

Some book pages for experiments in OCR postcorrection

OCR workshop at LMU, 14./15. September 2015

2015-09-10



<http://creativecommons.org/licenses/by-nc-sa/4.0/>

Books and sources

- for further experimentation, some [book excerpts](#) (ca. 100 pages each) can be found in the data directory of this workshop
 - preprocessed page images
 - ground truth (incomplete for Hobbes and Zonaras)
 - OCR output from ABBYY and Tesseract
- sources for the scans of complete volumes are given below
- many thanks to Kay Würzner (Grenzboten), Federico Boschetti (Zonaras) and Haide Friedrich-Salgado (Hobbes) for providing us with ground truth
- Goethe: [Wahlverwandtschaften](#), vol. 1, 1809
 - [Text](#) in Deutsches Textarchiv
- [Die Grenzboten](#), 1841
- Hobbes: [Leviathan](#), Latin edition, 1668
- Zonaras: [Epitome historiarum](#), 1870