These two programs take and parse the drugs.com drug interaction information, creating a searchable format for the drugs.com interactions then searching that format as many times as the user needs.

Setup:
- Python 3.6+
- Internet Connection
- Install pip

Install Libraries from pip, open the Terminal/Command line and input
{ pip install requests beautifulsoup4 }

Ensure you have all three files in the same directory
- drug-interactions.xml
- interactionScraperHTML.py
- InteractionXMLParser.py

How to Use:
The code is currently set to be ran then return the interactions for acebutolol and taurine, these are manually coded into the main() function of InteractionXMLParser. This was done for testing purposes as I knew they were valid inputs on the 1st XML file.
Both programs have commented out code that can allow for command line functionality and inputting drug names, feel free to alter the code or ask me to change it.

Data Structure:
InteractionXMLParser
Variables:
xml_file(string)
- Path to the xml_file
drug1, drug2 (strings)
- The two drugs we are parsing for
self.interactions[] (list)
- All the retrieved interaction URLS from the XML file
results[] (list)
- The matching URL from the XML file

Functions:
__init__(xml_file_path)

- Initializes the xml parsers and calls the parse_xml() function

parse_xml()
- Reads and extracts the URLs from the XML files, stores them to self.interactions

normalize_drug_name(drug_name)
- Makes drug names lowercase and removes spaces

extract_drugs_from_url(url)
- Parses URL for drug names and removes numeric IDs to leave with a string

check_interaction(drug1, drug2)
- Searches for self.interactions for matching drug URLs and returns the matching ones

search_drug_interactions(drug1, drug2)
- Prints the formatted results of the other functions for the drug interactions

InteractionScraperHTML
Variable:
self.url (string)
- URL to scrape
self.html_content (string)
- Raw HTML data
self.soup (BeautifulSoup Obj)
- Parsed HTML Data
self.interaction_data (dictionary)
- All extracted data

Function:
fetch_page() - downloads HTML given to the method
extract_drug_names() - Finds drug names in headers
extract_interaction_severity() - Uses drugs.com 3 danger levels for interaction
extract_interaction_descrition() - Scrapes the paragraph of found by extract_drug_names()
extract_professional_info() - Scrapes info used by medical professionals
extract_references() - Scrapes data references on site
scrape_all() - Runs all commands and saves the data
print_summary() – Prints all the data

save_to_json(outputfile) - Overwrite the output file with the new json file (can be removed)