

European Summer School for Digital Humanities

Text Mining with Canonical Text Services

Practical Tasks 0

CTS Requests and Access to the Virtual Machine

CTS URNs

Are these CTS URNs valid? Why?

urn:cts:demo:goethe.faust.de:35.1

Urn:cts:demo:goethe.faust.de:35.1

urn:cts:Demo:goethe.faust.de:35.1

urn:cts:demo:goethe.faust.de:invalid35.1

urn:cts:demo:goethe.faust.de:not-valid35.1

urn:cts:demo:goethe.faust1-2.de:valid35.1

urn:cts:goethe.faust1.de:valid35.1

urn:cts:demo:goethe.faust.de:

Request Information from CTS

Learn how you can request data from the PBC data set that is available at <http://cts.informatik.uni-leipzig.de/pbc/cts>.

The following questions and tasks can be used as a guideline:

Which and how many documents are part of the data set?

How can you find valid static CTS URNs?

How can you request text passages?

Request a text passage using spans of CTS URNs.

Request a text passage using sub passage notation with an increasing index value.

Request a text passage using spans of CTS URNs and sub passage notation.

How are the documents structured?

How can you request a human readable interpretation of a given CTS URN?

More advanced requests

Which function can be used to request the complete text(!) of a document?

The information of which function is not included in GetPassagePlus?
GetCapabilities

How could you rebuild the inner and outer structure hierarchy that is managed by a CTS?

Preparation for the Workshop

Check whether or not the server on your Virtual Machine is running and online.

If your server is not running, connect to the Virtual Machine using SSH and start the server.
The server binaries are located in tomcat/bin in the home directory of user esu.

Create a local document using your preferred office program. During the workshop, the document will be needed again.