

Inventorying, Transcribing, Collating, and Printing: Basic Components of a Virtual Platform for Scholarly Editing

Stefan Büdenbender
Trier University
bued2101@uni-trier.de

Frank Queens
Trier University
queens@uni-trier.de

Martin Sievers
Trier University
sievers@uni-trier.de

Joshgun Sirajzade
Trier University
sirajzad@uni-trier.de

The Trier Center for Digital Humanities is currently involved in the creation of a number of digital editions, be it through preparing and providing data or through the developing of software solutions.

Combining modules from earlier projects with two recent tools developed for the Arthur Schnitzler Edition [Arthur Schnitzler: Digitale historisch-kritische Edition (Werke 1905–1931)], we can now cover the basic functionalities of a virtual platform for scholarly editing. From the first assessment and inventory of textual witnesses, through their transcription, down to the comparison of the resulting texts and their printing, all steps are carried out in a tool chain built upon the technical infrastructure of the database environment FuD.

Technical Infrastructure

The research network and database system (Forschungsnetzwerk und Datenbanksystem) FuD forms the technical backbone of the platform, offering a set of features for decentralized collaborative work. On the one hand it allows the inventory (metadata capture, grouping) and commentary (creation of indexes) of the material, on the other hand it provides a database environment which manages all created content and thus represents the intersection between the individual modules. Although these have XML interfaces and can be run independently, the access to a central database facilitates collaboration and has advantages when dealing with concurrent hierarchies and structures transcending document borders.

Transcription of textual witnesses: Transcribo

The transcriptions are established in Transcribo, a graphical editor developed for the Schnitzler Edition. Thus, it is able to communicate with FuD, to handle large image files without delay and to support an extensive set of specific annotations. The user interface sets the digital facsimile (generally the scanned physical witness) at the centre. This appears in double form, always providing the original for close examination while all processing steps take place on a slightly attenuated duplicate. This arrangement accommodates the use of multiple monitors and above all saves time-consuming jumping back and forth between image and editor window. Thus, field markings in rectangular or polygonal shape can be set topographically exactly around the graphic unit to be transcribed, and the transcribed text can then be entered directly. For the processing of typescripts an additional OCR with options for image enhancement (such as contrast and colour adjustment) is integrated, providing a raw transcription as a basis for further editing.

More fundamentally, each transcribed element can be provided with a comment and each relevant philological or genetic phenomenon can be annotated in a uniform way. Here, context menus are used, offering a set of options which can be adapted to a project's specific requirements.

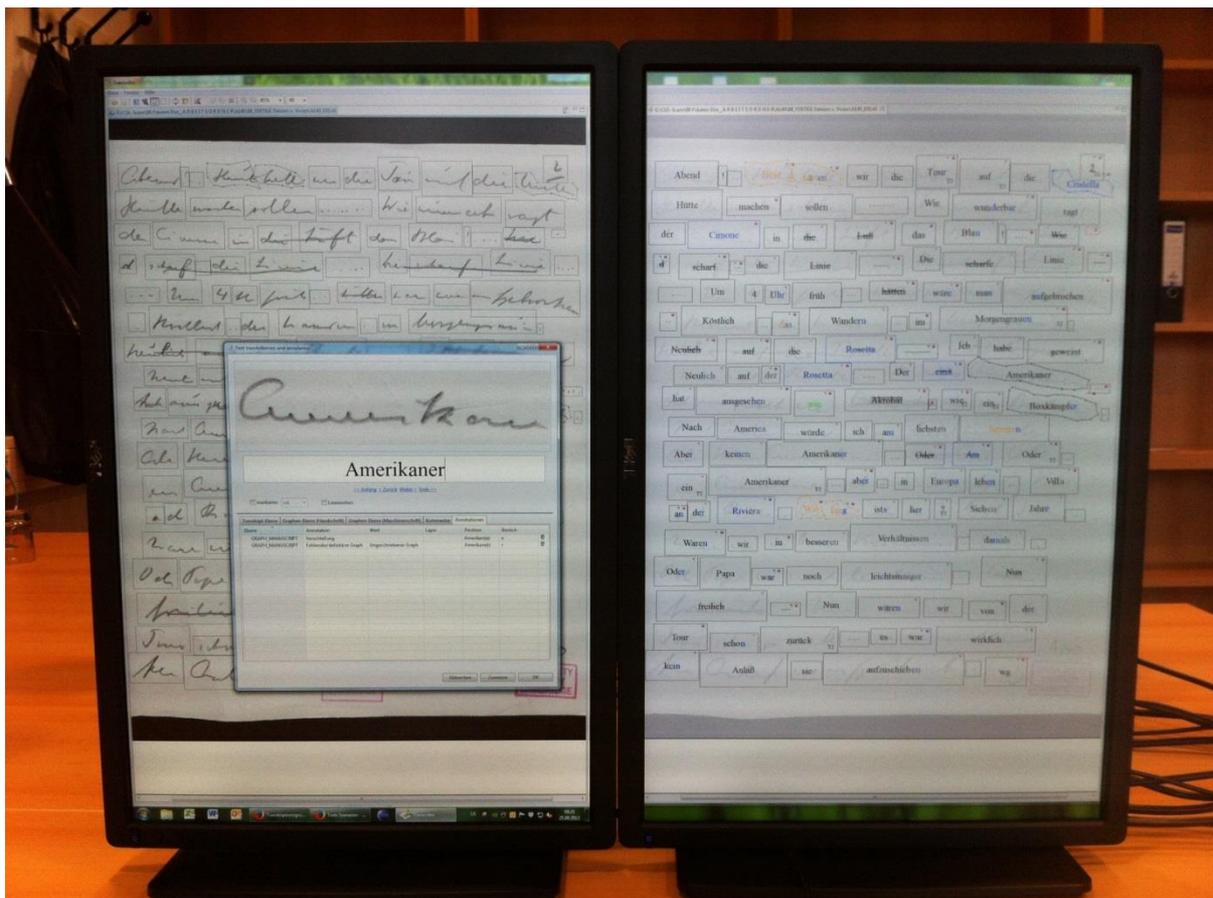


Figure 1: Transcribo

Collation

Additionally, we are developing a graphic environment for textual comparison, as existing solutions had proven to be inadequate for a number of reasons. Firstly, the variation between the textual witnesses can at times be quite significant, with the challenge being to spot the rare similarities rather than recording differences. Secondly, it is often necessary to compare a large number of witnesses simultaneously. Finally, we want to visualise the results in different contexts, with varying granularity. The comparison process is therefore divided into two phases: at first, larger linguistic or structural units (depending on the text: sentences, paragraphs, speeches, scenes...) are matched, regardless of their position in the overall text. This is then checked by a philologist, corrected if necessary, and then handed on for detailed comparison.

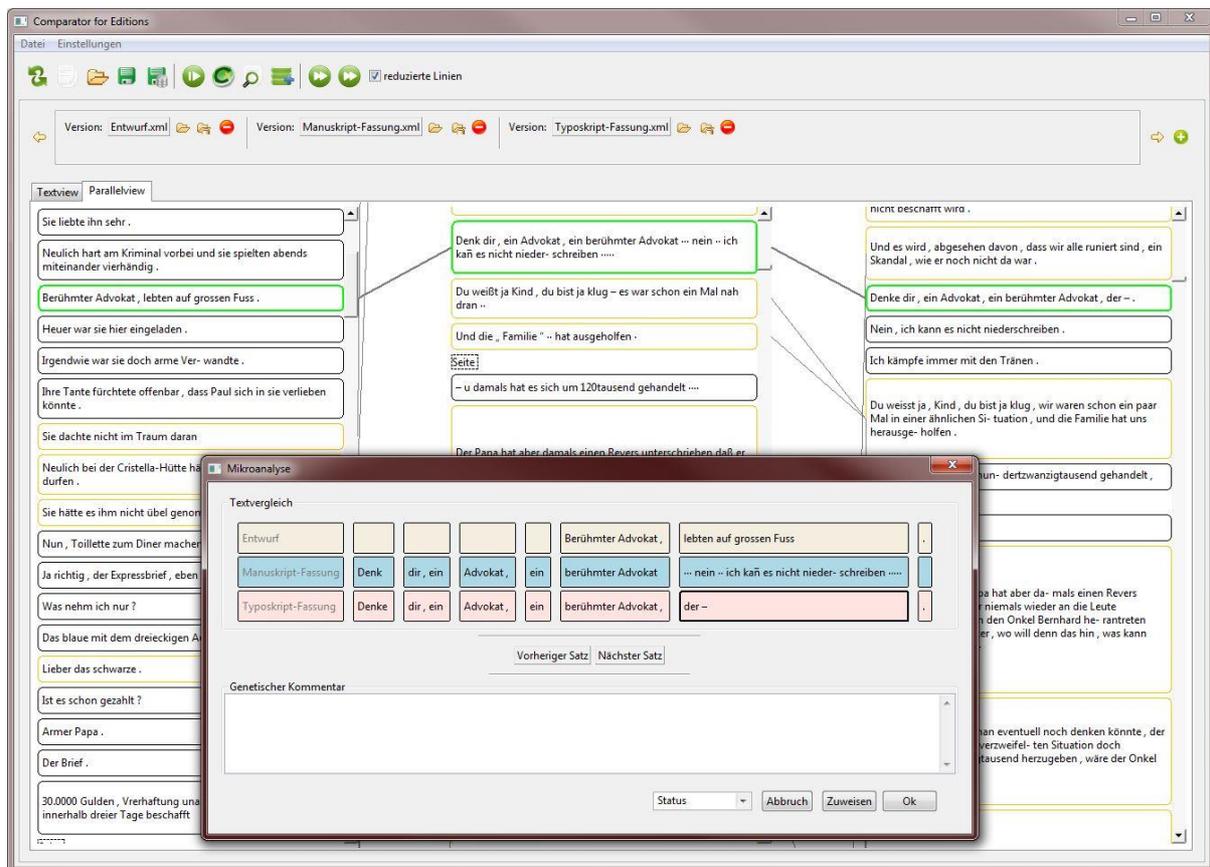


Figure 2: Collation tool

Outlook: export data in high-quality for printing

The platform is to be completed by an interface to the open source typesetting system TeX. Using the programming language Tcl, data stored in FuD is converted into a LaTeX file. In conjunction with an appropriate template (LaTeX class file), a high-quality PDF file can thus be generated on the fly. Critical apparatuses can be added as well as bibliographies or different indexes. For single projects this has already been done. For the next major release of FuD, a general PDF output will be available for all projects.