Week 5 Lab Assignment Goals

- More advanced BeautifulSoup

Step 0 : Ensure that you have the wikipedia and BeautifulSoup modules

- Type 'pip3 install wikipedia'
- For windows, if pip3 isn't working, use 'py -3 -m pip install wikipedia'

Step 1 : Create a GitHub repository

- Go to https://classroom.github.com/assignment-invitations/2241f3247656093c9b725b99f2b6570c
- Accept the assignment invite

Step 2 : Get starter code onto your machine

- Like last week, clone the assignment repository onto your machine
- Open a Terminal window or command prompt and 'cd' to the cloned directory

Step 3 : Retrieving and Parsing Wikipedia Pages - Part 0

- Open part1.py
- Use the wikipedia module to retrieve the HTML of the Harry Potter Wikipedia page (For installing wikipedia module, please refer to Week01 Lab instruction)
    - Hint: look at the page() and html() functions
- Create a BeautifulSoup object using this HTML
    - Hint: refer to last week's lab assignment

Step 4 : Retrieving and Parsing Wikipedia Pages - Part 1

- Write a function print_section_titles() that accepts a Wikipedia page BeautifulSoup object and prints out all the section titles in a numbered list
    - Hint 1: 'Inspect' the Wikipedia page in your browser to see how section titles are created in HTML
    - Hint 2: use find_all() to retrieve tags by CSS class
    - Hint 3: use get_text()
- Call print_section_titles() for the Harry Potter page
- The first few lines of your output should look like this:

```
Section Titles
1 Plot
2 Early years
3 Voldemort returns
4 Supplementary works
5 Harry Potter and the Cursed Child
```

Step 5 : Retrieving and Parsing Wikipedia Pages - Part 2

- Write a function print_references() that accepts a Wikipedia page BeautifulSoup object and prints out all references in a numbered list
- Call print_references() for the Harry Potter page
- The first few lines of your output should look like this:

```
References
1 "Harry Potter and the Cursed Child to be eighth book". BBC News.
```

```
2 Peter Svensson (27 March 2012). "Harry Potter breaks e-book lockdown".
Yahoo. Retrieved 29 July 2013.
3 Allsobrook, Dr. Marian (18 June 2003). "Potter's place in the literary
canon". BBC News. Retrieved 15 October 2007.
```

Step 6 : Make it interactive!

- Write a function interactive_wiki() that
    - Asks the user to enter a search term
    - Prints the list of page titles for that search term and asks the users to choose the desired page
    - Prints the section titles and references for the user's chosen page
- Your output should resemble this:

```
> python part1.py
Enter search term: harry potter
0 Harry Potter
1 Harry Potter (film series)
2 Lego Harry Potter
3 Harry Potter (character)
4 Harry Potter in amusement parks
5 Harry Potter and the Deathly Hallows
6 Harry Potter and the Chamber of Secrets
7 The Wizarding World of Harry Potter
8 Harry Potter and the Cursed Child
9 Harry Potter fandom

Select the page you want: 2

Section Titles
1 Information
2 Harry Potter and the Philosopher's Stone (2001–2002)
3 Harry Potter and the Chamber of Secrets (2002–2003)
...

References
...
6 LEGO Harry Potter Products
7 http://www.brickset.com/browse/themes/?theme=Harry%20Potter
```

Step 7 : Commit code to GitHub

- Commit and push all your code to GitHub