

PapyroLogos: Large data and little resources
by Holger Essler, Rodney Ast

Greek papyrology has a relatively large, uniform central database, containing 32440 documentary and 1765 literary papyri with digital images and line-by-line transcriptions. This seems to provide excellent conditions for training AI for HTR. The greatest challenge, however, lies in the fragmentary state of preservation of the originals, which makes layout analysis and line recognition difficult because the regular arrangement is often disturbed by holes and other types of damage. The same also applies to the writing, which is often incomplete due to abrasion, fading etc. In addition, the material varies greatly in content, layout and palaeography and many textual witnesses are short. This presentation describes the various approaches and steps we have taken so far using the eScriptorium platform to create a usable HTR model, the difficulties we faced working with very little financial support, and our plans for the immediate future.