

WEBDATA201: Collecting Web Data

Web scraping is a technique for extracting information from websites. This can be done manually but it is usually faster, more efficient and less error-prone if it can be automated.

Web scraping allows you to convert non-tabular or poorly structured data into a usable, structured format, such as a .csv file or spreadsheet. But scraping is about more than just acquiring data: it can help you track changes to data online, and help you archive data. In short, it's a skill worth learning.

So join us for this web scraping workshop to learn web scraping, using the researcher-focused training modules from the highly regarded Software Carpentry Foundation.

Learning Outcomes

- The concept of structured data
- The use of XPath queries on HTML document
- How to scrape data using browser extensions
- How to scrape using Python and Scrapy
- How to automate the scraping of multiple web pages

Prerequisites

A good knowledge of the basic concepts and techniques in Python. Consider taking our [Introduction to Programming in Python](#) course to come up to speed beforehand.