

Prática: Web Scraping com Python p/ Ciência de Dados (II)

Eduardo Prasniewski

Pessoalmente já tenho uma boa experiência com web scraping pois trabalhei como analista de aquisição de dados, porém usávamos a biblioteca Scrapy, do python também, acredito que seja um framework mais robusto que o BeautifulSoup. Na vídeo aula ele explica os conceitos básicos de HTML, tags, classes, atributos etc. Primeiramente passa um simples exemplo com um HTML estático local para passar alguns comandos básicos de filtro xml. Esse código não implementei pois não vi muita necessidade.

Por fim, é passado um código de web scraping na "vida real", requisitando dados de um site de empregos, e fazendo assim uma filtragem de dados. Isso se encaixa muito na base da pirâmide vista no video do módulo anterior. Após o processamento de dados o cada index é separado em um único arquivo. Vale ressaltar que esse tipo de scraping vale somente para sites que já vem carregados com os dados na requisição, páginas que possuem lazy load (consumindo dados após o site carregar através de uma API) não se encaixam nesse tipo de scraping, e isso é muito comum nas aplicações web modernas.