**Summary**

This Crawler is able to return the first 5 topics from the Cochrane Library review collection and return a text file that you name with the first 25 articles from the search. I was only able to return all articles from one page because I couldn’t figure out how to navigate to the next page. The url from searching doesn’t change so I assume it’s the javascript doing get requests. I would appreciate feedback on how to orient this part in getting the remaining articles or another way of extracting data from this site. Was there an API avaible for this?

**Design**

So I made 4 Classes. The Article Class was made to model what fields I needed to extract from the website and to write into the text file. The topic class was for the console options when selecting which topic to crawl. The topic builder is my hard coded class that builds topics and fills in their respective URL’s field and name of the topic. The Final class is the Main class in which I did the flow of the application logic.

**Flow**

* So first we initialize a simple scanner for user input.
* Then we initialize a topic builder instance which has a private build method in it’s constructor.
* We create a string builder and create a string to return to the console the options the user can choose from.
* Use the scanner to get user choice.
* Next we check if it’s an integer and if it is return the topic instance associated with the choice
* Build the urlConnection and if success return the html code
* Now we look for the html body tag that contains the articles in
* All the html code of the article body returns in one line
* Create a new article instance
* Find from the line the required fields and populate article instance
* Add the article object to a previously initialized List
* Create a new file writer
* Iterate through article list and write each article object state
* Add new lines and pipes accordingly
* And finally close the File writer and shoot out the file