Etosha CLI

The command line interface is one of the core components of the project. Dataset related tasks can be done manually or by using a so called shell action via Oozie workflows.

The CLI tool can work in stand alone mode, or in cluster mode. In cluster mode the data is first stored in HDFS for long term persistence and a Fuseki service (managed by Cloudera Manager) is used to expose the metadata and for SPARQL queries.

You can start the CLI with this commend script:

$ ./ecl.sh $ARG1 $ARG2 … $ARGn

The script is located in:

/GITHUB/ETOSHA.WS/etosha/etosha-parent/bin/

Note: On MacOS you will find this issue:

Exception in thread "main" java.io.IOException: Mkdirs failed to create /var/folders/3k/nyrd1dwj4d953l71ys0hnl680000gn/T/hadoop-unjar2822957638110687808/META-INF/license

We have to use shade-plugin for Maven to build an Uber-JAR excluding the content of the folder META-INF/license.

Quickstart Guide

In this section we assume, you have Linux box or VM with Java 8.

1. Clone the repository from Github:

$ git clone <https://github.com/kamir/etosha.git>