Framenet Tools

Release 0.0.1

Aug 27, 2019

Contents:

Provides functionality to find Frame Evoking Elements in raw text and predict their corresponding frames. Furthermore possible spans of roles can be found and assigned. Models can be trained either on the given files or on any annotated file in a supported format (For more information look at the section formats).

Find it on GitHub: framenet tools

CHAPTER 1

Installation

- Clone repository or download files
- Enter the directory
- Run: `pip install -e .`

1.1 Setup

- `framenet_tools download` acquires all required data and extracts it , optionally `--path` can be used to specify a custom path; default is the current directory. NOTE: After extraction the space occupied amounts up to around 9GB!
- `framenet_tools convert` can now be used to generate the CoNLL datasets This function is analogous to pyfn and simply propagates the call.
- `framenet_tools train` trains a new model on the training files and saves it, optionally `--use_eval_files` can be specified to train on the evaluation files as well. NOTE: Training can take a few minutes, depending on the hardware.

For further information run `framenet_tools --help`

1.1.1 Alternative

Alternatively conversion.sh provides a also the ability to convert FN data to CoNLL using pyfn. In this case, manually download and extract the FrameNet dataset and adjust the path inside the script.

CHAPTER 2

Usage

The following functions both require a pretrained model, which can be generated using `framenet_tools train` as explained previously.

- Stages:The System is split into 4 distinct pipeline stages, namely:
  - 1 Frameevoking element identification
  - 2 Frame identification
  - 3 Spanidentification (WIP)
  - 4 Role identification (WIP)

Each stage can individually be trained by calling it e.g. `--frameid`. Also combinations of mutliple stages are possible. This can be done for every option. NOTE: A usage of `evaluate` or `predict` requires a previous training of the same stage level!

- `framenet_tools predict --path [path]` annotates the given raw text file located at `--path` and prints the result. Optionally `--out_path` can be used to write the results directly to a file. Also a prediction can be limited to a certain stage by specifying it (e.g. `--feeid`). NOTE: As the stages build on the previous ones, this option represents a upper bound.
- `framenet_tools evaluate` evaluates the F1-Score of the model on the evaluation files. Here, evaluation can be exclusively limited to a certain stage.

2.1 Logging

Training automatically logs the loss and accuracy of the train- and devset in TensorBoard format.

- `tensorboard --logdir=runs` can beused to run TensorBoard and visualize the data.

2.2 Formats

Currently support formats include:

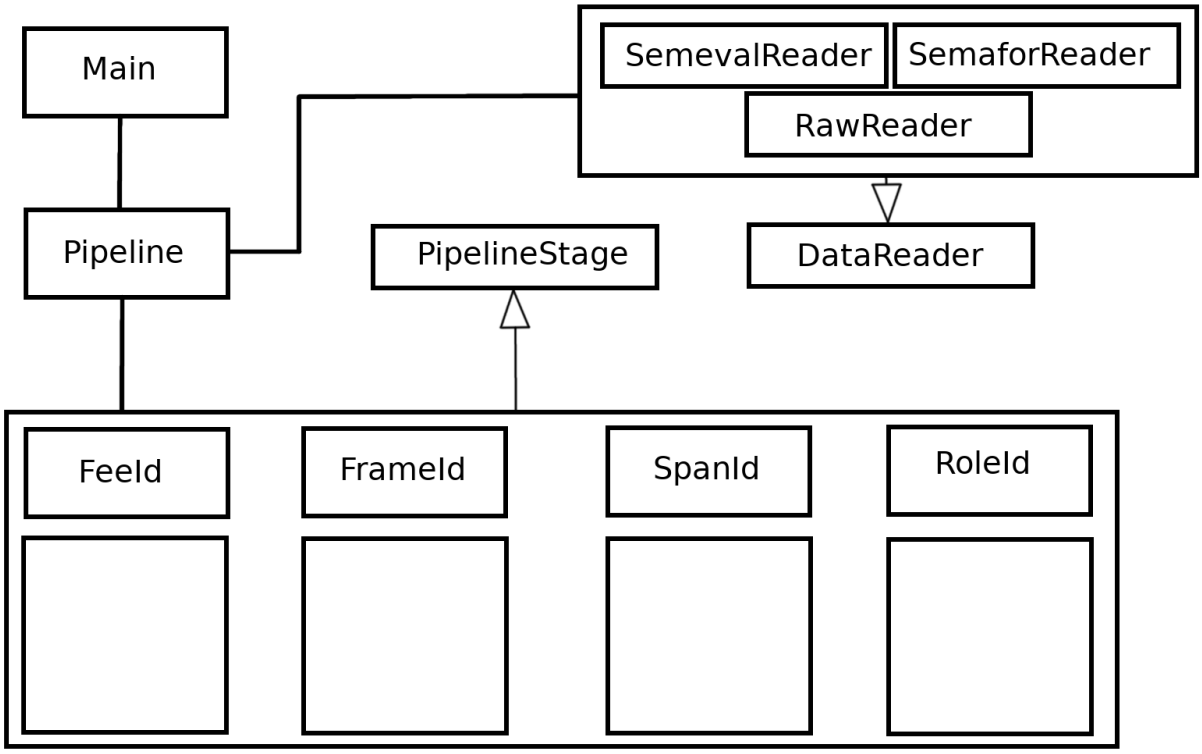
- Raw text
- SEMEVAL XML: the format of the SEMEVAL 2007 shared task 19 on frame semantic structure extraction
- SEMAFOR CoNLL: the format used by the SEMAFOR parser

NOTE: If the format is not supported, pyfn might be providing a conversion.

CHAPTER 3

Documentation

3.1 Architecture



The complete source code is available on GitHub.

3.2 Code Documentation

3.2.1 framenet_tools package

Subpackages

framenet_tools.data_handler package

Submodules

framenet_tools.data_handler.annotation module

framenet_tools.data_handler.frame_embedding_manager module

framenet_tools.data_handler.rawreader module

framenet_tools.data_handler.reader module

framenet_tools.data_handler.semaforreader module

framenet_tools.data_handler.semevalreader module

framenet_tools.data_handler.word_embedding_manager module

Module contents

framenet_tools.fee_identification package

Submodules

framenet_tools.fee_identification.feeidentifier module

Module contents

framenet_tools.frame_identification package

Submodules

framenet_tools.frame_identification.frameidentifier module

framenet_tools.frame_identification.frameidnetwork module

Module contents

framenet_tools.role_identification package

framenet_tools.main module

framenet_tools.pipeline module

framenet_tools.pipelinestage module

Module contents

CHAPTER 4

Indices and tables

- genindex
- modindex
- search