1. **Function Overview**

This is a simple crawler in Java to crawl and download all mails for a particular year and download in corresponding folder.

For example it can download mails from http://mail-archives.apache.org/mod\_mbox/maven-users/ for a particular year say 2015.

This domain and year is configurable.

1. **Basic Specification**
   1. **Usage**

Following are the steps to use the crawler

1. Clone the git repository <https://github.com/akash007ganga/WebCrawler>
2. Build repository using command : mvn clean install package
3. Extract the following zip file

../target/WebCrawler-0.0.1-SNAPSHOT-dist.zip

1. Configure appropriate domain, search year, download location in

../WebCrawler-0.0.1-SNAPSHOT/conf/crawler.properties

1. Run crawler using following command : java -jar WebCrawler-0.0.1-SNAPSHOT.jar
2. output will be deposited in configured download location