

***** WebCrawler.Java *****

```
package cu.cs.cpsc215.project1;
```

```
import java.io.File;
import java.io.IOException;
import java.net.MalformedURLException;
import java.net.URL;
import java.net.URLConnection;
```

```
/**
```

```
 * WebCrawler
```

```
 * Classes Required:
```

```
     WebElement interface
```

```
     WebCrawler class, The Working Class
```

```
     WebPage class, impl WebElement
```

```
     WebFile class, impl WebElement
```

```
     WebImage class, impl WebElement
```

```
     DownloadRepository interface
```

```
     Repository, impl DownloadRepository (Singleton class)
```

```
 * WebCrawler:
```

```
     Main Driver In Program - Take in only 3 arguments from command line
         <URL>, <Depth of crawl>, <Save Directory Loc>
```

```
     Uses of each given Param:
```

```
     <ULR> -> save to string -> Param for, WebPage(StringOfULR)
```

```
     <Depth> -> save to int -> Objective, crawl into pages of set depth
```

```
         Find uses when developing:
```

```
         Max crawl depth first?
```

```
         recursion?
```

```
         Loop with <Depth> as Param?
```

```
     <Dir Loc> -> save to string -> Main Directory for saving
```

```
         Will create a Files and Images folder.
```

```
 *
```

```
 * Personal Notes:
```

```
 *         The save directory needs to already exist or you will get an error.
```

```
 *         This project was especially hard, because of my unfamiliarity with
Java, Jsoup, HTML...
```

```
 *         but I feel like I produced something that actually "fits" the
requirements
```

```
 *         I want to improve this, so future implementations may include my
first version,
```

```
 *         which I attempted to make my own parser.
```

```
 *
```

```
 *         in all, I hope this works...
```

```

*
* @author cfitt
*
*/

public class WebCrawler {

    private static String url;
    private static int depthCrawl;
    private static String savePath;

    public static void main(String[] args) throws IOException{//Check If number
of arguments is correct
        if(args.length != 3) {
            System.err.println("Please Submit args in format: <URL> <int
DepthofCrawl> <Save Directory>");
            return;
        } else {

            try {
                url = args[0];
                URL testUrl = new URL(url);
                URLConnection urlConnection = testUrl.openConnection();
                urlConnection.connect();
            } catch (MalformedURLException e) {
                System.err.println("Not a valid URL");
                return;
            } catch (IOException e) {
                System.err.println("Could not open URL");
                return;
            }
        }
    }

    //-----
    try {
        depthCrawl = Integer.parseInt(args[1]);
    } catch (NumberFormatException e) {
        System.err.println("Argument is not of int type");
        System.err.println("Please Submit args in format: <URL> <int
DepthofCrawl> <Save Directory>");
        return;
    }

    //-----
    savePath = args[2];
    File path = new File(savePath);
    if(!path.exists()) {
        System.err.println("File Path Invalid");
    }
}

```

```

        System.err.println("Please specify of ammend the File Path within
arguments");
        System.err.println("Please Submit args in format: <URL> <int
DepthofCrawl> <Save Directory>");
        return;
    }
//-----
    System.out.println("Attempting to Crawl on " + url);
    WebPage initPage = new WebPage(url);
    initPage.crawl(depthCrawl);
    try {
        DownloadRepository.getInstance().download(savePath);
    } catch (IOException e) {
        System.err.println("");
    }
}
System.out.println("Finished Crawling");
}
}

```