Visual Text Analysis in Digital Humanities

S. Jänicke[1], G. Franzini[2], M. F. Cheema[1] and G. Scheuermann[1]

[1]Image and Signal Processing Group, Department of Computer Science, Leipzig University, Germany
{stjaenicke, faisal, scheuermann}@informatik.uni-leipzig.de
[2]Göttingen Centre for Digital Humanities, University of Göttingen, Germany
gfranzini@gcdh.de

Abstract
In 2005, Franco Moretti introduced Distant Reading to analyse entire literary text collections. This was a rather revolutionary idea compared to the traditional Close Reading, which focuses on the thorough interpretation of an individual work. Both reading techniques are the prior means of Visual Text Analysis. We present an overview of the research conducted since 2005 on supporting text analysis tasks with close and distant reading visualizations in the digital humanities. Therefore, we classify the observed papers according to a taxonomy of text analysis tasks, categorize applied close and distant reading techniques to support the investigation of these tasks and illustrate approaches that combine both reading techniques in order to provide a multi-faceted view of the textual data. In addition, we take a look at the used text sources and at the typical data transformation steps required for the proposed visualizations. Finally, we summarize collaboration experiences when developing visualizations for close and distant reading, and we give an outlook on future challenges in that research area.

Keywords: digital humanities, survey, visual text analysis, close reading, distant reading

ACM CCS: H.5.2 [Information Interfaces and Presentation]: User Interfaces–Evaluation/methodology

1. Introduction

Traditionally, humanities scholars carrying out research on a specific or on multiple literary work(s) are interested in the analysis of related texts or text passages. But the digital age has opened possibilities for scholars to enhance their traditional workflows. Enabled by digitization projects, humanities scholars can nowadays reach a large number of digitized texts through web portals such as Google Books [Goo15] and Internet Archive [Arc15]. Digital editions exist also for ancient texts; notable 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. Some of these questions can be answered with existent algorithms and tools provided by the computer science domain, but for other humanities questions scholars need to formulate 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) for texts [TEI15], and to aggregate, digitize and deliver data. In the last years, visualization techniques have gained more and more importance when it comes to analysing data. For example, Saito [SOI10] introduced her 2010 digital humanities conference paper 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'.

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 them. In other words, to visualize them.

This survey observes text analysis tasks of humanities scholars— for example, literary scholars, historians and philologists—and the

visualization techniques that have been developed in order to support these tasks. By providing a text analysis task taxonomy, categorizing applied close and distant reading techniques and outlining strategies that combine close and distant reading visualizations, we present an overview suitable for visualization scholars facing related digital humanities text analysis tasks. We further investigate the following questions:

- What are the used text sources and which data transformations are applied in order to investigate text analysis research questions with close and distant reading visualizations?
- Which experiences are reported regarding collaborations between visualization experts and humanities scholars?
- What are future challenges for visualization scholars concerning visual text analysis to further improve the support for humanities scholars?

1.1. Relation to the previous article

The focus of the previous version of this survey [JFCS15] was to illustrate the diversity of applied visualization techniques that support the close and distant reading of texts in digital humanities applications—enriched with used visualization tools, collaboration experiences of visualization researchers working together with humanities scholars and future challenges.

This survey extension aims to give visualization scholars new to the field of digital humanities an adequate overview of related works in order to support carrying out successful digital humanities projects. As an application domain for information visualization, these projects gain their motivation from humanities research questions on texts. Therefore, we introduce a taxonomy that groups the papers into classes of text analysis tasks in order to guide visualization researchers with similar tasks to related works that provide close and distant reading solutions. In addition, we list text sources, and we take a closer look at data transformations, which are substantial steps in order to afford designing valuable visualizations. Furthermore, we extended the collaboration experiences and future challenges sections. Finally, this survey considers 22 more related works.

2. Means of Visual Text Analysis

Close reading and *distant reading* are the prior means for the visual analysis of textual sources in digital humanities scenarios. While the close reading of a text is a traditional method that has its roots in antiquity when Aristotle close read the works of Plato [McC15], distant reading is a rather novel idea that was introduced by Franco Moretti at the beginning of the 21th century. In contrast to Moretti, Jockers uses the terms *micro-* and *macro-analysis* instead of close and distant reading [Joc13]. Inspired by micro- and macro-economics, he focuses on quantitative literary text analysis using statistical analysis methods. As the methods we analysed are more 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.
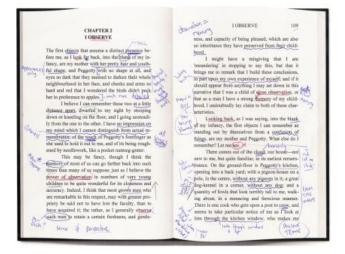

Figure 1: *Traditional close reading of the second chapter of Charles Dickens' David Copperfield (figure reproduced with permission from Kehoe et al. [KG13]).*

2.1. Close reading

The close reading of a text became a fundamental method in literary criticism in the 20th century [Haw00]. 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 1. In this example, the scholar used various methods to annotate various features of the source text, for example, the usage of different colours (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 2 shows a straightforward approach of visualizing various scholars' annotations of a digital edition [KG13] within the web-based environment eMargin [eMa15]. There, colours are used to highlight different text features, and a pop-up window lists the comments of collaborating scholars. In Section 8.1, we outline different approaches to support text analysis with 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

Figure 2: *Digital close reading of the second chapter of Charles Dickens' David Copperfield with eMargin (figure reproduced with permission from Kehoe et al. [KG13]).*

Figure 3: *Distant reading example shows the structure of and the themes in Jack Kerouac's On the Road (figure reproduced with permission from Posavec [Pos07]).*

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:

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

But, the proposed methods and the intention of distant reading are controversial in the humanities [GH11a, CRS*14], as they quantify and abstract texts at an expense of reflecting the actual ambiguity and complexity of literary forms [Dru11, Mar12]. However, many works in the digital humanities domain are based upon Moretti's idea. Figure 3 shows Posavec's Literary Organism [Pos07]—a distant reading of Jack Kerouac's *On the Road* in the form of a tree. While 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 branching structure represents the ordered hierarchy of content objects from chapters down to words, and themes are drawn with different colours. In Section 8.2, we present a list of different distant reading techniques developed for a wide range of text analysis research questions in the digital humanities.

2.3. Combining close and distant reading

In his digital humanities collaborations, Correll worked together with literary scholars who were unfamiliar with distant reading views [CAA*14]. Here, providing links to close reading was an important method to support distant reading hypotheses. During our literature research, we also discovered a multitude of works involving close reading and interfaces, which provide distant reading visualizations that allow to interactively drill down to specific
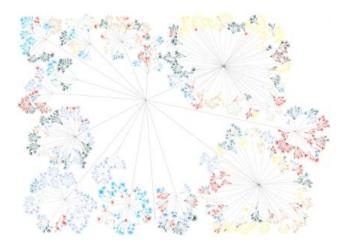
portions of the data. This suggests that direct access to the source texts is indeed very important for humanities scholars when working with visualizations. For example, Bradley [Bra12] asks whether it is 'possible to develop a visualization technique that does not destroy the original text in the process'. Similarly, Beals [Bea14] asks, 'In an age where distant reading is possible, is close reading dead?' 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. Visual Text Analysis

Figure 4 reflects a typical visual text analysis process in digital humanities workflows. The text analysis task at hand affects all steps of this process. First, the text sources are selected in accordance to the research task. Next, it is important to apply appropriate data transformations in order to design close and/or distant reading visualizations that are beneficial to investigate the given text analysis task. But, a humanities scholar often refers to both the visualizations and the underlying texts in order to gain insights concerning the research question. While the next section defines the scope of this survey, the following sections explain the various components of the visual text analysis process in detail. Section 5 lists used text sources, and Section 6 provides a brief overview of data transformation techniques. In Section 7, we provide a taxonomy of text analysis tasks. Applied close and distant reading techniques are outlined in Section 8.
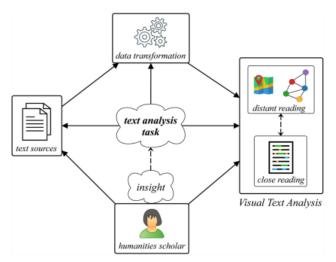
Figure 4: *Visual text analysis process.*

4. Scope

In order to generate the research paper pool of this survey, we used the publication year of Franco Moretti's book on distant reading techniques 'Graphs, Maps, Trees' [Mor05], 2005, as a starting point. We manually scanned through the major related visualization and digital humanities journals and conference proceedings in order to generate an appropriate snapshot of existent research on visualizations for close and distant reading. To receive a processable list of related works, we required to narrow the scope of our survey.

First, we only considered information visualization publications. For example, we did not include papers from human–computer interaction issues [e.g. *ACM Conference on Human Factors in Computing Systems (CHI)*]. Table 1 lists the number of related information visualization papers examined. We also considered approaches developed for other data types, where at least one related use case was provided for a cultural heritage dataset. The *TVCG* journal table entry includes all found papers presented at the *IEEE Symposium on Information Visualization (InfoVis)* as well as seven papers 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)*, or in the *IEEE Computer Graphics and Applications* journal.

Second, we included related works from the major digital humanities realms into our survey. Thereby, we decided to consider the proceedings of the *Annual Conference of the Alliance of Digital Humanities Organizations*, which yields a suitable snapshot of the research conducted in that field, although only short papers are contained. The 60 related works presented at this major digital humanities conference (see Table 2) indicate the importance of close and distant reading visualizations for text analysis tasks. In 2014, a total of 345 individual papers were presented at the conference, of which 33 thematize information visualization techniques for cul-

Table 1: *Visualization papers examined.*

Journal/Proceedings	#Papers
IEEE Transactions on Visualization and Computer Graphics (TVCG)	16
IEEE Symposium on Visual Analytics Science and Technology (VAST)	6
Computer Graphics Forum	5
Proceedings of the International Conference on Information Visualization Theory and Applications (IVAPP)	2
Information Visualisation Journal	1

Table 2: *Digital humanities papers examined.*

Journal/Proceedings	#Papers
Proceedings of the Annual Conference of the Alliance of Digital Humanities Organizations	60
Literary and Linguistic Computing/Digital Scholarship in the Humanities	18
Digital Humanities Quarterly	6

tural heritage data (9.6%), including 14 papers (4.1%) about close and/or distant reading relevant for our survey. The collection is completed by related papers published in two major digital humanities journals, *Literary and Linguistic Computing* (*Digital Scholarship in the Humanities* as of 2015) and *Digital Humanities Quarterly*.

Third, we did not consider works published in humanities realms such as *Shakespeare Quarterly* or *American Literary History*.

4.1. Considered research papers

In order to be considered for our survey, a paper needed to fulfill the following requirements:

- **Textual data:** The visualization is a solution for research questions on an arbitrary text corpus, 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 have 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 modelled in Wikipedia), we took into account visualizations for newspaper collections. This decision includes some visualization papers that do not directly relate to the humanities. But the proposed techniques are based upon or tested with contemporary newspaper collections, which are indeed part of 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 analysing stylistic features among the textual contents of novels, the unrelated graph visualization in [Fin10] uses Amazon recommendations to determine relationships between books.

- **No traditional charts:** In the digital humanities, the word *visualization* is frequently used. Traditional charts displaying statistical information such as line or bar charts are also labelled as visualizations. Based on the information visualization definitions given by Card [CMS99] and the UIUC DLI Glossary [UIU98], we only considered papers that provided computer-supported, non-traditional visual representations of abstract data. In contrast, we do not require interactive methods as humanities scholars often gain valuable 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 the close and/or distant reading of texts.

5. Text Sources

The focus of interest of the papers in our collection includes four major text types.

- **Single literary texts** often motivate the development of close and distant reading techniques, for example, Adam Smith's *The Wealth of Nations* [BJ14], Gertrude Stein's *The Making of Americans* [CDP*07], Herodotus' *Histories* [BPBI10] or *The Castle of Perseverance* [Pet14].
- **POI collections** are text corpora containing several works by a particular author (e.g. William Faulkner novels [DNCM14]), politician (e.g. the papers of Thomas Jefferson [Kle12] or the Kissinger Collection [Kau15]) or any other 'person of interest' (POI), for example, alchemical manuscripts written by Isaac Newton [WH11] or statements about Emily Carr [HSC08].
- **Text editions** are collections of variants of the same source text, for example, research questions concerning German translations of Shakespeare's *Othello* [GCL*13], English translations of the *Bible* [JGBS14] or reprints of Nathaniel Hawthorne's short story *The Celestial Railroad* [Cor13].
- **Text collections** contain a multitude of texts, which usually belong to a particular typology. Among others, this includes the collection of poems [Wil15a], biographies [Boo13], tales [WJ13a], novels [Ede14], medieval texts [JW13] or news articles [KLB14].

This variety of source texts reflects the diversity of text analysis research questions raised by humanities scholars. On the other hand, it suggests the requirement of multi-farious close and distant reading techniques.

While more and more digital editions and archives are made available online [GTW13, FMT15], distant reading studies are few and far between due to the limited number of available text collections [Und15]. It follows that the more and the larger the accessible text collections, the bigger the scope for close and distant reading

analyses and the higher the probability to generate new research challenges for visualization scholars. The issue with this seemingly logic conclusion, however, is the unavoidable and enormous amount of manual work required in order to assemble [Und15], encode and upload [Wei15] such text collections to the web for others to generate derivative outputs. Some of the papers provide information about used publicly available data sources, which we list in Table 3.

6. Data Transformation Aspects

Many papers in our collection do not provide sufficient information about applied pre-processing steps to transform the given textual data into the visualization's input format. Occasionally, a visualization directly processes annotated text in XML format (e.g. [Pie10, Boo13, BGHJ*14]). In particular, many techniques are based upon documents in the XML-based TEI format (e.g. [BHW11, CGM*12]) which has become the humanities leading technology to map the structure of a digital text edition [Sin13]. XSLT stylesheets are basic ways to transform the TEI encoded information into a meaningful visualization of an individual text (e.g. [Cor13, Pie13, HKTK14]). But most distant reading techniques require more sophisticated pre-processing steps—n brief summary of common data transformation approaches is shown below.

Tokenization and normalization are rather rudimentary natural language processing methods first applied to segment raw text sources. The then determinable frequency distribution of words is a valuable basis for various tasks (e.g. stylometric analyses [CEJ*14, Ede14]), and can be clearly visualized in the form of tag clouds to support the exploration of word statistics (e.g. [Bea08, GTAHS15, JBR*15]). Vector space models can be used to list term frequencies per text and support a variety of text analysis tasks (e.g. [DFM*08, GCL*13, KOTM13]). On the other hand, counting n-grams allows to draw more specific statements about a text corpus (e.g. [CDP*07, Bea12, MH13]). But tokenization and normalization are also necessary steps for the data transformation methods described below.

Sequence alignments are computed when investigating research questions concerning textual re-use (e.g. [BGHE10, JGBS14]) and for the analysis of similarities and differences among various text editions (e.g. [WJ13b, JGF*15]). In such scenarios, one typically applies the Gothenburg model, which includes tokenization, normalization, alignment, analysis and visualization [Got15]. One of the web-based tools implementing this model is CollateX [HDvHM*15].

Part-of-speech (POS) tagging is a frequently applied preprocessing technique to automatically annotate the words of a corpus according to their POS category. The use of tools like the Stanford POS tagger [TKMS03] is a mandatory basis for investigating diverse research questions. Typically, words and their relationships are explored (e.g. [KKL*11, MH13]) or linguistic patterns are extracted from a corpus (e.g. [Mur11, RFH14]). Furthermore, POS tagging is used to analyse phonetic features [CTA*13] or for research questions concerning stylometry [KO07].

Named entity recognition (NER) is the practice of extracting named entities such as places or persons from texts. Pre-processing steps like POS tagging can be applied to automatically list named entity candidates [GH11b]. With the help of

Table 3: *Publicly available data sources.*

Data source	URL	Used by
Digital libraries		
Perseus Digital Library	http://www.perseus.tufts.edu/hopper/	[BPBI10]
Project Gutenberg	https://www.gutenberg.org/	[CTA*13, GCL*13, KO07, GTAHS15]
TextGrid Repository	https://textgridrep.de/	[TFK15, JKH*15]
HathiTrust	https://www.hathitrust.org/	[AGZH15]
POI collections		
The Swinburne Project	http://swinburnearchive.indiana.edu/swinburne/	[WH11]
The Chymistry of Isaac Newton Project	http://webapp1.dlib.indiana.edu/newton/	[WH11]
The Papers of Thomas Jefferson	http://rotunda.upress.virginia.edu/founders/TSJN.html	[Kle12]
Internet Shakespeare Editions	http://internetshakespeare.uvic.ca/	[RSDCD*13]
Bob Gibson Collection of Speculative Fiction	http://contentdm.ucalgary.ca/cdm/search/collection/gcsf	[HFM16]
Literary collections		
The Poetess Archive	http://idhmc.tamu.edu/poetess/	[FS11, CGM*12]
Folklore and Mythology Electronic Texts	http://www.pitt.edu/~dash/folktexts.html	[RFH14]
The Skaldic Project	http://www.abdn.ac.uk/skaldic/	[Wil15a]
Eighteenth Century Collection Online	http://quod.lib.umich.edu/e/ecco/	[JOL*15]
MySword (Bible Collection)	http://www.mysword.info/	[JG15]
Political texts		
1641 Depositions	http://www.1641.tcd.ie/	[ÓML14]
Digital National Security Archive	http://nsarchive.chadwyck.com/	[Kau15]
PoliInformatics	http://poliinformatics.org/data/	[Poi15]
VroniPlag Wiki	http://de.vroniplag.wikia.com/	[RPSF15]
News archives		
Europe Media Monitor	http://emm.newsbrief.eu/	[KBK11]
Projekt Deutscher Wortschatz	http://corpora.uni-leipzig.de/	[KLB14]
Scientific paper collections		
Thompson Reuters Web of Knowledge	http://wokinfo.com/	[ARR*12]
JSTOR	http://www.jstor.org/	[OGH15, MSR*15]
Collected project corpora		
ANNIS	https://korpling.german.hu-berlin.de/annis3/	[KZ14]
Trading Consequences	http://tradingconsequences.blogs.edina.ac.uk/about/the-corpus/	[HAC*15]
Other sources		
British National Corpus	http://www.natcorp.ox.ac.uk/	[Bea08, ARLC*13]
Documenting the American South	http://docsouth.unc.edu/	[CTA*13]
Metadata Offer New Knowledge	http://monk.library.illinois.edu/	[VCPK09]
The Internet Movie Script Database	http://www.imsdb.com/	[HPR14]

lexicons, named entities are subsequently classified. For example, the Pleiades gazetteer [Ple15] is used for the extraction of ancient place names [EJ14], and DBPedia [LIJ*14] supports the discovery of commodities in [HAC*15]. The Stanford Named Entity Recognizer [FGM05] is a popular tool for automatic named entity extraction. The manual collection of named entities is not uncommon and guarantees the highest precision (e.g. [JW13, Wil15b]). Occasionally, named entities are already present when using annotated TEI files as data sources (e.g. [BB15a, TFK15]).

Topic modeling algorithms are fundamental for topic-related analyses of text collections. The Latent Dirichlet allocation is the most often applied topic model [BNJ03]. It requires a pre-defined number of topics, which are determined automatically based on the words contained in the text corpus (e.g. [AKV*14, BJ14]). The topic model can be used to cluster texts thematically [Wol13]—as happens with text classification methods (e.g. [PSA*06, DFM*08])—or to define the similarity among the texts of a corpus [Joc12]. When temporal metadata are provided, the change of topics can be analysed (e.g. [ARR*12, CLWW14]).

Semi-automatic approaches reflect the importance of integrating the humanities scholar into the data transformation process. For instance, the scholar's knowledge is required when manually generating or validating a training set to produce an appropriate data mining classifier (e.g. [PSA*06, KKL*11, KJW*14]). Other methods include semi-automatic alignments [GZ12] and the annotation of TEI documents (e.g. [Tót13, OGH15]). Sometimes, even the visualization entirely depends on manually collected data through crowdsourcing (e.g. [WMN*14, RPSF15, HFM16, JFS16]).

Table 4: *Classification of papers according to the taxonomy of text analysis tasks.*

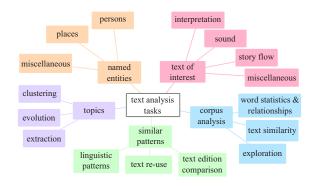|		
Named entities	*Places*	[MBL*06], [Tra09], [BPBI10], [GH11b], [JW13], [DNCM14], [EJ14], [GDMF*14], [ÓML14], [AGZH15], [BB15b], [HAHT*15], [Wil15a]
	Persons	[Cob05], [CSV08], [BDF*10], [RD10], [BHW11], [Kle12], [Boo13], [KOTM13], [OKK13], [Tót13], [KLB14], [Pet14], [BB15a], [Poi15], [TFK15], [Wil15b], [JFS16]
	Miscellaneous	[AGL*07], [HAHB15], [HAC*15], [OGH15]
Topics	*Extraction*	[AKV*14], [BJ14], [ESK14], [JOL*15], [Kau15], [MSR*15]
	Evolution	[CLT*11], [KBK11], [ARR*12], [DWS*12], [CLWW14], [Kau15]
	Clustering	[PSA*06], [DFM*08], [OST*10], [Wol13], [HFM16]
Similar patterns	*Linguistic patterns*	[CDP*07], [WV08], [vHWV09], [Mur11], [GZ12], [MLSU13], [BGHJ*14], [KZ14], [RFH14], [JKH*15]
	Text re-use	[BGHE10], [WH11], [Wol13], [HCC14], [JGBS14], [RARC*15], [RPSF15], [ZNMS15]
	Text edition comparison	[JRS*09], [Cor13], [GCL*13], [WJ13b], [JG15], [JGF*15], [MRMK15], [PMMR15]
Text of interest	*Interpretation*	[Pie10], [Arm14], [HKTK14], [WMN*14]
	Sound	[FS11], [CGM*12], [ARLC*13], [CTA*13], [Pie13], [Ben14], [MLCM16]
	Story flow	[RRRG05], [LWW*13], [RSDCD*13], [HPR14]
	Miscellaneous	[MFM13], [GWFI14], [KJW*14], [PBD14]
Corpus analysis	*word statistics & relationships*	[Bea08], [CVW09], [VCPK09], [LRKC10], [Bea11], [KKL*11], [Bea12], [MH13], [WJ13a], [GTAHS15], [JBR*15]
	text similarity	[RRRG05], [KO07], [EX10], [WH11], [Joc12], [CEJ*14], [Ede14]
	exploration	[HSC08], [CWG11], [JHSS12], [FKT14]

Figure 5: *Taxonomy of text analysis tasks.*

7. Taxonomy

There has been extensive research done in developing taxonomies for information visualization in the last decades. Unfortunately, these taxonomies were either too general (e.g. [BM13, RAW*15]) or too specific (e.g. [LPP*06, KKC15]) to be used for our paper collection. Therefore, we defined a taxonomy focusing the underlying text analysis tasks in the digital humanities domain (see Figure 5). A detailed classification of papers is given in Table 4. Papers focusing a single text analysis task are grouped to a single—the best fitting— category. The few works providing visualization methods for two text analysis tasks each appear in two categories [RRRG05, WH11, Wol13, Kau15]. The taxonomy consists of five major categories:

The analysis of ***named entities*** is a common text analysis task of humanities scholars that is supported with close and distant reading visualizations. When extracting *places*, fictional or reported geographies of a single text or a whole collection can be explored. The extraction of *persons* is required to analyse social networks of

individuals or of characters in a story. *Miscellaneous* tasks focus on other (e.g. encyclopedia entries [HAHB15]) or on multiple named entities (e.g. commodities and locations [HAC*15]).

The analysis of ***topics*** inherent in a text corpus supports text analysis tasks that require both close and distant reading techniques. Popular tasks are topic *extraction*s, so that major topics in the source texts can be tracked and topic-related text passages can be discovered. The presence of temporal data allows for the analysis of topical *evolution*, and on the basis of the found topics, a topical *clustering* of a corpus is possible.

An analysis of ***similar patterns*** that includes the discovery, the alignment and the visualization of similar text segments among the texts of a given collection is a typical text analysis task in the digital humanities. Dependent on the length of patterns, we divide the tasks belonging to that category into three sets. While the analysis of *linguistic patterns* concerns short phrases, *text re-use* analysis focuses on determining deliberately re-used text segments (e.g. quotes or plagiarized passages). In papers grouped to the category *text edition comparison*, the humanities scholar is more interested in analysing the variants between the text editions.

Some tasks focus an individual literary work, which we call ***text of interest***. Sophisticated close reading techniques are often applied but are sometimes coupled with distant reading representations of the textual content. The underlying research tasks vary from visualizing text *interpretation*s to the analysis of *sound* of literary works (mostly poems), and to the *story flow* analysis of a given source text. *Miscellaneous* tasks, for example, support the thorough analysis [KJW*14] or enhance the close reading [GWFI14] of a literary work.

The final category are ***corpus analysis*** tasks. Usually, distant reading visualizations are used to explore text corpora containing a

high number of texts. A major research task is the analysis of *word statistics & relationships* among them. Further interests concern *text similarity* between the texts of a corpus, others require platforms for corpus *exploration* to facilitate knowledge discovery.

8. Applied Visualization Techniques

This section provides an overview of close and distant reading visualizations examined in the papers in our collection. In addition, we outline strategies for combining close and distant reading visualizations that facilitate a multi-faceted analysis of the underlying textual data. Table 5 shows a distribution of the applied techniques according to the text analysis task taxonomy (see Table 4). For some research tasks, favoured visualization methods stand out. A detailed overview about what techniques are suited for which text analysis tasks is given below.

8.1. 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.

As can be seen in Table 5, the application of close reading techniques is particularly important when analysing *similar patterns*. For *text edition comparison*, close reading is necessary to discover occurring similarities and differences, and a close reading of similar *linguistic patterns* or *text re-use* patterns helps to analyse the contexts in which these patterns were used. When focusing a *text of interest*, close reading techniques are applied to illustrate various text features. Other research tasks concerning *named entities*, *topics* and *corpus analysis* rather investigate generic features of a corpus, and apply mostly distant reading techniques. Still, close reading is sometimes helpful to connect a computationally gained distant view with the underlying source texts.

While 10 visualizations provide only **plain** close reading views without additional information, 56 visualizations attend to the matter of enhancing the close reading capabilities of the humanities scholars. To visualize such additional information for a great variety of purposes, the researchers made use of the techniques listed below.

Colour is the visual attribute most often used to display the features of textual entities and it is applied in different ways. In most cases, a coloured background is used to express various types of information about a single word or an entire phrase (Figure 2). The tool Serendip [AKV*14] varies the transparency of background colours to encode the importance of individual words (Figure 6 a). Font colour is also frequently used for this purpose (Figure 6 b left). Coloured circumcircles (Figure 6 c) around words are used only once [MLCM16]. 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, colouring is a suitable method to express a great variety of textual features. Among other purposes, colouring
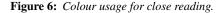

(a) Colored backgrounds and backgrounds with varying transparency (Figure provided by Alexander et al. and based on [AKV*14]).

(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 in the Poem View (left) and connections in the Path View (right) highlight rhyme sets in poems (Figure produced with Poemage [Poe16] based on McCurdy et al. [MLCM16]).

Figure 6: *Colour usage for close reading.*

is used for the analysis of *similar patterns*, for example, to mark common words (e.g. [JRS*09, Mur11]) and aligned text segments (e.g. [ZNMS15, RPSF15]) in parallel texts, or when exploring a *text of interest*, for example, to highlight the automated or manual classification of words or phrases (e.g. [KJW*14, WMN*14]), or to visualize various sound patterns in poems (e.g. [CTA*13, Ben14]).

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 [JGF*15, JG15], which is a directed acyclic graph that is used for *text edition comparison* as it visualizes differences and similarities among text variants, font size encodes the number of occurrences of a word in all editions (Figure 9 a). 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 6 b right). In [CWG11], varying font size is used

Table 5: *Applied close and distant reading techniques according to text analysis tasks.*

	Close Reading					Distant Reading						Combining		
	Plain	Color	Font size	Glyphs	Connections	Heat maps	Tag clouds	Maps	Timelines	Graphs	Miscellaneous	Bottom-up	Top-down & Bottom-up	Top-down
named entities														
places	1	2				1		12	5	1		1		2
persons	2					1		2	1	15	1		1	1
miscellaneous		2					1	1	1	3		1		1
topics														
extraction		2				2	4		1	1	1			1
evolution	1						3	1	6					1
clustering	1	2					1	1		2	1			3
similar patterns														
linguistic patterns		8			1	4	1			5		1	2	4
text re-use		6			2	4					3			3
text edition comparison	1	5	2		4	4				1		2	2	1
text of interest														
interpretation		2	1	1						1				
sound		3		3	2	3								
story flow	1	1							2	1	1			2
miscellaneous	1	1			1	1	1			1	1		1	1
corpus analysis														
word statistics & relationships		3				2	8			4	1		1	2
text similarity						2				5				
exploration	2	1	1			2	1	1	2	1				3
	10	38	4	5	9	26	20	18	18	41	9	5	8	25

Figure 7: *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]).*
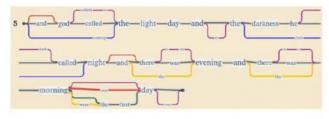

Figure 8: *Close reading example with glyphs illustrating the hermeneutic structure of a poem (figure provided by Piez and based on [Pie10]).*

(a) Variant graph for seven English translations of *Genesis 1:5* connecting subsequent words displayed with variable font size (Figure based on [JGF*15]).

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

Figure 9: *Connections used for text alignment.*

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 colouring or varying font size. All examples we found enhance the close reading of a *text of interest*, mostly poems. In [ARLC*13], phonetic units are drawn atop each word using colour to classify phonetic types (Figure 7). 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 8. Goffin explores the placement and design of so-called word-scale visualizations, which are small glyphs enriching the base text with additional information [GWFI14]. For example, the background colour of words contained in digital copies speaks for Optical Character Recognition (OCR) certainty. Furthermore, small interactive bar charts illustrate variants of observed words.

Connections aid to illustrate the structure among textual entities most often applied to support text analysis tasks concerning *similar patterns*. One usage of connections in close reading is to highlight subsequent words in a variant graph to track variation among text editions [BGHE10]. As shown in Figure 9(a), coloured links can help to identify certain editions [JG15, JGF*15]. Other approaches juxtapose the texts of various editions and visually linkrelated text passages [WJ13b, HKTK14, JGBS14], as instantiated in Figure 9(b). Furthermore, connections can also be used to

visualize sentence structure [KZ14]. Two close reading visualizations use connections for the analysis of *sound* in poems. While Abdul-Rahman [ARLC*13] illustrates phonetic and semantic relations within poems (Figure 7), McCurdy [MLCM16] draws paths between words of a poem sharing the same tones (assonances) to highlight sonic patterns (Figure 6 c).

8.2. 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, for example, 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.

An overview of applied distant reading visualizations according to the text analysis task taxonomy is given in Table 5. The overall usage of such techniques suggests their importance for nearly all text analysis tasks in digital humanities, even when the close reading of a text is more important, for example, when focusing a *text of interest* or analysing *similar patterns*.
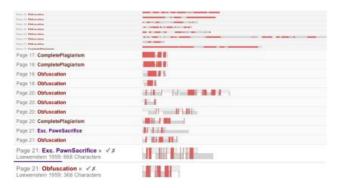
Figure 10: *Heat map highlighting potential plagiarized text passages in a PhD thesis (figure reproduced with permission from Riemann et al. [RPSF15]).*

Within our research papers collection, we found 132 visualizations providing a distant reading view of a given text corpus. We extracted and grouped various approaches found to visualize summarized information into the six following categories.

Heat maps or block matrices are often used to highlight text snippets, especially, when analysing *similar patterns* and in *corpus analysis* tasks. Thereby, a heat map may reflect structural elements of a text [JRS*09, VCPK09] or the structure of an entire corpus [CDP*07, Mur11, BGHJ*14]. In such scenarios, the colouring of rectangular blocks helps to analyse the distribution of specific textual patterns [CWG11, MH13, JKH*15]. Another example is the usage of heat maps to show relationships among various texts in a corpus. The similarity for each tuple of texts within the corpus can be determined by counting similar text passages, and the result can be visualized as a heat map [GCL*13, FKT14], for example, to highlight the similarity between Shakespearean plays [RRRG05]. Heat maps are also applied to visualize similarities or differences among text editions [JG15, PMMR15], or to highlight re-used passages between the texts of a corpus [JGBS14, RARC*15, ZNMS15]. For the analysis of potentially plagiarized texts, socalled Difflines reveal structural differences between several suspicious text fragments and their alleged originals in a Focus+Context view [RPSF15], an example is shown in Figure 10. A further heat map variant are fingerprinting techniques as introduced in [KO07] in order to visualize characteristic textual features of literary works. In text analysis tasks concerning *named entities*, heat maps can be used to analyse *places* mentioned in texts [AGZH15], or to reveal interpersonal relationships between characters in prose literature [OKK13]. For the analysis of *topics*, Alexander *et al.* [AKV*14] propose two matrix representations. The RankViewer illustrates the ranking of words belonging to topics and the CorpusViewer shows relations to certain topics for each document of a corpus. Heat maps are also used in [MSR*15] to display 'high-level summaries' of topic modeling results. Finally, heat maps are used to analyse a ***text of interest*** [KJW*14], for example, to visualize the similarity [CTA*13] or the flow [FS11, Ben14] of *sound* in poems.

Tag clouds are intuitive visualizations to encode the frequency of words within a selected section, a whole document or an entire text corpus by using variable font size. Tag clouds are therefore a suit-
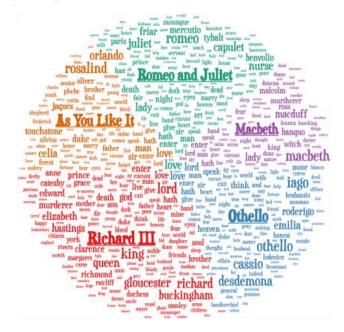
Figure 11: *Comparing the tags in five Shakespeare plays (figure based on [JBR*15]).*

able method for *corpus analysis* tasks [VCPK09, Bea12, FKT14, GTAHS15]. TagPies—a tag cloud arranged in a pie chart manner—support the comparative analysis of the co-occurrences of search terms [JBR*15], an example of which is shown in Figure 11. Other approaches visualize the temporal evolution of tags in tag clouds, either listing tags per time period [CVW09] or by attaching a time graph to each tag [LRKC10]. Beaven uses tag clouds to illustrate collocational relationships of a single word [Bea08] and to compare the collocates between two words [Bea11]. Tag clouds are also applied when analysing *topics*, for example, by displaying a topic's characteristic tags [BJ14, ESK14, JOL*15, MSR*15] (an example can be seen Figure 16 b) or by summarizing the major tags for certain time periods [CLT*11, CLWW14]. The usage of tag clouds to explore the classification of speculative fiction anthologies [HFM16] is shown in Figure 13. Tag clouds are rarely applied for the analysis of *named entities* [HAC*15] (see Figure 12 a) or when focusing a ***text of interest*** [KJW*14]. In some of the above mentioned works, tag colouring is used to express additional information such as the temporal evolution of a word's significance or the classification of tags.

Maps are widely used to display the geospatial information contained in a text. Most often, maps support the analysis of ***named entities*** extracted from a text or an entire corpus. Two works illustrate the geographical areas which are associated to *persons* [Wil15b], for example, by mapping the places of activity of musicians manually extracted from musicological literature in order to support the geospatial comparison of musicians' activity regions [JFS16]. But usually, *places* mentioned in texts are analysed. 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

(a) Linked views for the exploration of commodity trading (Figure reproduced with permission from Hinrichs et al. [HAC*15]).

(b) Colored shapes encode emotions about London places (Figure reproduced with permission from Heuser et al. [HAHT*15]).

(c) Fictional map of *Yoknapatawpha County* and related places (Figure reproduced with permission from Dye et al. [DNCM14]).

Figure 12: *Maps supporting distant reading.*

source text(s). Some approaches use thematic [ÓML14] or density maps [GH11b, BB15b] for this purpose, but the usage of glyphs in the form of circles is more frequent [Tra09, DWS*12, HAC*15, Wil15a] as it simplifies the interaction with individual plotted places (e.g. see Figure 12 a). In [HAHT*15], circles are used to map discrete London places occurring in fictional literature, and polygons represent wider spaces such as neighbourhoods or districts of London. The colouring of shapes indicates collected emotions to these places (see Figure 12 b). In [JW13], various glyphs encode various types of places occurring in medieval texts. Two works that 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 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]. The visualization of Faulkner's fictional *Yoknapatawpha County* includes various means of geographic mapping [DNCM14]: on the one hand, the imagined geography and, on the other, the placenames displayed on the geographic levels region, nation and world (Figure 12 c). In addition to named entity analysis, maps are used for the analysis of *topics* [DFM*08, GDMF*14] and *corpus exploration* [JHSS12].

Timelines are appropriate techniques to visualize historical text corpora carrying various types of temporal information. Often, timelines support the analysis of *named entities*. In [JW13], uncertain datings of medieval texts are visualized to support the analysis of cited *places*. Sometimes, the temporal information about events reported in a text need to be extracted in order to visualize (fictional) calendars [DNCM14, GDMF*14, ÓML14, HAC*15] (e.g. see Figure 12 a). For the exploration of placenames in Herodotus' *Histories*, a timeline is used to show where certain placenames occur in the text [BPBI10]. Timelines also support the temporal analysis of *topics* [HFM16], for example, the exploration of events in news articles [ESK14] (see Figure 16 b). Furthermore, timelines are used for *corpus analysis* tasks [JHSS12]. 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 14). Streamgraphs are popular techniques that produce aesthetic visualizations and allow to track the *evolution* of *topics* over time [BW08], thus
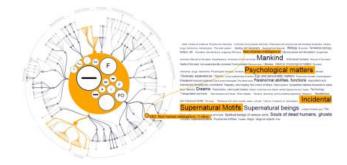
Figure 13: *Exploring the tree structure and representative tags of a novel classification (figure provided by Hinrichs based on [HFM16]).*

generating enhanced versions of the timeline metaphor. Such visualizations are often based on newspaper sources [CLT*11, KBK11, DWS*12, CLWW14] or political text archives [Kau15, Poi15] to support the analysis of contemporary topic changes. Using a research paper pool [ARR*12], the changing importance of research topics can be explored. Streamgraphs may also be used to support the analysis of storylines in a *text of interest* to illustrate plot evolution and changing locations [LWW*13]. Based upon Hollywood screenplays, the tool ScripThreads visualizes action lines of movie characters [HPR14].

Graphs are valuable visualizations for all text analysis tasks. They are most often applied 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 tree [HFM16] (see Figure 13) or a network [WH11]. Proximity can be used to express the similarity of these texts [EX10, MRMK15], for example, based upon stylistic closeness [Joc12, CEJ*14]. For instance, Eder visualizes a network of poetic texts based on stylistic features [Ede14], thereby highlighting connected editions and sequels (Figure 15 a). Graphs can also be used to illustrate topics (as nodes) and proximities of topic models applied to text corpora [Kau15]. In other works, nodes represent the words of a corpus and links are drawn to reflect semantic relationships [AGL*07, RFH14, OGH15] or co-occurrences [VCPK09, KKL*11, WJ13a,
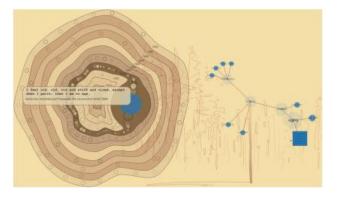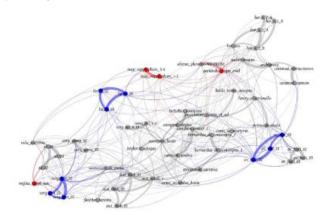
Figure 14: *A combination of an abstract timeline view and a tree in EMDialog (figure provided by Hinrichs based on [HSC08]).*

(a) Network of 55 poetic texts. Imitated texts marked in blue, sequels marked in red (Figure reproduced with permission from Eder [Ede14]).

(b) Excerpt from the social network in Mikhail Bulgakov's *Master and Margarita* (Figure provided by Maciej Ceglowski and reproduced with permission from Coburn [Cob05]).

(c) Thomas Jefferson's social relationships (Figure reproduced with permission from Klein [Kle12]).

Figure 15: *Graphs supporting distant reading.*

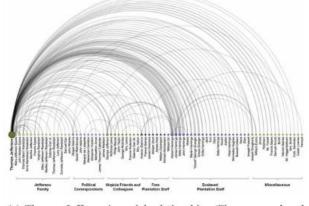BB15b]. In [HAHB15], a graph visualizes cross references in historic encyclopedia by linking related entries. Further applications are the visualization of scene changes and character movements in Shakespearen plays [RRRG05], as well as the display of conceptual [Arm14], contextual [HSC08] or multi-lingual [GZ12] information. Phrase nets connect textual entities that appear in the form of a user-specified relation (syntactic or lexical) [vHWV09]. All aforementioned works apply force-directed algorithms for the placement of nodes. Radial graphs can be used to unveil the relationships of words within poems [MFM13], or, again, to highlight the similarity among texts and, in this case, as nodes radially grouped by authors [Wol13]. The Word Tree [WV08], also used in [MH13], visualizes sentences sharing the same beginning in the form of a tree. In contrast to the variant graph, a technique that supports close reading of textual editions, the Word Tree is a distant reading technique as it dissolves the order of sentences. Finally, we found a method that visualizes plain event trigraphs extracted through phrase mining algorithms and thus providing metaphors to display uncertain information [MLSU13]. When analysing *named entities*, graphs are the means of choice to visualize the relationships between *people* in the form of social networks. Such representations are widely applied in the digital humanities to illustrate the relationships between characters in literary texts [CSV08, Tót13, BB15a, TFK15]. In these graphs, the size of a node can be used to encode the frequency of a character name in the text [BHW11, Pet14], the thickness of an edge [Cob05] (Figure 15 b) or the proximity of the nodes [Poi15, JFS16] can serve to reflect the strength of a relationship and edge style can be used to classify the type of relationship [KOTM13]. As per the aforementioned works, Kochtchi uses a force-based graph layout to visualize social networks automatically extracted from newspaper articles [KLB14]. In contrast, radial layouts and parallel coordinates are used in [Boo13]. For the visualization of Thomas Jefferson's social relationships (Figure 15 c), the nodes placed on a vertical axis are connected with arcs [Kle12]. Riche proposed a layout for Euler diagrams, which can also be utilized to visualize relationships between characters extracted from Shakespearean texts [RD10]. Finally, GeneaQuilts smartly visualizes large genealogies extracted from literary texts such as the Bible [BDF*10].
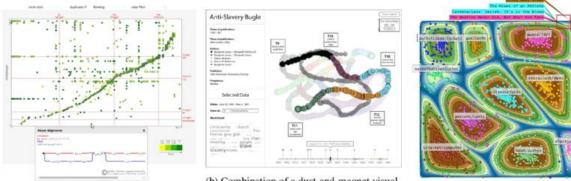
Miscellaneous methods also produce beneficial results for certain text analysis tasks, most often for the exploration of *similar patterns*. In [JGBS14], an interactive dot plot interface is used to visualize and explore patterns of text re-use between two texts (Figure 16 a). In [GCL*13], a parallel coordinates and a dot plot view, which is used for filtering purposes, visualizes the similarity of parallel text sections. Sankey diagrams are used to compare the categories of words contained in two books [HCC14], and to highlight

(a) Interactive dot plot visualizes text re-use patterns between two Arabic texts (Figure based on [JGBS14]).

(b) Combination of a dust-and-magnet visualization, a timeline and a tag cloud for browsing historical newspapers (Figure reproduced with permission from Eisenstein et al. [ESK14]).

(c) Topological landscape visualizes thematic clusters in the New York Times corpus (Figure based on [OST*10]).

Figure 16: *Miscellaneous distant reading methods.*

plagiarized text passages when juxtaposing a PhD thesis to potential sources [RPSF15]. 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. Two miscellaneous methods are applied to analyse ***topics***. For the exploratory thematic analysis of historical newspaper archives [ESK14], an application of the dust-and-magnet metaphor [YMSJ05] yielded useful results (Figure 16 b). Another topical analysis technique uses a landscape metaphor to visualize the topology-based clustering of articles taken from the New York Times Corpus [OST*10] (Figure 16 c). Various methods were also developed to support the analysis of a ***text of interest***. The tool PlotVis allows users to model and interact with XML-encoded literary narratives in three dimensions [PBD14]. A further complex tool named 'Simulated Environment for Theatre (SET)' supports the story flow simulation of theatrical plays [RS-DCD*13]. It consists of various two-dimensional (2D) interfaces illustrating the 'line of action' and a 3D interface populated by character avatars. For the analysis of *word statistics & relationships*, tree maps are used to illustrate the occurrences of adjectives in fairy tales in [WJ13a]. The Column Explorer introduced in [JFS16] supports the analysis of ***named entities***, in that case by comparatively visualizing biographical profiles of musicians.

8.3. Techniques for combining close and distant reading

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 coordinated 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. Table 5 orders the remaining 37 remaining techniques, which are outlined in detail below, according to the given text analysis tasks in three groups.

Bottom-up methods focus primarily on close reading, especially, when focusing ***similar patterns***. 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, heat map) based on that selection. Another bottom-up approach supports the semi-automatic alignment of early new high German text variants [MRMK15]. A graph displaying the similarities between text editions is updated as annotations are collected in close reading sessions. 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. Two approaches provide bottom-up strategies to support the analysis of ***named entities***. When annotating literary texts [AGZH15], *places* related to Edinburgh are marked, and a linked heat map that displays the distribution of all annotations is accordingly updated. In [OGH15], the user explores automatically tagged named entities of scientific papers in close reading mode. After editing, a graph reflecting contained entities and relationships among them is generated.

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. Some of these approaches support the analysis of ***similar patterns***. In [JRS*09] and [PMMR15], the user can switch between heat map (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. Other text analysis tasks also benefit from the combination of top-down and bottom-up approaches. A typical use case are visual analytics methods. The VarifocalReader [KJW*14] hierarchically visualizes a document with the help of distant views (structural overview, tag clouds) and close reading techniques (use of colour, 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 multi-faceted 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.

Top-down strategies support nearly all types of text analysis tasks. They are mostly applied to combine 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 [HSC08, DWS*12, MFM13, RSDCD*13, FKT14, Wil15a, Wil15b, HFM16]. Observed words or text patterns are often highlighted in the close reading view by way of colouring [VCPK09, GZ12, Wol13, AKV*14, HPR14, HAC*15, JBR*15, JKH*15]. Various colours can thereby illustrate word categories [CDP*07], for example, types of toponyms in the Herodotus Timemap [BPBI10] 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 [JGBS14], a grid-based heat map visualizes similarities between the 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. Similarly, the navigation between distant plagiarism overviews and the close reading of plagiarized passages is organized in [ZNMS15] and [RPSF15]. A distant reading visualization illustrating the variance of verses among multiple Bible editions provides distant views as heat maps on various text hierarchy levels (entire Bible, book, chapter) [JG15]. In the chapter view, the close reading of individual verses is possible. 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 colouring and sizing lines.

9. 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. Some projects reveal valuable insights into collaboration experiences. Excellent design studies are outlined by Abdul-Rahman *et al.* [ARLC*13], McCurdy *et al.* [MLCM16] and Hinrichs *et al.* [HFM16]. All applications were successfully presented in visualization and digital humanities issues. Other publications also share important experi-

ences. A collective overview of the gained insights regarding various aspects of the development phase are outlined below.

Project start. The beneficial, initial decision of carrying out a user-centered design study [Mun09] is reported in various works (e.g. [ARLC*13, JFS16]). This leads to a very close collaboration between researchers of the different fields, which helps to avoid gearing the development of a visualization into false directions. Further important tasks at the beginning of a digital humanities project are discussions about the research questions and perspectives for which a visualization, be it for close or distant reading, can be beneficial [JGBS14]. 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. A typical problem of digital humanities projects is reported in [MLCM16]. The 'initial conversations [between visualization and humanities scholars] were broad and open-ended', also, because the humanities scholars 'did not have specific goals' in mind. Furthermore, the humanities scholars were sceptical that visualization can support their research, and there was also an 'anxiety that the computer would inhibit the qualitative experience of the poetic encounter'. After humanities scholars presented examples of interesting features and computer scientists 'established methods for computationally detecting and analyzing the devices that most interested them', a common project basis and tasks had been be generated. In such circumstances, special workshops can also 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. The generation of a text corpus is often an enduring humanities scholars' task that begins with a project launch [HFM16]. As a consequence, visualization researchers start with a small training set, and should therefore design a visualization as flexible as possible in order to enable potential changes of humanities scholars' research interests, and to avoid limitations. In the best case, the text corpus to be analysed is already available in digital form and a precise research question is at hand, as outlined in [JFS16].

Iterative development of prototypes. The involvement of humanities scholars in various stages of the development is necessary to ensure creating an intuitive visualization that will be used. 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 [JGBS14]. 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]. Geßner [JGF*15] stated that such a process finally helps to gain an intuitive result 'even for the inexperienced, maybe sceptical user'. When designing a profiling system for musicians [JFS16], a frequent interdisciplinary get-together was important for the visualization researchers to communicate their own concerns and to iteratively redesign the underlying mathematical basis (similarity measures)—thereby ensuring that aspects of data

transformation retained comprehensible for the collaborating musicologists. That the scholarly exchange is of particular importance if the textual data source evolves throughout the project time, is outlined by Hinrichs *et al.* [HFM16]. On the one hand, the archival work of humanities scholars when working on the dataset may further develop hypotheses that trigger new visualization ideas, and on the other hand, a visualization has the potential of changing the humanities scholars' research processes and their perspective on a text collection. For the development of Poemage [MLCM16], frequent meetings helped visualization researchers in understanding the problem space and engaged literary scholars in working with the visualization, and finally, in developing 'an interface that reflected their interests, aesthetics, and values'. The authors also document that the departure from wellestablished design principles such as regarding 'ambiguity as a fundamental source of insight' or 'not restricting the tool to avoid clutter' was necessary in raising the value of the visualization. Another example 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 provide 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, HFM16]. 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 Varifocal-Reader [KJW*14], the user liked to view 'the digitized image of a book's page and mentioned that this would increase his trust in the approach'. The metaphor of a digitized text is also used when comparing various English translations of the Bible [JGF*15], 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 analysing larger datasets or, in other words, more texts and in more languages [GCL*13]. In [HAC*15], the attached labelling 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'. As stated in [JFS16], future extensions of visualizations usually require efforts from both computer scientists and humanities scholars. Scientists involved in [HKTK14] stated that collaborative work helped to reactivate and to regenerate traditional literary methodologies rather than abandon them. The turn from initial scepticism when starting the digital humanities project to enthusiasm when using the resultant visualization is reported in [MLCM16]. In a long-term evaluation, Hinrichs *et al.* summarize the potential of their developed visualization for the collaborating humanities scholars [HFM16]. Like in their case study on fiction literature, a visualization should be able (1) to confirm existing hypotheses, (2) to refine humanities scholars' research questions, (3)

to offer new ways of answering research questions, (4) to negotiate quantitative and qualitative interpretation of the underlying text corpus and (5) to trigger new research questions. Other visualization researchers share similar experiences gained in evaluation sessions with humanities scholars [CWG11, ARLC*13, GCL*13, HAC*15, MLCM16]. Such an example is given by Vuillemot *et al.* [VCPK09]. When working on Gertrude Stein's *The Making of Americans* with POSVis, 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].

10. Future Challenges

Throughout our work on this survey, we marked major challenges in the digital humanities where the visualization community can contribute valuable research.

Adapting existing visualization techniques. For some of the text analysis research questions posed in the digital humanities, the adaption of existent techniques proposed in visualization research papers is beneficial. A positive example is the *Trading Consequences* project [HAC*15]. Involved visualization scholars designed a system inspired by VisGets [DCCW08] and made use of Parallel Tag Clouds [CVW09]. Both visualization techniques were not primarily developed for digital humanities data, but they were beneficially adapted to support humanities scholars. Occasionally, new techniques for close and distant reading are designed while appropriate, sophisticated visualizations unrelated to digital humanities data already exist. For future research tasks, the inclusion of these visualizations into the workflows of humanities scholars could lead to faster hypotheses generation due to the limited time for development. As an example, the Sequence Surveyor, which provides a dendrogram to explore genomic structures [ADG11], could support future research. Each leaf of the dendrogram 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 phonetic patterns. Other possible adaptations of existing visualization techniques for digital humanities research can be found in the previous version of this survey [JFCS15].

Novel techniques for close reading. Various publications outline that close reading benefits from visualization, for example, by highlighting crowdsourcing statistics [KG13, WMN*14] or displaying information about textual features and structure [ARLC*13, JGF*15] alongside the source text. Although close reading is an essential task for humanities scholars, in most cases only simple visualization techniques, such as colour 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 proposed 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 glyphbased approaches for close reading in comparison to using colour 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, for example, 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, JGF*15], 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 place-names 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, for example, the exact locations of 'Sparta' and 'Atlantis' have yet to be discovered. Another form of uncertainty is defined by contextual information, for example, 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 multi-farious manners, for example, 1450, before 1450, after 1450, around 1450, 15th century, first half of the the 15th century. 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 multi-farious 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. 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].

Qualitative studies. The number of projects that include visualization components as valuable means of text analysis indicates the potential of visualization to support digital humanities research. Some scholars suggest the role of visualization as providers of new perspectives on the texts that facilitate text comprehension and hypothesis generation. For example, humanities scholars involved in the development of the PoemViewer [ARLC*13] mentioned that 'they would not likely look for insight from the tool itself ... they would look for enhanced poetic engagement, facilitated by visualization'. Hinrichs *et al.* [HFM16] state that 'information visualizations ... are not a means to an end but a starting point to explore, interpret, and discuss literary collections'. Similarly, Sinclair [SRR13] argues that 'a visualization that produces a single output for a given body of material is of limited usefulness; a visualization that provides many ways to interact with the data, viewed from different perspectives, is better; a visualization that contributes to new and emergent ways of understanding the material is best'. Comprehensive case studies that scientifically debate the actual influence and impact of visualization could further specify its role and further strengthen its value as part of humanities research.

11. 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 that support a number of text analysis tasks. We classified the papers providing visualizations for historical texts according to our proposed taxonomy for text analysis tasks in digital humanities,
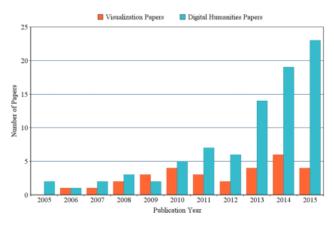
Figure 17: *Papers published in visualization and digital humanities communities by year.*

categorized applied close and distant reading techniques and analysed methods of combining both views to allow for multi-faceted 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. Figure 17 shows the temporal distribution of the papers in our collection. The trend of related works published within the digital humanities reflects the increasing value of close and distant reading visualizations for text analysis tasks in the recent years. Until now, the visualization community did not notice or consider these needs. The reason for this may lie in the obstacles encountered in publishing application papers with a digital humanities background because the often demanded quantitative evaluations are hard to perform due to the usually limited number of collaborating humanities scholars.

To strike a balance between our discussed shortcomings, we listed future challenges to support humanities scholars' tasks with close and distant reading. Developing solutions could provide beneficial contributions to 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.

References

[ADG11] ALBERS D., DEWEY C., GLEICHER M.: Sequence surveyor: Leveraging overview for scalable genomic alignment visualization. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (Dec. 2011), 2392–2401.

[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* (University of Illinois, Urbana-Champaign, USA, June 2007).

[AGZH15] ALEX B., GROVER C., ZHOU K., HINRICHS U.: Palimpsest: Improving assisted curation of loco-specific literature. In *Proceedings of the Digital Humanities* (University of Western Sydney, Australia, July 2015).

[AKV*14] ALEXANDER E., KOHLMANN J., VALENZA R., WITMORE M., GLEICHER M., Serendip: Topic model-driven visual exploration of text corpora. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)* (Oct. 2014), pp. 173–182.

[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* (University of Maryland, College Park, Maryland, USA, June 2009).

[Arc15] Internet Archive: http://www.archive.org (2015) (Retrieved 2015-01-09).

[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 B. Preim, P. Rheingans, and H. Theisel (Eds.). *Computer Graphics Forum*, vol. 32, Wiley Online Library (2013), pp. 381–390.

[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* (University of Lausanne – EPFL, Switzerland, July 2014).

[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* (University of Hamburg, Germany, 2012).

[BB15a] BESHERO-BONDAR E.: Visualizing the digital Mitford Project's prosopography data. In *Proceedings of the Digital Humanities* (University of Western Sydney, Australia, July 2015).

[BB15b] BESHERO-BONDAR E.: World-view from poetic structure: An 'Anti-Social' network analysis of Robert Southey's and Eramus Darwin's epic poems. In *Proceedings of the Digital Humanities* (University of Western Sydney, Australia, July 2015).

[BDF*10] BEZERIANOS A., DRAGICEVIC P., FEKETE J., BAE J., WATSON B.: GeneaQuilts: A system for exploring large genealogies. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (Nov. 2010), 1073–1081.

[Bea08] BEAVAN D.: Glimpses though the clouds: Collocates in a new light. In *Proceedings of the Digital Humanities* (University of Oulu, Finland, June 2008).

[Bea11] BEAVAN D.: ComPair: Compare and visualise the usage of language. In *Proceedings of the Digital Humanities* (Stanford University, USA, June 2011).

[Bea12] BEAVAN D.: DiaView: Visualise cultural change in diachronic corpora. In *Proceedings of the Digital Humanities* (University of Hamburg, Germany, July 2012).

[Bea14] BEALS M.: TEI for close reading: Can it work for history? http://tinyurl.com/nvdndsb (2014) (Retrieved 2015-01-09).

[Ben14] BENNER D. C.: The sounds of the Psalter: Computational analysis of soundplay. *Literary and Linguistic Computing 29*, 3 (2014), 361–378.

[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* (University of Lausanne – EPFL, Switzerland, July 2014).

[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* (King's College London, UK, July 2010).

[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* (University of Lausanne – EPFL, Switzerland, July 2014).

[BHW11] BINGENHEIMER M., HUNG J.-J., WILES S.: Social network visualization from TEI data. *Literary and Linguistic Computing 26*, 3 (2011), 271–278.

[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.

[BM13] BREHMER M., MUNZNER T.: A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2376–2385.

[BNJ03] BLEI D. M., NG A. Y., JORDAN M. I., Latent Dirichlet allocation. *The Journal of Machine Learning Research 3* (2003), 993–1022.

[Boo13] BOOTH A.: Documentary social networks: Collective biographies of women. In *Proceedings of the Digital Humanities* (University of Nebraska-Lincoln, USA, July 2013).

[Boy13] BOYLES N.: Closing in on close reading. *Educational Leadership 70*, 4 (2013), 36–41.

[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* (King's College London, UK, July 2010).

[Bra12] BRADLEY A. J.: Violence and the digital humanities text as Pharmakon. In *Proceedings of the Digital Humanities* (University of Hamburg, Germany, July 2012).

[BW08] BYRON L., WATTENBERG M.: Stacked graphs: Geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (Nov. 2008), 1245–1252.

[CAA*14] CORRELL M., ALEXANDER E., ALBERS D., SARIKAYA A., GLEICHER M.: Navigating reductionism and holism in evaluation. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization* (New York, NY, USA, 2014), BELIV '14, ACM, pp. 23–26.

[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* (University of Illinois, Urbana-Champaign, USA, June 2007).

[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* (University of Lausanne – PFL, Switzerland, 2014).

[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* (University of Hamburg, Germany, July 2012).

[CL13] COLES K., LEIN J. G.: Solitary mind, collaborative mind: close reading and interdisciplinary research. In *Proceedings of the Digital Humanities* (University of Nebraska-Lincoln, USA, July 2013).

[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. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (Dec. 2011), 2412–2421.

[CLWW14] CUI W., LIU S., WU Z., WEI H.: How hierarchical topics evolve in large text corpora. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (Dec. 2014), 2281–2290.

[CMS99] CARD S. K., MACKINLAY J. D., SHNEIDERMAN B.: *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, Burlington, Massachusetts, USA, 1999.

[Cob05] COBURN A.: Text modeling and visualization with network graphs. In *Proceedings of the Digital Humanities* (University of Victoria, British Columbia, Canada, June 2005).

[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).

[CPV09] CLEMENT T., PLAISANT C., VUILLEMOT R.: The story of one: Humanity scholarship with visualization and text analysis. In *Proceedings of the Digital Humanities* (University of Maryland, College Park, Maryland, USA, June 2009).

[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* (University of Lausanne – EPFL, Switzerland, July 2014).

[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.

[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.

[CVW09] COLLINS C., VIEGAS F., WATTENBERG M., Parallel tag clouds to explore and analyze faceted text corpora. In *IEEE Symposium on Visual Analytics Science and Technology, 2009. VAST 2009* (Oct. 2009), pp. 91–98.

[CWG11] CORRELL M., WITMORE M., GLEICHER M.: Exploring collections of tagged text for literary scholarship. *Computer Graphics Forum 30*, 3 (2011), 731–740.

[DCCW08] DÖRK M., CARPENDALE S., COLLINS C., WILLIAMSON C.: VisGets: Coordinated visualizations for web-based information exploration and discovery. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (Nov. 2008), 1205–1212.

[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).

[DNCM14] DYE D. J., NAPOLIN J. B., CORNELL E., MARTIN W.: Digital Yoknapatawpha: Interpreting a Palimpsest of place. In *Proceedings of the Digital Humanities* (University of Lausanne – EPFL, Switzerland, July 2014).

[Dru11] DRUCKER J.: Humanities approaches to graphical display. *Digital Humanities Quarterly 5*, 1 (2011).

[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 *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)* (Oct 2012), pp. 93–102.

[Ede14] EDER M.: Stylometry, network analysis, and Latin literature. In *Proceedings of the Digital Humanities* (University of Lausanne – EPFL, Switzerland, July 2014).

[EJ14] EVANS C., JASNOW B.: Mapping Homer's catalogue of ships. *Literary and Linguistic Computing 29*, 3 (2014), 317–325.

[eMa15] eMargin: http://eMargin.bcu.ac.uk/ (2015) (Retrieved 2015-01-09).

[ESK14] EISENSTEIN J., SUN I., KLEIN L. F.: Exploratory thematic analysis for historical newspaper archives. In *Proceedings of the Digital Humanities* (University of Lausanne – EPFL, Switzerland, July 2014).

[EX10] ESTEVA M., XU W.: Finding stories in the archive through paragraph alignment. In *Proceedings of the Digital Humanities* (King-s College London, UK, July 2010).

[FGM05] FINKEL J. R., GRENAGER T., MANNING C.: Incorporating non-local information into information extraction systems by Gibbs sampling. In *Proceedings of the 43rd Annual Meeting on*

Association for Computational Linguistics (2005), Association for Computational Linguistics, pp. 363–370.

[Fin10] FINN E.: The social lives of books: Mapping the ideational networks of Toni Morrison. In *Proceedings of the Digital Humanities* (King's College London, UK, July 2010).

[FKT14] FANKHAUSER P., KERMES H., TEICH E.: Combining macro- and microanalysis for exploring the construal of scientific disciplinarity. In *Proceedings of the Digital Humanities* (University of Lausanne – EPFL, Switzerland, July 2014).

[FMT15] FRANZINI G., MAHONY S., TERRAS M.: A catalogue of digital editions. In *Scholarly Digital Editions: Theory, Practice and Future Perspectives*. E. Pierazzo and M. J. Driscoll (Eds.). Open Book Publishers (Cambridge, United Kingdom, 2015).

[FS11] FORSTALL C., SCHEIRER W. J.: Visualizing sound as functional N-grams in Homeric Greek poetry. In *Proceedings of the Digital Humanities* (Stanford University, USA, June 2011).

[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 14*, 4 (2015), 273–288.

[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* (University of Lausanne – EPFL, Switzerland, July 2014).

[Geo15] GeoNames Gazetteer: http://www.geonames.org/ (2015). (Retrieved 2015-01-10).

[GH11a] GOODWIN J., HOLBO J., *Reading Graphs, Maps, Trees: Responses to Franco Moretti*. Parlor Press, Anderson, SC, 2011.

[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.

[GO12] GIBBS F., OWENS T.: Building better digital humanities tools: Toward broader audiences and user-centered designs. *Digital Humanities Quarterly 6*, 2 (2012).

[Goo15] Google Books: https://books.google.com/ (2015) (Retrieved 2015-01-09).

[Got15] Gothenburg model: http://wiki.tei-c.org/index.php/Textual_Variance (2015) (Retrieved 2015-10-06).

[GTAHS15] GRAY S. J., TERRAS M., AMMANN R., HUDSON-SMITH A.: Textal: Unstructured text analysis workflows through interactive smartphone visualisations. In *Proceedings of the Digital Humanities* (University of Western Sydney, Australia, July 2015).

[GTW13] GOODING P., TERRAS M., WARWICK C.: The myth of the new: Mass digitization, distant reading, and the future of the book. *Literary and Linguistic Computing 28*, 4 (2013), 629–639.

[GWFI14] GOFFIN P., WILLETT W., FEKETE J.-D., ISENBERG P.: Exploring the placement and design of word-scale visualizations. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (Dec. 2014), 2291–2300.

[GZ12] GEDZELMAN S., ZANCARINI J.-C.: HyperMachiavel: A translation comparison tool. In *Proceedings of the Digital Humanities* (University of Hamburg, Germany, July 2012).

[HAC*15] HINRICHS U., ALEX B., CLIFFORD J., WATSON A., QUIGLEY A., KLEIN E., COATES C. M.: Trading consequences: A case study of combining text mining and visualization to facilitate document exploration. *Digital Scholarship in the Humanities 30*, 1 (2015), i50–i75.

[HAHB15] HEUSER R., ALGEE-HEWITT M., BENDER J.: Knowledge networks, juxtaposed: Disciplinarity in the Encyclopédie and Wikipedia. In *Proceedings of the Digital Humanities* (University of Western Sydney, Australia, July 2015).

[HAHT*15] HEUSER R., ALGEE-HEWITT M., TRAN V., LOCKHART A., STEINER E.: Mapping the emotions of London in fiction, 1700-1900: A crowdsourcing experiment. In *Proceedings of the Digital Humanities* (University of Western Sydney, Australia, July 2015).

[Haw00] HAWTHORN J.: *A Glossary of Contemporary Literary Theory*. Oxford University Press, Oxford, United Kingdom, 2000.

[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* (University of Lausanne – EPFL, Switzerland, July 2014).

[HDvHM*15] HAENTJENS DEKKER R., VAN HULLE D., MIDDELL G., NEYT V., VAN ZUNDERT J.: Computer-supported collation of modern manuscripts: CollateX and the Beckett Digital Manuscript Project. *Digital Scholarship in the Humanities 30*, 3 (2015), 452–470.

[Hen14] HENSELER C.: Minecraft anyone? Encouraging a new generation of computer scientists and humanists. http://tinyurl.com/lk58xlv (2014). (Retrieved 2015-01-09).

[HFM16] HINRICHS U., FORLINI S., MOYNIHAN B.: Speculative practices: Utilizing InfoVis to explore untapped literary collections. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (Jan. 2016), 429–438.

[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).

[Hoc04] HOCKEY S.: The history of humanities computing. In *A Companion to Digital Humanities*, S. Schreibman, R. Siemens and J. Unsworth (Eds.). Blackwell Publishing Ltd, Malden, MA, USA (2004), pp. 3–19.

[HPR14] HOYT E., PONTO K., ROY C.: Visualizing and analyzing the Hollywood screenplay with ScripThreads. *Digital Humanities Quarterly 8*, 4 (2014).

[HSC08] HINRICHS U., SCHMIDT H., CARPENDALE S.: EMDialog: Bringing information visualization into the museum. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (Nov. 2008), 1181–1188.

[Jas01] JASINSKI J.: *Rhetoric and Society: Sourcebook on Rhetoric: Key Concepts in Contemporary Rhetorical Studies*, vol. 4. Sage Publications, Thousand Oaks, California, 2001.

[JBR*15] JÄNICKE S., BLUMENSTEIN J., RÜCKER M., ZECKZER D., SCHEUERMANN G.: Visualizing the results of search queries on ancient text corpora with tag pies. *Digital Humanities Quarterly* (2016), to appear.

[JFCS15] JÄNICKE S., FRANZINI G., CHEEMA M. F., SCHEUERMANN G.: On close and distant reading in digital humanities: A survey and future challenges. In *Eurographics Conference on Visualization (EuroVis): STARs* (2015), Borgo R., Ganovelli F., Viola I., (Eds.), The Eurographics Association.

[JFS16] JÄNICKE S., FOCHT J., SCHEUERMANN G.: Interactive visual profiling of musicians. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (Jan. 2016), 200–209.

[JG15] JÄNICKE S., GESSNER A.: A distant reading visualization for variant graphs. In *Proceedings of the Digital Humanities* (University of Western Sydney, Australia, July 2015).

[JGBS14] JÄNICKE S., GESSNER A., BÜCHLER M., SCHEUERMANN G., Visualizations for text re-use. *GRAPP/IVAPP* (2014), 59–70.

[JGF*15] JÄNICKE S., GESSNER A., FRANZINI G., TERRAS M., MAHONY S., SCHEUERMANN G.: TRAViz: A visualization for variant graphs. *Digital Scholarship in the Humanities 30*, suppl 1 (2015), i83–i99.

[JHSS12] JÄNICKE S., HEINE C., STOCKMANN R., SCHEUERMANN G.: Comparative visualization of geospatial-temporal data. In *GRAPP/IVAPP* (2012), pp. 613–625.

[JKH*15] JOHN M., KOCH S., HEIMERL F., MÜLLER A., ERTL T., KUHN J.: Interactive visual analysis of German poetics. In *Proceedings of the Digital Humanities* (University of Western Sydney, Australia, July 2015).

[Joc12] JOCKERS M.: Computing and visualizing the 19th-century literary Genome. In *Proceedings of the Digital Humanities* (University of Hamburg, Germany, July 2012).

[Joc13] JOCKERS M. L.: *Macroanalysis: Digital Methods & Literary History*. University of Illinois Press, Champaign, Illinois, 2013.

[JOL*15] JÄHNICHEN P., OESTERLING P., LIEBMANN T., HEYER G., KURAS C., SCHEUERMANN G.: Exploratory search through interactive visualization of topic models. In *Proceedings of the Dig-

ital Humanities (University of Western Sydney, Australia, July 2015).

[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* (University of Maryland, College Park, Maryland, USA, June 2009).

[JW13] JÄNICKE S., WRISLEY D. J.: Visualizing uncertainty: How to use the fuzzy data of 550 medieval texts? In *Proceedings of the Digital Humanities* (University of Nebraska-Lincoln, USA, July 2013).

[Kau15] KAUFMAN M.: 'Everything on Paper Will Be Used Against Me': Quantifying Kissinger. In *Proceedings of the Digital Humanities* (University of Western Sydney, Australia, July 2015).

[KBK11] KRSTAJIC M., BERTINI E., KEIM D.: CloudLines: Compact display of event episodes in multiple time-series. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (Dec. 2011), 2432–2439.

[KG13] KEHOE A., GEE M.: eMargin: A collaborative textual annotation tool. *Ariadne 71* (July 2013).

[KJW*14] KOCH S., JOHN M., WORNER M., MULLER A., ERTL T.: VarifocalReader: In-depth visual analysis of large text documents. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (Dec. 2014), 1723–1732.

[KKC15] KERRACHER N., KENNEDY J., CHALMERS K.: A task taxonomy for temporal graph visualisation. *IEEE Transactions on Visualization and Computer Graphics 21*, 10 (Oct. 2015), 1160–1172.

[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* (Stanford University, USA, June 2011).

[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.

[Kle12] KLEIN L. F.: Social network analysis and visualization in 'The Papers of Thomas Jefferson'. In *Proceedings of the Digital Humanities* (University of Hamburg, Germany, July 2012).

[KO07] KEIM D., OELKE D., Literature fingerprinting: A new method for visual literary analysis. In *IEEE Symposium on Visual Analytics Science and Technology, 2007. VAST 2007.* (Oct. 2007), pp. 115–122.

[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.

[KZ14] KRAUSE T., ZELDES A.: ANNIS3: A new architecture for generic corpus query and visualization. *Literary and Linguistic Computing 31*, 1 (2014), 118–139.

[LIJ*14] LEHMANN J., ISELE R., JAKOB M., JENTZSCH A., KONTOKOSTAS D., MENDES P. N., HELLMANN S., MORSEY M., VAN KLEEF P., AUER S., BIZER, C.: DBpedia: A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web Journal 5* (2014), 1–29.

[LPP*06] LEE B., PLAISANT C., PARR C. S., FEKETE J.-D., HENRY N.: Task taxonomy for graph visualization. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization* (New York, NY, USA, 2006), BELIV '06, ACM, pp. 1–5.

[LRKC10] LEE B., RICHE N., KARLSON A., CARPENDALE S.: SparkClouds: Visualizing trends in tag clouds. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (Nov. 2010), 1182–1189.

[LWW*13] LIU S., WU Y., WEI E., LIU M., LIU Y.: StoryFlow: Tracking the evolution of stories. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (Dec. 2013), 2436–2445.

[Mar12] MARCHE S.: Literature is not data: Against digital humanities: http://www.lareviewofbooks.org/article.php?id=1040 (2012) (Retrieved 2015-01-09).

[MBL*06] MEHLER A., BAO Y., LI X., WANG Y., SKIENA S.: Spatial analysis of news sources. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (Sept. 2006), 765–772.

[McC15] MCCABE M. M.: *Platonic Conversations*. Oxford University Press, Oxford, United Kingdom, 2015.

[MFM08] MENESES L., FURUTA R., MALLEN E.: Exploring the biography and artworks of Picasso with interactive calendars and timelines. In *Proceedings of the Digital Humanities* (University of Oulu, Finland, June 2008).

[MFM13] MENESES L., FURUTA R., MANDELL L.: Ambiances: A framework to write and visualize poetry. In *Proceedings of the Digital Humanities* (University of Nebraska-Lincoln, USA, July 2013).

[MH13] MURALIDHARAN A., HEARST M. A.: Supporting exploratory text analysis in literature study. *Literary and Linguistic Computing 28*, 2 (2013), 283–295.

[MLCM16] MCCURDY N., LEIN J., COLES K., MEYER M.: Poemage: Visualizing the sonic topology of a poem. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (Jan. 2016), 439–448.

[MLSU13] MILLER B., LI F., SHRESTHA A., UMAPATHY K.: Digging into human rights violations: Phrase mining and trigram visualization. In *Proceedings of the Digital Humanities* (University of Nebraska-Lincoln, USA, July 2013).

[Mor05] MORETTI F.: *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso, (London and New York, July 2005).

[Mor13] MORETTI F.: *Distant Reading*. Verso, (London and New York, 2013).

[MRMK15] MEDEK A., RITTER J., MOLITOR P., KÖSSER S.: Interactive similarity analysis of early new high German text variants. In *Proceedings of the Digital Humanities* (University of Western Sydney, Australia, July 2015).

[MRO*12] MACEACHREN A., ROTH R., O'BRIEN J., LI B., SWINGLEY D., GAHEGAN M.: Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (Dec. 2012), 2496–2505.

[MSR*15] MONTAGUE J., SIMPSON J., ROCKWELL G., RUECKER S., BROWN S.: Exploring large datasets with topic model visualizations. In *Proceedings of the Digital Humanities* (University of Western Sydney, Australia, July 2015).

[Mun09] MUNZNER T.: A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 921–928.

[Mur11] MURALIDHARAN A.: A visual interface for exploring language use in slave narratives. In *Proceedings of the Digital Humanities* (Stanford University, USA, June 2011).

[NMG*13] NOWVISKIE 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.

[OGH15] ODAT S., GROZA T., HUNTER J.: Extracting structured data from publications in the Art Conservation Domain. *Digital Scholarship in the Humanities 30*, 2 (2015), 225–245.

[OKK13] OELKE D., KOKKINAKIS D., KEIM D. A.: Fingerprint matrices: Uncovering the dynamics of social networks in prose literature. In *Computer Graphics Forum*, B. Preim, P. Rheingans, and H. Theise (Eds.). (2013), vol. 32, Wiley Online Library, pp. 371–380.

[Ome15] Omeka: http://www.omeka.org/ (2015) (Retrieved 2015-01-10).

[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* (University of Lausanne – EPFL, Switzerland, July 2014).

[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 *2010 IEEE Symposium on Visual Analytics Science and Technology (VAST)* (Oct. 2010), pp. 91–98.

[Pal02] PALEY W. B.: TextArc: Showing word frequency and distribution in text. In *Poster Presented at IEEE Symposium on Information Visualization* (2002), vol. 2002.

[PBD14] PE NA E., BROWN M., DOBSON T.: On metaphor in text visualization prototypes. In *Proceedings of the Digital Humanities* (University of Lausanne – EPFL, Switzerland, July 2014).

[Per15] *Perseus Digital Library*: R. Gregory (Ed.). Crane. Tufts University. http://www.perseus.tufts.edu (2015) (accessed March 19, 2015).

[Pet14] PETERSON N.: Visualization as a bridge to close reading: The audience in the castle of perseverance. In *Proceedings of the Digital Humanities* (University of Lausanne – EPFL, Switzerland, July 2014).

[PHI15] PHI Latin Texts: http://latin.packhum.org/ (2015) (Retrieved 2015-01-09).

[Pie10] PIEZ W.: Towards hermeneutic markup: An architectural outline. In *Proceedings of the Digital Humanities* (King's College London, UK, July 2010).

[Pie13] PIEZ W.: Markup beyond XML. In *Proceedings of the Digital Humanities* (University of Nebraska – Lincoln, USA, July 2013).

[Ple15] Pleiades: A community-built gazetteer and graph of ancient places: http://pleiades.stoa.org/ (2015) (Retrieved 2015-10-06).

[PMMR15] PÖCKELMANN M., MEDEK A., MOLITOR P., RITTER J.: _CATview_: Supporting the investigation of text genesis of large manuscripts by an overall interactive visualization tool. In *Proceedings of the Digital Humanities* (University of Western Sydney, Australia, July 2015).

[Poe16] Poemage: http://www.sci.utah.edu/nmccurdy/Poemage/ (2016) (Retrieved 2016-03-18).

[Poi15] POIBEAU T.: Generating navigable semantic maps from social sciences corpora. In *Proceedings of the Digital Humanities* (University of Western Sydney, Australia, July 2015).

[Pos07] POSAVEC S.: Literary organism http://www.stefanieposavec.co.uk/ (2007) (Retrieved 2015-01-09).

[Pro15] Project Gutenberg: http://www.gutenberg.org/ (2015) (Retrieved 2015-01-09).

[PSA*06] PLAISANT C., SMITH M. N., AUVIL 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* (The Sorbonne, Paris, France, July 2006).

[RARC*15] ROE G., ABDUL-RAHMAN A., CHEN M., GLADSTONE C., MORRISSEY R., OLSEN M.: Visualizing text alignments: Image processing techniques for locating 18th-century commonplaces. In *Proceedings of the Digital Humanities* (University of Western Sydney, Australia, July 2015).

[RAW*15] RIND A., AIGNER W., WAGNER M., MIKSCH S., LAMMARSCH T.: Task cube: A three-dimensional conceptual space of user tasks in visualization design and evaluation. *Information Visualization* (Dec. 2015).

[RD10] RICHE N., DWYER T.: Untangling Euler diagrams. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (Nov. 2010), 1090–1099.

[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.

[RPSF15] RIEHMANN P., POTTHAST M., STEIN B., FROEHLICH B.: Visual assessment of alleged plagiarism cases. *Computer Graphics Forum 34*, 3 (2015), 61–70.

[RRRG05] RUECKER S., RAMSAY S., RADZIKOWSKA M., GALEY A.: Interface design. In *Proceedings of the Digital Humanities* (University of Victoria, British Columbia, Canada, June 2005).

[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).

[SDW11] SLINGSBY A., DYKES J., WOOD J.: Exploring uncertainty in geodemographics with interactive graphics. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (Dec. 2011), 2545–2554.

[Shn96] SHNEIDERMAN B., The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of Visual Languages* (1996), pp. 336–343.

[Sin13] SINGER K.: Digital close reading: TEI for teaching poetic vocabularies. *Journal of Interactive Technology and Pedagogy 3* (2013).

[SOI10] SAITO S., OHNO S., INABA M.: A platform for cultural information visualization using schematic expressions of cube. In *Proceedings of the Digital Humanities* (King's College London, UK, July 2010).

[SRR13] SINCLAIR S., RUECKER S., RADZIKOWSKA M.: *Information Visualization for Humanities Scholars*, Modern Language Association, New York, 2013.

[TEI15] TEI Consortium (Eds.): *TEI P5: Guidelines for Electronic Text Encoding and Interchange. 2.8.0. 2015-04-06.* TEI Consortium. http://www.tei-c.org/Guidelines/P5/ (2015) (Retrieved 2015-10-09).

[TFK15] TRILCKE P., FISCHER F., KAMPKASPAR D.: Digital network analysis of dramatic texts. In *Proceedings of the Digital Humanities* (University of Western Sydney, Australia, July 2015).

[TKMS03] TOUTANOVA K., KLEIN D., MANNING C. D., SINGER Y.: Feature-rich part-of-speech tagging with a cyclic dependency net-

work. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1* (2003), Association for Computational Linguistics, pp. 173–180.

[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.

[Tra09] TRAVIS C.: Patrick Kavanagh's Poetic Wordscapes: GIS, literature and Ireland, 1922-1949. In *Proceedings of the Digital Humanities* (University of Maryland, College Park, Maryland, USA, June 2009).

[UIU98] *University of Illinois at Urbana-Champaign Digital Libraries Initiative: UIUC DLI Glossary*, 1998. http://archive.is/qOUJz (Retrieved 2016-03-18).

[Und15] UNDERWOOD T.: A dataset for distant-reading literature in English, 1700-1922. http://tinyurl.com/nu26zr7 (2015) (Retrieved 2015-10-09).

[VCPK09] VUILLEMOT R., CLEMENT T., PLAISANT C., KUMAR A., What's being said near 'Martha'? Exploring name entities in literary text collections. In *IEEE Symposium on Visual Analytics Science and Technology, 2009. VAST 2009.* (Oct. 2009), pp. 107–114.

[vHWV09] VAN HAM F., WATTENBERG M., VIEGAS F.: Mapping text with phrase nets. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (Nov. 2009), 1169–1176.

[VWF09] VIEGAS F., WATTENBERG M., FEINBERG J.: Participatory visualization with Wordle. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (Nov. 2009), 1137–1144.

[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.

[Wei15] WEIGEL M.: Graphs and Legends: Raymond Williams tried to save culture from a priestly elite. Can the same be said of the digital humanities?. http://www.thenation.com/article/graphs-and-legends/ (2015) (Retrieved 2015-10-09).

[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* (Stanford University, USA, June 2011).

[Wil15a] WILLS T.: Relational data modelling of textual corpora: The Skaldic Project and its extensions. *Digital Scholarship in the Humanities 30*, 2 (2015), 294–313.

[Wil15b] WILSON E. A.: Building The Early Modern Digital University: Using social network analysis and digital visualization tools to bring the early modern network of networks (EMNON)

to life. In *Proceedings of the Digital Humanities* (University of Western Sydney, Australia, July 2015).

[WJ13a] WEINGART S., JORGENSEN J.: Computational analysis of the body in European fairy tales. *Literary and Linguistic Computing 28*, 3 (2013), 404–416.

[WJ13b] WHEELES D., JENSEN K.: Juxta commons. In *Proceedings of the Digital Humanities* (University of Nebraska-Lincoln, USA, July 2013).

[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.

[Wol13] WOLFF M.: Surveying a corpus with alignment visualization and topic modeling. In *Proceedings of the Digital Humanities* (University of Nebraska-Lincoln, USA, July 2013).

[WV08] WATTENBERG M., VIEGAS F.: The word tree, an interactive visual concordance. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (Nov. 2008), 1221–1228.

[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.

[ZNMS15] ZAHORA T., NIKULIN D., MEWS C. J., SQUIRE D.: Deconstructing Bricolage: Interactive online analysis of compiled texts with factotum. *Digital Humanities Quarterly 9*, 1 (2015).