## JOHN — Interactive

## <3 figures>

## Interactive Visual Analysis of German Poetics

John, M., Koch, S., Heimerl, F., Müller, A., Ertl, T. and Kuhn, J.

Information visualization plays an important role in analyzing and understanding large amounts of quantitative data, a task getting evermore important in our age of big data. The increasing volumes of digital data sources being created and the growing use of quantitative methods for their analysis in the digital humanities additionally motivates the application of information visualization approaches in this particular discipline. It is no surprise that visualization techniques, including network visualization, trees, and all kinds of diagrams and charts, rise in popularity in the digital humanities as an elegant means to convey knowledge and insights buried deep inside large heaps of quantitative data.

Information visualization does not have to stop at the stage of creating static images, which it is still largely used for. Adding interaction can be a game changer in many respects. Visual representations are often not authoritative, since they were created based on quantitative data, which in turn may have been derived through automatic processing steps. Such automatic processing can be a source of uncertainties and errors that are then reflected in visual representation. Even with an entirely noiseless data source, such as manually prepared data, inadequate mappings to visual representations may lead to misinterpretations. For those two reasons, uncritical interpretation of such visual representations is likely to cause problems.

Interaction can be used to let users navigate to the data sources of visualizations and make the methods abstracting this data more transparent for users to understand and control their accuracy and validity. For text documents as the data source, interaction helps to bridge the gap between distant and close reading, to use Moretti’s terminology (2005).

Thinking further along these lines, the next logical step would be the introduction of facilities for updating, steering, and improving the analytic processes through interactive feedback on the visual representation. With the mentioned interaction methods achieving a new level of control, the development of a visual analytics approach offering such possibilities comes into reach.

In the following we illustrate our effort towards this goal for the analysis of text documents with approaches developed in the context of the project ePoetics—Corpus Analysis and Visualization of German Poetics Towards an ‘Algorithmic Criticism’.1

## Interactive Visualization Concepts

We developed two interactive visualization methods to support literary analysis with the help of advanced navigation concepts. Both visualize text documents at different levels of abstraction, including the distribution of various details. These include search terms, available annotations, and automatically extracted information, e.g., named entities or derived concepts. With the simultaneous depiction of overview and detail, visually emerging patterns can be readily perceived and the reason for their occurrence can be identified quickly.

While the approaches were developed with the goal of offering both overview and detail at the same time, they differ in their scope of application. The first one provides a multi-level visual abstraction of a single literary work (Koch et al., 2014). It supports researchers in analyzing text from different ‘distances’, acknowledging the inherent hierarchical structure of a document. The second approach offers just two abstraction levels but is designed to let researchers compare an aspect of interest between different texts (John et al., 2014).

## Data

The ePoetics project aims at analyzing and visualizing a digitized collection of very specific historic documents, namely 20 selected German poetics from the years 1770 to 1960 (Richter, 2010). Poetics are secondary sources including discussions and criticism of literary texts that can be regarded as one of the building blocks of modern literary studies. The corpus was digitized through a double-keying process, and layout, structure, and contained literary concepts were manually annotated.

## Approaches

The multi-level visualization approach enables a researcher to browse and inspect documents based on the inherent structure of the document, e.g., chapters, subchapters, pages, and lines (Figure 1). It encompasses several views to show abstractions of the text, results of search requests, and annotations appropriate to the user’s needs. These views can be attached to each level, as is most adequate for the task at hand. The hierarchical visualization and the navigation concept are based on the SmoothScroll approach (Wörmer and Ertl, 2013). It provides analysts with means to navigate through the visualization and to keep track of the current position across all aggregation levels.

As an interactive visual approach to literary text analysis, it combines the concepts of distant and close reading. Analysts are provided with a complete overview of the document and visual representations of different levels of abstraction that depict aggregated information about text passages of different lengths.

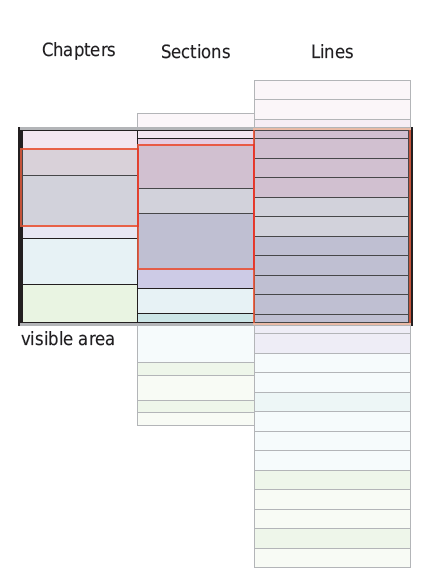


Figure 1. Schematic representation of a three-layer SmoothScroll view. The left layer displays a coarse view of the entire text document, the right layer a detailed but clipped view of individual lines. Highlights indicate which portion of the less detailed layers correspond to the section visible on the detail layer (see Koch et al., 2014).

Several abstract views can be attached to each layer, including word clouds, bar charts, and pictograms. The word clouds highlight prominent concepts and serve as a summarization of a text passage. Exploring these concepts may lead to further analyses and is conducive to the development of new hypotheses and ideas. Bar charts show the number of occurrences of annotations—e.g., persons, quotes, or search terms in a section of the text—while pictograms display the location on a small multiple of the respective page. This gives users an overview of the distribution annotations and helps them to find passages for further analysis. An additional feature is to display facsimiles of the original pages. This gives the analysts immediate access to all non-textual information. This may include handwritten text or pictures, which will typically look different when converted to a digital text format.

The approach assists in literary text analysis with a flexible combination of different visual abstractions that can be adapted to summarize all the information that is important for the task at hand. An analysis example is shown in Figure 2. In the depicted configuration, the term ‘Wallenstein’ (3rd word cloud) arouses the analyst’s interest. After selecting ‘Wallenstein’ the bar charts and pictograms show the distribution of this term and let her navigate quickly to all relevant text passages. In this case, additional literature related to ‘Wallenstein’ is of particular interest to the analyst. As all citations are automatically identified and highlighted on multiple levels of abstraction, she can easily find and analyze them, and thus quickly gain a valuable resource for further studies.

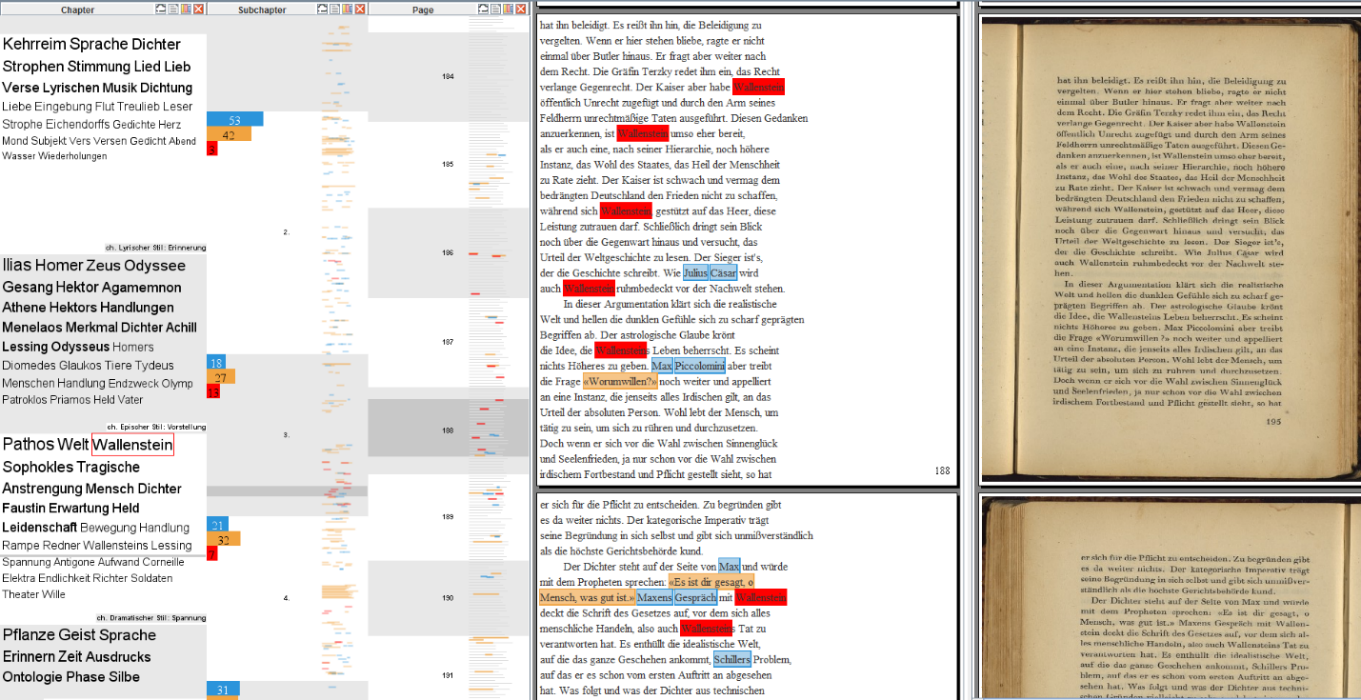


Figure 2. Emil Staiger’s ‘Grundbegriffe der Poetik’ divided into layers, showing chapters (word clouds), subchapters (bar charts and pictograms), pages (pictograms), lines of text, and scanned images of the actual pages.

The second approach we want to discuss is designed to support and expedite the analysis of multiple texts in parallel. The implementation is depicted in Figure 3. Three selected text sources are displayed next to each other as a ribbon. Users can select text passages for analysis, which open up a text box for the document at the corresponding position. Green bars show the position of the selected annotations. The selected text passages are placed next to each other for easy comparison. The scrollbars allow navigation through the documents. For further analysis, users can search for arbitrary terms whose distribution is displayed as orange bars on the ribbons.

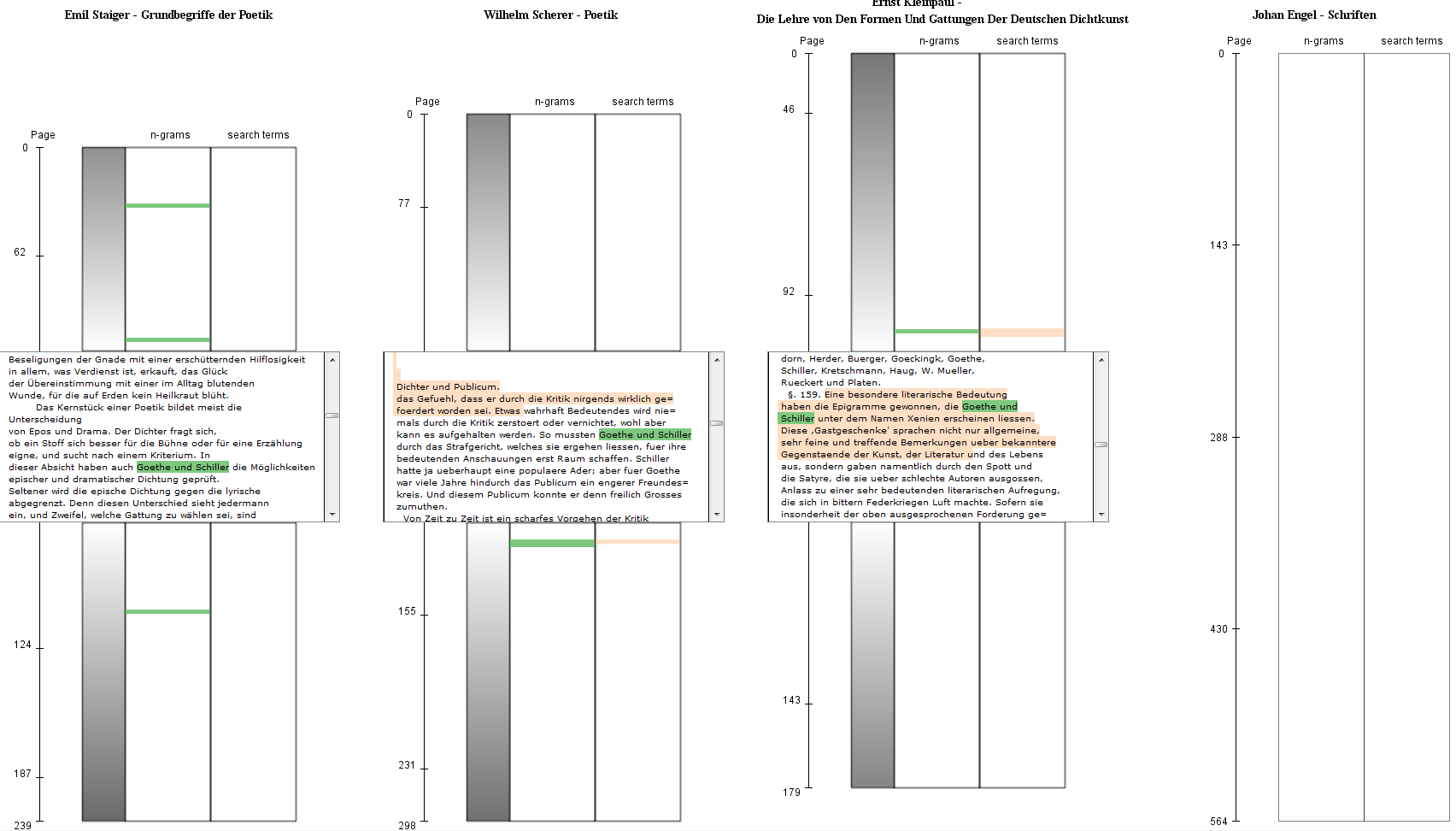


Figure 3. Three selected texts are displayed. Each of the sources has a scale indicating the page number of the original documents to the left, a ribbon showing the position of the selected annotations (green bars), and another for the search results (orange bars).

The focus + context technique of this approach supports a smooth switching between distant and close reading of scholarly sources. It allows the comparison of text passages across different documents to contrast and compare texts from, e.g., different authors or different periods of time. The overview visualization enables the comparison of multiple documents on an abstract level with respect to the distribution annotations and terms chosen by the analyst, while the text boxes allow for free navigation through the single texts and make it easy to find interesting passages for comparison.

## Future Work

The approaches have been evaluated through expert feedback that suggests that they are both effective for literary analysis. For future work, we are planning a comprehensive study of their effectiveness compared to existing text analysis approaches and are aiming at examining different analysis strategies accommodated by the approaches.

As previously mentioned as a general goal, we are currently working towards a comprehensive analysis approach, including automatic processing of text and other data that can be steered and adapted interactively according to users’ needs. We have already conducted multiple experiments to include such methods into our approaches, and we plan to extend these techniques for various new applications.

We believe that the analysis of other types of text and literary works can benefit considerably from the visualization and interactive techniques developed in the ePoetics project.

**Note**

1. ePoetics (www.epoetics.de) is an ongoing research collaboration between the University of Stuttgart and the Technical University Darmstadt funded by the German Federal Ministry of Education and Research (BMBF).

**References**

**John, M., Heimerl, F., Müller, A. and Koch, S**. (2014). A Visual Focus+Context Approach for Text Comparison Tasks. In *Proceedings of VisLR Workshop, LREC 2014*.

**Koch, S., John, M., Wörner, M. and Ertl, T**. (2014). VarifocalReader—In-Depth Visual Analysis of Large Text Documents. In *Transactions on Visualization and Computer Graphics,* **20**(12), IEEE, pp. 1723–32. **Moretti, F.** (2005). *Graphs, Maps, Trees: Abstract Models for a Literary History.* Verso Books.

**Richter, S**. (2010). *A History of Poetics: German Scholarly Aesthetics and Poetics in International Context, 1770–1960.* De Gruyter.

**Wörner, M. and Ertl, T.** (2013). SmoothScroll: A Multi-Scale, Multi-Layer Slider. In *Computer Vision, Imaging and Computer Graphics—Theory and Applications,* Communications in Computer and Information Science 274, Heidelberg: Springer, pp. 142–54.