Pepper Grinder User Guide

Version 0.1.8 (2018-05-30 11:07)

Stephan Druskat

Abstract

Pepper Grinder (TraCES Edition) provides a graphical user interface for Pepper, a conversion framework for linguistic data, and allows conversion from the custom TraCES JSON format to the ANNIS format.

Requirements

In order to convert sucessfully between the TraCES JSON format and the ANNIS format, corpora must be split over separate directories. Each directory must contain the corpus' JSON file (`corpusEA.json`) as well as the respective annotation files (`corpusDEA.ann` and `corpusTEA.ann`).

Only one corpus can be converted at any one time with this version of Pepper Grinder.

Installation

Extract the ZIP archive to a destination of your choice.

Usage

Run Pepper Grinder by double-clicking on the executable (`pepper-grinder` or `pepper-grinder.exe`, depending on the operating system you use).

Using the *Browse* button, browse for a folder containing a corpus (cf. Requirements for details). Start the conversion process by clicking on *Run conversion*.

Once the corpus has been converted (may take a few minutes), the folder "output" will contain a sub-folder with the corpus name, which contains a folder with the output format name, which contains a folder with the conversion timestamp, which contains the ANNIS source files, and an archive of these files with the name of the corpus and the ending `.zip`. Use the ZIP file to upload the corpus to ANNIS.

Memory issues

When converting large files, you may run into problems with memory and/or Java garbage collection.

If you encounter out of memory errors, please try adding the following line to `pepper-grinder.ini`, just before the line `-XX:-UseGCOverheadLimit`:

```
-Xmx1024m
```

In this case, the memory available on the heap would be 1GB. You may adjust this to a maximum value near your physically available RAM (e.g., `4096` for 4GB of memory), but you should never set the heap size to a value larger than your RAM as this may cause further errors.

Note that per default, Pepper Grinder is set to ignore limits on garbage collection. The respective setting in `pepper-grinder.ini` is `-XX:-UseGCOverheadLimit`. If you run into trouble with garbage collection, simple delete this line and see if the situation improves.

How to cite Pepper Grinder

If you use data converted with Pepper Grinder in your work, you are required to cite the following software.

- Stephan Druskat. Pepper Grinder (v0.1.8). Zenodo. https://doi.org/10.5281/zenodo.1255895.
- Stephan Druskat. GeTaModules (v0.9.1). Zenodo. https://doi.org/10.5281/zenodo.1255867.

License

Pepper Grinder is licensed under the Apache License, Version 2.0.