de nombres o funciones.

**Mi** recomendación sobre el uso de ésta librería para aplicarla en el Proyecto de scraping sobre la página web TripAdvisor, y aclarando que mi opinión es ampliamente de principiante en el tema, creo que se podría utilizar porque hay mucha información en la web a cerca de su uso, es fácil de aplicar y aprender, muy similar a otras librerías.

## Bibliografía

*Cómo analizar y extraer información de HTML utilizando Pyquery.* (s.f.). Obtenido de [trspos.com/panse-raspar-html-usando-pyquery/](https://trspos.com/panse-raspar-html-usando-pyquery/)

Kaubre, V. (25 de 01 de 2023). *Cómo analizar HTML con PyQuery: tutorial de Python.* Obtenido de <https://oxylabs.io/blog/parse-html-python-pquery>

PyQuery. (s.f.). Obtenido de <https://www.pyquery.org/>

ZUL, A. D. (28 de julio de 2022). *defullstack*. Obtenido de <https://defullstack.com/web-scraping-con-pyquery/>