

Medium(dot)com Web Crawler with Django

Timeline-

DATE	WORK
03/06	Understanding requirements of the project
04/06	Learning the ethical and legal implications of scraping.
04/06	Determining which technology stack will be fastest and easiest to implement a web crawler
04/06	Decided upon python with Django and started with the config
05/06	Learning beautiful soup Learning to locate elements using in beautifulsoup
05/06	Encountering Infinite scrolling website problem.
06/06	Understanding use of pagination, Ajax, JQuery in websites for autoscrolling.
06/06	Finding a solution with selenium. Discarded due to it being inelegant and unusable.
07/06	Deciding to Implement the solution using Scrapy library. Learning Scrapy.
07/06	Locating medium(dot)com archives and implementing scraping on archives
08/06	Learning Scrapy and Django integration. Storing Scraped Data in sqlite.
08/06	Implementing makeshift Webpage to display data. Extracting Data from RDBMS and displaying it on a webpage.