UNIVERSITY OF
COPENHAGEN

Text Tonsorium

Demo and Discussion of Existing and New Features

Bart Jongejan (bartj@hum.ku.dk), University of Copenhagen, Denmark

The Text Tonsorium (TT)

TT is an open source web application for Natural Language Processing (NLP) that is publicly available at **https://clarin.dk/clarindk/toolstexton.jsp**, and that can be even installed on your laptop.

The TT is to NLP as a salon de beauté (=tonsorium) is to installing artificial hair integrations:
- Users provide input and state their wishes, but do not need to know how to generate the desired result.
- The TT automatically assembles NLP tools into a viable workflow design and then enacts it with the user's input.
- The absence of human workflow architects makes the TT flexible and reliable.

Front page

https://clarin.dk/clarindk/tools-texton.jsp

Improvements

Explanations:
Left: A goal that cannot be fulfilled.
Below: An explanation why the goal cannot be fulfilled.

Your goal cannot be fulfilled with the currently integrated tools. Try to make your goal less specific.
(There is no tool chain leading from feature "Format" = "TEIP5" to "Format" = "RTF".)

In this stall

I demonstrate the latest features, which include:
✂ the support for seven recently added languages: Albanian, Asturian, Catalan, Faroese, Galician, Irish Gaelic, Manx, Scottish Gaelic, Welsh,
✂ the improved user interface: production of helpful tips for users when they specify their goal too narrowly,
✂ the improved user experience: reduction of the time the user has to wait between user interactions.

The purpose of the demo is to gather **feedback** on the current features and to discuss possible future features.

A **tonsorium** in Brooklyn, NY
© Bart Jongejan

Even though Text Tonsorium's user interface has been improved in the past year, response from users with different technical background is still desirable in order to further improve the system.
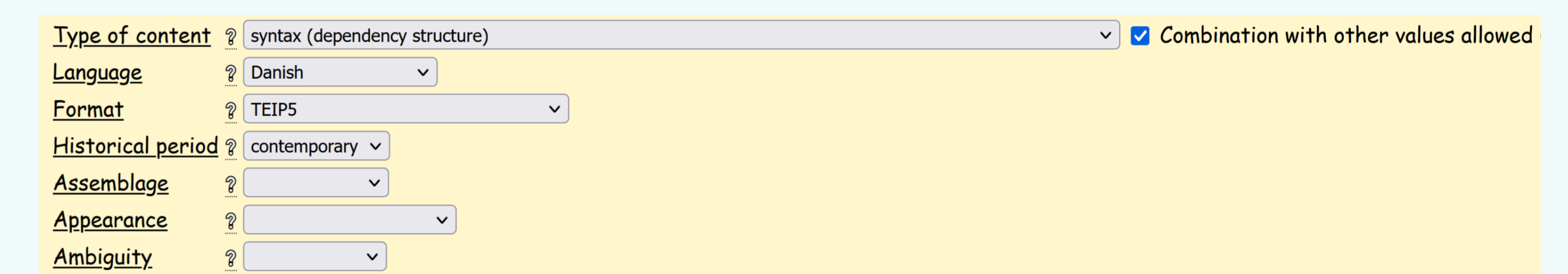
Parallel processing

You can upload hundreds of inputs to the Text Tonsorium at the same time, and they will be processed in parallel, CPUkernels allowing.

Toolbox

The Text Tonsorium has NLP and other tools for many types of content, file formats and languages.

Improvements
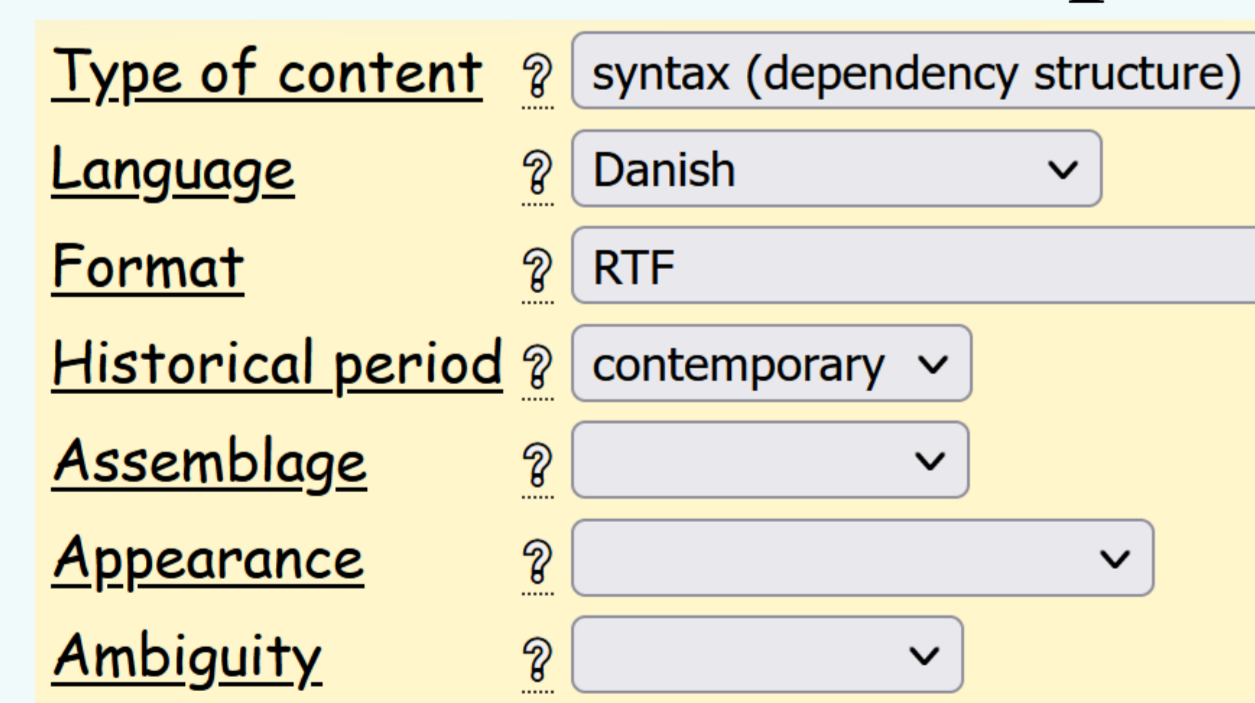
Faster and more user-friendly
Above: Users only partially specify their goal.
Below: The next step offers viable refinement options.
Advantages: easier for users and quicker rendering

Links