Web Scraping

Alex Liu

Web Scraping (overview)

- Systematically retrieve data displayed on a website
- Types of websites
  - Basic - html
  - Dynamic - javascript
- Tools
  - "Inspect" the webpage (e.g. [link](#))
  - Create a css selector ([link](#))
  - Get the attribute (e.g. text, link, class name, etc.)

Web Scraping (process)

1. Get the url of website you'd like to scrape
2. Explore the website code
   a. Right-click → "inspect" (Chrome)
   b. Build a "Selector" to collect the data you want
      i. Understand attributes
      ii. Understand chaining selections
3. Use BeautifulSoup to get the data
   a. Create a "Soup" with the url
   b. Get the Object with `soup.select(selector)`
   c. Get the correct Attribute (e.g. the text)

Exercise

- url: https://chess-results.com/tnr669871.aspx?lan=1
- First, practice making the css selector
  - Questions
    - For the table of players, how would you select the 9th row?
    - How would you select the element that contains the title of the tournament?
  - Tools
    - ctrl-f in the html inspector
    - css-selector building (link)
- Second, place the selector into soup.select() to get data and format it