

## Faceted Search for Digital Humanities Tools in CLARIN

Jan Odijk (Utrecht University, Utrecht) and Daan Broeder (Meertens Institute, Amsterdam)

Type: poster and demo

We will present a system of faceted search for Digital Humanities software created in the context of CLARIN in the Netherlands and CLARIAH WP3. It is based on an earlier version of faceted search that we created and that can be viewed at the CLARIN-NL Portal:

<http://portal.clarin.nl/clarin-resource-list-fs> . In the new system, which is still under development but for which we aim to have a first version in March 2018, the information on tools in the CLARIN-NL portal has been transposed into CLARIN-compatible CMDI records (Broeder et al. 2012), in accordance with a profile specifically developed for this purpose (clarin.eu:cr1:p\_1342181139640), an extension of a profile developed earlier (Westerhout & Odijk 2013). In addition, new metadata records have been added (currently covering 73 pieces of software). We defined a new list of query facets, the most important ones of which are *Research Phase* (to which phase(s) in a research project is this tool relevant), *tool task* (what does the tool do?), *research domain* (for which research domain(s) is the tool relevant). Given CLARIN's focus on language and linguistics there is a specific facet for *linguistic subjects*. Other prominent facets include *input language* (what languages does the tool apply to?) and *application type* (web or desktop, application or service, etc.). There is also a facet for *national infrastructure project* so that one can easily see (in the Netherlands context) which tools were delivered by the CLARIN-NL project, which ones by CLARIAH, and which ones by other projects. The faceted search system is dynamically generated on the basis of the CMDI metadata records: changes in or additions to these metadata records will become visible in the faceted search interface by regular harvesting of the facets from the CMDI records.

The metadata contain links to the applications (most are web applications), or to locations where the software can be obtained and these are visible in the faceted search, so that a researcher can apply the application immediately once (s)he has found its description. The faceted search is accompanied by a text that explains which phases in a research project we distinguish and that contains a link for each research phase to a query for the relevant tools. In this way we aim to guide the researcher to the right selection of tools for his/her research project.

It is a further aim to extend this faceted search for tools beyond the Netherlands to CLARIN as a whole, and to make it part of the CLARIN *Virtual Language Observatory*. To that end we have experimented with (semi-automatically) enriching other CMDI metadata descriptions for software, in particular the *WebLichtWebserviceProfile* (clarin.eu:cr1:p\_1320657629644, which has 287 instances), and we are in close contact with the VLO developers. Possibly some initial results of this cooperation can be shown as well.

## References

Broeder, D., Windhouwer, M., Van Uytvanck, D., Goosen, T., & Trippel, T. (2012). CMDI: a Component Metadata Infrastructure. Proceedings of LREC 2012: 8th International Conference on Language Resources and Evaluation. Istanbul, Turkey.

Westerhout, E. & J. Odijk (2013), 'Metadata for tools: creating a CMDI profile for tools', presentation held at CLIN 2013, Enschede, Jan 18, 2013. [[pdf](#)]