

In this project, you are to create a web crawler class. See Chapter 12 in your book for an example Web Crawler to get you started. I highly encourage you to use anything available to you from Maven repository, it will make this project much easier to implement.

Two notes before we discuss what to do with the class:

1. Make sure that you sleep for at least 0.05 seconds between hitting each link. This is to make sure that you do not ultimately DDOS any site that you wish to crawl. To get your program to sleep, check out this documentation from oracle:  
<https://docs.oracle.com/javase/tutorial/essential/concurrency/sleep.html>
2. [\(Links to an external site.\)](#)
3. For the sake of not traversing advertisement sites, we will be using wikipedia links.

This class needs to do two things:

1. Have a function that traverses 1000 wikipedia links.
2. Have a function that counts words, that is, every time you see a specific word, increment a number associated with that word. Note: Your function should count words only, not HTML elements / attributes.

For example:

```
<p id="myId">
```

This is a paragraph.

```
</p>
```

`<p>` and `</p>` are HTML elements and should not be counted. `id="myId"`, from inside the `<p>` tag, is an HTML attribute and should not be counted. "This is a paragraph." is pure text output to the screen and should be counted as 4 separate words.

I do not necessarily care how you implement this project so long as you have a class that, at the end of your 1000 wikipedia link traversal your class:

- prints out the title of each page you've traversed (there is a `<title>` HTML element that may exist on your site)
- prints out a list of words encountered. Again, html elements / attributes should not be in this list (i.e. no `'<p>'` or `'<div>'`s should be in your list).