

WEB SCRAPING RESEARCH

Seems as if you can't directly read PDFs and may be some legal rules around doing so, but some libraries do have the option to download PDFs

Some options:

Selenium

- Can't directly load asynchronously, fundamentally synchronous
 - There are some other ways to do indirectly maybe? Conditions that make the program wait, or potentially using some javascript?
- Allows for user interactions
- Can't directly open PDFs, but could download them, allowing us to use an additional libraries to scrape text from them. (PDFMiner? Camelot?)

MechanicalSoup

- Can be loaded asynchronously but not guaranteed it can be done directly, may need additional libraries
- Allows for user interactions
- Similar to Selenium, allows for triggering PDF downloads

Playwright

- Supports asynchronous loading!
- Allows for user interactions
- Can download PDFs as well

Puppeteer (Pyppeteer)

- Also supports asynchronous loading
- Allows for user interactions
- Supports PDF downloading

I would recommend we use Playwright or Pyppeteer because they meet our project requirements better than any of the other libraries (I believe). Both have the code to get them started in our project.

Personally, I would choose Playwright because the documentation is more approachable and seems to be more thorough