

Wouter Vandueren - CaseStudy: Webscraper

Github:

Repo indeed: <https://github.com/WouterV2207/CaseStudyIndeed>

Repo youtube: <https://github.com/WouterV2207/CaseStudyYoutube>

Repo hbvl: <https://github.com/WouterV2207/CaseStudyHBVL>

Wat en hoe heb ik mijn oplossing gemaakt:

Ik ben begonnen met op te zoeken wat en hoe je te werk gaat met selenium. Vervolgens heb ik een voorbeeld opgezocht van een webscraper die ook gemaakt is met selenium. Daarna ben ik begonnen met de webscraper van youtube omdat we hiervoor al een voorbeeld kregen bij de uitleg over de case study. Nadat ik klaar was met de webscraper van youtube ben ik begonnen aan de webscraper van indeed. Voor deze webscraper heb ik veel info opgezocht en veel verschillende dingen geprobeerd om de link van de job te kunnen verkrijgen wat me helaas niet gelukt is. Bij mijn zelfgekozen site heb ik gekozen voor de internetsite van hbvl. Hier heb ik gekozen om de krantenartikels te sorteren op uploaddatum vandaag en als data de krantenartikels en uploaddatum te nemen. Ik heb ook in elke webscraper ter controle laten staan dat de data ook in de console wordt geprint. De data wordt ook weggeschreven naar een csv bestand. Ik heb elke webscraper in een aparte github repo geupload.

Youtube link (demonstratie):

<https://www.youtube.com/watch?v=S8x8kLmS0II>

Hoe het werkt:

Wordt uitgelegd in de video

Hoofdbronnen:

<https://www.youtube.com/watch?v=wtH4i7CPg1M>

<https://www.youtube.com/watch?v=gRoMR3NcpPQ&t=75s>

<https://vainolo.com/2020/09/30/scraping-dynamic-web-pages-with-c-and-selenium/>

<https://stackoverflow.com/questions/63231164/indeed-web-scraping-python-selenium-beautifulsoup>