
Transkribus Best-Practices – Using Handwritten Text Recognition (HTR) for Specific Purposes

Tobias Hodel*^{1,2}

¹University of Zurich – Switzerland

²State Archives of Zurich – Switzerland

Abstract

Text is at the heart and center of research in the humanities. In order to use "text" as "data" we not only need digital facsimiles but properly recognized text. Thanks to the use of machine learning, text recognition has recently leaped not only in the recognition of early prints (16th to 18th century) but also in the recognition of handwriting. Within project READ (finished in June 2019) the platform Transkribus has been developed, allowing for scholars to produce their own recognition models and correct as well as annotate derived text.

Transkribus not only provides high quality recognition but also capable search tools. Depending on the needs of researchers and students, Transkribus facilitates access to texts from different times and locations.

Bearing from examples from the State Archives of Zurich as well as from a variety of other projects, the talk will showcase how different use cases benefited from using Transkribus not only for the recognition of texts, but also from semantic annotations of visual features as well as different search engines (such as fulltext databases or search algorithms comparing variants of recognized text).

The paper will thus be instructive for scholars to decide how to approach Transkribus as a virtual research environment for their goals, depending on the planned outcomes.

*Speaker