<https://www.youtube.com/watch?v=y3nUE-RNETY>

So first of all I installed a chome extension which is “selector gadget” then it appeared up in the chrome’s toolbar.

To get css or xpath of certain item all you need is to click on the item and all similar items will get also highlighted. For irrelevant ones you can remove them simply by clicking on them in which a red box will show once you hover on them.

Copy the css selector they gave you and add it to your script by

response.css(‘coppied selector’)

open on your cmd scrapy shell by: scrapy shell url

example scrapy shell <https://www.amazon.com/s?k=hair+dryer>

then you can view the browser by typing: view(response)

and get values of response by: response.css(‘coppied selector’)

Notes:

To prevent amazon from blocking your browser while scraping🡪 install pip install scrapy-user-agents and copy the below code to your settings.py

To check the source of this code go to pypi and check the

“Configuration

Turn off the built-in UserAgentMiddleware and add RandomUserAgentMiddleware.

In Scrapy >=1.0::

DOWNLOADER\_MIDDLEWARES = {

'scrapy.downloadermiddlewares.useragent.UserAgentMiddleware': None,

'scrapy\_user\_agents.middlewares.RandomUserAgentMiddleware': 400,

}

Syntax:

.extract() 🡪 removes tags from css selector

response.css(‘coppied selector::text’ ---> get text from span

response.css(‘coppied selector’).css(‘coppied selector::text’).extract()🡪 to access nested selectors

response.css(‘coppied selector’).css(‘coppied selector::attr(src)’).extract()🡪 to get href