Plugins required: python, bs4, beautifulsoup, requests, selenium

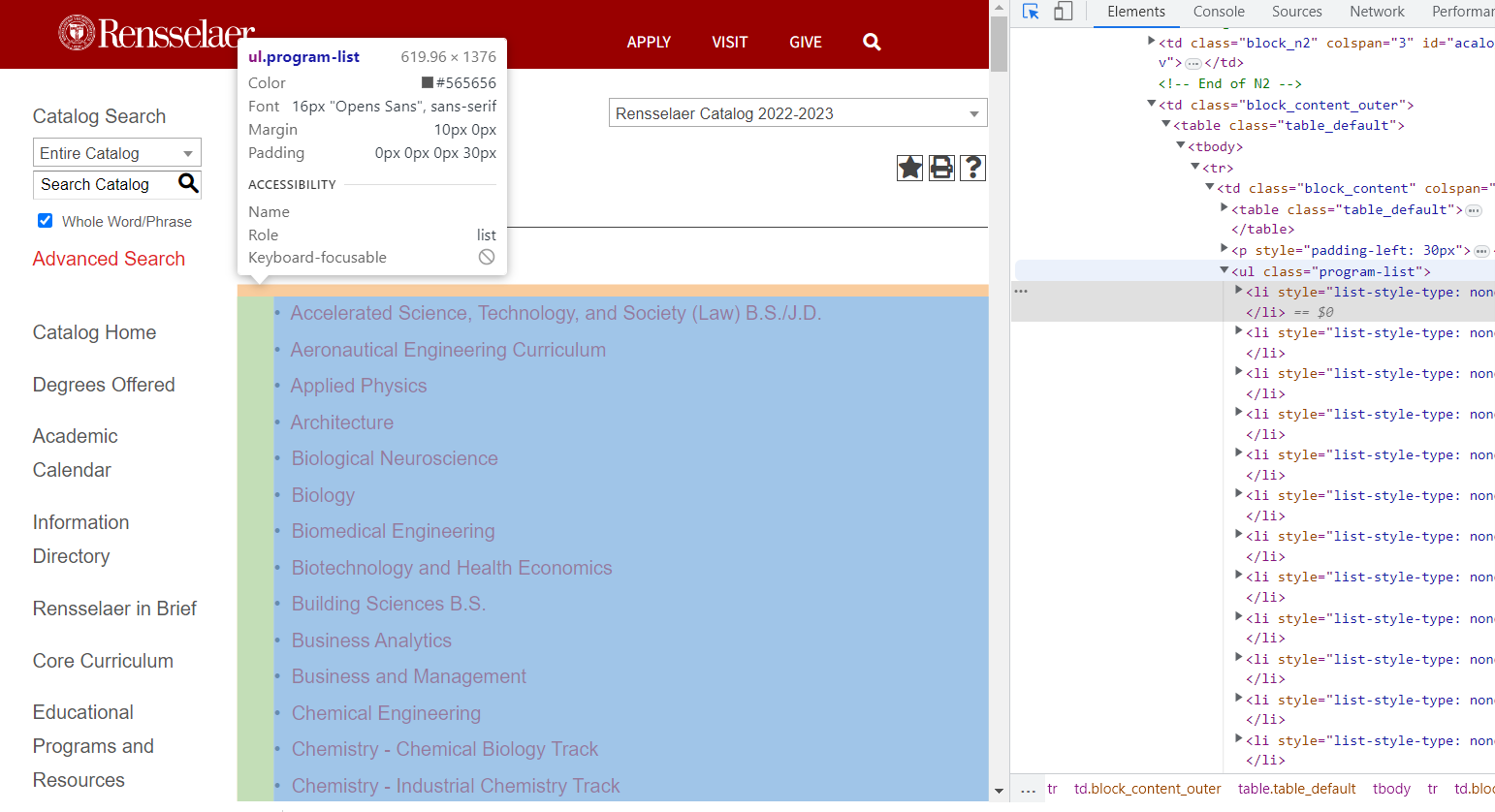
pip install selenium

pip install requests

pip install bs4

pip install beautifulsoup4

1. Go to RPI catalog -> Programs: <http://catalog.rpi.edu/content.php?catoid=24&navoid=604>
2. Create html\_raw.txt file Copy the HTML code under program-list ul (example: URL2022FallRaw):



1. cd to \yacs.n\rpi\_data\Degree\_Tamplates, run html\_parser.py on rawfile, this creates majorURLlist.txt (example: majorURLlist2022FA.txt)
2. cd to \yacs.n\rpi\_data\Degree\_Tamplates, run major\_scrap on majorURLlist, this creates a organized majordata.txt (example: majorData2022.txt). An error throws at the end of the program but its ok.
3. cd to \yacs.n\rpi\_data\Degree\_Tamplates\new run to\_json\_2022.py on majordata. This generates the majordata.json file (example: majorData2022.json). Due to different format among majors in RPI catalog. Some bugs need to be fixed manually.
4. In \yacs.n\rpi\_data\Degree\_Tamplates\new\Environmental Science the ienv.py is a selenium scraper dealing with special case. For 2022, Environmental Science URL is kicked from majorURLlist2022FA.txt and added back in json afterwards.