

European Summer School for Digital Humanities

Text Mining with Canonical Text Services

Practical Tasks 5

Managing the Canonical Text Miner

First Steps

The CTSTM Admin Tool is available as a deployable .warfile. It can be installed and deinstalled without any impact on the published CTSTM instances. If it breaks for any reason, you can simply delete and redeploy the .war file. After deploying the .war file, the password and user name is both "cts". The CTSTM template may have to be updated after deployment.

Access the pre installed CTSTM Admin Tool using the user name "cts" and the password "cts".

Change the password to something else. Since it is a training workshop, it is advised to choose a simple password. There is no password recovery process. If the password is forgotten, the tool has to be re installed.

Deploy a fresh CTSTM instance with the name "myfirstCTSTM".

Rename it to a name of your choosing.

Deploy a new Canonical Text Miner

Deploy a fresh CTSTM instance with a name of your choosing

Familiarize yourself with the parameters in the tab *Data Import*.

Configure the import process according to given information and publish the pre configured data set.

Change the configuration in such a way that data from the TED Talk Transcripts (CTS Namespace urn:cts:ted) is imported. Limit the data to a maximum of 10 documents in a language of your choosing. Deactivate the calculation of ngrams.

Check whether or not your deployed Canonical Text Miner instance is available online.

Experiment with different configurations.

If you brought your own data

Build a new data set based on your documents in a newly deployed instance of the Canonical Text Miner.