

ConlluEditor: a fully graphical editor for Universal dependencies treebank files

Johannes Heinecke

Orange / TGI / Data & IA / DESKIÑ

Universal Dependencies Workshop 2019



Editing CoNLL-U files

- readability (on one line, sub-trees easy to grasp, all important information presented)
- browsing
- searching (for forms, lemmas, U/XPOS, deprels, sequence of all of these)
- modifying any CoNLL-U column in graphic mode (including enhanced relations)
- trees, not flat graphs
- abstraction from underlying format
- git support
- svg/ \LaTeX export

Editing

- full graphical editor for basic and enhanced dependency relations
- word edit (form, lemma, UPOS, XPOS, features, misc-column)
- autocompletion (UPOS, XPOS, deprels, using user provided lists of valid labels)
- editing multitoken words ([1-2] ...) and empty nodes ([5.1] ...)
- comment editing
- support for right-to-left scripts like Arabic or Hebrew
- split and join words (to correct bad tokenization), split and join sentences (to modify sentence segmentation)
- undo/redo
- git support (add/commit every n edits, default: 1)
- export of dependency graphs as .svg-image, X_{LaTeX} -code (for the tikz-dependency package or the `deptree.sty` X_{LaTeX} style), sd-parse, CoNLL-U

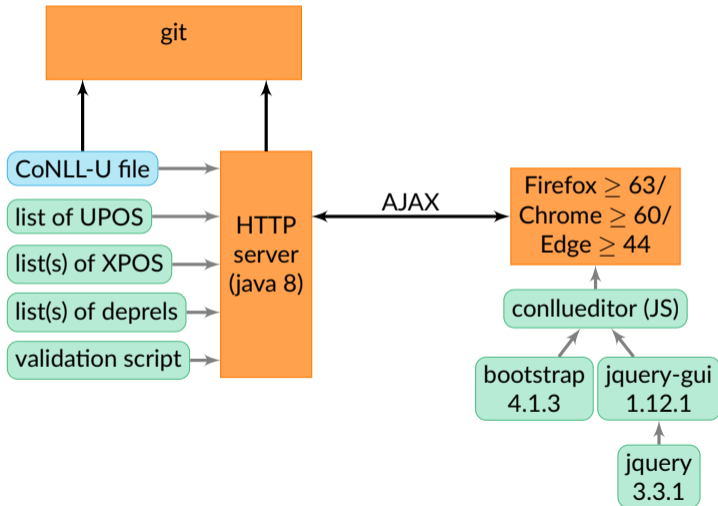
Searching

- regex search functions (including sequences of tokens, sub-graphs and comments)

Validation

- validation: indicates undefined UPOS, XPOS, dependency relations (based on lists given to the server)
- prohibition of invalid (cyclic) trees
- normalisation of token ids (first column, from 1 to n , taking into account multitoken words, empty words and heads)
- validation with external script (such like UD's `validate.py`) on the current sentence
- limited multi-user support: as long as two users do not edit the same sentence

Architecture



Let's see the real thing

- configure validator (in `valid.conf`):

```
$HOME/UD/tools/validate.py --lang cy --max-err 0 --level 5 {FILE}
```

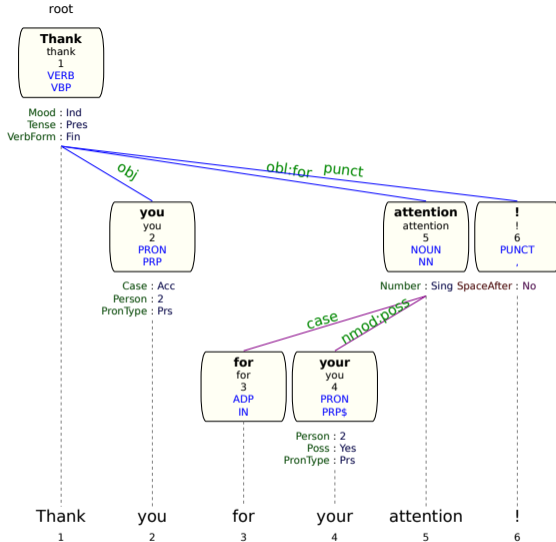
- start server

```
$CE/bin/conlluedit.sh -r \  
--deprels deprel.cy,deprel.ud \  
--UPOS cpos.ud \  
--XPOS xpos.cy \  
--validator valid.conf \  
demo.conllu 5555
```

- open navigator

- `http://localhost:5555`
- `demo`

The screenshot shows the ConlluEditor web interface in a browser window. The address bar shows `localhost:5555`. The page title is `Conllu-U Editor`. The main content area displays a dependency parse tree for the sentence: "Ieri notte hanno dormito all'aperto, senza acqua e senza cibo." The tree is rooted at "ROOT" and branches into "advmod" (Ieri), "advmod" (notte), "advmod" (hanno), "advmod" (dormito), "advmod" (all'aperto), "advmod" (senza), "advmod" (acqua), "advmod" (e), "advmod" (senza), and "advmod" (cibo). Each node contains the word, its index, and its part-of-speech tag. The interface also includes a search bar, navigation buttons, and a text input field at the bottom.



<https://github.com/Orange-OpenSource/conlleditor>
johannes.heinecke@orange.com