Visual Text Analytics in Context of Digital Humanities

Mennatallah El-Assady, Valentin Gold, Markus John, Thomas Ertl, and Daniel Keim

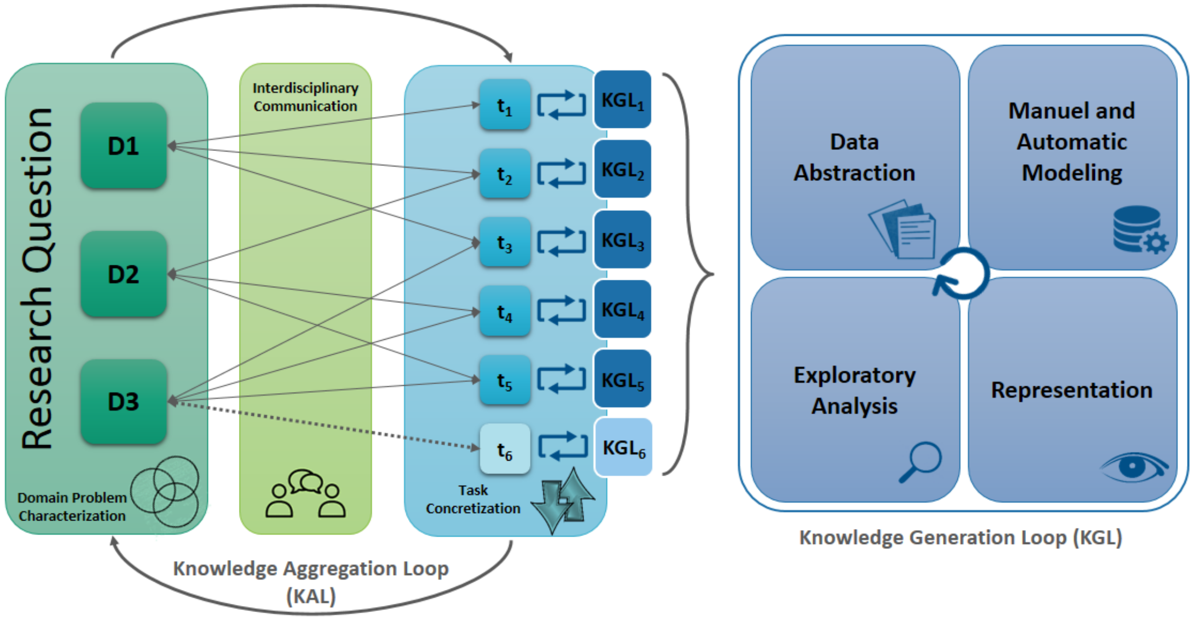

Figure 1. Conceptual workflow of the problem-solving process in digital humanities projects.The overall research question is subdivided into domain-specific questions, based on domain knowledge. In the Knowledge Aggregation Loop (KAL), single concrete tasks are derived from unsolved domain problems, at any given time in the project. Each task can be solved using a concrete implementation of some of the components in the Knowledge Generation Loop (KGL). After a task is solved, the feedback of the KAL enhances the domain knowledge and can result in further tasks that are tackled at a later point in time.

Abstract— Digital Humanities (DH) research brings together scholars from different disciplines to work on tackling a common research challenge. Hence, DH-projects have to overcome common challenges of multi-disciplinary research, such as methodological differences or communication issues. However, in contrast to interdisciplinary collaborations from related fields of science, in the digital humanities, a gap between qualitative scholarship traditions and quantitative data-driven research has to be bridged to achieve a common goal. With this position paper, we aim at starting a discussion between the various involved disciplines in the digital humanities on how to approach the problem-solving process in DH-projects. Based on our experience in different visual text analytics projects and extensive two years of discussions, we propose an abstract conceptual workflow as a best practice for digital humanities projects.

1 Introduction

Projects in the field of digital humanities are often characterized by collaborations between various scientific disciplines. One of the main challenges to scholars working in these projects is applying a procedural framework addressing a common research interest while still focusing on their own domain-specific research questions. DH-projects need to bridge the gap between various scientific cultures, most importantly the gap between qualitative research traditions and quantitative data approaches [RR12, 72]. For instance, scholars in the humanities often apply the interpretative methodology while the emerging field of digital humanities supports quantitative visual and statistical approaches supporting the interpretation and evaluation of specific hypotheses. These hypotheses are derived from the common research interest and address both domain-independent and domain-specific causal relations. To foster successful collaborations, in this paper, we propose an abstract conceptual framework supporting scholars collaborating in DH-projects to initiate a reflexive process between the common research interest and the domain-specific questions.

In general, the computational turn in the humanities has resulted in a fundamental change of the research methodologies in all its subfields [Ber11, 1]. In our understanding, however, a digital humanities project is not just the mere application of known computational methods to solve research problems in the humanities but is a collaboration of different fields in the humanities and computer science which strives to solve research questions in a process that encompasses scientific contributions for all involved disciplines.

In this position paper, we propose a conceptual framework that models the process of problem solving and collaboration within DH-projects. We derive this framework from our discussions with researchers from various disciplines of the humanities and computer science. Moreover, this framework has been revised in several prototyping cycles over the last two years, in an attempt to model the lessons learned from two visual text analytics projects into the pipeline. With the proposed framework, we contribute to establishing a *best practice* to formulate epistemic scientific foundations for the digital humanities [KSD15, 31].

- *The authors of this paper are listed in alphabetical order*
- *Correspondance to Mennatallah El-Assady, Valentin Gold, and Daniel Keim: <firstname.lastname>@uni-konstanz.de; Correspondance to Markus John and Thomas Ertl: <firstname.lastname>@vis.uni-stuttgart.de*

2 RELATED WORK

Since our presented paper is concerned with collaborative research in the digital humanities, we first briefly review literature in this area and describe two frameworks in more detail. Next, we summarize visual analytics approaches which have been developed for text and document visualization in the context of digital humanities.

2.1 Collaborative Research in the Digital Humanities

Recent developments in collaborative research highlight the specific challenges in DH-projects [DM12, Gol12] and propose (visual) frameworks to address these challenges. For instance, based on [vFB+14], [KSD15] propose the *New Visual Hermeneutics* – a methodological framework supporting knowledge generation for unstructured text relying on information visualizations. This framework separates four distinctive stages: In the first stage, the raw data is sampled and preprocessed. Second, text-mining and information-retrieval algorithms are applied to analyze the preprocessed data. In the third stage, visualizations are presented to the scholars allowing in the fourth stage the hermeneutic interpretation of the data. Each of the four stages is connected interdependently and support scholars to interpret the analyzed phenomenon from various angles. Within this framework, a special focus is put on the applied algorithms as they are seen as having a transformative power on the interpretations of the data, cf. [Ram03, Ram11].

Furthermore, frameworks have been proposed addressing the issues of interdisciplinary communication, e.g. [DM11, DM12], and defining the values of digital humanities [Spi12]. To foster collaboration, Simon et al. [SMKS15] suggest to improve the common understanding between the disciplines by establishing a Liaison – experts that mediate the knowledge and tasks between the domains. A Liaison should have obtained domain knowledge in various disciplines and contributes to the project by capturing the problem complexity or the mental model. In general, visual evidence is seen as a powerful tool to achieve mutual understanding [RR12, 70]. Consequently, our proposed framework includes a visual component supporting collaborative research.

2.2 Visual Text Analytics

In recent years several techniques for visual text analytics were introduced to support humanities scholars and experts from other domains. For example, word clouds [WV08], topological landscapes [WTP+95], pixel based overview [KO07], or clustering and classification techniques [HKBE12] were developed to support analysts in exploring and analyzing text corpora on abstract levels. Normally, humanities scholars read and analyze a text in a sequential manner by so-called close reading. In contrast to this, in the DH, there is a recent trend to develop methods that facilitate distant reading. Moretti [Mor05] introduced the idea of distant reading, which abstracts the text by providing visualizations such as graphs, maps or trees. There are quite a number of visual text analytics approaches in the field of digital humanities that support close and distant reading [JFCS15]. [JGBS14] and [ARRO+16] propose several techniques for the visual representation and comparison of reused text in document collections in order to facilitate humanities scholars in discovering intertextual similarities. [VCPK09] and [JLK+16] present approaches which support the character analysis in novels by providing several interrelated visual abstraction views that enable the interactive switching to the corresponding text passage in the document. [CWG11] and [KJW+14] propose approaches that combine close and distant reading by providing both, the distribution of findings and the possibility to inspect a specific text passage in detail. [TKK11] and [JHSS12] present interactive systems that support analysts in exploring and understanding geospatial-temporal data. [EAGA+16] introduced an approach to analyze speaker behavior patterns in multi-party conversations.

3 CONCEPTUAL WORKFLOW

We propose a generic pipeline for modeling a joint and collaborative workflow of research projects in the field of digital humanities, with a focus on projects that have a visual text analysis component. This process pipeline is subdivided into two parts; the Knowledge Aggregation Loop (KAL) and the Knowledge Generation Loop (KGL). Both loops play an important role in addressing common research interests. However, the KAL is particularly relevant to DH-projects. In general, the KGL is an abstract module that can be adapted to the concrete problems and tasks that are being addressed.In the following sections, both loops and their interactions are explained in more detail.

One of the key characteristics of projects in the field of digital humanities is the common research interest of participating researchers from different disciplines [Hay12] . However, in contrast to other areas in which collaborative research is conducted, digital humanities projects bring together disciplines that do not share a broad technical, epistemological, methodological, or conceptual common ground. Hence, these researchers do not speak the same language to refer to a common concept and have diverging mental models towards the research challenges of the project. In order to identify mutual research interests and to communicate the concepts and terms of each discipline for a better understanding, the Knowledge Aggregation Loop is essential. Hereby, starting with a mutual research question, each domain identifies relevant sub-questions and sub-challenges that revolve around but are not limited to, the interests of their specific discipline. These sub-questions are subsequently disaggregated to build a set of concrete tasks, which themselves become the starting points for the second loop.

3.1 Knowledge Aggregation Loop

Before proposing a solution to the common research question, all participating domains have to formulate a set of conceptual tasks that contribute to the solution of their domain-specific research challenges, as illustrated in Figure 1. To derive these concrete tasks, and in accordance to [SMKS15], a layer of *Interdisciplinary Communication* is introduced to the Knowledge Aggregation Loop. Given our experiences in project collaborations, this layer is essential for the success of DH-projects as it supports the identification of challenges and fosters mutual understanding. In the following sections, we describe each of the components of the KAL in more detail.

Domain Problem Characterization

The starting point of the model is the high-level research interest that is typically coming from the application domain or that is derived from the data. This general research interest defines the research question and the interdisciplinary goal of the project. In most circumstances, the research question originates from one of the participating disciplines. For instance, in the VisArgue project – which is introduced in more detail in section 4.2, the overall research question was introduced by the domain of political science and is defined by the theory of deliberation.

In our understanding, the high-level research interest is to be understood as a general question that defines the scope of the project. In contrast to natural and –to some extent– social sciences, research questions in the (digital) humanities allow a broader specification. Hence, also the "hermeneutics of digging into data" is per se a valuable research question. However, every research question needs to be *clear* (i.e. allows only one interpretation), *focused* (i.e. concentrates on specific aspects), *concise* (i.e. based on rigorous conceptualization), *complex* (i.e. includes multiple layers) and *arguable* (i.e. addresses a research gap).

A solution to the question posed by the application domain typically exceeds the capabilities of this discipline. Therefore, to tackle this challenge, the expertise of other domains is required. Since the research questions for a digital humanities project are usually complex generic constructs, the involvement of multiple disciplines is needed to construct a system with the ability to solve this problem. During the course of the project, all participating domains aim at learning, explaining, and deriving conclusions about the research question in order to answer the high-level research interest at the final step using the knowledge generated during the construction of a solution.

Interdisciplinary Communication

To bridge the gap between the disciplines and develop a true cross-discipline solution, each domain has to find relevant research questions and tasks within the overall project research question. In Fig.1, this is illustrated by three sub-questions (D1, D2, D3) that result from the mutual research question. The tasks (t_1, t_2, ..., t_5) are derived from these questions and evolve through the interdisciplinary communication of concrete problems. Hence, derived from the overall research question, each discipline might pose various hypotheses that help to answer relevant aspects of the research question.

In our point of view, an essential factor for a successful collaboration is the layer of interdisciplinary communication. In general, this layer includes–but is not limited to–the adaptation of similar terminology, intensive debates on mental models, the adaptation of different perspectives, and the creation of a common ground. To foster digital humanities collaborations, projects can also adopt one of the many proposed modes of interdisciplinary communication, see e.g. [Hol13].

Task Concretization

When the goal and high-level tasks of the project and all involved researchers are clearly defined and a similar terminology is adopted, the next step for the project is to derive smaller specific work packages and concrete tasks that can be solved independently, contributing to the solution of the overall research challenge. These tasks are defined by breaking down the domain specific research questions to measurable observable constructs. The abstract tasks (constructs) might apply to only one discipline–e.g. defining the corpus of data to be analyzed–or to more than one discipline–e.g. the application of a natural linguistic preprocessing pipeline annotating theoretical constructs. Over the course of the project, the KAL supports the generation of new tasks over time and the accumulation of new domain knowledge through the feedback from the solution of each concrete task.

In general, two approaches can be used to define the tasks. First, the tasks can be deduced from theoretical components of the high-level research interest. By using this ”top-down approach”, empirical measures directly refer to the theory defining the research question. The tasks are restricted to theoretical elements of the research question. On the other hand, a ”bottom-up approach” might also be used to contribute to answering the research question of the project. By adopting this approach, the tasks might, on a first sight, be unrelated to the research question and–consequently–also to the theory that defines the scope of the project. Only on a later stage, certain tasks might prove useful to the research question. This approach allows for a broader and non-restricted application of methods.

3.2 Knowledge Generation Loop

The Knowledge Generation Loop represents an interactive and iterative process and consists of several steps. These interchangeable steps can include Data Abstraction, Modeling, Representation, and Exploratory Analysis. Hence, this loop can be extended and adapted to concrete problems and tasks. During this iterative process, analysts attempt to find evidence for existing hypotheses and learn new knowledge about the specific domain problem. Collected insights can lead to developing new research ideas and form hypotheses, as well as to help to solve the project research question.

Data Abstraction

Normally an analysis starts with data and analysis tasks. For the analysis, domains bring in their data (e.g. books or dialogues), knowledge about the data, as well as analysis systems and methods to handle this data. The data to be analyzed may come from different domains with various formats or the data can be inconsistent and noisy. Therefore it is necessary to preprocess the data in order to perform any data modeling methods. Data preprocessing steps can involve automatic tasks such as cleaning, integration, transformation, reduction, or feature extraction. However, there are data preprocessing methods which cannot be automatized. They can be operationalized for machine processing only in stages. For example, the digitization of a corpus in the form of image representation and as digital full texts – recorded and corrected by a „double keying“ process.

Data Modeling

Data modeling refers to a group of processes in which multiple sets of data and tasks can be combined to solve the Unit of Analysis (UoA). UoA is the first step in deciding how the given data will be analyzed. It is the major entity that is being analyzed (e.g. the “what” or “who” that is being studied). To find and extract interesting information and knowledge from text, natural language processing (NLP) techniques are the methods of choice for automatic analysis. Representative tasks can include for example concept and entity extraction, text classification, text clustering, or sentiment analysis. Since NLP methods are normally trained on specific text corpora, such as newspaper or journal article texts, they do not provide entirely correct results for texts that differ from the training corpus. Therefore, a manual annotation can play an important role in this step. For example, they can be used as training data for the automatic methods to improve their accuracy or in cases which are too complex for an automatic processing.

Representation

Visual representations can map data to visual variables to highlight features such as similarities or extracted information and identify patterns. As aforementioned, Moretti introduced the idea of distant reading by providing visual abstractions as overviews. These visual representations can support humanities scholars in analyzing and understanding complex information. Furthermore, information visualization can be enhanced and extended from only static images to interactive visual analytics systems, through human interaction with the visualization and the data model. This is useful for uncovering insightful patterns and information, in addition to evaluating and refining the underlying data model. Since the visual representations are often derived through automatic processing, which can be a source of uncertainties, interactive visualization can support analysts in identifying outliers and correcting data or modeling errors. In summary, knowledge can be gained from the interaction between the visualization, the data model, and the analyst.

Exploratory Analysis

As mentioned above visualization and interaction can play an important role in analyzing and understanding data. To generate knowledge, the Exploratory Analysis is based on the exploration and verification loop by Sacha et al. [SSS+14]. The exploration loop illustrates how users analyze the data. Analysts explore the data for example by interacting and analyzing several views to search for findings. During this process, the analysts can gain further insights, which can support them in solving different tasks or defining new ones as a starting point for a further analysis. The verification loop [SSS+14] is integrated within the exploration loop and guides analysts to confirm hypotheses.

4 Project Examples

To demonstrate the application of the proposed generic pipeline on real-world research projects, in the following subsections, two eHumanities project are described using the different components of the pipeline.

4.1 ePoetics

The eHumanities project ePoetics deals with an area which is essential for the humanities: poetics and aesthetics from 1770 until 1960. These works record thinking and writing about literature and contain systematic basic knowledge about literary studies and philosophical aesthetics. The project goal is to process the collection as a digital corpus and analyze as well as visualize it by means of information technological methods and hermeneutic procedures. To achieve these objectives, ePoetics is a research collaboration and consists of following disciplines. The literature scientists bring in the selected 20 poetics and the hermeneutic expertise, as well as research questions, such as ”how does the author of a poetic describe the concept poet” or ”what is the essence of a poetic and what is addressed by it”. The computer philologists are responsible for the digitization of the poetics which allows analyzing the texts by quantitative methods. Whereas the computational linguists focus on the automatic extraction of terms and concepts. Last, the computer scientists support the humanities by providing interactive approaches that enable them in annotating, exploring and understanding complex relationships.

At the beginning of the project, each of disciplines defined research questions based on the overall research questions from the literature scientists. For example, the computer philologists had the goal that annotations of the corpus are conformable to the conventions of the Text Encoding Initiative and that the corpus will be integrated into virtual research environments such as TextGrid or DARIAH. The computational linguists and computer scientists were interested in developing new methods or improving existing approaches to support text analysis.

In the task concretization stage of the project, the primary goal was to find out how hermeneutic and algorithmic processes can inspire, complement, and support each other. Therefore, several specific tasks were defined after long discussions. For example, the literary scholars decide which text characteristics should be annotated and provide through manual annotations the basis for automatic extraction of these characteristics. The computer philologists assist the literary scholars with their knowledge in developing the annotation guidelines and by providing annotation tools. Whereas the computational linguists have the goal to help the humanities with automatic extractions of terms and concepts. The computer scientists provide interactive visual approaches that enable fast access to the initial text sources to quickly identify and remove errors induced by the algorithmic processing.

To illustrate the usefulness of the Knowledge Generation Loop, we focus on one of the tasks in more detail. For example, the literary scholars are interested in which works of literary authors are named and cited in the different poetics. The computer linguists developed regular expressions to detect and extract quotes in the text. However, we noticed that text inside quotation marks does not always constitute a quotation from another work. For example, quotation marks are used to emphasize certain words that are important in the current context. As next step, the computer scientists developed an interactive approach that visualizes the automatically extracted annotations including their confidence value. It highlights annotations with a high uncertainty and enables users to correct or confirm these classifications and trigger a retraining to improve the classifier. On one hand, the achieved results contribute to answering overall research question from the literary scholars. On the other hand, the trained classifier and the interactive approach can be applied to further texts and thus helps to solve the specific research question of the both disciplines.

4.2 VisArgue

The VisArgue project is an eHumanities research project aiming at developing a framework for the measurement and exploration of the degree of deliberation within political discourses. Deliberation is a political science theory claiming to resolve public conflicts through extensive discourse and civil dialogue. Hereby, the discourse should be based on the principle of deliberative communication, i.e. the rational exchange of arguments to convince the opponents. For the measurement of the degree of deliberation, linguistic and statistical patterns in the conversations are analyzed. To accomplish these tasks and tackle the research challenges that emerge, a team of computational linguists, information scientists, and political scientists came together. The computational linguists bring in the expertise for analyzing the syntactic and semantic structures of deliberative communication. Whereas the computer and information scientists model the data and compute characteristic features relevant for deliberation. These are visualized in a later step to enable a detailed exploratory analysis of political discussions.

During the project finding phase each of the involved disciplines could derive domain-specific research questions. Since the overall research question was derived from the political science domain, the complex challenge of measuring the deliberative quality of political discourse had to be broken-down to relevant research questions for all involved disciplines. For instance, the computational linguists derived the questions whether deliberative patterns can be detected by analyzing syntactic and semantic structures. The information scientists were interested in visualizing deliberative patterns to assist the other disciplines in exploring and determining the deliberative quality of communication.

In the next step, several specific tasks were formulated that contribute to aggregate the knowledge that is generated while the tasks are being conducted. The tasks were derived using a top-down as well as a bottom-up approach. For instance, the dimensions of deliberation had to be extracted relying on the political science theories of deliberation. On the contrary, the computational linguists mostly relied on a data-driven bottom-up approach. Even though political science theories had not included some of the linguistic elements that were proposed by the semantic analysis of the communication, the proposed elements proved to be useful to explore and explain the degree of deliberation. Hence, both approaches–top-down and bottom-up–were combined. Finally, the information scientists derived the specific task of data representation and visualization as well as software engineering.

To demonstrate the useful application of the KGL, we explain one of the tasks in more detail. For deliberation to be determined, it is important to know the attitudes of speakers to the topics. To analyze these stances, several tasks were derived: First, the data needs to be represented on an abstract level allowing to derive relevant speaker stances. The data can also be used as training data to evaluate the automatic algorithms that are developed. Second, based on semantic, syntactic, and morphologic patterns, rules are deduced to automatically identify speaker attitudes. Finally, the data is exported to an XML-based representation of communication to allow further analysis of the data. As a result of these tasks, further tasks can be specified. For instance, the evaluation of the developed algorithms and the visual representation of the XML-annotated communication. Moreover, the link to the derived dimensions of deliberation needs to be done. Overall, the results of the tasks contribute to answering the research questions of all involved disciplines as well as the overall research interest.

4.3 Lessons Learned

To improve the problem-solving process of future DH-projects, we briefly summarize two important lessons learned. First, given our experiences, to successfully collaborate in DH-projects, extensive interdisciplinary communication is fundamental. This collaborative dialogue supports merging and adapting mental models and identify potential conflicts in terminology or concepts at an early stage of the project. Both projects–ePoetics and VisArgue– benefited from the fact that most members of the project teams shared a common workspace, as most members were part of the same university. The physical proximity allowed discussing smaller issues on a more informal ad-hoc basis. We would, therefore, like to emphasize the importance of a collaborative project management arrangement that allows team members to share a physical or virtual workspace. In general, in both projects, the collaboration between the visualization scientists and computational linguists has proven to be much easier than for the other involved disciplines. This is due to the fact that these two disciplines already share a set of mental models and common vocabulary.

Second, we propose to define clear and coherent research questions and to communicate these questions to all disciplines. In our experience, it is important to start with less complex tasks and successively add complexity. This also helps to determine the merits of interdisciplinary collaborations. Moreover, both, questions and tasks should be designed to be achieved within a given timeframe. To realize our project milestones and advance with the solution of the overall research question, having a realistic time management was crucial.

5 Conclusion

In this paper, we proposed a conceptual workflow designed for digital humanities projects. The focus of the pipeline is on projects processing textual data and with a strong emphasize on the merits of visualizations. We see this workflow as a proposal of best practice and a contribution to the discussion of the standardization of conceptual project management pipelines for DH-projects. However, we acknowledge that the proposed workflow is based on our experiences within two eHumanities projects and numerous extensive discussions and is not representative of the full range of possible projects in the field of digital humanities.

To complete the proposed pipeline, several steps could be made. For instance, we could attempt to apply the workflow to other digital humanities projects, with a different focus than visual text analytics, to examine the generalizability of this conceptual pipeline. Alternatively, we could conduct a survey of various digital humanities projects to derive at a more general best practice.

Acknowledgements

We would like to thank Stefan Alscher, Michael Bender, Steffen Koch, Nils Reiter, and Svenja Simon for intense and interesting discussions. After having discussed the conceptual workflow several times, our mental models seemed to converge.

References

[ARRO+16] A. Abdul-Rahman, G. Roe, M. Olsen, C. Gladstone, R. Whaling, N. Cronk, R. Morrissey, and M. Chen. Constructive visual analytics for text similarity detection. *Computer Graphics Forum*, pages n/a–n/a, 2016.

[Ber11] David M Berry. The Computational Turn: Thinking about the Digital Humanities. *Culture Machine*, 12:1–22, 2011.

[CWG11] M. Correll, M. Witmore, and M. Gleicher. Exploring collections of tagged text for literary scholarship. *Computer Graphics Forum*, 30(3):731–740, 2011.

[DM11] Marilyn Deegan and Willard McCarty. *Collaborative research in the digital humanities*. Ashgate, Farnham, Surrey, England; Burlington, VT, 2011.

[DM12] Marilyn Deegan and Willard McCarty. *Collaborative Research in the Digital Humanities: A Volume in Honour of Harold Short, on the Occasion of his 65th Birthday and His Retirement, September 2010*. Digital Research in the Arts and Humanities Series. Ashgate, 2012.

[EAGA+16] Mennatallah El-Assady, Valentin Gold, Carmela Acevedo, Christopher Collins, and Daniel A. Keim. ConToVi: Multi-Party Conversation Exploration using Topic-Space Views. *Computer Graphics Forum*, 2016.

[Gol12] Matthew K. Gold, editor. *Debates in the Digital Humanities*. University of Minnesota Press, Minneapolis, MN, 2012.

[Hay12] Katharine N. Hayles. How We Think: Transforming Power and Digital Technologies. In David M. Berry, editor, *Understanding Digital Humanities*, chapter 3, pages 42–66. Palgrave Macmillan, New York, NY, 2012.

[HKBE12] Florian Heimerl, Steffen Koch, Harald Bosch, and Thomas Ertl. Visual Classifier Training for Text Document Retrieval. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2839–2848, 2012.

[Hol13] J. Britt Holbrook. What is interdisciplinary communication? Reflections on the very idea of disciplinary integration. *Synthese*, 190(11):1865–1879, 2013.

[JFCS15] Stefan Jänicke, Greta Franzini, Muhammad Faisal Cheema, and Gerik Scheuermann. On close and distant reading in digital humanities: A survey and future challenges. In *Eurographics Conference on Visualization – STARs*, EuroVis '15. The Eurographics Association, 2015.

[JGBS14] Stefan Jänicke, Annette Geßner, Marco Büchler, and Gerik Scheuermann. Visualizations for text re-use. In *Proceedings of the 5th International Conference on Information Visualization Theory and Applications*, IVAPP '14, pages 59–70. Scitepress, 2014.

[JHSS12] Stefan Jänicke, Christian Heine, Ralf Stockmann, and Gerik Scheuermann. Comparative visualization of geospatial-temporal data. In *Proceedings of the 3rd International Conference on Information Visualization Theory and Applications*, IVAPP '12, pages 613–625. Scitepress, 2012.

[JLK+16] Markus John, Steffen Lohmann, Steffen Koch, Michael Wörner, and Thomas Ertl. Visual analytics for narrative text - visualizing characters and their relationships as extracted from novels. In *Proceedings of the 7th International Conference on Information Visualization Theory and Applications*, IVAPP '16, pages 27–38. Scitepress, 2016.

[KJW+14] Steffen Koch, Markus John, Michael Wörner, Andreas Müller, and Thomas Ertl. VarifocalReader – In-depth visual analysis of large text documents. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1723–1732, 2014.

[KO07] D.A. Keim and D. Oelke. Literature fingerprinting: A new method for visual literary analysis. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, VAST '07, pages 115–122, 2007.

[KSD15] Roxana Kath, Gary S. Schaal, and Sebastian Dumm. New Visual Hermeneutics. *Zeitschrift für germanistische Linguistik*, 43(1):27–51, 2015.

[Mor05] Franco Moretti. *Graphs, maps, trees: abstract models for a literary history*. Verso, 2005.

[Ram03] Stephen Ramsay. Special Section: Reconceiving Text Analysis Toward an Algorithmic Criticism. *Literary and Linguistic Computing*, 18(2):167–174, 2003.

[Ram11] Stephen Ramsay. *Reading Machines: Toward an Algorithmic Criticism*. University of Illinois Press, Urbana, 2011.

[RR12] Bernhard Rieder and Theo Röhle. Digital Methods: Five Challenges. In David M. Berry, editor, *Understanding Digital Humanities*, chapter 4, pages 67–84. Palgrave Macmillan, London, 2012.

[SMKS15] Svenja Simon, Sebastian Mittelstädt, Daniel A. Keim, and Michael Sedlmair. Bridging the Gap of Domain and Visualization Experts with a Liaison. In E. Bertini, J. Kennedy, and E. Puppo, editors, *Eurographics Conference on Visualization (EuroVis) - Short Papers*. The Eurographics Association, 2015.

[Spi12] Lisa Spiro. "This Is Why We Fight": Defining the Values of the Digital Humanities. In Matthew K. Gold, editor, *Debates in the Digital Humanities*, pages 16–35. University of Minnesota Press, Minneapolis, MN, 2012.

[SSS+14] Dominik Sacha, Andreas Stoffel, Florian Stoffel, Bum Chul Kwon, Geoffrey Ellis, and Daniel A Keim. Knowledge generation model for visual analytics. *IEEE transactions on visualization and computer graphics*, 20(12):1604–1613, 2014.

[TKK11] Spyridon Tsipidis, Alexandra Koussoulakou, and Kostas Kotsakis. Geovisualization and archaeology: supporting excavation site research. In *Advances in Cartography and GIScience. Volume 2*, pages 85–107. Springer, 2011.

[VCPK09] R. Vuillemot, T. Clement, C. Plaisant, and A. Kumar. What's being said near "martha"? exploring name entities in literary text collections. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology, 2009*, VAST '09, pages 107–114, 2009.

[vFB+14] Tatiana von Landesberger, Sebastian Fiebig, Sebastian Bremm, Arjan Kuijper, and Dieter W. Fellner. Interaction Taxonomy for Tracking of User Actions in Visual Analytics Applications. In Huang Weidong, editor, *Handbook of Human Centric Visualization*, pages 653–670. Springer, New York, NY, 2014.

[WTP+95] J.A. Wise, J.J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, and V. Crow. Visualizing the non-visual: spatial analysis and interaction with information from text documents. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 51–58, 1995.

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