

Processing, Analysing and Visualising Language Data using solutions prepared in CLARIN-PL LTC

Tomasz Walkowiak, Marcin Pol, Maciej Piasecki

CLARIN-PL

Wroclaw University of Science and Technology, Poland
{tomasz.walkowiak, marcin.pol, maciej.piasecki}@pwr.edu.pl

Demo/poster abstract

Language Technology Centre (LTC) uses technology infrastructure developed by CLARIN project which is a complex computer system that enables combining language tools with language resources into processing workflows. The processing tools are developed as microservices connected to each other using solutions like RabbitMQ broker¹ and Citrix Server Virtualization. Researchers use complete workflow submitting data using modified version of DSpace² or NextCloud³ repository system. CLARIN-PL LTC focuses on research tools based on LT.

The core functionality of the LTC is that the researchers have the possibility to create and develop NLP projects. All tools and applications are inter-connected to give users ability to set up a project for a particular research task or a specific experiment. To use resources and tools of LTC users have to set up a free account in our D-Space

repository or login via federation identity using shibboleth – a system of distributed authorisation. Once logged, users can upload their files to one of the storage systems: NextCloud or DSpace. After that the data can be processed by NLP services and pushed to different CLARIN-PL applications. They have been primarily developed to support researchers and students from H&SSs, but also can be found in other areas of science and technology.

The achieved high degree of flexibility means that it will be possible to easily customize projects to fit a variety of needs, from small ones to large research collaborations. Researchers can process large corpora of text and easily publish or share results. So far, CLARIN-PL is focused mainly on Polish, but most of the applications are able to work with English and German. We continue expanding our tools to support multi-linguality to much broader extent.

¹ <https://www.rabbitmq.com>

² <http://www.dspace.org/>

³ <https://nextcloud.com/>