# **ReadMe**

***Software Used:***

* Luna Eclipse IDE
* JDK 8u31
* Crawler from crawler4j open source code.

***Installation:***

* Download the JDK 8 and run the .exe (jdk-8u31-windows-x64.exe) file.
* Eclipse is an Open source IDE to run the Java programs. Luna is a package of eclipse which is compatible with current JDK version. Run the downloaded eclipse file.
* Import the file system of the crawler4j in to eclipse environment. Download all the dependent jars like 'org.apache.http','slf4j logger', 'com.sleepycat.je.' etc. These jars are identified by the import statements in the programs. Configure the build path for these external jars.

***Compilation and execution:***

* To redirect the output to the log file instead of console a .properties file as below is configured and set into the class path as shown;
* Place the below properties file in to a folder. From Eclipse follow the navigation; Run As----- > Run Configurations ----- > Class Path----- >Advanced----- > Add External Folder------ > Select the folder of the properties file.

*Properties File:*

log4j.appender.file=org.apache.log4j.RollingFileAppender

log4j.appender.file.maxFileSize=908KB

log4j.appender.file.maxBackupIndex=10

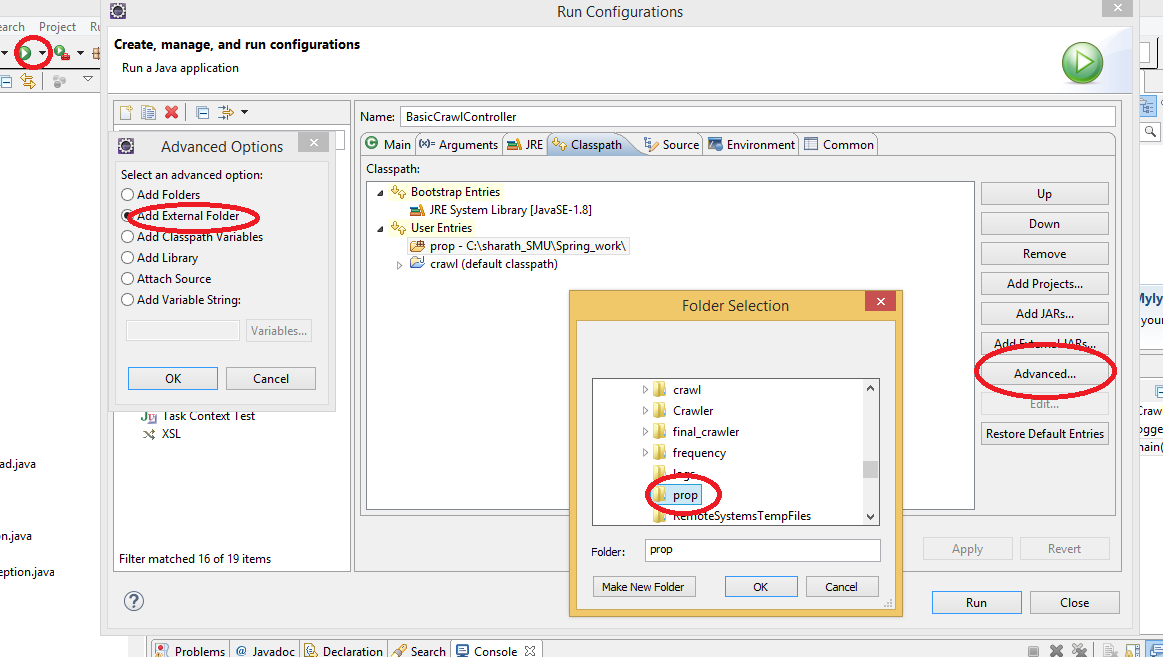
log4j.appender.file.File=C:/sharath\_SMU/Spring\_work/logs/crawl.out

log4j.appender.file.threshold=debug

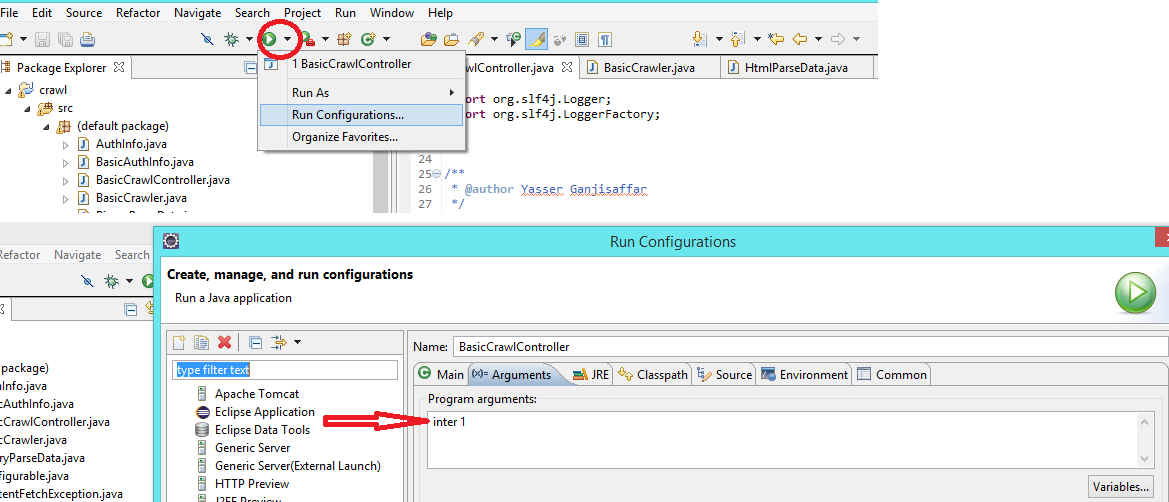
log4j.appender.file.layout=org.apache.log4j.PatternLayout

log4j.appender.file.layout.ConversionPattern=%d{ABSOLUTE} %5p %c{1}:%L - %m%n

log4j.rootLogger=debug,file



* The execution of the crawler starts from the BasicCrawlController.java.The crawler requires two parameters as input which are provided as arguments. The two parameters are 1) Name of the folder for intermediate crawled data, 2) Number of concurrent threads used for crawling. Here, inter is the folder name and only one thread is used to crawl.
* In Eclipse the input parameters are configured for the first time through Run As--->Run Configurations---->Arguments.



* Once the configurations are done. Click RUN to compile and execute the program.