Arcade documentation

1. **JavaParser** is used to analyze Java classes or jar files and generate the dependencies between classes. This is required to run the clustering algorithms. The tool takes four parameters:
   a. Classes directory path or path of the jar file. This is the input to the tool.
   b. Path of the .rsf file name, provide here the path of the Rsf file name to be created. The is one of the outputs of this tool.
   c. Path of feature vector file name, provide here the path of the Rsf file name to be created. The is one of the outputs of this tool.
   d. Package prefix to filter dependencies. This helps to filter out un-needed classes or dependencies.

An example: `java –jar arcade_core_JavaParser.jar hadoop–hdfs.jar /Path/To/RSF/HDFS2.rsf /Path/To/FeatureVector/HDFS2.fv "org.apache.hadoop.hdfs"`

2. **Clusterer** is used to execute three of the clustering algorithms WCA, Limbo and ARC. To execute these algorithms, you need to provide certain parameters. Some of these are mandatory and the other are optional. Some parameters require certain values, and others are flags. Below, I list and explain each parameter:

Mandatory parameters:

- **algo** specifies which algorithm you will run: *WCA*, *ARC*, *LIMBO*
- **language** specifies the programming language: *C or java*
- **deps** specifies the path to the .rsf file that contains the dependencies (generated by the parser).
- **measure** specifies the similarity measure desired. This should be compatible with each algorithm*:* JS, SCM, UEM, UEMNM, IL, ARCIL, ARCUEM, ARCUEMNM, WJS, PKG. WCA is compatible with UEM and UEMNM. Limbo is compatible with IL, ARC is compatible with ARCIL, ARCUEM, ARCUEMNM.
- **projname** *is the name of the subject system.*
- ***projversion*** *is the version of the subject system. You can put here the commit ID.*
- ***projpath*** *is the path to place the clustering output.*

Optional parameters:

- **stop** is a stopping criterion: PRESELECTED, ARCHSIZEFRACTION. The default is preselected.
- **stopthreshold** specifies the number of clusters to stop. The *default is 50.0.*
- **serial** *is Serialization criterion. (ARCHSIZE*, *ARCHSIZEMOD*, *STEPCOUNT*, *ARCHSIZEFRACTION), default archsize*
- **Serialthreshold** *: Parameter value of the serialization criterion. (Double number e.g. 50.0)*

- **packageprefix***: Package prefix to include in the analysis. If C, empty string. Used if analysis not on file level.*
- **artifacts***: Path to directory containing auxiliary artifacts. (Used For ARC)*
- **reassignversion***: Reassign DocTopics (for ARC)*
- **printdots***: Print DOT outputs for each cluster (flag), default false*
- **filelevel***: Dependency type (flag), default true, if false it works on package base.*

Example:

```
java –Xmx14024m –jar arcade_core_clusterer.jar algo=WCA
language=java deps=/Path/To/RSF/HDFS.rsf measure=UEM projname=HDFS
projversion=123456 projpath=/Output/Folder stopthreshold=50
packageprefix="org.apache.hadoop.hdfs"
```

Please note that the -Xmx14024m used to specify the memory size to be consumed by the jar file.

3. **ACDC** is used to execute the ACDC algorithm and it takes two parameters:
   - path to rsf file input (generated by the parser).
   - path to rsf file output.

Example: `java –jar arcade_core–ACDC2.jar /Path/To/RSF/HDFS.rsf /Path/To/Output/HDFS_ACDC.rsf`

4. Pkg is used to cluster classes based on their packages. This can be used as a baseline to compare it with the results of clustering algorithms. It takes the following parameters:
   - ***depspath*** specifies the path to the .rsf file that contains the dependencies (generated by the parser).
   - ***projectpath*** *is the path to place the clustering output.*
   - ***projectname*** *is the name of the subject system.*
   - ***projectversion*** *is the version of the subject system. You can put here the commit ID.*
   - ***language*** specifies the programming language: *C or java*
   - ***filelevel*** is a flag to specify if you work on package level or file level. So, it should be true or false.

   Example:
   ```
   java –jar arcade_core–Pkg.jar depspath=/Path/To/RSF/HDFS.rsf
   projectpath=/Output/Folder projectname=HDFS projectversion=123456
   language=java filelevel=false
   ```

5. A2a and Cvg are used to calculate metrics to evaluate the difference between two architectures. They take two parameters for the paths of the rsf files (output of clustering algorithms) for the two architectures.