

First Steps with ANNIS.local

(Draft version of May 31st, 2007)

0. Overview

This is a quick introduction to ANNIS, which describes basic functionalities of the web application. More detailed descriptions and a tutorial will follow.

1. Browsing and Visualizing Corpus Data

For browsing the data, click 'Data' in the left menu bar.

An overview about the corpora in ANNIS shows up, together with information about the tag sets and the size of the corpora.

For browsing the data in a single corpus, click onto 'Browse' of this corpus.

A list with the documents of this corpus – together with its first words - opens up.

For visualizing a corpus document, click onto its name in the corpus view.

The annotation view for this corpus document opens up.

The annotation view

The annotation view consists of several components, which can be clicked away if wished:

1. **The discourse view** provides a view onto the discourse. Clicking onto a word causes focusing this part of the discourse also in the detailed annotation views.
2. **The detailed annotation views** provide access to the annotations of the corpus. Each view consists of a line with the textual representation of the primary data (e.g. the transcription) and the annotations. Usually, the annotations refer directly to the textual data representation, overspanning the annotated items. There are several modi for visualizing these annotations.

2. Searching the Data

Free query formulation

In the menu bar at the left hand side, there is a free text field for the formulation of queries. Knowledge of the ANNIS Query Language is helpful. A short overview about the query operators can be found in the Menu bar 'Info' with the link 'Quick Reference'.

Query formulation in the corpus query interface

A prototype of a query interface for corpora is available. Choose 'Query' in the Corpus Data overview to query a specific corpus.

3. Data management

We encourage the use of the PAULA-format. But other formats are also supported: MMAX, TIGER-XML, RS3, EXMARALDA.

Corpus data

Data is imported from the source data files, which are stored on the filesystem in the folder of the web application. Relevant folders include:

- annotation directory: annis.local/annodoks
- annotation archive directory: annis.local/annoarchive
- type definition directory: annis.local/annotypen

Importing Corpus data

There are two possibilities for importing corpus data:

1) Using the Import Interface for ZIP archives

Choose 'Data Management' in the Menu bar and click 'Import' for registering new corpus data. Select the format of your data and specify the name of your subcorpus. Then, upload the ZIP-archive with your corpus data.

Your data will be unpacked and stored in the respective folder in the annotation directory ('annodoks').

2) Depositing the data directly on the file system in the annotation directory

You may also copy your (unpacked/unzipped!) data directly to the annotation directory. Choose the respective format folder for your files (paula, rs3, tiger, exmaralda, mmax) and create a folder of the form: ats-YOUR-CORPUS, where YOUR-CORPUS specifies the name of your corpus in ANNIS. Deposit your data into this folder.

Removing Corpus data from ANNIS

Corpus data can be removed from ANNIS as well. Currently (May 2007), only whole subcorpora can be removed:

- Click onto 'Data management' in the left hand menu. Choose 'remove' for the (sub)corpus to be removed.
- On the file system, you find the removed corpus data in the annotation archive directory at "annis_local/annoarchive"

Michael Götze, 31st of May 2007