Web Crawling

1. Crawling Target
2. Crawling tool

In this project, we use two libraries of Python as crawler tools, namely Beautiful Soup and Selenium. Beautiful Soup is a Python library that can extract data from HTML or XML files. It can use the converter you like to implement the document navigation, find and modify the document. Selenium was originally born for testing. Today, it has become a good tool for crawlers. It controls the browser and browses the web like humans which is Ideal for crawling dynamic web pages.

1. Crawling Environment

Language: Python 3

IDE: Jupiter Notebook

Library: Beautiful Soup, Selenium

Device: MacBook Pro (13-inch, 2017, Four Thunderbolt 3 Ports)

OS: MacOS Mojave 10.14.3

CPU: 3.1 GHz Intel Core i5

Memory: 8 GB 2133 MHz LPDDR3

1. Crawling Step

(1)

(2)