

Cochrane Library Crawler Documentation

I have provided a briefing about this program below :

The main class of the program is "ApplicationMain.java", this class controls the entire program. And it invoked a method "Collection()" from the "GetData" class.

```
package com.job.vantage;

import org.springframework.boot.autoconfigure.SpringBootApplication;

@SpringBootApplication
public class ApplicationMain {

    public static void main(String[] args) {
        GetData.Collection();
    }

}
```

In the "GetData" class I import two classes and create objects in the "Collection()" method, under this method I import two libraries, such as "BufferedWriter" and "Jsoup".

```
import java.io.BufferedWriter;
import java.io.IOException;
import org.jsoup.Jsoup;
import org.jsoup.nodes.Document;
import org.jsoup.nodes.Element;
import org.jsoup.select.Elements;

public class GetData {
    public static boolean Collection() {
```

The three objects contain all the properties of their class.

One class is "GetElements", it works to get the document page and element list connecting to the base URL using Jsoup.

```
1 public class GetElements {
2
3     public Document GetDocument_page(String BaseUrl) throws IOException {
4         return Jsoup.connect(BaseUrl).get();
5     }
6
7     public Elements get_list_Elements(String BaseUrl, String element) throws IOException {
8         Document document = GetDocument_page(BaseUrl);
9         return document.select(element);
10    }
11
12 }
13 }
```

The second class is "Urls_and_components". I declare some required variables in the class and throw them in the constructor with some static and logical values. I use "ArrayList" and "hashmap" to get specific moth-based data. For data invoked from others, I create some getters.

```
public Urls_and_components(){
    baseUrl = "https://www.cochranelibrary.com/cdsr/reviews/topics";
    topics_list = "li[class=browse-by-list-item]";
    fileName = "cochrane-reviews.txt";
    pagelist = "ul[class=pagination-page-list]";
    url_slicer = "&p_id=";
    reviews_list = "div[class=search-results-item-body]";
    library_url = "http://onlinelibrary.wiley.com";
    authors = "div[class=search-result-authors]";
    date = "div[class=search-result-date]";

    months = Arrays.asList("january", "february", "march", "april", "may", "june", "july", "august", "september", "october", "november", "december");
    map = new HashMap<>();

    for(int i = 0 ;i < months.size(); i++){
        map.put(months.get(i) , ( i+1)+"");
    }
}
```

The other class is "GetFile". By importing "BufferedWriter" in the "GetFile" class I create a txt file for writing the entire data using the program. And export the "GetFile" class by an object in the "GetData" class.

```
public class GetFile {
    private String fileName;

    public GetFile(String fileName) {
        this.fileName = fileName;
    }

    public BufferedWriter create_file() throws FileNotFoundException {
        File fout = new File(fileName);
        FileOutputStream fos = new FileOutputStream(fout);
        return new BufferedWriter(new OutputStreamWriter(fos));
    }
}
```

Finally, I run the program's main class ("ApplicationMain.java") and want to get all data properly. After some time I check the "Cochrane-reviews.txt" file. I got all data that I want.

