

# On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges

S. Jänicke<sup>1</sup>, G. Franzini<sup>2</sup>, M. F. Cheema<sup>1</sup> and G. Scheuermann<sup>1</sup>

<sup>1</sup>Image and Signal Processing Group, Department of Computer Science, Leipzig University, Germany

<sup>2</sup>Göttingen Centre for Digital Humanities, University of Göttingen, Germany

---

## Abstract

*We present an overview of the last ten years of research on visualizations that support close and distant reading of textual data in the digital humanities. We look at various works published within both the visualization and digital humanities communities. We provide a taxonomy of applied methods for close and distant reading, and illustrate approaches that combine both reading techniques to provide a multifaceted view of the data. Furthermore, we list toolkits and potentially beneficial visualization approaches for research in the digital humanities. Finally, we summarize collaboration experiences when developing visualizations for close and distant reading, and give an outlook on future challenges in that research area.*

---

## 1. Introduction

Traditionally, humanities scholars carrying out research on a specific or multiple literary work(s) are interested in the analysis of related texts or text passages. But the digital age has opened possibilities for the scholars to enhance their traditional workflows. Enabled by digitization projects, humanities scholars can nowadays reach a vast amount of digitized texts through web portals such as Google Books [Goo15] and Internet Archive [Arc15]. Even for ancient texts, large digital editions exist; popular examples are PHI [PHI15] and the Perseus Digital Library [Per15].

This shift from reading a single book “on paper” to the possibility of browsing many digital texts is one of the origins and principal pillars of the digital humanities domain, which helps to develop solutions to handle vast amounts of cultural heritage data – text being the main data type. In contrast to the traditional methods, the digital humanities allow to pose new research questions on cultural heritage datasets. Thereby, existent algorithms and tools provided by the computer science domain are used, but for various research questions in the humanities, scholars need to invent new methods in collaboration with computer scientists.

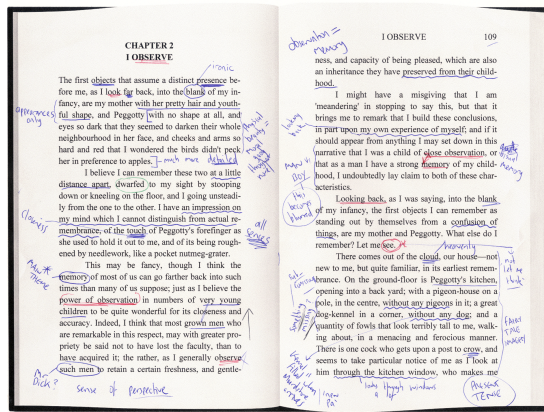
Developed in the late 1980s [Hoc04], the digital humanities primarily focused on designing standards to represent cultural heritage data such as the Text Encoding Initiative (TEI) [TEI15] for texts, and to aggregate, digitize and deliver data. In the recent years, visualization techniques have

gained more and more importance when it comes to analyzing the provided data. For example, Saito introduced her digital humanities conference paper 2010 with: “In recent years, people have tended to be overwhelmed by a vast amount of information in various contexts. Therefore, arguments about ‘Information Visualization’ as a method to make information easy to comprehend are more than understandable.” [SOI10].

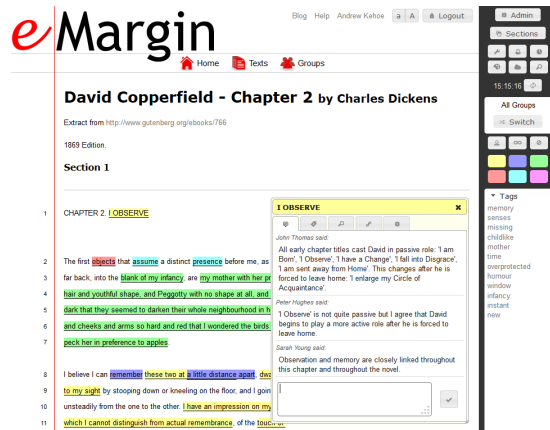
A major impulse for this trend was given by Franco Moretti. In 2005, he published the book “Graphs, Maps, Trees” [Mor05], in which he proposes the so-called *distant reading* approaches for textual data, which steered the traditional way of approaching literature towards a completely new direction. Instead of reading texts in the traditional way – so-called *close reading* –, he invites to count, to graph and to map or, in other words, to visualize them.

This state-of-the-art report surveys close and distant reading visualization techniques that have been developed to support humanities scholars working with literary texts, for example literary scholars, historians and philologists. But humanities scholars from many other fields also apply the typical literary criticism techniques close and distant reading. We present a taxonomy and classify the surveyed approaches based on underlying characteristics. We further investigate the following questions:

- Which experiences are reported regarding collaborations between visualization experts and humanities scholars?



(a) Traditional close reading.



(b) Digital close reading with eMargin.

Figure 1: Examples of close reading of the second chapter of Charles Dickens' *David Copperfield* (Figures reproduced with permission from Kehoe et al. [KG13]).

- Which existing visualization approaches developed for other research fields can be also used in the humanities to support close and distant reading?
- What are future challenges for text visualization concerning close and distant reading to further improve the support for humanities scholars?

## 2. A Definition of Close and Distant Reading

While the close reading of a text is a traditional method in literary criticism that developed in the middle of the 20th century [Haw00], distant reading is a rather novel idea that was introduced by Franco Moretti at the beginning of the 21st century. In contrast to Moretti, Jockers uses the terms *micro-* and *macroanalysis* instead of close and distant reading [Joc13]. Inspired by micro- and macroeconomics, he focuses on quantitative literary text analysis using statistical analysis methods. As the methods we analyzed are rather related to visualization, we decided to use the traditional, more common terms close and distant reading, but we also considered related works using different terminologies. This section introduces close and distant reading techniques and draws a line from the digital humanities to information visualization by combining both techniques.

### 2.1. Close Reading

Close reading is a fundamental method in literary criticism. Nancy Boyles [Boy13] defines it as follows: “Essentially, close reading means reading to uncover layers of meaning that lead to deep comprehension.” In other words, close reading is the thorough interpretation of a text passage by the determination of central themes and the analysis of their development. Moreover, close reading includes the analysis of (1) individuals, events, and ideas, their development

and interaction, (2) used words and phrases, (3) text structure and style, and (4) argument patterns [Jas01]. The result of a traditional close reading approach is shown in Figure 1a. In this example, the scholar used various methods to annotate various features of the source text, e.g., the usage of different colors (blue, red, green) and underlining styles (straight or wavy lines, circles). Furthermore, numerous thoughts are written next to the corresponding sentences. Although most humanities scholars are trained in this traditional approach of close reading, today's large availability of digitized texts and of digital editions through web portals like Google Books [Goo15] or Project Gutenberg [Pro15], opens up new possibilities for close reading, and especially for sustainable and collaborative annotation.

Figure 1b shows a straightforward approach of visualizing various scholars' annotations of a digital edition [KG13] within the web-based environment eMargin [eMa15]. There, colors are used to highlight different text features, and a pop-up window lists the comments of collaborating scholars. In Section 4.1 we outline different approaches to support close reading by visualizing supplementary human- or computer-generated information.

### 2.2. Distant Reading

While close reading retains the ability to read the source text without dissolving its structure, distant reading does the exact opposite. It aims to generate an abstract view by shifting from observing textual content to visualizing global features of a single or of multiple text(s). Moretti [Mor13] describes distant reading as “a little pact with the devil: we know how to read texts, now let's learn how not to read them.” In 2005, he introduces his idea of distant reading [Mor05] with three examples using:

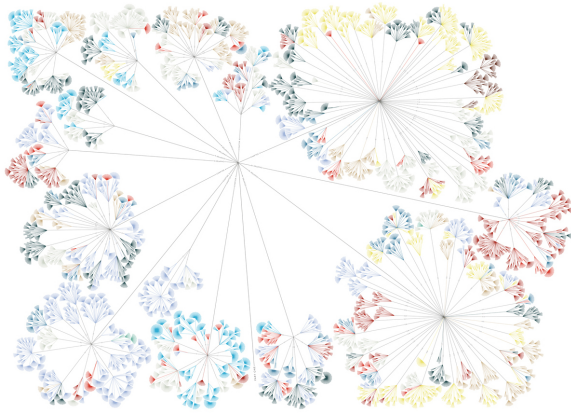


Figure 2: Distant reading example shows the structure of and the themes in Jack Kerouac's "On the Road" (Figure reproduced with permission from Posavec [Pos07]).

- *graphs* to analyze genre change of historical novels,
- *maps* to illustrate geographical aspects of novels, and
- *trees* to classify different types of detective stories.

Although the proposed methods and the intention of distant reading are controversial in the humanities [GH11a, Mar12, CRS\*14], many works in the digital humanities domain are based upon Moretti's idea. Figure 2 shows Posavec's Literary Organism [Pos07] – a distant reading of Jack Kerouac's *On the Road* in the form of a tree. Although being a non-interactive infographic, Posavec's approach perfectly illustrates the idea behind Moretti's distant reading, as it turns away from the traditional close reading by providing an abstract view of a literary text. The shown branching structure represents the ordered hierarchy of content objects from chapters down to words, and themes are drawn with different colors. In Section 4.2 we present a list of different techniques of distant reading developed for a wide range of research questions in the digital humanities.

### 2.3. Combining Close and Distant Reading

During our literature research (see Section 3), we discovered a multitude of works involving close reading and interfaces, which provide distant reading visualizations that allow to interactively drill down to specific portions of the data. This suggests that the direct access to the source texts is important for humanities scholars when working with visualizations. For example, Bradley asks whether it is "possible to develop a visualization technique that does not destroy the original text in the process" [Bra12]. Similarly, Beals asks: "In an age where distant reading is possible, is close reading dead?" [Bea14] Coles et al. argue that distant reading visualizations cannot replace close reading, but they can direct the reader to sections that may deserve further investigation [CL13].

When distant reading views are interactively used to switch to close reading views, the Information Seeking Mantra "Overview first, zoom and filter, details-on-demand" [Shn96] is accomplished. It follows that an important task for the development of visualizations is to provide an overview of the data that highlights potentially interesting patterns. A drill down on these patterns for further exploration is the bridge between distant and close reading.

### 3. Scope & Considered Research Papers

We used the publication year of Moretti's book on distant reading techniques "Graphs, Maps, Trees" [Mor05], 2005, as a starting point to manually scan through all related journals and conference proceedings in order to generate a snapshot of existent research on distant and close reading. We looked at visualization and digital humanities papers spanning a development period of ten years. In order to be considered for our state-of-the-art report, a paper needed to fulfill the following requirements:

- **Textual data:** The visualization is a solution for research questions on an arbitrary text corpus, either a small text unit such as a poem, a large text unit such as a book, or a whole text collection. For example, we did not include a timeline visualization of Picasso's works [MFM08] as it is based upon artworks.
- **Cultural heritage:** The underlying textual data has a historical value. While not considering approaches dealing with texts extracted from social media or wiki systems (e.g., a social network visualization of philosophers presented in [AL09], which is based upon relationships modeled in Wikipedia), we took visualizations for newspaper collections into account. This decision includes some visualization papers without an obvious relation to the humanities. But the proposed techniques are based upon or tested with contemporary newspaper collections, which are indeed part of the cultural heritage.
- **No straightforward metadata visualization:** We only considered papers that provide a visualization that is based upon the inherent textual content. We omitted methods that only use the texts' associated metadata. An example is given by two graphs displaying relationships between texts. Whereas the related network graph presented in [Ede14] is determined by analyzing stylistic features among the textual contents of novels, the unrelated graph visualization in [Fin10] uses Amazon recommendations to determine relationships between books.
- **No basic charts:** In the digital humanities, the word *visualization* is frequently used. Basic charts displaying statistical information are also labeled as visualizations. Based on the definition of information visualization given by Card [CMS99], we only considered papers that provided computer-supported visual representations of abstract data. In contrast to Card, we do not require interactive methods as humanities scholars often gain valu-

Journal/Proceedings	#Papers
IEEE Transactions on Visualization and Computer Graphics (TVCG)	12
IEEE Symposium on Visual Analytics Science and Technology (VAST)	6
Computer Graphics Forum	4
Proceedings of the International Conference on Information Visualization Theory and Applications (IVAPP)	2
Information Visualisation Journal	1

Table 1: Visualization papers examined.

able insights using non-interactive visualizations. However, most proposed techniques implement certain means of interaction.

Altogether, in order to be considered, a paper required a research question on textual data of historical interest emphasizing content rather than metadata and supporting close and/or distant reading of texts.

Table 1 lists the number of related visualization papers examined. We also considered approaches developed for other data types, where at least one use case was provided for a dataset fulfilling the above listed conditions. The *TVCG* journal table entry includes all found papers presented at the *IEEE Symposium on Information Visualization (InfoVis)* and six papers were presented at the *IEEE Symposium on Visual Analytics Science and Technology (VAST)*. Likewise, the *Computer Graphics Forum* entry includes all related papers also contained in the proceedings of the *Joint Eurographics-IEEE VGTC Symposium on Visualization (EuroVis)*. No related papers were found in the proceedings of the *IEEE Pacific Visualization Symposium (PacificVis)* and of the *International Conference on Information Visualisation (IV)*, and the *IEEE Computer Graphics and Applications* journal.

In Table 2, we see the number of papers within the digital humanities domain fulfilling our given requirements. The 49 related works presented at the major digital humanities conference, the *Annual Conference of the Alliance of Digital Humanities Organizations*, indicates the importance of visualizations for close and distant reading. In 2014, a total of 345 individual papers were presented at the conference, of which 33 thematize information visualization techniques for cultural heritage data (9.6%), including 14 papers (4.1%) about close and/or distant reading relevant for our state-of-the-art report. The annual conference proceedings is a suitable snapshot of the research done in that field. The collection is completed by related papers published in two major digital humanities journals, *Literary and Linguistic Computing* and *Digital Humanities Quarterly*.

Figure 3 shows the temporal distribution of the papers within our collection. The trend of related works published within the digital humanities reflects the increasing value

Journal/Proceedings	#Papers
Proceedings of the Annual Conference of the Alliance of Digital Humanities Organizations	49
Literary and Linguistic Computing	14
Digital Humanities Quarterly	4

Table 2: Digital humanities papers examined.

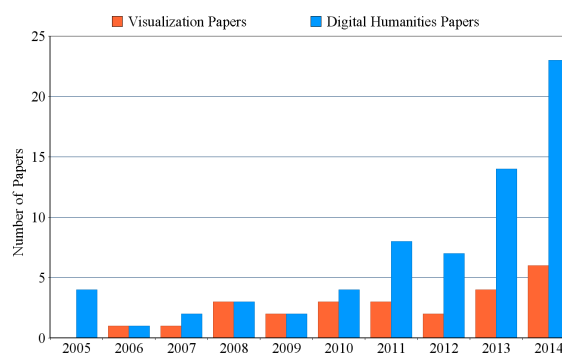


Figure 3: Papers published in visualization and digital humanities communities by year.

of visualizations for close and distant reading in the recent years. But this tendency is only faintly visible for visualization papers. After classifying and analyzing the research papers under investigation, we discuss future challenges, which may be valuable to the visualization community.

#### 4. Taxonomy

This section provides a taxonomy of close and distant reading visualizations found in the papers of our collection. But we first outline the various types of used source texts and diverse data transformation aspects that altogether generate the basis of the proposed visualizations.

**Source texts.** Single literary texts are often the motivation for developing close and distant reading techniques, e.g., Adam Smith's *The Wealth of Nations* [BJ14], Gertrude Stein's *The Making of Americans* [CDP\*07], Herodotus' *Histories* [BPB10], or *The Castle of Perseverance* [Pet14]. Sometimes, an underlying text corpus consists of several works of an author (e.g., the papers of Thomas Jefferson [Kle12] or William Faulkner novels [DNM14]) or different editions of the same literary text (e.g., Shakespeare's *Othello* [GCL\*13] or the *Bible* [JGBS14b]). But most distant reading visualizations are based on large text collections, e.g., biographies [BHW11, Boo13], tales [WJ13a], novels [Ede14], medieval texts [JW13], or news article collections [Wea08, OST\*10, KKL\*11]. This variety of source texts reflects the diversity of research questions raised by humanities scholars, and on the other hand it suggests the

requirement of multifarious close and distant reading techniques.

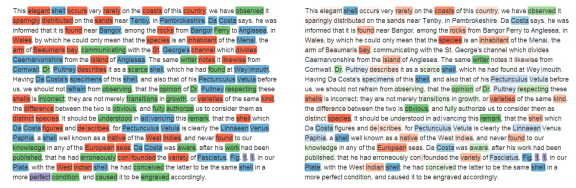
**Data transformation.** Many papers of our collection do not provide sufficient information about applied preprocessing steps to transform the given textual data into the visualization's input format. Sometimes, a visualization directly processes annotated text in XML format [Pie10, Boo13, BGHJ\*14]. Particularly, many techniques are based upon TEI documents, which is a standard digital humanities format [Cay05, BHW11, CGM\*12]. Tokenization is often the first preprocessing step. It is especially required to align certain text passages when analyzing parallel texts [WJ13b, JGBS14a], or to determine the similarity of re-used text passages [BGHE10, JGBS14b]. But most distant reading techniques require more sophisticated preprocessing steps. Often, the frequencies of single words or n-grams are determined as the basis to compute the similarity among texts of a corpus [Bea12, Joc12, Ede14]. Other techniques take advantage of topic modeling approaches to cluster texts or to determine similarities among them [DWS\*12, AKV\*14, BJ14]. Vector space models are a further popular method to represent documents and compute similarities [DFM\*08, EX10, KOTM13]. Different strategies are applied when extracting geospatial information from texts. Automated approaches use gazetteers [EJ14, HACQ14], and sometimes humanities scholars manually extract placenames occurring in texts and collect corresponding geographical coordinates [JHSS12, JW13]. In [WMN\*14], the close reading visualization of texts solely depends on manually collected data through crowdsourcing. Some semi-automatic scenarios include the humanities scholar's knowledge, who manually generates or validates a training set to produce an appropriate data mining classifier [PSA\*06, KKL\*11].

In the following, we list all close and distant reading techniques found. Furthermore, we outline strategies for combining close and distant reading visualizations to facilitate a multifaceted analysis of the underlying textual data. Finally, we classify all papers of our collection in dependency on used source texts, intended purpose and provided close and/or distant reading techniques.

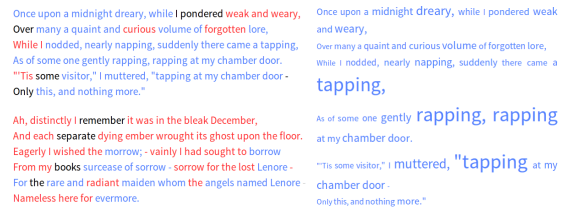
**4.1. Close Reading Techniques**

A visualization that allows to close read a text requires that the structure of the text be retained in order to facilitate a smooth analysis. With additional information in the form of manual annotations or of automatically processed features of textual entities or relationships among them, a plain text can be transformed into a comprehensive knowledge source.

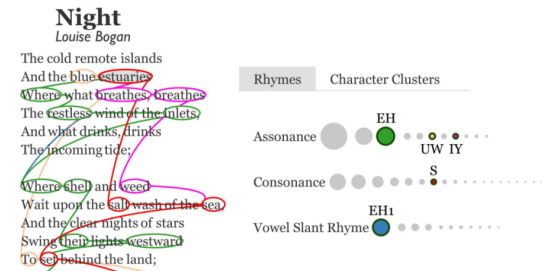
While eight visualizations provide only plain close reading views without additional information, 38 works of our research paper collection developed enhanced close reading approaches. To visualize such additional information for a great variety of purposes, the researchers made use of the techniques listed below.



(a) Colored backgrounds and backgrounds with varying transparency (Figure provided by Alexander et al. and based on [AKV\*14]).



(b) PRISM uses color to highlight the classification of words and font size to encode the number of annotations (Figures under CC BY 3.0 license based on [WMN\*14]).



(c) Circumcircles and connections highlight assonance, consonance, and slant rhyme sets (Figure reproduced with permission from Coles et al. [CMLM14]).

Figure 4: Color usage for close reading.

**Color** is the visual attribute most often used to display the features of textual entities and it is applied in various ways. In most cases, a colored background is used to express various types of information about a single word or an entire phrase (Figure 1b). The tool Serendip [AKV\*14] varies the transparency of background colors to encode the importance of individual words (Figure 4a). Font color is also frequently used for this purpose (Figure 4b left). Colored circumcircles (Figure 4c) around words are used only once [CMLM14]. When displaying digital editions of literary texts, insertions are underlined. This might be the reason that this metaphor of underlining words is also rarely used to enhance close reading [CWG11]. Overall, coloring is a suitable method to express a great variety of textual features. Among other purposes, coloring is used to highlight the automated or manual classification of words or phrases [KJW\*14, WMN\*14], to mark common words in parallel texts [JRS\*09, Mur11] or to visualize various sound patterns in poems [CTA\*13, Ben14].

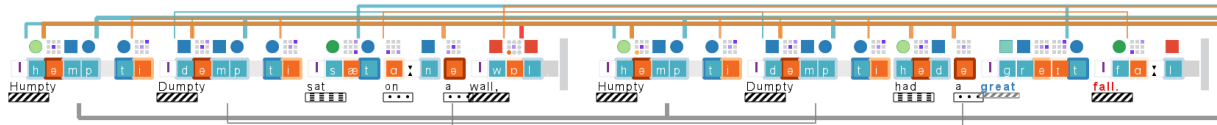


Figure 5: Poem Viewer uses glyphs to encode phonetic units, and connections show phonetic and semantic relationships (Figure reproduced with permission from Abdul-Rahman et al. [ARLC\*13]).

**Font size** is another method of visualizing features of textual entities. Adopted from tag cloud design [VWF09], this metaphor serves best to highlight the significance or weight of a textual entity in relation to the given text or corpus. In the design of a variant graph [JGBS14a], which is a directed acyclic graph visualizing differences and similarities among various editions of a text, font size encodes the number of occurrences of a word among all editions (Figure 7a). Within the web-based tool PRISM [WMN\*14], users collaboratively group the words of literary texts into different categories. The collected statistics are used to display the number of annotations of each word by variable font size (Figure 4b right). In [CWG11], varying font size is used to visualize the importance of text passages according to the user’s preferences.

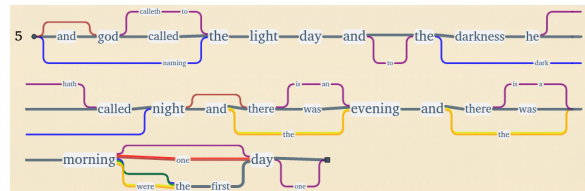
**Glyphs** attached to individual textual entities are convenient techniques to visualize abstract annotations that are hardly expressible with plain coloring or varying font size. Most examples we found enhance the close reading of poems. In [ARLC\*13], phonetic units are drawn atop each word using color to classify phonetic types (Figure 5). Additionally, pictograms illustrate phonetic features. The Myopia Poetry Visualization tool uses rectangular blocks to visualize poetic feet and the spoken length of syllables [CGM\*12]. For the visualization of a poem’s hermeneutic structure, Piez deploys glyphs in the form of rectangular and circular

maps [Pie10, Pie13]. An example is given in Figure 6. Goffin explores the placement and design of so-called word-scale visualizations, which are small glyphs enriching the base text with additional information [GWF14]. For example, the background color of words contained in digital copies speaks for OCR certainty. Furthermore, small interactive bar charts illustrate variants of observed words.

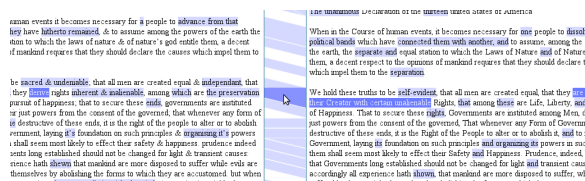
**Connections** aid to illustrate the structure among textual entities. One usage of connections in close reading is to highlight subsequent words in a variant graph to track variation among various text editions [BGHE10]. As shown in Figure 7a, colored links can help to identify certain editions [JGBS14a]. Other approaches juxtapose the texts of various editions and visually link related text passages [WJ13b, HKTK14, JGBS14b], as instantiated in Figure 7b. Connections can also be used to visualize sentence structure [KZ14] or the phonetic and semantic relations within poems [ARLC\*13] (Figure 5). For example, Coles draws arcs between words of a poem sharing the same tones (assonances) to support the analysis of occurring sonic patterns [CMLM14] (Figure 4c).



Figure 6: Close reading example with glyphs illustrating the hermeneutic structure of a poem (Figure provided by Piez and based on [Pie10]).



(a) Variant graph for seven English translations of *Genesis 1:5* connecting subsequent words displayed with variable font size (Figure based on [JGBS14a]).



(b) Juxta commons supports close reading of two parallel texts by connecting related text passages (Figure provided by Wheelers et al. and based on [WJ13b]).

Figure 7: Connections used for text alignment.

## 4.2. Distant Reading Techniques

A visualization that displays summarized information of the given text corpus facilitates distant reading. The process of transforming such information into complex representations can be based upon a large variety of data dimensions, e.g., various types of metadata of textual entities, automatically processed or manually retrieved relationships between textual elements, or quantitative and qualitative statistics about unstructured textual contents.

81 research papers of our collection attend to the matter of providing a rather abstract distant reading view of a given text corpus. We extracted and grouped various approaches found to visualize summarized information into the seven following categories.

**Structure** overviews illustrate the hierarchy of an individual text or an entire corpus. A common method is the usage of rectangular blocks to display the structural elements of a text [Cay05, JRS\*09, VCPK09]. In [BGHJ\*14], bars with variable length are used to visualize certain statements about Stuttgart 21 from different speakers. Another approach is the use of pictograms to illustrate a distant view of the structure of a text [KJW\*14]. In digital humanities, the XML-based TEI format has become the humanities leading technology to map the structure of a digital text edition [Sin13]. Figure 8 shows that hierarchical treemaps are effective methods for the visualization of TEI document structures [Pie05].



Figure 8: Hierarchical treemap visualizes the TEI document structure (Figure reproduced with permission from Piez [Pie05]).

**Heat maps** or block matrices are often used to highlight textual patterns. An example is the usage of these techniques to show relationships among various texts in a corpus. The similarity for each tuple of texts within the corpus can be determined by counting the number of similar text passages, and the result can be visualized as a heat map [GCL\*13, FKT14], e.g., to highlight similar Shakespearean plays [RRRG05]. In [JGBS14b], heat maps are used to highlight systematic and repetitive text re-use among the books of the Bible (Figure 9). Other works use heat maps to illustrate patterns within single texts, e.g., to highlight the occurrence and frequency of requested text patterns [CDP\*07, CWG11, Mur11, MH13]. Alexander et al. propose two matrix representations [AKV\*14]. The RankViewer illustrates the ranking of words belonging to topics and the CorpusViewer shows relations to certain topics for each document of a corpus. Fingerprinting techniques, as introduced in [KO07], visualize characteristic tex-

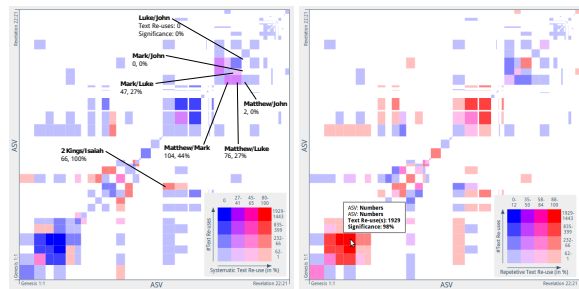


Figure 9: Heat maps highlight systematic (left) and repetitive (right) text re-use between Bible books (Figure reproduced with permission from Jänicke et al. [JGBS14b]).

tual features of literary works, or can be further used to reveal interpersonal relationships between characters in prose literature [OKK13]. Weaver uses a matrix calendar view to indicate the occurrence of certain events [Wea08]. Finally, heat maps are used to visualize the similarity [CTA\*13] or the flow [FS11, Ben14] of sound in poems.

**Tag clouds** are intuitive visualizations to encode the number of occurrences of words within a selected section, a whole document or an entire text corpus by using variable font size [VCPK09, FKT14]. By applying significance measures, the visualization can be limited to displaying only characteristic tags [ESK14, HACQ14, KJW\*14] (examples can be seen in Figure 10a and Figure 15b). Tag clouds can also summarize the major tags for certain time periods [CLT\*11, Bea12, CLWW14] or topics inherent in a text corpus [BJ14]. Beaven used tag clouds to illustrate collocational relationships of a single word [Bea08] and to compare the collocates between two words [Bea11]. In some of the mentioned works, tag coloring is used to express additional information such as classifications or the temporal evolution of a word's significance.

**Maps** are widely used to display the geospatial information contained in a text. Many approaches project the place-names mentioned in a text or in an entire corpus onto maps. With the help of contemporary (e.g., GeoNames [Geo15]) and historical gazetteers (e.g., Pleiades [Ple15]), the extracted placenames can be enriched with geographical coordinates, and their visualization on a map supports the analysis of the (fictional) geographic space described in the source text(s). Some approaches use thematic [DFM\*08, ÓML14] or density maps [GHI1b, GDMF\*14] for this purpose, but the usage of glyphs in the form of circles is more frequent [Tra09, DWS\*12, HACQ14] as it simplifies the interaction with individual plotted places (e.g., see Figure 10a). In [JW13], various types of glyphs encode various types of places occurring in medieval texts (Figure 10b). Two works, which focus on mapping the geographical knowledge of ancient Greek authors, draw connections between glyphs to illustrate travel routes [EJ14] or to highlight the strength of

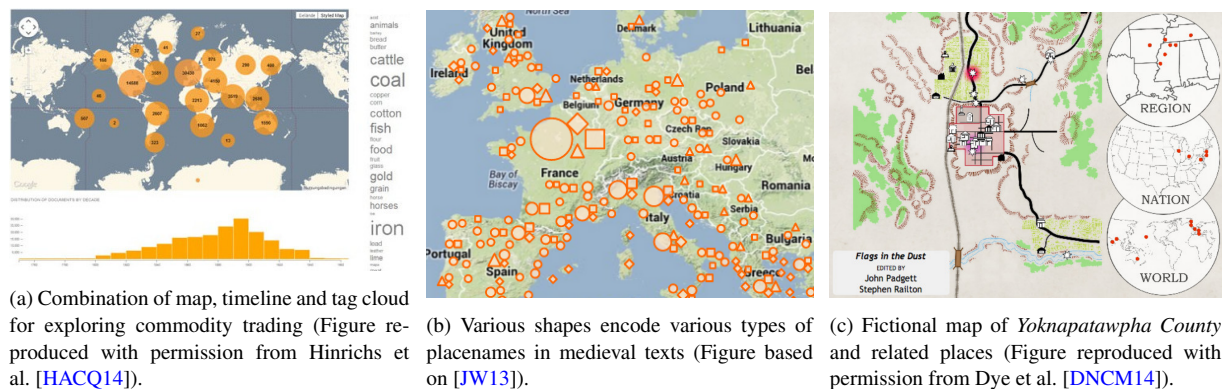


Figure 10: Maps supporting distant reading.

the relationship between placenames, which is reflected by the number of co-occurrences [BPBI10]. In contrast to the previous works, the geospatial metadata associated with individual corpus texts (text creation timestamp) can be used for mapping [MBL\*06, JHSS12], e.g., to analyze geographic patterns of political activity [Wea08]. The visualization of Faulkner's fictional *Yoknapatawpha County* includes various means of geographic mapping [DNM14]: on the one hand, the imagined geography and, on the other, the placenames displayed on the geographic levels *region*, *nation* and *world* (Figure 10c).

**Timelines** are appropriate techniques to visualize historical text corpora carrying various types of temporal information. One approach is the straightforward use of the text's metadata. This supports the temporal analysis of a word's usage in ancient Greek texts [JHSS12], the visualization of uncertain datings of medieval texts through cited toponyms [JW13], or the exploration of events in news articles [Wea08, ESK14] (e.g., see Figure 15b). Sometimes, the temporal information about events reported in

a text needs to be extracted in order to visualize (fictional) calendars [DNM14, GDMF\*14, HACQ14, ÓML14] (e.g., see Figure 10a). For the exploration of placenames in Herodotus' *Histories*, a timeline is used to show where certain placenames occur in the text [BPBI10]. A somewhat abstract timeline view is shown in [HSC08]. Here, a so-called tree cut section, whereby each ring represents a decade, visualizes statements from and about Emily Carr's life and work (Figure 11). Streamgraphs are popular techniques that produce aesthetic visualizations and allow to track the evolution of themes over time [BW08], thus generating enhanced versions of the timeline metaphor. Such visualizations are often based on newspaper sources to support the analysis of contemporary topic changes [CLT\*11, KBK11, DWS\*12, CLWW14]. Using a research paper pool [ARR\*12], the changing importance of research topics can be explored. Streamgraphs may also be used to visualize storylines or to illustrate plot evolution and changing locations within literary texts [LWW\*13]. Based upon Hollywood screenplays, the tool ScripThreads visualizes action lines of movie characters [HPR14].

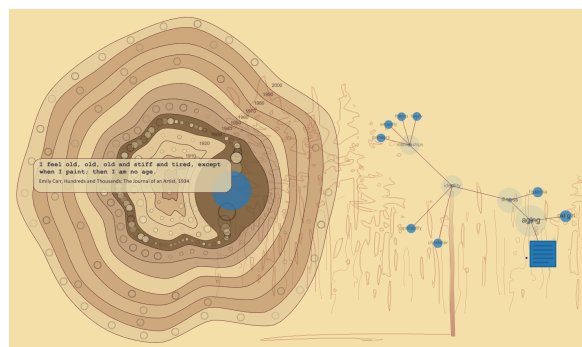


Figure 11: A combination of an abstract timeline view and a tree in EMDialog (Figure provided by Hinrichs based on [HSC08]).

**Graphs** are the most often applied method to visualize certain structural features of a text corpus. A common usage is the visualization of relationships between the texts (represented as nodes) of a corpus in the form of a network [Gal11, WH11]. Proximity can be used to express the similarity of these texts based upon similar paragraphs [EX10] or stylistic closeness [Joc12, CEJ\*14]. For example, Eder visualizes a network of poetic texts based on stylistic features [Ede14], thereby highlighting connected editions and sequels (Figure 13). In other works, nodes represent the words of a corpus and links are drawn to reflect semantic relationships [AGL\*07, RFH14] or co-occurrences [KKL\*11, WJ13a]. Further applications are the visualization of scene changes and character movements in Shakespearean plays [RRRG05], as well as the display of conceptual [Arm14], contextual [HSC08] or multilingual [GZ12] information. Phrase nets connect textual enti-



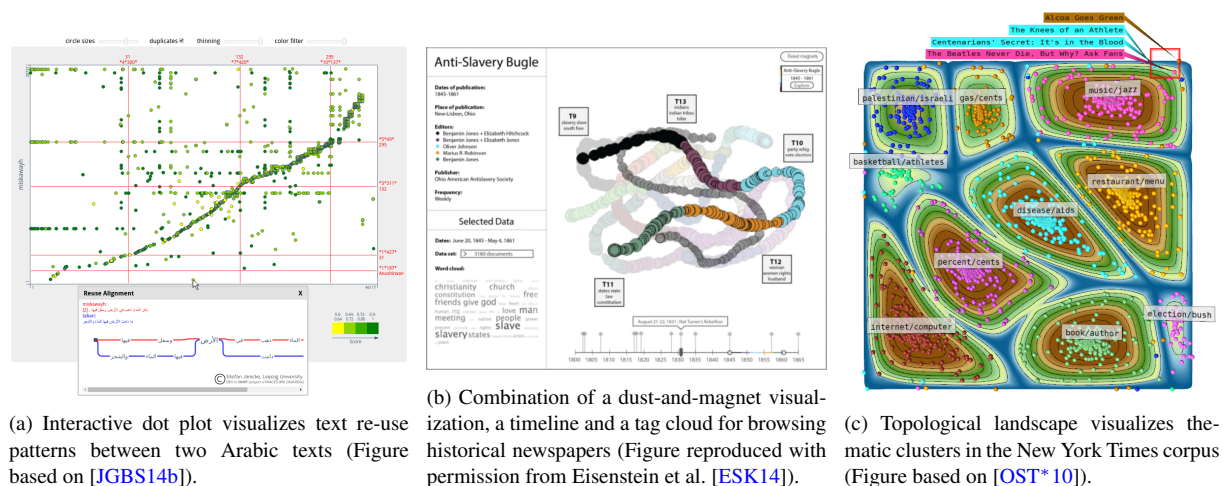


Figure 15: Miscellaneous distant reading methods.

**Miscellaneous** methods also produce beneficial results for certain research questions. In [JGBS14b], an interactive dot plot interface is used to visualize and explore patterns of text re-use between two texts (Figure 15a). In [GCL\*13], a parallel coordinates and a dot plot view, which is used for filtering purposes, visualizes the similarity of parallel text sections. For the exploratory thematic analysis of historical newspaper archives [ESK14], an application of the dust-and-magnet metaphor [YMSJ05] yielded useful results (Figure 15b). The tool PlotVis allows users to model and interact with XML-encoded literary narratives in 3D [PBD14]. A complex tool for the simulation of theatrical plays consists of various 2D interfaces illustrating the “line of action” and a 3D interface populated by character avatars [RSDCD\*13]. The categories of words contained in two books can be compared with Sankey diagrams [HCC14]. Tree maps can be used to illustrate the occurrences of adjectives in fairy tales [WJ13a]. For the analysis of repetitions in Gertrude Stein’s *The Making of the Americans* [CDP\*07], parallel coordinates visualize the frequency of phrases across sections, and TextArc [Pal02] is used to explore the repetition of individual words. Finally, the topology-based clustering of articles taken from the New York Times Corpus (Figure 15c) can be visualized using a landscape metaphor [OST\*10].

#### 4.3. Techniques for Combining Close and Distant Reading Visualizations

Most of the visualizations we found provide either a close or a distant reading of a text corpus. Still, an important feature for literary scholars when working with distant reading visualizations is direct access to source texts or, in other words, close reading. Among the papers in our collection providing close and distant reading, some visualizations combine both techniques – most often in the form of coor-

dated views [WBWK00]. We do not consider the methods in [WH11, CTA\*13, Ben14, BJ14] as the presented visualizations for close and distant reading serve different purposes, and are not connected to one another. The 26 remaining techniques are grouped into the three categories outlined below.

**Top-down** strategies are mostly applied when combining close and distant reading visualizations. Such methods implement the Information Seeking Mantra in its original meaning. Initially, a distant view on the textual data is shown, the user can often manipulate the visualization by means of filtering and zooming, and finally retrieve the details-on-demand by clicking on a potentially interesting data item. In some cases, the texts are simply shown at the end of the information seeking pipeline [Cay05, HSC08, DWS\*12, MFM13, RSDCD\*13, FKT14]. Observed words or text patterns are often highlighted in the close reading view by way of coloring [VCPK09, GZ12, Wol13, AKV\*14, HACQ14, HPR14]. Various colors can thereby illustrate word categories [CDP\*07], e.g., toponym types in the Herodotus Timemap [BPB110] or topological cluster information [OST\*10]. In some systems, close reading is more closely related to the preceding distant reading. In [BGHJ\*14], the connection between close and distant reading is achieved by zooming. The distant view, a structural overview, highlights certain patterns, and zooming allows the close reading of individual passages. In [JGBS14b], a grid-based heat map visualizes similarities between texts of a corpus, and clicking on a grid cell opens a close reading view showing the corresponding two texts juxtaposed with connections between related text passages. The CorpusSeparator presented in [CWG11] is a distant view used to generate a weighted tag list (dependent on corpus statistics). Based upon these weights, the close reading view of a text (illustrated with Shakespeare’s *A Midsummer Night’s Dream*) is manipulated by coloring and sizing lines.

**Bottom-up** methods are rarely applied. The major focus in such scenarios is the source text and, therefore, close reading. In [GCL\*13], the user selects a desired text passage in Shakespeare's *Othello*, which is shown in various German translations. Distant reading visualizations are processed (parallel coordinates view, dot plot view, heatmap) based on that selection. Although text coloring can be applied to filtering mechanisms, the literary scholar's interest in this scenario is the comparison of text editions on a textual level. In [Mur11], the literary scholar selects a certain phrase during the close reading process. Next, that phrase is searched within the text corpus and the phrase's distribution is shown in the form of a heat map.

**Top-down & bottom-up** approaches taken within one visualization entity allow for switching between close and distant reading while taking into account manipulations of the preceding view. In [JRS\*09], the user can switch between structural overview (distant reading) and text view (close reading). A side-by-side navigation between source text (close reading) and a distant reading graph showing the relationships among textual entities is illustrated in [WV08, RFH14]. Here, textual entities can be selected in both the graph and the text, triggering mutual updates. A typical use case for the combination of top-down and bottom-up behaviors are visual analytics methods. The Varifocal-Reader [KJW\*14] hierarchically visualizes a document with the help of distant views (structural overview, tag clouds) and close reading techniques (use of color, digital copy), thus supporting hierarchical navigation. In close reading mode, automatically acquired classifications of textual entities can be manually modified, which subsequently affects distant views. The same applies to social networks automatically extracted from newspaper articles [KLB14]. The user browses the graph, opens close reading views associated with individual nodes and annotates the source text, which, again, affects the distant view and is used for classifier training. WordSeer [MH13] allows for a multifaceted perusal of a text corpus. For selected textual entities, several close and distant reading views can be used to browse the corresponding source texts. Within the close reading views, the user can group words into classes, which can then be used as a starting point for text corpus analysis.

#### 4.4. Paper Classification

Table 3 shows the classification of all research papers based on the underlying source text(s), intended purpose and implemented technique(s). Visualizations allowing for both distant and close reading are classified in relation to their main purpose. The works providing solutions for various purposes each appear in two categories [RRRG05, WH11, CTA\*13, WJ13a, Ben14, BJ14].

**Single Text Analysis.** The main purpose of such a visualization is the analysis of an individual literary work, e.g., eight of the classified works support close or distant reading

of single poems. Whereas *enhanced text view* techniques focus solely on visualizing additional information within the text flow, *abstract text views* support the distant reading of a single text. Some methods provide visualizations for *both* close and distant reading of a given text.

**Parallel Text Analysis.** This category lists all techniques aiming to visualize the similarities and differences between various editions of a literary work or between texts containing similar passages. Here, the number of individual texts is limited. Visualizations for *section alignments* often display only two texts next to each other. In contrast, the number of editions is potentially high for *sentence alignments* focusing on small textual entities, which are displayed by connecting subsequent words.

**Corpus Analysis.** Unlike the previous categories, the visualizations listed here are based upon an entire corpus containing a high number of texts. One goal is to show automatically processed *statistics about textual entities*, often visualized in the form of tag clouds or heat maps. The means of choice for the other subcategories are graphs that display *relationships between texts* based on the similarities of the textual contents, and *relationships between textual entities* such as words or *social networks*, showing the relations between characters appearing in the texts. Although some social networks are generated for individual texts, we place this category here as its general idea is similar to that of showing relationships between textual entities. Also, many approaches visualize geospatial and temporal information extracted from text(s). The category *space and time* lists visualizations that display placenames extracted from texts on maps in dependency on time, a value that is often inherent in digital humanities data. Other approaches only use information about *space* or *time*.

## 5. Collaboration Experiences

Within our collection, we examined papers about the research experiences reported by visualization researchers in order to provide suggestions that might help visualization scholars new to the field of digital humanities to develop successful visualizations. Not many projects reveal insights into collaboration experiences. An excellent procedure of developing an interface to analyze the sound in poems is given by Abdul-Rahman [ARLC\*13], who successfully presented her method at visualization and digital humanities conferences. Based upon a user-centered design study [Mun09], she reports insights "of a fascinating collaboration between computer scientists and literary scholars". Other publications also share important experiences. Some of the gained insights regarding various aspects of the development phase are outlined below.

**Project start.** One of the most important, initial tasks of a digital humanities project are discussions about the research questions and perspectives for which a visual-

			Close Reading					Distant Reading						
			Plain	Color	Font size	Glyphs	Connections	Structure	Heat maps	Tag clouds	Maps	Timelines	Graphs	Miscellaneous
Single Text Analysis	enhanced text views	[Pie10], [CGM*12], [Pie13], [GWFI14]				x								
		[PSA*06], [CTA*13], [Ben14], [BJ14]		x										
		[ARLC*13]				x	x							
	both	[WMN*14]		x	x									
		[VCPK09], [BGHJ*14], [KJW*14]		x				x		x				
		[WJ13b], [CMLM14], [KZ14]		x			x							
		[Cay05]	x					x						
		[CDP*07]		x						x				
		[WV08]		x									x	
		[MFM13]	x											x
abstract text views	[RSDCD*13]	x											x	
	[KO07], [FS11], [CTA*13], [OKK13], [Ben14]								x					
	[Pie05]						x							
		[PBD14]											x	
Parallel Text Analysis	section alignments	[WH11], [HKTK14]		x										
		[Cor13], [WJ13b]		x			x							
		[JRS*09]		x				x						
		[GCL*13]		x						x				
	sentence alignments	[JGBS14b]		x			x		x					x
[BGHE10]			x			x								
		[JGBS14a]			x	x								
Corpus Analysis	statistics for textual entities	[Bea08], [Bea11], [Bea12], [BJ14]								x				
		[WJ13a], [HCC14]												x
		[CWG11]		x	x				x					
		[Mur11]		x					x					
	relationships between texts	[FKT14]	x						x	x				
		[EX10], [Gal11], [WH11], [Joc12], [CEJ*14], [Ede14]												x
		[RRRG05]								x				
		[OST*10]		x										x
	relationships between textual entities	[Wol13]		x										x
		[RRRG05], [AGL*07], [vHWV09], [KKL*11], [MLSU13], [WJ13a], [Arm14]												x
		[GZ12], [RFH14]		x										x
		[MH13]		x						x				x
	social networks	[AKV*14]		x						x				
		[Cob05], [CSV08], [BDF*10], [RD10], [BHW11], [Kle12], [Boo13], [KOTM13], [Tôt13], [Pet14]												x
		[KLB14]	x											x
	space and time	[JHSS12], [JW13], [DNM14], [GDMF*14], [ÓML14]										x	x	
		[Wea08]								x		x	x	
		[BPBI10]		x								x	x	
		[DWS*12]	x								x	x	x	
	space	[HACQ14]		x							x	x	x	
[MBL*06], [DFM*08], [Tra09], [GHI1b], [EJ14]														
time	[KBK11], [ARR*12], [LWW*13]												x	
	[CLT*11], [CLWW14]									x		x		
	[HSC08]	x									x	x		
	[DWS*12]	x								x	x	x		
	[ESK14]									x		x	x	
		[HPR14]		x								x		

Table 3: Classification of research papers according to provided visualization techniques.

ization, be it for close or distant reading, can be beneficial [JGBS14b]. These discussions include the analysis of the data features [VCPK09] as well as the setup of regular project meetings to work on and extend a collaborative idea. Special workshops can help computer scientists and humanities scholars get acquainted with each others' tasks, mindsets and workflows [ARLC\*13]. Abdul-Rahman reports the importance of visualization researchers participating in poetry readings and in-depth discussions with literary scholars to discover "a variety of interesting problems that might be subject to visualization solutions". Also, a small corpus generated for literary scholars was helpful for Abdul-Rahman to examine research questions without the aid of existent visualizations.

**Iterative development of prototypes.** The beneficial involvement of humanities scholars in various stages of the visualization development is reported in a number of publications. For example, regular face-to-face sessions between computer scientists and humanities scholars can help to identify problems and potential enhancements of the prototype design [JGBS14b]. Such a session should be composed of a demonstration and trials of the visualization prototype as well as intense discussions in order to gather the levels of detail and complexity that a visualization should ideally reach [ARLC\*13]. Gefner stated that such a process finally helps to gain an intuitive result "even for the inexperienced, maybe sceptical user" [JGBS14a]. A further experience is given by scholars involved in the development of Neatline [NMG\*13], which is based upon Omeka [Ome15], a content management system for online digital collections. The stepwise development of Neatline led to advancements of Omeka itself, thus benefiting a far wider audience than originally anticipated.

**Evaluating visualizations with humanities scholars.** The evaluation sessions provided important insights into design, intuitiveness, the utility of visualizations and into potential enhancements. A number of humanities scholars working with the visualizations suggested further enhancements, some of which strengthen the importance of close reading solutions [CWG11]. For example, when similar close and distant views were provided, "users stressed that it is preferable to see the actual words" rather than abstract overviews [JRS\*09]. When working with the VarifocalReader, the user liked to view "the digitized image of a book's page and mentioned that this would increase his trust in the approach" [KJW\*14]. The metaphor of a digitized text is also used when comparing various English translations of the Bible [JGBS14a], which "reminds the user that it is a book to be read, not just some string of letters". Although developed for museum visitors, the importance of aesthetic appeal to engage in information exploration was reported in [HSC08]. The fact that visualizations should be designed to meet humanities work practices is mentioned in [BDF\*10]. Some humanities scholars also mentioned issues or limitations with the presented tools. For instance,

the need to confirm temporary results by analyzing larger datasets or, in other words, more texts and in more languages [GCL\*13]. In [HACQ14], the attached labeling was a crucial issue. The authors resumed the requirement of a visual representation "to be clear in order to make visualizations a valid research tool". Scientists involved in [HKTK14] stated that collaborative work helped to reactivate and to regenerate traditional literary methodologies rather than abandon them. Finally, the utility of visualization in the humanities is corroborated by the fact that, already during the evaluation phase, many literary scholars make surprising discoveries [HACQ14], generate new hypotheses or suggest further usage scenarios for the tools [CWG11, ARLC\*13, GCL\*13].

When working on Gertrude Stein's *The Making of Americans* with POSVis [VCPK09], the collaborating literary scholar could generate substantial knowledge about the usage of the word *one*. This led to a publication she presented at the digital humanities conference 2009 [CPV09]. A statement, taken from [ARLC\*13], finally points out the value of visualization for humanities scholars, who mentioned that "they would not likely look for insight from the tool itself ... they would look for enhanced poetic engagement, facilitated by visualization."

## 6. Visualization Solutions Used in the Digital Humanities

In this section, we take a look at toolkits mainly developed by visualization researchers, used for various purposes in digital humanities applications. Furthermore, we list potentially beneficial visualization techniques, which, however, are not currently related to the humanities domain.

### 6.1. Visualization Tools in the Digital Humanities

While several visualizations presented within the digital humanities community are created with the help of standard technologies such as HTML, JavaScript, SVG, or geographic information systems (GIS), many other visualizations make use of existent toolkits to transform textual data into visual metaphors. In most cases, these toolkits are used to visualize graphs based on a given dataset. For that purpose the following libraries developed by information visualization scientists are cited: Bostock's D<sup>3</sup>-Data Driven Documents [BOH11] and its predecessor Protovis [BH09], Heer's Prefuse [HCL05] and Viegas' ManyEyes [VWvH\*07]. Another popular tool in the digital humanities community, frequently used to draw graphs, is Bastian's Gephi [BHJ\*09]. Besides D<sup>3</sup>, Neatline [NMG\*13] and GeoTemCo [JHS13] are the basis for visualizations of geospatial data. Other used libraries are the InfoVis toolkit [Fek04], D2K [DUY\*05], FeatureLens [DZG\*07] and TextArc [Pal02].

Solomon questions the benefit of a toolkit that "has its roots ... in other contexts and contains its own embedded goals, methodologies, and ideological underpin-

nings” [Sol13] in reference to ManyEyes, which was not originally developed for the digital humanities domain. Although the results presented in the corresponding papers indeed illustrate the utility of visualization toolkits – especially the usage of D<sup>3</sup>, which has markedly increased in 2013 and 2014 – some digital humanities research questions require the invention of new tools, e.g., PoemViewer [ARLC\*13] and ProseVis [CDP\*07, CTA\*13], both used for the visualization of poem features. Another example is given by a method that provides a design for variant graphs [JGBS14a]. The visualization is comparable to the Word Tree [WV08], but instead of only displaying text patterns that share the same beginning in the form of a tree, it is capable of visualizing a directed acyclic graph reflecting the variance among textual editions. The Word Graph [RGP\*12] proposes a similar method, but its design is not usable to support those humanities scholars who want to track individual and to compare multiple text editions in these graphs.

Despite the fact that complex visualization toolkits cannot provide out-of-the-box solutions for all research questions in the digital humanities, the following section lists potentially beneficial methods for a number of purposes.

## 6.2. Applicable Visualization Techniques for Digital Humanities Research

Originally, Parallel Tag Clouds were designed to analyze the development of text features in legal documents over time [CVW09]. Later, the Trading Consequences project used the proposed idea to visualize mentioned locations with their frequencies per year [HACQ14]. This is one example that illustrates the potential utility of existent visualization techniques for digital humanities purposes. Bearing in mind our main classification categories, we scanned through works published in the visualization community that might address research questions in the digital humanities. 13 related visualization ideas and techniques are outlined below.

**Single Text Analysis.** The DAViewer, a visualization system for exploring discourse texts, displays the discourse’s hierarchy in the form of a dendrogram, and the texts can be read in an overview panel [ZCCB12]. In this way, it provides an effective combination of close and distant reading, which could be adapted to visualize tragedies such as Shakespeare’s *Hamlet*. The Sequence Surveyor also provides a dendrogram to explore genomic structures [ADG11]. Each leaf shows a heat map illustrating genome distributions. This metaphor could be used to visualize both the rhyme structure of a poem in dendrogram form and the heat maps displaying sound patterns. Another visualization suitable for humanities data is DocuBurst, a technique that visualizes hierarchical summaries of a text in circular patterns [CCP09].

**Parallel Text Analysis.** MizBee is a multiscale synteny browser that visualizes synteny relationships with the help of linked views on various levels (genome, chromosome,

and block levels) [MMP09]. Such a multilevel visualization could also help to enhance current visualizations displaying similar text passages among textual editions. The provided circular overview illustrating relationships could be applied in the form of a distant view showing similarity patterns between many text editions, a substantial improvement to current solutions working with a limited (mostly two) number of editions.

**Corpus Analysis.** Much research has been done in graph visualization. In particular, improvements regarding the readability of and interaction on large graphs [ZBDS12] are also relevant for handling large graphs in humanities applications. Other works focus on improving the means of interaction for large social network graphs [HB05, PS06]. Often, digital humanities data contains more facets than can be visualized. An adaptation of the approach visualizing set relations in graphs [XDC\*13] could be used to display further data facets. The dynamic access to texts within a given corpus is also an important feature for browsing purposes; Overview, a tool used by investigative journalists for text mining, supports a systematic search on hierarchically clustered contents based on similarity [BISM14]. This could also help humanities scholars improve the navigation of related text passages. Many visualization techniques for geospatial-temporal data can be adapted for textual humanities data. VisGets [DCCW08] is one such example, already mentioned as an inspiration when designing linked views within the Trading Consequences project [HACQ14]. Other methods focus on visualizing migration patterns [Guo09, BSV11] in an intuitive manner and could represent suitable techniques to visualize large travel-log collections. A Spark-Cloud [LRKC10], which visualizes trends in tag clouds, could be a beneficial visualization for the exploration of term usages in ancient text corpora. For example, generation, development and extinction of synonyms could all be tracked over time.

## 7. Future Challenges

The aforementioned techniques tailored for the digital humanities may solve some issues. Yet, there are still major challenges in the digital humanities where the visualization community can contribute valuable research.

**Novel techniques for close reading.** Various publications outline that close reading benefits from visualization, e.g., by highlighting crowdsourcing statistics [KG13, WMN\*14] or displaying information about textual features and structure [ARLC\*13, JGBS14a] alongside the source text. Although close reading is an essential task for humanities scholars, in most cases only simple visualization techniques, such as color coding textual entities, are provided. Few works attend to the matter of enhancing close reading in a beneficial manner. For example, the work on word scale visualizations is a promising technique [GWFI14] from which many humanities scholars may profit. But despite the pro-

posed annotations of individual words with statistics or of country names with polygons, the concept needs to be expanded to annotating other kinds of named entities. For example, providing supplementary information about (1) acting persons and their relationships, (2) artifacts mentioned in texts, or (3) occurring references could be interesting features for humanities scholars. Future work in visualization should include the development of design methods to meet such use cases, and studies that measure the benefit of glyph based approaches for close reading in comparison to using color or font size to express certain text features.

**Visualizing transpositions in parallel texts.** When observing similarities and differences among various editions of a text, one focus is to detect transpositions of textual entities. Such transpositions may occur on various text hierarchy levels, e.g., changed word order, modified argumentation structures, or even when exchanging whole paragraphs or sections. Although suitable methods exist for the first two hierarchy levels (words, sentences) [WJ13b, JGBS14a], there are no visualization techniques capable of coherently visualizing transpositions on all hierarchy levels by combining means of close and distant reading.

**Geospatial uncertainty.** Many visualizations deal with placenames extracted from literary texts to illustrate the geographical knowledge of a particular era. Here, various mapping issues arise [JW13]. Texts may contain placenames of varying granularity (e.g., country, region, city) or type (e.g., points for cities, polygons for areas, polylines for rivers) or even fictional placenames, which are hard to represent. Furthermore, placenames can themselves carry uncertainty of varying degrees, e.g., the exact locations of “Sparta” and “Atlantis” have yet to be discovered. Another form of uncertainty is defined by contextual information, e.g., expressions like “in London” and “close to London” cover various geospatial ranges. The development of a design space providing solutions to visualize these various types of geospatial uncertainty is one of the current primary challenges in digital humanities. Such a design space could be built upon the ideas of MacEachren for visualizing geospatial uncertainty [MRO\*12].

**Temporal uncertainty.** The visualization of temporal uncertainty is an equally important future task. Such uncertainties occur, for instance, when dating cultural heritage objects, such as historical manuscripts [JW13, BESL14]. Temporal metadata, in fact, can be provided in multifarious manners, e.g., 1450, before 1450, after 1450, around 1450, 15th century, first half of the the 15th century, etc. One can try to transform such temporal formats into machine-parsable time ranges, but the visualization of such uncertainties is a crucial issue as it comprises considerable risks of misinterpretation. Applying methods capable of visualizing temporal uncertainty as proposed by Slingsby [SDW11] can be a first step, but their utility for humanities applications needs to be investigated.

**Reconstructing workflows with visualization.** In two visualization papers, authors related situations where, during their conducted case studies, humanities scholars mentioned the importance of visualization features that emulate the scholar’s workflow. In [KJW\*14], users liked the display of digital copies as this builds trust in the visualization. When working with genealogy visualizations [BDF\*10], historians “insisted on redundant representation of gender ... that is consistent with their current practices.” Both situations illustrate the future challenge of inventing visualization techniques for digital humanities applications that the humanities scholar can easily adapt. An important task for the computer scientist is not only to incorporate a scholar’s workflow when designing the visualization, but to also communicate all aspects of data transformation, so that a scholar is able to generate trustworthy hypotheses. The importance of this issue is documented in [GO12].

**Usability studies.** Although the utility of most visualizations considered within our survey was illustrated by usage scenarios, we found little evidence about conducted usability studies to, for example, justify taken design decisions. The number of humanities scholars participating in such studies is potentially very small due to the multifarious research interests scholars may have on a large body of texts belonging to different eras and genres. Generating a user study format that caters for the interests of many different scholars is required to gain valuable insights into guidelines for designing visualizations for the digital humanities. When it comes to tool building, in fact, the digital humanities community poses interesting and complex challenges by virtue of its interdisciplinary nature. It embraces a wider range of disciplines, so the techniques it offers should address the larger scope. It also welcomes contrasting mindsets, methods and cultures, thus complicating communication. While sharing similar logical and analytical methods, computer scientists tend towards problem solving, humanities scholars towards knowledge acquisition and dissemination [Hen14]. No one community should operate in subservience to the other but together, complementing each others’ approaches. For these reasons and in this context, specialist terminology, assumptions and technical barriers should all be avoided. It is in this sense that tool usability should be understood not only as improved functionality or aesthetics but as a transparent guide to utility [GO12].

## 8. Conclusion

Computer scientists and humanities scholars seemingly do not have many things in common. Although they share some methodologies, they are geared towards different goals. But the digital age created a platform that brings people from two research areas together: the digital humanities.

During our survey, we had the opportunity to take a look at various fascinating digital humanities projects proposing visualization techniques supporting close and distant reading

of texts. In combination with papers providing visualizations for historical texts, we classified close and distant reading techniques and analyzed methods of combining both views to allow for multifaceted data analyses. In the process, we derived insights into a research area that requires the design of intuitive interfaces, but visualizations for textual data as part of the cultural heritage are rarely published in the visualization community. Therefore, we listed future challenges to support humanities scholars with close and distant reading. Developing solutions could provide beneficial contributions for both research fields. Furthermore, we outlined collaboration experiences reported by visualization researchers working in the field of digital humanities as a means of singling out the important ingredients for a successful project.

### Acknowledgments

The authors like to thank Judith Blumenstein for fruitful discussions on various digital humanities matters. This research was funded by the German Federal Ministry of Education and Research.

### References

- [ADG11] ALBERS D., DEWEY C., GLEICHER M.: Sequence Surveyor: Leveraging Overview for Scalable Genomic Alignment Visualization. *Visualization and Computer Graphics, IEEE Transactions on* 17, 12 (Dec 2011), 2392–2401. 14
- [AGL\*07] AUVIL L., GROIS E., LLORÀ X., PAPE G., GOREN V., SANDERS B., ACS B.: A Flexible System for Text Analysis with Semantic Networks. In *Proceedings of the Digital Humanities 2007* (2007). 8, 12
- [AKV\*14] ALEXANDER E., KOHLMANN J., VALENZA R., WITMORE M., GLEICHER M.: Serendip: Topic Model-Driven Visual Exploration of Text Corpora. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on* (Oct 2014), pp. 173–182. 5, 7, 10, 12
- [AL09] ATHENIKOS S. J., LIN X.: WikiPhiloSofia: Extraction and Visualization of Facts, Relations, and Networks Concerning Philosophers Using Wikipedia. In *Proceedings of the Digital Humanities 2009* (2009). 3
- [Arc15] Internet Archive, 2015. <http://www.archive.org> (Retrieved 2015-01-09). 1
- [ARLC\*13] ABDUL-RAHMAN A., LEIN J., COLES K., MAGUIRE E., MEYER M., WYNNE M., JOHNSON C. R., TREFETHEN A., CHEN M.: Rule-based Visual Mappings—with a Case Study on Poetry Visualization. In *Computer Graphics Forum* (2013), vol. 32, Wiley Online Library, pp. 381–390. 6, 11, 12, 13, 14
- [Arm14] ARMASELU F.: The Layered Text. From Textual Zoom, Text Network Analysis and Text Summarisation to a Layered Interpretation of Meaning. In *Proceedings of the Digital Humanities 2014* (2014). 8, 12
- [ARR\*12] ARAZY O., RUECKER S., RODRIGUEZ O., GIACOMETTI A., ZHANG L., CHUN S.: Mapping the Information Science Domain. In *Proceedings of the Digital Humanities 2012* (2012). 8, 12
- [BDF\*10] BEZERIANOS A., DRAGICEVIC P., FEKETE J., BAE J., WATSON B.: GeneaQuilts: A System for Exploring Large Genealogies. *Visualization and Computer Graphics, IEEE Transactions on* 16, 6 (Nov 2010), 1073–1081. 9, 12, 13, 15
- [Bea08] BEAVAN D.: Glimpses through the clouds: collocates in a new light. In *Proceedings of the Digital Humanities 2008* (2008). 7, 12
- [Bea11] BEAVAN D.: ComPair: Compare and Visualise the Usage of Language. In *Proceedings of the Digital Humanities 2011* (2011). 7, 12
- [Bea12] BEAVAN D.: DiaView: Visualise Cultural Change in Diachronic Corpora. In *Proceedings of the Digital Humanities 2012* (2012). 5, 7, 12
- [Bea14] BEALS M.: TEI for Close Reading: Can It Work for History?, 2014. <http://tinyurl.com/nvdnds> (Retrieved 2015-01-09). 3
- [Ben14] BENNER D. C.: "The Sounds of the Psalter: Computational Analysis of Soundplay". *Literary and Linguistic Computing* 29, 3 (2014), 361–378. 5, 7, 10, 11, 12
- [BESL14] BINDER F., ENTRUP B., SCHILLER I., LOBIN H.: Uncertain about Uncertainty: Different ways of processing fuzziness in digital humanities data. In *Proceedings of the Digital Humanities 2014* (2014). 15
- [BGHE10] BÜCHLER M., GESSNER A., HEYER G., ECKART T.: Detection of Citations and Textual Reuse on Ancient Greek Texts and its Applications in the Classical Studies: eAQUA Project. In *Proceedings of the Digital Humanities 2010* (2010). 5, 6, 12
- [BGHJ\*14] BÖGEL T., GOLD V., HAUTLI-JANISZ A., ROHRDANTZ C., SULGER S., BUTT M., HOLZINGER K., KEIM D. A.: Towards visualizing linguistic patterns of deliberation: a case study of the S21 arbitration. In *Proceedings of the Digital Humanities 2014* (2014). 5, 7, 10, 12
- [BH09] BOSTOCK M., HEER J.: Protovis: A graphical toolkit for visualization. *Visualization and Computer Graphics, IEEE Transactions on* 15, 6 (2009), 1121–1128. 13
- [BHJ\*09] BASTIAN M., HEYMANN S., JACOMY M., ET AL.: Gephi: an open source software for exploring and manipulating networks. *ICWSM 8* (2009), 361–362. 13
- [BHW11] BINGENHEIMER M., HUNG J.-J., WILES S.: Social network visualization from TEI data. *Literary and Linguistic Computing* 26, 3 (2011), 271–278. 4, 5, 9, 12
- [BISM14] BREHMER M., INGRAM S., STRAY J., MUNZNER T.: Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool for Investigative Journalists. *Visualization and Computer Graphics, IEEE Transactions on* 20, 12 (Dec 2014), 2271–2280. 14
- [BJ14] BINDER J. M., JENNINGS C.: Visibility and meaning in topic models and 18th-century subject indexes. *Literary and Linguistic Computing* 29, 3 (2014), 405–411. 4, 5, 7, 10, 11, 12
- [BOH11] BOSTOCK M., OGIEVETSKY V., HEER J.: D3 – Data-Driven Documents. *Visualization and Computer Graphics, IEEE Transactions on* 17, 12 (Dec 2011), 2301–2309. 13
- [Boo13] BOOTH A.: Documentary Social Networks: Collective Biographies of Women. In *Proceedings of the Digital Humanities 2013* (2013). 4, 5, 9, 12
- [Boy13] BOYLES N.: Closing in on Close Reading. *Educational Leadership* 70, 4 (2013), 36–41. 2
- [BPBI10] BARKER E., PELLING C., BOUZAROVSKI S., ISAKSEN L.: Mapping the World of an Ancient Greek Historian: The HESTIA Project. In *Proceedings of the Digital Humanities 2010* (2010). 4, 8, 10, 12
- [Bra12] BRADLEY A. J.: Violence and the Digital Humanities Text as Pharmakon. In *Proceedings of the Digital Humanities 2012* (2012). 3

- [BSV11] BUCHIN K., SPECKMANN B., VERBEEK K.: Flow Map Layout via Spiral Trees. *Visualization and Computer Graphics, IEEE Transactions on* 17, 12 (Dec 2011), 2536–2544. 14
- [BW08] BYRON L., WATTENBERG M.: Stacked Graphs – Geometry & Aesthetics. *Visualization and Computer Graphics, IEEE Transactions on* 14, 6 (Nov 2008), 1245–1252. 8
- [Cay05] CAYLESS H.: DocScapes: Visualizing Document Structures with SVG. In *Proceedings of the Digital Humanities 2005* (2005). 5, 7, 10, 12
- [CCP09] COLLINS C., CARPENDALE S., PENN G.: Docuburst: Visualizing document content using language structure. In *Computer Graphics Forum* (2009), vol. 28, Wiley Online Library, pp. 1039–1046. 14
- [CDP\*07] CLEMENT T., DON A., PLAISANT C., AUVIL L., PAPE G., GOREN V.: 'Something that is interesting is interesting them': Using Text Mining and Visualizations to Aid Interpreting Repetition in Gertrude Stein's The Making of Americans. In *Proceedings of the Digital Humanities 2007* (2007). 4, 7, 10, 12, 14
- [CEJ\*14] CRAIG H., EDER M., JANNIDIS F., KESTEMONT M., RYBICKI J., SCHÖCH C.: Validating Computational Stylistics in Literary Interpretation. In *Proceedings of the Digital Humanities 2014* (2014). 8, 12
- [CGM\*12] CHATURVEDI M., GANNOD G., MANDELL L., ARMSTRONG H., HODGSON E.: Myopia: A Visualization Tool in Support of Close Reading. In *Proceedings of the Digital Humanities 2012* (2012). 5, 6, 12
- [CL13] COLES K., LEIN J. G.: Solitary Mind, Collaborative Mind: Close Reading and Interdisciplinary Research. In *Proceedings of the Digital Humanities 2013* (2013). 3
- [CLT\*11] CUI W., LIU S., TAN L., SHI C., SONG Y., GAO Z., QU H., TONG X.: TextFlow: Towards Better Understanding of Evolving Topics in Text. *Visualization and Computer Graphics, IEEE Transactions on* 17, 12 (Dec 2011), 2412–2421. 7, 8, 12
- [CLWW14] CUI W., LIU S., WU Z., WEI H.: How Hierarchical Topics Evolve in Large Text Corpora. *Visualization and Computer Graphics, IEEE Transactions on* 20, 12 (Dec 2014), 2281–2290. 7, 8, 12
- [CMLM14] COLES K., MEYER M., LEIN J. G., MCCURDY N.: Empowering Play, Experimenting with Poems: Disciplinary Values and Visualization Development. In *Proceedings of the Digital Humanities 2014* (2014). 5, 6, 12
- [CMS99] CARD S. K., MACKINLAY J. D., SHNEIDERMAN B.: *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999. 3
- [Cob05] COBURN A.: Text Modeling and Visualization with Network Graphs. In *Proceedings of the Digital Humanities 2005* (2005). 9, 12
- [Cor13] CORDELL R.: "Taken Possession of": The Reprinting and Reauthorship of Hawthorne's "Celestial Railroad" in the Antebellum Religious Press. *Digital Humanities Quarterly* 7, 1 (2013). 12
- [CPV09] CLEMENT T., PLAISANT C., VUILLEMOT R.: The Story of One: Humanity scholarship with visualization and text analysis. In *Proceedings of the Digital Humanities 2009* (2009). 13
- [CRS\*14] CHRISTIE A., ROSS S., SAYERS J., TANIGAWA K., TEAM I.-M. R.: Z-Axis Scholarship: Modeling How Modernists Write the City. In *Proceedings of the Digital Humanities 2014* (2014). 3
- [CSV08] CIULA A., SPENCE P., VIEIRA J. M.: Expressing complex associations in medieval historical documents: the Henry III Fine Rolls Project. *Literary and Linguistic Computing* 23, 3 (2008), 311–325. 9, 12
- [CTA\*13] CLEMENT T., TCHENG D., AUVIL L., CAPITANU B., BARBOSA J.: Distant Listening to Gertrude Stein's 'Melanctha': Using Similarity Analysis in a Discovery Paradigm to Analyze Prosody and Author Influence. *Literary and Linguistic Computing* 28, 4 (2013), 582–602. 5, 7, 10, 11, 12, 14
- [CVW09] COLLINS C., VIEGAS F., WATTENBERG M.: Parallel Tag Clouds to explore and analyze faceted text corpora. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on* (Oct 2009), pp. 91–98. 14
- [CWG11] CORRELL M., WITMORE M., GLEICHER M.: Exploring collections of tagged text for literary scholarship. *Computer Graphics Forum* 30, 3 (2011), 731–740. 5, 6, 7, 10, 12, 13
- [DCCW08] DÖRK M., CARPENDALE S., COLLINS C., WILLIAMSON C.: VisGets: Coordinated Visualizations for Web-based Information Exploration and Discovery. *Visualization and Computer Graphics, IEEE Transactions on* 14, 6 (Nov 2008), 1205–1212. 14
- [DFM\*08] DYENS O., FOREST D., MONDOU P., COOLS V., JOHNSTON D.: Information visualization and text mining: application to a corpus on posthumanism. In *Proceedings of the Digital Humanities 2008* (2008). 5, 7, 12
- [DNM14] DYE D. J., NAPOLIN J. B., CORNELL E., MARTIN W.: Digital Yoknapatawpha: Interpreting a Palimpsest of Place. In *Proceedings of the Digital Humanities 2014* (2014). 4, 8, 12
- [DUY\*05] DOWNIE J. S., UNSWORTH J., YU B., TCHENG D., ROCKWELL G., RAMSAY S. J.: A Revolutionary Approach to Humanities Computing?: Tools Development and the D2K Data-Mining Framework. In *Proceedings of the Digital Humanities 2005* (2005). 13
- [DWS\*12] DOU W., WANG X., SKAU D., RIBARSKY W., ZHOU M.: LeadLine: Interactive visual analysis of text data through event identification and exploration. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on* (Oct 2012), pp. 93–102. 5, 7, 8, 10, 12
- [DZG\*07] DON A., ZHELEVA E., GREGORY M., TARKAN S., AUVIL L., CLEMENT T., SHNEIDERMAN B., PLAISANT C.: Discovering interesting usage patterns in text collections: integrating text mining with visualization. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management* (2007), ACM, pp. 213–222. 13
- [Ede14] EDER M.: Stylometry, network analysis, and Latin literature. In *Proceedings of the Digital Humanities 2014* (2014). 3, 4, 5, 8, 9, 12
- [EJ14] EVANS C., JASNOW B.: Mapping Homer's Catalogue of Ships. *Literary and Linguistic Computing* 29, 3 (2014), 317–325. 5, 7, 12
- [eMa15] eMargin, 2015. <http://eMargin.bcu.ac.uk/> (Retrieved 2015-01-09). 2
- [ESK14] EISENSTEIN J., SUN I., KLEIN L. F.: Exploratory Thematic Analysis for Historical Newspaper Archives. In *Proceedings of the Digital Humanities 2014* (2014). 7, 8, 10, 12
- [EX10] ESTEVA M., XU W.: Finding Stories in the Archive through Paragraph Alignment. In *Proceedings of the Digital Humanities 2010* (2010). 5, 8, 12
- [Fek04] FEKETE J.: The infovis toolkit. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on* (2004), IEEE, pp. 167–174. 13

- [Fin10] FINN E.: The Social Lives of Books: Mapping the Ideational Networks of Toni Morrison. In *Proceedings of the Digital Humanities 2010* (2010). 3
- [FKT14] FANKHAUSER P., KERMES H., TEICH E.: Combining Macro- and Microanalysis for Exploring the Construal of Scientific Disciplinarity. In *Proceedings of the Digital Humanities 2014* (2014). 7, 10, 12
- [FS11] FORSTALL C., SCHEIRER W. J.: Visualizing Sound as Functional N-Grams in Homeric Greek Poetry. In *Proceedings of the Digital Humanities 2011* (2011). 7, 12
- [Gal11] GALEY A.: Approaching the Coasts of Utopia: Visualization Strategies for Mapping Early Modern Paratexts. In *Proceedings of the Digital Humanities 2011* (2011). 8, 12
- [GCL\*13] GENG Z., CHEESMAN T., LARAMEE R. S., FLANAGAN K., THIEL S.: ShakerVis: Visual analysis of segment variation of German translations of Shakespeare's Othello. *Information Visualization* (2013). 4, 7, 10, 11, 12, 13
- [GDMF\*14] GREGORY I., DONALDSON C., MURRIETA-FLORES P., RUPP C., BARON A., HARDIE A., RAYSON P.: Digital approaches to understanding the geographies in literary and historical texts. In *Proceedings of the Digital Humanities 2014* (2014). 7, 8, 12
- [Geo15] GeoNames Gazetteer, 2015. <http://www.geonames.org/> (Retrieved 2015-01-10). 7
- [GH11a] GOODWIN J., HOLBO J.: *Reading graphs, maps, trees: responses to Franco Moretti*. Parlor Press, Anderson, SC, 2011. Book, Whole. 3
- [GH11b] GREGORY I. N., HARDIE A.: Visual GISting: bringing together corpus linguistics and Geographical Information Systems. *Literary and Linguistic Computing* 26, 3 (2011), 297–314. 7, 12
- [GO12] GIBBS F., OWENS T.: Building Better Digital Humanities Tools: Toward broader audiences and user-centered designs. *Digital Humanities Quarterly* 6, 2 (2012). 15
- [Goo15] Google Books, 2015. <https://books.google.com/> (Retrieved 2015-01-09). 1, 2
- [Guo09] GUO D.: Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data. *Visualization and Computer Graphics, IEEE Transactions on* 15, 6 (Nov 2009), 1041–1048. 14
- [GWF14] GOFFIN P., WILLETT W., FEKETE J.-D., ISENBERG P.: Exploring the Placement and Design of Word-Scale Visualizations. *Visualization and Computer Graphics, IEEE Transactions on* 20, 12 (Dec 2014), 2291–2300. 6, 12, 14
- [GZ12] GEDZELMAN S., ZANCARINI J.-C.: HyperMachiavel: a translation comparison tool. In *Proceedings of the Digital Humanities 2012* (2012). 8, 10, 12
- [HACQ14] HINRICHS U., ALEX B., CLIFFORD J., QUIGLEY A.: Trading Consequences: A Case Study of Combining Text Mining & Visualisation to Facilitate Document Exploration. In *Proceedings of the Digital Humanities 2014* (2014). 5, 7, 8, 10, 12, 13, 14
- [Haw00] HAWTHORN J.: *A glossary of contemporary literary theory*. Oxford University Press, 2000. 2
- [HB05] HEER J., BOYD D.: Vizster: visualizing online social networks. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on* (Oct 2005), pp. 32–39. 14
- [HCC14] HSIANG J., CHEN L., CHUNG C.-H.: A glimpse of the change of worldview between 7th and 10th century China through two leishu. In *Proceedings of the Digital Humanities 2014* (2014). 10, 12
- [HCL05] HEER J., CARD S. K., LANDAY J. A.: Prefuse: a toolkit for interactive information visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (2005), ACM, pp. 421–430. 13
- [Hen14] HENSELER C.: Minecraft Anyone? Encouraging A New Generation of Computer Scientists and Humanists, 2014. [http://www.huffingtonpost.com/christine-henseler/minecraft-anyone-encourag\\_b\\_5629708.html](http://www.huffingtonpost.com/christine-henseler/minecraft-anyone-encourag_b_5629708.html) (Retrieved 2015-01-09). 15
- [HKTK14] HOWELL S., KELLEHER M., TEEHAN A., KEATING J.: A Digital Humanities Approach to Narrative Voice in The Secret Scripture: Proposing a New Research Method. *Digital Humanities Quarterly* 8, 2 (2014). 6, 12, 13
- [Hoc04] HOCKEY S.: The history of humanities computing. *A companion to digital humanities* (2004), 3–19. 1
- [HPR14] HOYT E., PONTO K., ROY C.: Visualizing and Analyzing the Hollywood Screenplay with ScripThreads. *Digital Humanities Quarterly* 8, 4 (2014). 8, 10, 12
- [HSC08] HINRICHS U., SCHMIDT H., CARPENDALE S.: EMDialog: Bringing Information Visualization into the Museum. *Visualization and Computer Graphics, IEEE Transactions on* 14, 6 (Nov 2008), 1181–1188. 8, 10, 12, 13
- [Jas01] JASINSKI J.: *Rhetoric and Society: Sourcebook on Rhetoric: Key Concepts in Contemporary Rhetorical Studies*, vol. 4. Sage Publications, 2001. 2
- [JGBS14a] JÄNICKE S., GESSNER A., BÜCHLER M., SCHEUERMANN G.: 5 Design Rules for Visualizing Text Variant Graphs. In *Proceedings of the Digital Humanities 2014* (2014). 5, 6, 12, 13, 14, 15
- [JGBS14b] JÄNICKE S., GESSNER A., BÜCHLER M., SCHEUERMANN G.: Visualizations for Text Re-use. *GRAPP/IVAPP* (2014), 59–70. 4, 5, 6, 7, 10, 12, 13
- [JHS13] JÄNICKE S., HEINE C., SCHEUERMANN G.: GeoTemCo: Comparative Visualization of Geospatial-Temporal Data with Clutter Removal Based on Dynamic Delaunay Triangulations. In *Computer Vision, Imaging and Computer Graphics. Theory and Application*. Springer, 2013, pp. 160–175. 13
- [JHSS12] JÄNICKE S., HEINE C., STOCKMANN R., SCHEUERMANN G.: Comparative Visualization of Geospatial-temporal Data. In *GRAPP/IVAPP* (2012), pp. 613–625. 5, 8, 12
- [Joc12] JOCKERS M.: Computing and Visualizing the 19th-Century Literary Genome. In *Proceedings of the Digital Humanities 2012* (2012). 5, 8, 12
- [Joc13] JOCKERS M. L.: *Macroanalysis: Digital Methods & Literary History*. University of Illinois Press, 2013. 2
- [JRS\*09] JONG C.-H., RAJKUMAR P., SIDDIQUIE B., CLEMENT T., PLAISANT C., SHNEIDERMAN B.: Interactive Exploration of Versions across Multiple Documents. In *Proceedings of the Digital Humanities 2009* (2009). 5, 7, 11, 12, 13
- [JW13] JÄNICKE S., WRISLEY D. J.: Visualizing Uncertainty: How to Use the Fuzzy Data of 550 Medieval Texts? In *Proceedings of the Digital Humanities 2013* (2013). 4, 5, 7, 8, 12, 15
- [KBK11] KRSTAJIC M., BERTINI E., KEIM D.: CloudLines: Compact Display of Event Episodes in Multiple Time-Series. *Visualization and Computer Graphics, IEEE Transactions on* 17, 12 (Dec 2011), 2432–2439. 8, 12
- [KG13] KEHOE A., GEE M.: eMargin: A Collaborative Textual Annotation Tool. *Ariadne* 71 (July 2013). 2, 14

- [KJW\*14] KOCH S., JOHN M., WORNER M., MULLER A., ERTL T.: VarifocalReader – In-Depth Visual Analysis of Large Text Documents. *Visualization and Computer Graphics, IEEE Transactions on* 20, 12 (Dec 2014), 1723–1732. 5, 7, 11, 12, 13, 15
- [KKL\*11] KIM H., KANG B.-M., LEE D.-G., CHUNG E., KIM I.: Trends 21 Corpus: A Large Annotated Korean Newspaper Corpus for Linguistic and Cultural Studies. In *Proceedings of the Digital Humanities 2011* (2011). 4, 5, 8, 12
- [KLB14] KOCHTCHI A., LANDESBERGER T. V., BIEMANN C.: Networks of Names: Visual Exploration and Semi-Automatic Tagging of Social Networks from Newspaper Articles. *Computer Graphics Forum* 33, 3 (2014), 211–220. 9, 11, 12
- [Kle12] KLEIN L. F.: Social Network Analysis and Visualization in 'The Papers of Thomas Jefferson'. In *Proceedings of the Digital Humanities 2012* (2012). 4, 9, 12
- [KO07] KEIM D., OELKE D.: Literature Fingerprinting: A New Method for Visual Literary Analysis. In *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on* (Oct 2007), pp. 115–122. 7, 12
- [KOTM13] KIMURA F., OSAKI T., TEZUKA T., MAEDA A.: Visualization of relationships among historical persons from Japanese historical documents. *Literary and Linguistic Computing* 28, 2 (2013), 271–278. 5, 9, 12
- [KZ14] KRAUSE T., ZELDES A.: ANNIS3: A new architecture for generic corpus query and visualization. *Literary and Linguistic Computing* (2014). 6, 12
- [LRKC10] LEE B., RICHE N., KARLSON A., CARPENDALE S.: SparkClouds: Visualizing Trends in Tag Clouds. *Visualization and Computer Graphics, IEEE Transactions on* 16, 6 (Nov 2010), 1182–1189. 14
- [LWW\*13] LIU S., WU Y., WEI E., LIU M., LIU Y.: StoryFlow: Tracking the Evolution of Stories. *Visualization and Computer Graphics, IEEE Transactions on* 19, 12 (Dec 2013), 2436–2445. 8, 12
- [Mar12] MARCHE S.: Literature is not Data: Against Digital Humanities, 2012. <http://www.lareviewofbooks.org/article.php?id=1040> (Retrieved 2015-01-09). 3
- [MBL\*06] MEHLER A., BAO Y., LI X., WANG Y., SKIENA S.: Spatial Analysis of News Sources. *Visualization and Computer Graphics, IEEE Transactions on* 12, 5 (Sept 2006), 765–772. 8, 12
- [MFM08] MENESES L., FURUTA R., MALLIN E.: Exploring the Biography and Artworks of Picasso with Interactive Calendars and Timelines. In *Proceedings of the Digital Humanities 2008* (2008). 3
- [MFM13] MENESES L., FURUTA R., MANDELL L.: Ambiances: A Framework to Write and Visualize Poetry. In *Proceedings of the Digital Humanities 2013* (2013). 9, 10, 12
- [MH13] MURALIDHARAN A., HEARST M. A.: Supporting exploratory text analysis in literature study. *Literary and Linguistic Computing* 28, 2 (2013), 283–295. 7, 9, 11, 12
- [MLSU13] MILLER B., LI F., SHRESTHA A., UMATHY K.: Digging into Human Rights Violations: phrase mining and trigram visualization. In *Proceedings of the Digital Humanities 2013* (2013). 9, 12
- [MMP09] MEYER M., MUNZNER T., PFISTER H.: MizBee: A Multiscale Synteny Browser. *Visualization and Computer Graphics, IEEE Transactions on* 15, 6 (Nov 2009), 897–904. 14
- [Mor05] MORETTI F.: *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso, July 2005. 1, 2, 3
- [Mor13] MORETTI F.: *Distant reading*. Verso, 2013. 2
- [MRO\*12] MACEACHREN A., ROTH R., O'BRIEN J., LI B., SWINGLEY D., GAHEGAN M.: Visual Semiotics & Uncertainty Visualization: An Empirical Study. *Visualization and Computer Graphics, IEEE Transactions on* 18, 12 (Dec 2012), 2496–2505. 15
- [Mun09] MUNZNER T.: A nested model for visualization design and validation. *Visualization and Computer Graphics, IEEE Transactions on* 15, 6 (2009), 921–928. 11
- [Mur11] MURALIDHARAN A.: A Visual Interface for Exploring Language Use in Slave Narratives. In *Proceedings of the Digital Humanities 2011* (2011). 5, 7, 11, 12
- [NMG\*13] NOWVSKIE B., MCCLURE D., GRAHAM W., SOROKA A., BOGGS J., ROCHESTER E.: Geo-Temporal Interpretation of Archival Collections with Neatline. *Literary and Linguistic Computing* 28, 4 (2013), 692–699. 13
- [OKK13] OELKE D., KOKKINAKIS D., KEIM D. A.: Fingerprint Matrices: Uncovering the dynamics of social networks in prose literature. In *Computer Graphics Forum* (2013), vol. 32, Wiley Online Library, pp. 371–380. 7, 12
- [Ome15] Omeke, 2015. <http://www.omeka.org/> (Retrieved 2015-01-10). 13
- [ÓML14] Ó MURCHÚ T., LAWLESS S.: The Problem of Time and Space: The Difficulties in Visualising Spatiotemporal Change in Historical Data. In *Proceedings of the Digital Humanities 2014* (2014). 7, 8, 12
- [Ope09] OPENBIBLE.INFO: Phrase Net Bible Visualizations, 2009. <http://www.openbible.info/blog/2009/03/phrase-net-bible-visualizations/> (Retrieved 2015-01-09). 9
- [OST\*10] OESTERLING P., SCHEUERMANN G., TERESNIAK S., HEYER G., KOCH S., ERTL T., WEBER G.: Two-stage framework for a topology-based projection and visualization of classified document collections. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on* (Oct 2010), pp. 91–98. 4, 10, 12
- [Pal02] PALEY W. B.: TextArc: Showing word frequency and distribution in text. In *Poster presented at IEEE Symposium on Information Visualization* (2002), vol. 2002. 10, 13
- [PBD14] PEÑA E., BROWN M., DOBSON T.: On Metaphor in Text Visualization Prototypes. In *Proceedings of the Digital Humanities 2014* (2014). 10, 12
- [Per15] Perseus Digital Library, 2015. <https://www.perseus.tufts.edu/> (Retrieved 2015-01-09). 1
- [Pet14] PETERSON N.: Visualization As a Bridge to Close Reading: The Audience in The Castle of Perseverance. In *Proceedings of the Digital Humanities 2014* (2014). 4, 9, 12
- [PHI15] PHI Latin Texts, 2015. <http://latin.packhum.org/> (Retrieved 2015-01-09). 1
- [Pie05] PIEZ W.: SVG Visualization of TEI Texts. In *Proceedings of the Digital Humanities 2005* (2005). 7, 12
- [Pie10] PIEZ W.: Towards Hermeneutic Markup: An architectural outline. In *Proceedings of the Digital Humanities 2010* (2010). 5, 6, 12
- [Pie13] PIEZ W.: Markup Beyond XML. In *Proceedings of the Digital Humanities 2013* (2013). 6, 12
- [Ple15] Pleiades Gazetteer, 2015. <http://pleiades.stoa.org/> (Retrieved 2015-01-10). 7
- [Pos07] POSAVEC S.: Literary Organism, 2007. <http://www.stefanieposavec.co.uk/> (Retrieved 2015-01-09). 3

- [Pro15] Project Gutenberg, 2015. <http://www.gutenberg.org/> (Retrieved 2015-01-09). 2
- [PS06] PERER A., SHNEIDERMAN B.: Balancing Systematic and Flexible Exploration of Social Networks. *Visualization and Computer Graphics, IEEE Transactions on* 12, 5 (Sept 2006), 693–700. 14
- [PSA\*06] PLAISANT C., SMITH M. N., AUVEL L., ROSE J., YU B., CLEMENT T.: "Undiscovered Public Knowledge": Mining for Patterns of Erotic Language in Emily Dickinson's Correspondence with Susan Huntington (Gilbert) Dickinson. In *Proceedings of the Digital Humanities 2006* (2006). 5, 12
- [RD10] RICHE N., DWYER T.: Untangling Euler Diagrams. *Visualization and Computer Graphics, IEEE Transactions on* 16, 6 (Nov 2010), 1090–1099. 9, 12
- [RFH14] REITER N., FRANK A., HELLWIG O.: An NLP-based cross-document approach to narrative structure discovery. *Literary and Linguistic Computing* 29, 4 (2014), 583–605. 8, 11, 12
- [RGP\*12] RIEHMANN P., GRUENDL H., POTTHAST M., TRENMANN M., STEIN B., FROELICH B.: WORDGRAPH: Keyword-in-Context Visualization for NETSPEAK's Wildcard Search. *Visualization and Computer Graphics, IEEE Transactions on* 18, 9 (Sept 2012), 1411–1423. 14
- [RRRG05] RUECKER S., RAMSAY S., RADZIKOWSKA M., GALEY A.: Interface Design. In *Proceedings of the Digital Humanities 2005* (2005). 7, 8, 11, 12
- [RSDCD\*13] ROBERTS-SMITH J., DESOUZA-COELHO S., DOBSON T. M., GABRIELE S., RODRIGUEZ-ARENAS O., RUECKER S., SINCLAIR S., AKONG A., BOUCHARD M., HONG M., JAKACKI D., LAM D., KOVACS A., NORTHAM L., SO D.: Visualizing Theatrical Text: From Watching the Script to the Simulated Environment for Theatre (SET). *Digital Humanities Quarterly* 7, 3 (2013). 10, 12
- [SDW11] SLINGSBY A., DYKES J., WOOD J.: Exploring Uncertainty in Geodemographics with Interactive Graphics. *Visualization and Computer Graphics, IEEE Transactions on* 17, 12 (Dec 2011), 2545–2554. 15
- [Shn96] SHNEIDERMAN B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Visual Languages, Proceedings* (1996), pp. 336–343. 3
- [Sin13] SINGER K.: Digital Close Reading: TEI for Teaching Poetic Vocabularies. *Journal of Interactive Technology and Pedagogy* 3 (2013). 7
- [SOI10] SAITO S., OHNO S., INABA M.: A Platform for Cultural Information Visualization Using Schematic Expressions of Cube. In *Proceedings of the Digital Humanities 2010* (2010). 1
- [Sol13] SOLOMON D. R.: Theorizing Data Visualization: A Comparative Case-Study Approach. In *Proceedings of the Digital Humanities 2013* (2013). 14
- [TEI15] Text Encoding Initiative: TEI, 2015. <https://www.tei-c.org/> (Retrieved 2015-01-09). 1
- [Tót13] TÓTH G. M.: The computer-assisted analysis of a medieval commonplace book and diary (MS Zibaldone Quaresimale by Giovanni Rucellai). *Literary and Linguistic Computing* 28, 3 (2013), 432–443. 9, 12
- [Tra09] TRAVIS C.: Patrick Kavanagh's Poetic Wordscapes: GIS, Literature and Ireland, 1922-1949. In *Proceedings of the Digital Humanities 2009* (2009). 7, 12
- [VCPK09] VUILLEMOT R., CLEMENT T., PLAISANT C., KUMAR A.: What's being said near "Martha"? Exploring name entities in literary text collections. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on* (Oct 2009), pp. 107–114. 7, 10, 12, 13
- [vHWV09] VAN HAM F., WATTENBERG M., VIEGAS F.: Mapping Text with Phrase Nets. *Visualization and Computer Graphics, IEEE Transactions on* 15, 6 (Nov 2009), 1169–1176. 9, 12
- [VWF09] VIEGAS F., WATTENBERG M., FEINBERG J.: Participatory Visualization with Wordle. *Visualization and Computer Graphics, IEEE Transactions on* 15, 6 (Nov 2009), 1137–1144. 6
- [VWvH\*07] VIEGAS F., WATTENBERG M., VAN HAM F., KRISS J., MCKEON M.: ManyEyes: a Site for Visualization at Internet Scale. *Visualization and Computer Graphics, IEEE Transactions on* 13, 6 (Nov 2007), 1121–1128. 13
- [WBWK00] WANG BALDONADO M. Q., WOODRUFF A., KUCHINSKY A.: Guidelines for using multiple views in information visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (New York, NY, USA, 2000), AVI '00, ACM, pp. 110–119. 10
- [Wea08] WEAVER C.: Multidimensional visual analysis using cross-filtered views. In *Visual Analytics Science and Technology, 2008. VAST '08. IEEE Symposium on* (Oct 2008), pp. 163–170. 4, 7, 8, 12
- [WH11] WALSH J. A., HOOPER W.: Computational Discovery and Visualization of the Underlying Semantic Structure of Complicated Historical and Literary Corpora. In *Proceedings of the Digital Humanities 2011* (2011). 8, 10, 11, 12
- [WJ13a] WEINGART S., JORGENSEN J.: Computational analysis of the body in European fairy tales. *Literary and Linguistic Computing* 28, 3 (2013), 404–416. 4, 8, 10, 11, 12
- [WJ13b] WHEELS D., JENSEN K.: Juxta Commons. In *Proceedings of the Digital Humanities 2013* (2013). 5, 6, 12, 15
- [WMN\*14] WALSH B., MAIERS C., NALLY G., BOGGS J., TEAM P. P.: Crowdsourcing individual interpretations: Between microtasking and macrotasking. *Literary and Linguistic Computing* 29, 3 (2014), 379–386. 5, 6, 12, 14
- [Wol13] WOLFF M.: Surveying a Corpus with Alignment Visualization and Topic Modeling. In *Proceedings of the Digital Humanities 2013* (2013). 9, 10, 12
- [WV08] WATTENBERG M., VIEGAS F.: The Word Tree, an Interactive Visual Concordance. *Visualization and Computer Graphics, IEEE Transactions on* 14, 6 (Nov 2008), 1221–1228. 9, 11, 12, 14
- [XDC\*13] XU P., DU F., CAO N., SHI C., ZHOU H., QU H.: Visual analysis of set relations in a graph. In *Computer Graphics Forum* (2013), vol. 32, Wiley Online Library, pp. 61–70. 14
- [YMSJ05] YI J. S., MELTON R., STASKO J., JACKO J. A.: Dust & Magnet: multivariate information visualization using a magnet metaphor. *Information Visualization* 4, 4 (2005), 239–256. 10
- [ZBDS12] ZINSMAIER M., BRANDES U., DEUSSEN O., STROBELT H.: Interactive Level-of-Detail Rendering of Large Graphs. *Visualization and Computer Graphics, IEEE Transactions on* 18, 12 (Dec 2012), 2486–2495. 14
- [ZCCB12] ZHAO J., CHEVALIER F., COLLINS C., BALAKRISHNAN R.: Facilitating Discourse Analysis with Interactive Visualization. *Visualization and Computer Graphics, IEEE Transactions on* 18, 12 (Dec 2012), 2639–2648. 14

## Biographies

**Stefan Jänicke** graduated in Computer Science at Leipzig University, Germany, in 2009. He is currently working as a research assistant in the Image and Signal Processing Group at Leipzig University. Over the last years, he has gained experience in developing information visualization techniques within the following digital humanities projects: europeana-connect, HeyneDigital (both with SUB Göttingen), Deutsche Digitale Bibliothek (with Fraunhofer IAIS), eAQUA and eTRACES (both with Leipzig University). In the past two years he has had the opportunity to present his work at the annual digital humanities conference. He is currently involved in the digital humanities project eXChange, whose aim is to discover concept change in ancient corpora. His PhD topic investigates the utility of visualization techniques in the digital humanities.

**Greta Franzini** completed her Classics BA and Digital Humanities MA degrees at King's College London. Currently, she is completing her PhD at the UCL Centre for Digital Humanities and works as a researcher at the Göttingen Centre for Digital Humanities. Here, she is involved in research pertaining to the fields of text re-use and visualization, classics, manuscript studies and natural language processing. Previously, she worked for the Alexander von Humboldt Chair of Digital Humanities at Leipzig University.

**Muhammad Faisal Cheema** is a PhD candidate working in the Image and Signal Processing Group at the Department of Computer Science at Leipzig University, Germany. He is carrying out research on information visualization techniques for the digital humanities. He is currently involved in the digital humanities project eXChange hosted at Leipzig University. His research interests include visualization of textual and hierarchical data.

**Gerik Scheuermann** received his BS and MS degrees in mathematics in 1995 from the University of Kaiserslautern. In 1999, he received a PhD degree in computer science from the University of Kaiserslautern. During 1995-1997, he conducted research at Arizona State University. In 1999-2000 he worked as postdoctoral researcher at the Center for Image Processing and Integrated Computing (CIPIC) at the University of California, Davis. Between 2001 and 2004, he was an assistant professor for computer science at the University of Kaiserslautern. Today, he is a professor in the Computer Science Department of Leipzig University. His research topics include algebraic geometry, topology, Clifford algebra, image processing, graphics, visual analytics, information and scientific visualization, and application of visualization to the fields of medicine, fluid dynamics, text analysis and digital humanities. He is a member of the ACM, the IEEE and of the GI.