There are 10 bash files located at /MetricsExteractor/out/artifacts/MetricsExtractor\_jar. For each project, there are two bash files: one for ARC and one for packages. You basically need to run these scripts to generate the results, which are generated in a folder called “Output”.

The “Output” folder has three subfolders:

-ChangeHistoryData which includes the change history data for files.

-FileData which includes other metrics for files.

-SubsystemData which is the aggregated results for ARC clusters or packages. We are actually using the data in this folder as an input for the R scripts. The first two folders are files metrics.

Source code files are located in /MetricsExteractor/src/gmu/edu/sdalab/. The main method is MetricsExteractor/src/gmu/edu/sdalab/MQandSmellsPrediction/RunningEngine.java

The source code is fairly simple. You should be able to follow the logics from the main method. It basically parses the input files which are passed from the shell scripts and builds the output results that we use as an input for the R scripts.

The bash scripts includes 3 main group of parameters:

-Change history parameters: These are the parameters for connecting to Apache SVN repositories to extract the change history information.

-Understand C++ metrics: The input files that contain metrics that are generated using understand C++ tool e.g. LOC

-Arcade metrics: The remaining metrics that are generated using Arcade e.g. smell metrics.

For example these are the parameters in changesCamelArc.sh:

-"11/22/2011" : This is the release date of this version of the project.

-"/Users/Ehsan/Documents/Data/MSR/Projects/Camel/apache-camel-2.8.3-2011-11-22/workspace/ArcLowLevel.rsf" : This is the cluster information after using ARC. You can find that in the “RSFFiles” folder.

-<http://svn.apache.org/repos/asf/camel/trunk/> : This is the SVN address of repository of the project

-"/org/apache/camel/" : a prefix that I need for generating change history information.

-".\*CAMEL.\*" : Regular expression that I use in the code for calculating the number of bugs in each file (through scanning the commit logs of files and look for the ones that contains this regular expression)

-"/Users/Ehsan/Workspace/UndrestandC++Projects/Camel/apache-camel-2.8.3-src/camel.csv" : This file contains the metrics generated by Understand C++ tool

-"/Users/Ehsan/Workspace/ArcadeProjects/Camel/Results/ResultsLowLevel/smells/Output/apache-camel-2.8.3-src\_arc\_smells.txt" : This file that contains the smell metrics generated by Arcade

-"/Users/Ehsan/Workspace/ArcadeProjects/Camel/Results/ResultsLowLevel/clusters/ClusterMetrics/apache-camel-2.8.3-LowLevel\_arc\_clusters.txt" : This file contains the dependency based metrics that are generated by Arcade

-The next two parameters are the smell metrics and dependency metrics for the next release of the software

-"/Users/Ehsan/Workspace/ArcadeProjects/Camel/Results/ResultsLowLevel/mapping/apache-camel-2.8.3-2011-11-22.csv" – This is the mapping information for a cluster (kth release and k+1th release)

You first need to make sure that the input parameters in the bash files are set correctly. You need to update the address path of some the input files.

I put the input files in these directories:

-“UndrestandC++Projects”: This folder includes the input files for the metrics generated by understand c++ tool

-“ArcadeProjects” : The input files for ARC

-“JavaPackages” : The input files for packages