

Web-scraping for Social Scientists

Wednesday 5 June 2024

Dr Diarmuid McDonnell

There is an unprecedented amount of information on the internet that could usefully be harvested in order to build social science research datasets. This half-day course will showcase suitable techniques for web scraping. The value, logic and process of capturing data stored on websites will be described in detail, and practical examples and exercises will be demonstrated using the Python programming language.

It is most suited to empirical social science researchers but will be of value to researchers from a wide range of disciplines (e.g., digital humanities).

Programme

14:00 – 14:10	Welcome
14:10 – 14:30	Lecture 1 – Fundamentals of web scraping
14:30 – 15:15	Practical 1 – Scraping simple webpages
15:15 – 15:30	Break
15:30 – 16:15	Practical 2 – Scraping complex webpages
16:15 – 16:45	Exercise 1 – Web-scraping challenge
16:45 – 17:00	Conclusion