

User Manual

The basic functions of the database include browsing records with various display options, filtering records by different categories, and performing complex searches within the database content. The database also allows users to download the entire corpus or selected records.

Browsing Records

Upon entering the corpus, all records (i.e., transcript files) stored in the database are displayed in a clear table. For each transcript file, the table shows, in addition to the file name, the level and identifier of the exam, the task number, the source of the preliminary annotation, the annotator's code, and information on whether the transcript for that recording is canonical. The files in the table can be sorted by the values in a selected column. Records can also be filtered based on any substring in the file name by entering this substring in the "Search" text box located to the right above the table. Clicking on a specific file will display that file.

Viewing a File

The database allows users to view the transcripts of individual turns along with annotations and metadata, and to listen to the corresponding audio recordings. The nature of the displayed information varies according to the selected display mode, which can be switched at the bottom of the page below the transcript.

Text View Mode

Text View is the basic display mode that appears upon opening a file. At the top of the screen is a header with the title of the transcript and selected metadata. The transcript itself is displayed at the bottom, divided into turns. Each turn is marked with the speaker's identifier (EXAM_1 for the examiner and CAND_1 for the candidate).

This mode also allows users to view automatic morphological annotation and lemmatization. Hovering the cursor over a specific token will display the corresponding annotation in context. To display a selected attribute for all tokens in the transcript, controls located below the header can be used, which include the following buttons: - PoS: Displays parts of speech. - Tag: Shows morphological tags. - Features: Provides detailed morphological information. - Lemma: Displays base forms of words.

Waveform View Mode

At the top of the screen, there is an extended playback control for the recording, which displays a signal graph (i.e., waveform). Below it, the transcripts of individual turns are displayed. Clicking on a specific turn will play that turn.

Dependencies Mode

This mode displays syntactic annotation. When clicking on a specific turn, an automatically generated dependency tree is displayed, with details available via mouse hover. In the upper right corner of the tree is a \equiv button for additional display options for the tree. It is possible to arrange nodes by word order, display punctuation, or save an image of the tree in SVG format.

Filtering Records by Categories

By clicking on the *Browse* button in the left main menu, users can filter transcripts based on the values of individual categories. For example, it is possible to display only a list of canonical transcripts or transcripts from a specific annotator.

Searching

Searching within the corpus can be done on a page that appears after clicking the *Search* button in the left main menu. This page allows users to enter queries in CQL (Corpus Query Language) format. For example:

```
[upos = "NUM.*"] [lemma = "otázka"]
```

to find forms of the word *otázka* that are preceded by a numeral.

To facilitate searching, the TEITOK interface provides a query builder tool. This tool allows users to easily define simple queries in CQL through a form. Just click the *Query builder* icon, define your query, and then press the *Create query* button, which inserts the query into the CQL text box where it can be further edited if needed.

By default, TEITOK searches the entire corpus, which may contain multiple transcripts for a single recording. If you want to search only in the part of the corpus where each recording has only a single associated transcript, you must restrict the search to so-called canonical transcripts. For example:

```
[lemma = "situace"] :: match.text_canonical = "1"
```

searches for the lemma *situace* only in canonical transcripts.

Downloading

The entire corpus, including recordings and documentation, can be downloaded from the main menu on the left.

A specific transcript can be downloaded in *Text view* mode by clicking the *Download XML* button located at the bottom of the page.