Natural Language Processing for Historical Documents

9-11 September 2019 | Berlin

CONTRIBUTION TO CLARIN GOALS

The workshop is part of CLARIN's mission to provide and support NLP for research in the humanities and social sciences. Implementing and improving tagging and lemmatization for historical documents is key to improving access to text collections, and as a first step towards distributional semantics and 'big data' approaches, and towards enabling new types of research.

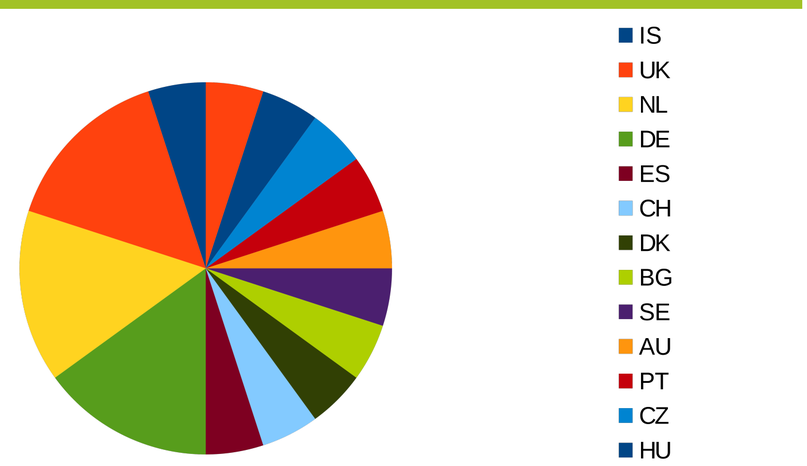
INVITED TALKS

Worksflows in Nederlab
Hennie Brugman, KNAW, Netherlands

DTA::CAB ~ Normalization in the *Deutsches Textarchiv*
Bryan Jurish, BBAW, Germany

Natural Language Processing for Historical Documents
Martin Wynne, Bodleian Libraries, University of Oxford, UK

PARTICIPANTS BY COUNTRY



RESULTS

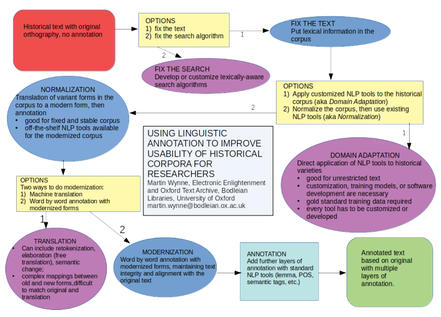
The main goal of the workshop was produce a **guide to software applications** for processing historical language varieties, a document which will help users to find, understand, choose and deploy natural language processing software applications for the annotation and analysis of texts in historical language varieties.

The workshop also generated a set of **recommendations** for candidate software applications for **integration** into the CLARIN infrastructure, via the CLARIN Language Resources Switchboard, and via the various web service orchestration platforms.

An outline plan for a **user involvement workshop** was formulated, focussing on helping researchers who are manually annotating data to create complete hand-crafted datasets which can serve as gold standard data for training and/or evaluation purposes. A proposal will be developed for a 'hackathon' or 'data carpentry' event on this topic in 2020.

FOLLOW UP

For more information about the workshop visit CLARIN Events page:
https://www.clarin.eu/event/2019/clarin-workshop-nlp-tools-historical-documents
https://sprache.hypotheses.org/1790



Martin Wynne
martin.wynne@bodleian.ox.ac.uk

Bryan Jurish
jurish@bbaw.de