Scraping a JavaScript Rendered Web Page with python and selenium

NOTE

Before starting to write any of the code to scrape a particular site there are certain permitions people needs to verify as part of **The Robots Exclusion Protocol**.\s\s

Read more about it from the wikipedia page

https://en.wikipedia.org/wiki/Robots_exclusion_standard

In short every site comes with a **robots.txt** file which specifies what can be crawled or scraped. So make sure you are not scraping anything that is not permitted.

Requirements

> *selenium binding for python* `pip install selenium`
>
> *python json module -> build in module no need to install*
>
> *BeautifulSoup for parsing html* `pip install beautifulsoup4`
>
> *Google Chrome browser https://pages.github.com/ Download for the url provided here based on your operating system.*
>
> *ChromeDriver(WebDriver for Chrome)*
>
> *http://chromedriver.chromium.org download from this url*

There is an Easy way to install python requirements with

requirements.txt.

Requirements files" are files containing a list of items to be installed using pip install like so

> `pip install -r requirements.txt`

In this example i am scraping world intellectual property organization brand database page. It has a table on the middle and right which is rendered using java script.

http://www.wipo.int/branddb/en/ this is the page we will be working on.

We will be getting all the info from the **Source** in **FILTER BY** section.

Output

Running the code in wipo_scrape.py will json output for the scraped data.