

Annotating Medieval Manuscript Layout

Hannah Busch

Trier Center for Digital Humanities
buschh@uni-trier.de

Jochen Graf

University of Cologne
jochen.graf@uni-koeln.de

Philipp Vanscheidt

Darmstadt University of Technology
vanscheidt@linglit.tu-darmstadt.de

In addition to traditional human research with digitized historical sources, computer-assisted possibilities of research evolve: while digitized medieval manuscripts can be handled with programs and tools for supporting critical editing, it is also feasible to analyze the images themselves with trained algorithms and establish classifications for detecting codicological information on larger corpora and to evaluate them statistically. New and in some cases more precise information can be added to the catalogue data. Especially concerning measurements algorithmic results might be more accurate and reproducible. The quantitative evaluation can help to find hidden relations between layout features, genre and historical developments in a corpus or between different corpora. Both ways are used in the project “eCodicology” (www.ecodicology.org).

Using digitized images and catalogue information of the “*Virtuelles Skriptorium St. Matthias*” (stmatthias.uni-trier.de), 470 medieval books from many different areas of human knowledge and culture like theology, philosophy, or literature, and from the 8th to the 18th century are used for analyzing the practice and the history of this library of a Benedictine Abbey near the city of Trier. “eCodicology” searches for patterns in the development of the medieval book, different genres and historical ruptures insofar as they became manifest in layout features and relations between parts of the page like text and image areas or margins.

On the one hand, in cooperation with the virtual research environment “TextGrid” (www.textgrid.de) the opportunity is given to use the digital facsimiles of 470 codices in critical editions. On the other hand, by using the Software Workflow for the Automatic Tagging of (Medieval Manuscript) Images (SWATI) mise-en-page elements can be discovered automatically. These two ways – the human researcher’s and the computer’s way – cannot be strictly separated. The results can be ambiguous from a codicological point of view. In cases of unstable features like the position of a rubric on the page or in complex cases like multiple glossed texts, a human annotation can be useful. The software “SemToNotes” allows to select, delete and edit automatically recognized areas of the digital image and to annotate them by using some kind of codicological terminology. Thus, it becomes possible to adapt

automatic results to research questions and use these results for the reconstruction of the complex relations between visual phenomena and semantic relevance in the history of the medieval book.