The scraping process:

R code: (websitename.R)

Working Directory: (website name folder)

1) Identify a website, save the URL of the homepage

2) Extract all the urls of each plant species “factsheet” or webpage into a list in r

3) Identify which data you would like to extract about the plant, then code R to open each url, download the specific elements you need and put them into a table format. At this stage, the table format is:

|  |  |  |  |
| --- | --- | --- | --- |
| study | species | trait\_name | value |
| (to identify the website) | The species as is written on the website | The trait as a heading written on the website, or summarised by me | The word used by the website to describe the trait |

4) Write the file out as a csv file titled the same as the website name.

Cleaning scraped data:

R code: (AAAv1.R)

Working Directory: (website name folder)

1) Take out synonyms included in the species name. Take out any special characters that did not download properly in the species name.

2) Address each of the unique values in the trait\_name folder and