

Creation and analysis of text corpora with the TXM platform

Ἀαῖα Ἀεῖαῖα ἰεῖαῖα- Ἰαῖαῖα

07.08.2012 ἰ.

Ἰῖαῖαῖα ἰῖαῖαῖα 07.08.2012 ἰ.

Ḷῖῖῖ ἰ Ḷῖῖῖ Ḷῖῖῖ.

Relevant topics from the preliminary list: 2. Specialized systems for processing full textdatabases; 7. Methods and tools of text processing, 1. XML and TEI technologies

Written by Serge Heiden, Alexei Lavrentiev

2 sessions including each one lecture (2 hours) and one hands-on practice (2 hours). Total duration: 8 hours.

Topics:

Introduction to TXM platform

–methodology
(textometry)

–desktop
vs portal

–basic
functions:

–documentary
lists, KWIC concordances...

–subcorpora
and partitions

–statistical:
specificity, collocates

–exporting
results

Data sources engineering with TXM

–sample

corpora (English and Russian)

–TXM
import modules (TXT, XML)

–raw and
XML text preparation

–proprietary
format conversion

–textual
units engineering (splitting, merging)

–metadata
editing (CSV, XPath)

–tagging
(part-of speech, lemma...)

Advanced TXM

–full-text
search engine CQL patterns

–statistical
analysis: factorial analysis (clustering), classification

–multi-facet
and parallel corpora

Advanced data sources engineering

–XML-TEI,
XML-TXM

–XSLT2

–Oxygen

–Groovy

References:

TXM Project: <http://textometrie.ens-lyon.fr/?lang=en>

TXM platform development site:

<http://sourceforge.net/projects/txm>

TXM demo portail:

<http://txm.risc.cnrs.fr/demo/?locale=en>