

The Mill – Interactive FIN-CLARIN Web Services

Mietta Lennes
FIN-CLARIN, University of Helsinki

Researchers in Digital Humanities need to manage, process and analyze increasing quantities of material that may contain text and/or speech in different languages. Efficient tools may already exist for various tasks, but in order to apply the tools, the user often needs to be familiar with programming and/or with command-line interfaces. Moreover, the output generated by one tool may need to be converted into suitable input for the next one, and it is not easy to keep track of all the processing steps.

The Language Bank of Finland (Kielipankki) is developing an online platform called Mylly, “the Mill”, for processing speech and language data in a graphical user interface that integrates different tools into a single workflow. Mylly provides tools and computational resources for processing material and for inspecting the results. Users can upload data, select menu commands to be applied to their material, and download intermediate results in convenient formats at any stage of the tool chain. It is also possible to save workflows for future use.

The Mylly service runs on CSC’s Taito cluster and is an instance of the open-source Chipster platform. Mylly is available for all academic users that are registered with a CSC user account.

Currently, the tools that are plugged into Mylly include a parser, morphological analyzers, generic finite-state technology, some R functions, and an automatic speech recognizer. In Mylly, it is also possible to perform searches directly from Kielipankki corpora over Korp API and to process the results further. Researchers are welcome to suggest other tools they may need for data processing, analysis and visualization. Technically, any tool that can be run on CSC’s Taito server may be suitable.

The Language Bank of Finland is a collection of corpora, tools and other services maintained by FIN-CLARIN, a consortium of Finnish universities and research organizations coordinated by the University of Helsinki. The technological infrastructure for the Language Bank of Finland is provided by CSC – IT Center for Science.

The Language Bank of Finland: <https://www.kielipankki.fi>

How to use Mylly – The Mill: <https://www.kielipankki.fi/support/mylly> (in English)
Myllyn käyttöohjeet: <https://www.kielipankki.fi/tuki/mylly> (suomeksi)