Web Scraping Reference

Requests

Operation	Purpose
`requests.get(url, timeout=10)`	Request a URL with a limit on waiting for connection or read activity
`response.status_code`	Read the HTTP status number
`response.raise_for_status()`	Raise an exception for an HTTP error
`response.text`	Read the decoded body as a string
`response.content`	Read the body as bytes

Beautiful Soup

Operation	Result
`soup.select(selector)`	List of matching tags in document order
`tag.select(selector)`	Matching descendants inside one selected tag
`tag.text`	Text inside a tag, including its nested elements
`tag['href']`	The value of the required `href` attribute
`tag.get('title', 'Untitled')`	The attribute value, or the supplied default when absent
`len(matches)`	Number of tags in a list of matches

JSON

Operation	Purpose
`json.dump(data, file, indent=2)`	Write Python data to a file as readable JSON
`json.load(file)`	Read JSON from a file and reconstruct its data
`json.loads(text)`	Parse a string containing JSON; the final `s` means string