Ethical and legal considerations of web scraping

Web Scraping to build social research data

Overview

Two key, interrelated questions:

Can you use web scraping in this instance?

Should you use web scraping in this instance?

Legal considerations

There is a lack of certainty around the legal basis of web scraping in the UK.

Web scraping may contravene the Terms of Service/Terms of Use/User Agreement of a website.

Web scraping is a means to an end: gathering useful data.

Ethical considerations

Web-scraping is a small part of a larger project which will almost certainly require ethical approval.

Is it ok to deceive the website by masking your identity? By pretending you are a web browser?

Are publicly available data 'fair game' in every instance?

What is the impact of your web scraping activities on the website itself (and ultimately its users)?

Ethical considerations

Tension in collecting data from the web: often want broad coverage rather than samples, but it is not good practice to collect egregious or indiscriminate amounts of data also.

What do we mean by public data? Hogan (2022): data you do not need to specifically identify or authenticate yourself in order to use.

Therefore the data minimisation principle is a good guide to collecting data from the web (take only what you need e.g., Twitter posts a week either side of a key event)

Guidance

Seek ethical approval from institution for research project more broadly.

Read and respect Terms of Service/Use/Agreement and *robots.txt* file.

Seek written permission from website for web scraping activities.

Clearly identify yourself when requesting web pages (e.g., {'User-Agent': 'Dr Diarmuid McDonnell).

Try not to circumvent technical barriers in place (e.g., by masking/rotating your IP address).

Just because the data are public does not mean they are yours! (Check licenses)

QUESTIONS

Exercise

You are interested in studying the discursive (language and arguments) elements of fundraising campaigns. Your colleague suggested collecting data from JustGiving, a well-known fundraising platform.

https://www.justgiving.com/crowdfunding

1. Select a fundraising topic
2. View a sample of fundraising pages
3. Devise a plan for collecting fundraising campaign details from these pages:
   1. Which details (story, donors, £)?
   2. Are there ethical issues? Can you and/or should you collect data?