**FACER-CD: API Usage-based Semantic Clone Detection**

**FACER-CD Replication Instructions**

The following FACER replication setup instructions are for Windows OS.

**Installation Pre-requisites:**

1. Java (<https://www.java.com/en/download/>)
2. xampp (<https://www.apachefriends.org/download.html>)
3. R (<https://cran.r-project.org/bin/windows/base/old/3.5.1/>)
4. Python (https://www.python.org/downloads/)

Configure path variables in your System's Environment Variables for \xampp\mysql\bin and \R-3.5.1\bin and \Java\jdk1.8.0\_73\bin Please Start the Apache and MySQL Services using the XAMPP Control Panel before proceeding.

**Creating FACER-CD database**

1. Navigate to FACER\_Replication\_Pack folder in the command prompt
2. Execute the following command to log in to mysql

<your root>\FACER\_Replication\_Pack>mysql -u root -p

Enter password:

When asked for password, simply press enter key. You will be logged into MariaDB terminal where you will create a database for FACER-CD.

4. Execute the following command to create a mysql database for FACER-CD

MariaDB [(none)]> create database <DB name>;

Query OK, 1 row affected (0.02 sec)

1. Enter \q to exit MariaDB environment

MariaDB [(none)]> \q

Bye

1. Execute the following command to create FACER database schema using provided script

<your root>\FACER\_Replication\_Pack>mysql -u root -p <DB name> < FACER\_schema.sql

Enter password:

**Running FACER-CD Program Analyzer**

To run program analyzer you need to supply three arguments; path to dataset folder(either BigCloneBenchSubsetDataset or AndroidDataset), the name of the database you created and the path to rt.jar inside your Java installation

<your root>\FACER\_Replication\_Pack>java -jar FACER\_CD\_Program\_Analyzer.jar <path to dataset folder> <DB name> "<your drive>:\\Program Files\\Java\\jre1.8.0\_144\\lib\\rt.jar"

As a result two csv files will be output in the ‘output’ folder.

**Running FACER-CD Clustering**

We have provided two R scripts to cluster using simple technique or density based technique. Run the following command if you want to use simple clustering technique:

<your root>\FACER\_Replication\_Pack>Rscript FACER\_CD\_SimpleClustering.R "output"

Instead of “output”, you can enter the complete path to the folder containing the CSV files generated in the previous step

As a result four new csv files will be created in the output folder.

**Getting the CSV files for Python Script**

We need to do a post-processing step to get the correct methodIDs for the clusters. For the last argument, we need to enter the path to either AndroidDataset\methods folder or BigCloneBenchSubsetDataset\java\_test\_subset\_complete\_codes folder

<your root>\FACER\_Replication\_Pack>java -jar FACER\_CD\_GetClusteringCSVFileForPython.jar <DB name> <Path to csv file created by R script including file name> <path to dataset folder containing methods>

After this, we will get a CSV file ClusteringResultsForPythonScript.csv in the output folder which will be used for the next step

**Getting the Performance Metrics and Confusion Matrices**

We need to run the following command to get the performance metrics for FACER-CD against a ground truth csv file. The ground truth files are in the GroundTruth folder for each dataset and you need to specify the file according to the dataset you parsed and obtained the clusters.

<your root>\FACER\_Replication\_Pack>python facer\_output\_evaluation.py <path to ground truth csv including file name> <path to csv file generated in the previous step ClusteringResultsForPythonScript.csv including file name> output/confusionmatrix.jpg