Teaching on the Intersection of Visualization and Digital Humanities

Stefan Jänicke
University of Southern Denmark, Odense, Denmark

Keywords: Teaching Visualization, Digital Humanities, Journalism.

Abstract: Visualization as a means to generate hypotheses and to communicate insights on digitized cultural heritage data sets has become more and more important in the recent years. While many digital humanities researchers transform their data in order to be processed with ready-to-use tools, others engage in interdisciplinary collaborations with visualization scholars aiming to design novel solutions and interactive visual interfaces as occurring data features are more carefully mapped to visual attributes. Many of those collaborations suffer from loss of time at the beginning of the project as scholars with diverse research backgrounds require to understand each others' mindsets, ways of thinking and research interests. This paper reports on three years of teaching visualization design for digital humanities projects. It provides an overview of theoretical course contents, practical training, collaboration setups—involving computer science, digital humanities as well as humanities students who experienced typical collaboration obstacles—and remarkable project results.

1 INTRODUCTION

For three years, I taught a course on information visualization for the digital humanities equivalent to five ECTS credits. Due to varying student groups throughout the years involving Masters students of the subjects computer science, digital humanities, humanities and journalism, the course included a balanced training of technical, domain-specific and interdisciplinary aspects of digital humanities research involving visualization for hypothesis verification and generation. In this paper, I want to share my experiences in teaching theoretical and practical contents of the course that may serve other teachers as guidelines in designing related courses on the intersection of visualization and digital humanities.

My course included a theoretical training discussed in Section 3 in the form of a lecture giving a broad overview of general visualization aspects, digital humanities as an interdisciplinary domain, means to improve close reading in a digital environment as well as typical distant reading techniques. The theoretical training was complemented with a practical training discussed in Section 4, in which the students carried out little visualization projects themselves. On the basis of related research questions and data sets, the students processed the data according to visualization ideas and developed Web-based applications used for interactive visual data analyses. The main focus is that participating students learn the challenges of interdisciplinary work, and, independent on their studied subjects, gain an understanding on research-related topics and strategies how to design and implement visualizations that appropriately reflect data features visually, serving to support related data analysis tasks.

2 RELATED WORKS

The main goal of my course is to prepare students for interdisciplinary, research-related project work. To achieve this goal, the course structure is mainly inspired by Healey's research-based teaching approach (Healey, 2005), who outlines different ways of connecting research and teaching that are reflected in different carried out activities in my course. According to Dohn and Bolin (Dohn and Dolin, 2015), research-based teaching is the best suited approach as (1) the content and form of teaching is largely focused on current research endeavors in digital humanities, (2) my own background is on the intersection of visualization and digital humanities, the course is designed on the pillars of the visual text analysis process (Jänicke et al., 2017) and includes my own research outcomes, and (3) students actively participate in current research, e.g., by carrying out projects driven by real-world research inquiries.

Jänicke, S.
Teaching on the Intersection of Visualization and Digital Humanities.
DOI: 10.5220/0008987101000109
In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - Volume 3: IVAPP*, pages 100-109
ISBN: 978-989-758-402-2; ISSN: 2184-4321

To serve the students with a fundamental theoretical basis of visualization, I drew suggestions from related works in the visualization field too (Dykes et al., 2010). The course structure is in line with Munzner's subdivions "core readings"—complying to my theoretical contents section—at the beginning, and "individual student presentations"—complying to my practical training section—to the end of the course. The advantage of students having done "the bulk of the readings before they must choose project topics" is especially reflected in the high quality of student project outcomes. As opposed to preferred literature (Spence, 2007; Ware, 2012) used by other teachers in the field (Kerren et al., 2008), I assess the rather application-driven approach to teaching visualization design principles (Munzner, 2014) as best suited for my interdisciplinary course.

3 THEORETICAL CONTENTS

The main objective is that students learn all relevant aspects of interdisciplinary digital humanities projects that aim to include visual interfaces as valuable instruments for humanities scholars to gain insights into their data, most often being text. Moreover, students should aquire the capability of implementing digital humanities projects, and to discuss project directions with—dependent on their own backgrounds—humanities or visualization scholars. The contents of my course are related to the visual text analysis process (Jänicke et al., 2017) as illustrated in Figure 1. Most importantly, projects should start with a central idea, the *text analysis task* that influences all further steps throughout the project. Throughout the course, most related works were discussed in the context of the visual *text* analysis process, but it is important to note that not all of them were based on raw textual data, i.e., some projects based their research on relational databases containing structured textual information. The text analysis task is typically directed towards a text source that is of interest for the humanities scholar. Data transformation and the developed visualizations strongly relate to this task and the underlying data. After a broad overview of the visual text analysis process and exemplary projects, the following contents composed the theoretical training of the course:

- **Visualization Design:** First, I gave an overview of general visualization guidelines based on Munzner's book "Visualization Analysis and Design" (Munzner, 2014). Other visualization text books can be used as a basis for teaching related courses, e.g., Colin Ware's "Information Visualization" (Ware, 2012) choosing a rather theoretical approach or Cao and Cui's "Introduction to Text Visualization" (Cao and Cui, 2016) for a mere application-driven approach. However, Munzner's practical approach of breaking down the visual design process into the questions *WHAT* (data is to be visualized), *WHY* (a visualization is necessary, in other words, what are the user tasks to be supported), and *HOW* (data can be visualized to support the intended user tasks) provides a well-suited, comprehensive overview of the major aspects related for the design process of visualizations in the context of digital humanities projects. The great opportunity is that students learn, independent on their academic backgrounds in computer science, digital humanities, social science or humanities, the same terminology for visual design eabling them to purposefully discuss project ideas. In addition, conveying data and task abstraction strategies is very important to shed a light onto particularities and commonalities among the many concrete projects discussed throughout the course. Finally, one major reason for teaching on the basis of Munzner's works is the nested model (Munzner, 2009). My own experiences (Jänicke et al., 2016; Jänicke and Wrisley, 2017; Khulusi et al., 2019) outline that carrying out digital humanities projects based on the nested model model significantly increases the likeliness that the developed tool serves the intended purpose because domain experts with their knowledge in the targeted research field are strongly involved in the development process. The interdisciplinary exchange necessary for successful digital humanities research students experience in the practical training.

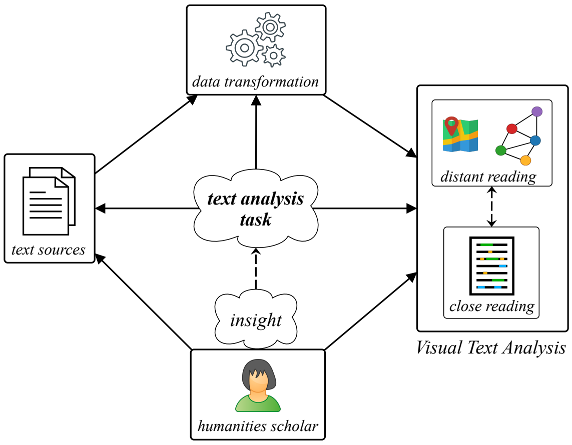


Figure 1: Visual text analysis process (Jänicke et al., 2017).

- **Digital Humanities:** Now that all students were supplied with a basic knowledge of visual design, we discussed the digital humanities as the focused domain. This included major historical developments in the field, e.g., Wladimir Jakowlewitsch Propp's "Morphology of the Folktale" published in 1928, in which he offers the first known quantitative view on a collection of Russian folktales (Propp, 2010), or Roberto Busa's "Index Thomisticus" project that started in 1948 with the lemmatization of Thomas Aquinas' works (Busa, 1980). The overview ended with a discussion of recent driving forces including Franco Moretti's "Graphs, Maps, Trees" (Moretti, 2005) who controversially promotes visual representations for texts, the Culturomics project (Michel et al., 2011) being one of the first attempts to quantitatively analyze digitized text collections in the field, and Matthew Jocker's "Macroanalysis" (Jockers, 2013) giving an overview of the statistical fundaments for such analyses.
- **Close Reading:** Munzner asks for mutual exchange of domain knowledge and research approaches in interdisciplinary projects (Munzner, 2009). As most students of the course are not familiar with the close reading technique (Burke, 2012; Mesmer and Rose-McCully, 2018) being a fundamental method in traditional humanities scholarship, one lecture was dedicated to it, and the students could perform a close reading themselves in order to reflect on the main idea of a text, to think about narrative style, vocabulary, syntax and the context in which a text was written. Contextualizing our own work (Cheema et al., 2016), I conveyed the capabilities of how visualization can support digital close reading (Kehoe and Gee, 2013) ranging from manual annotations to quantitative distant reading statistics.
- **Information Seeking Mantra:** On the bridge between close and distant reading, I discussed the main differences among both techniques using Shneiderman's mantra "Overview first, zoom and filter, details-on-demand" (Shneiderman, 1996). While distant reading or macroanalysis are the terms used by digital humanities scholars for *overview*, close reading or mircoanalyis relate to *details on demand*. For the continuum in between, zooming and meso reading (Jänicke and Wrisley, 2017) are the established terms. Many visualizations that have been designed implement the information seeking mantra as the humanities scholars are provided with a distant reading, and means of interaction lead them to interesting portions of the underlying data source.
- **Distant Reading:** Throughout the course, I discussed the main related distant reading techniques—heat maps, tag clouds, maps, timelines and graphs—with the students both in theory and practice. On the one hand, conceptual and technical details are outlined such as for tag clouds (Sinclair and Cardew-Hall, 2008; Viegas and Wattenberg, 2008) and timelines (Aigner et al., 2011; Bach et al., 2015; Brehmer et al., 2017), on the other hand, practical examples from related works underpin the value of distant reading techniques for (digital) humanities scholars (Jänicke et al., 2017). Related works were chosen based on their actual applicability to digital humanities research. For example, Collins' Parallel Tag Clouds (Collins et al., 2009) that have been applied in the Trading Consequences project (Hinrichs et al., 2015), was discussed. This way students get to know strategies how to adapt existing techniques to support research inquiries in digital humanities. Further, basics of drawing graphs (Tamassia, 2007)—being the most frequently occurring data structure in digital humanities research—were conveyed. My motivation to stress this topic is the rather low quality of graph visualization design in related works. Digital humanities researchers typically apply ready-to-use tools (Bastian et al., 2009) to generate visual output quickly, disregarding means to appropriately map graph-structural to visual features. The benefits of careful graph design were dicussed on the basis of our work on TRAViz (Jänicke et al., 2015), a graph visualization reflecting the features of aligned text editions on sentence level, which was compared to previous approaches (Haentjens Dekker et al., 2014; Andrews and Macé, 2013) using a standard graph drawing library. Finally, as uncertain information is often comprised in digital humanities data due to the long history the data refers to, current research endeavors to tackle this problem (Börner et al., 2019) were discussed in the course.

4 PRACTICAL TRAINING

One of the most important driving forces for the practical training was my own interdisciplinary background. When I started working in digital humanities projects as a computer scientist (Jänicke, 2016), me as well as my collaboration partners from the humanities needed to learn speaking the same language and to understand each others' research interests and ways of thinking. I admired that students get this ex-

Table 1: Projects carried out in three years involving students and researchers with backgrounds in the humanities [orange], digital humanities [green] and computer science [blue].

Project Topic	Perfect Setup?				R	F	H	Major Project Tasks
Merseburg Incantations	orange	orange	gray	gray	gray	gray	blue	data retrieval, data modeling, geovisualization
Witchcraft Trials	orange	orange	blue	blue	gray	gray	gray	digitization, data modeling, geo-temporal visualization
Movie Relations	blue	gray	gray	gray	gray	orange	gray	network visualization, optimization
Actor Biographies	blue	blue	gray	gray	gray	orange	gray	interface design, network visualization, statistical charts
Musical Instruments	blue	gray	gray	gray	orange	gray	gray	geo-temporal visualization, uncertainty mapping
Letter Exchange	blue	blue	gray	gray	gray	orange	gray	text processing/ranking, parallel tag cloud variant
Engineering Subjects	blue	gray	gray	gray	orange	gray	gray	interactive tag cloud (to derive meta-subjects)
Musical Genres	blue	gray	gray	gray	orange	gray	gray	interactive tag cloud (to guide to close reading)
Instrument Makers	blue	gray	gray	gray	orange	gray	gray	repository linking, geo-temporal visualization
Witchcraft Trials II	orange	blue	blue	gray	gray	gray	blue	data modeling, geo-temporal visualization
Movie Locations	blue	green	gray	gray	gray	orange	gray	interface design, geo-temporal visualization
Michelin Restaurants	orange	orange	orange	orange	gray	gray	blue	data retrieval, geovisualization, uncertainty mapping
Bundestag Politicians	orange	orange	orange	green	gray	gray	blue	data retrieval, parallel coordinates variant
Music Styles	orange	orange	orange	blue	gray	gray	gray	data retrieval, geovisualization
Fake News	orange	orange	orange	gray	gray	gray	blue	data retrieval, text processing, tag cloud adaption
Hip Hop Musicians	orange	orange	gray	gray	gray	gray	blue	data retrieval, data modeling, network graph adaption
Education Spending	orange	orange	gray	gray	gray	gray	blue	data retrieval, geovisualization
Solidary Pact	orange	orange	gray	gray	gray	gray	blue	data retrieval, geovisualization

perience before they actually commit to join a digital humanities project.

The theoretical part of the course covered half of the semester. In order to practically apply the gained knowledge, students required to implement small project ideas in groups. I outlined the frame of the practical training at semester start and asked the students to think about projects they like to work on. It was interesting to observe that the more time in the theoretical section passed, the more precise students were able to formulate project ideas using the learned technical terms.

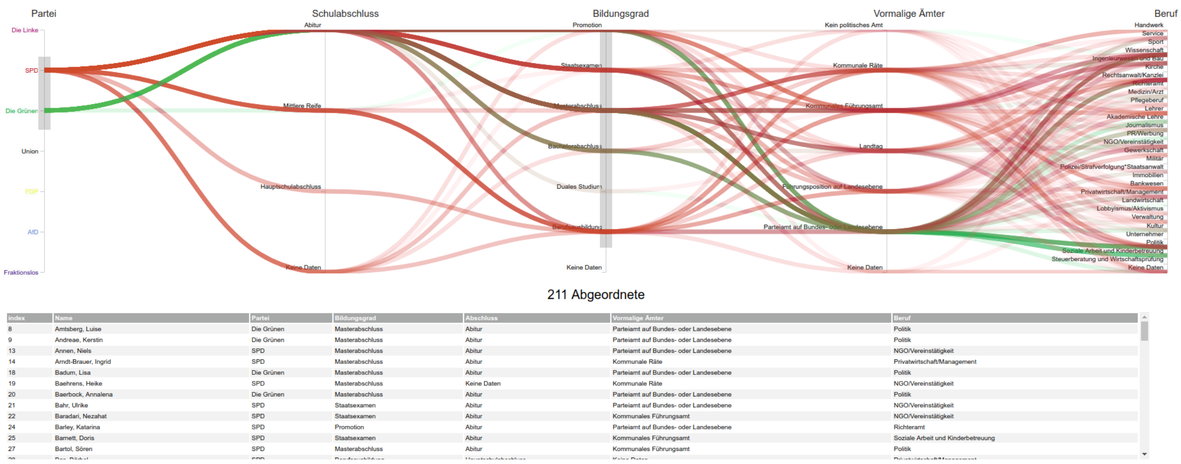
In three years, I faced very different student groups. In the first two years, only computer science and a few humanities students were allowed to take part in the course. In the last year, the ratio completely switched as the course was offered for the journalism and digital humanities Masters degrees for the first time. Thus, in all three years I had to improvise when setting up interdisciplinary projects involving visualization as well as humanities scholars. An overview of all projects carried out in three years is provided in Table 1.

4.1 The Perfect Setup

Only few projects involved students with a computer science (or digital humanities) and a humanities (or journalism or digital humanities) background alike. This way, the whole project was in the students' hands, and I just joined as a mediator if necessary. The results of those projects were usually limited as students faced coordination and communication issues. In one case, humanities students very carefully generated a database manually, while, on the other end, computer science students were waiting for real world data to start with. Another issue arose when only one student with a computer science background joined a group with more than one humanities scholar—the computer science student was expected to implement the project.

However, this constellation also brought forth remarkable results. A project, in which one digital humanities scholar with a decent computer science training and three journalism students participated, focused on the biographies of politicians of the German Bundestag as of June 2019. They were inspired by the New York Times visualization (NYT, 2019) reporting on the career paths of politicians in the U.S. congress. The students took the biographies from the Bundestag homepage[1] having a far easier understandable visualization as opposed to the New York Times in mind. They focused on different aspects of a career: political party, school degree, education degree, former political position and former practised profession. In order to extract those information for 709 politicians, they implemented rules to parse the crawled biographical texts accordingly. In order to browse the gained data, they adapted a parallel coordinates visualization variant that can be used to interactively search for patterns in the data. Figure 2 compares members of the SPD and Die Grünen Bundestag factions. One can find out that all members of Die Grünen had a position on federal or state level before joining the Bundestag. Instead, many SPD politicians having a position in the commune were elected for the Bundestag.

[1] https://www.bundestag.de/abgeordnete/biografien



index	Name	Partei	Bildungsgrad	Abschluss	Vormalige Ämter	Beruf
8	Amtsberg, Luise	Die Grünen	Masterabschluss	Abitur	Parteiamt auf Bundes- oder Landesebene	Politik
9	Andreae, Kerstin	Die Grünen	Masterabschluss	Abitur	Parteiamt auf Bundes- oder Landesebene	Politik
13	Annen, Niels	SPD	Masterabschluss	Abitur	Parteiamt auf Bundes- oder Landesebene	NGO/Vereinstätigkeit
14	Arndt-Brauer, Ingrid	SPD	Masterabschluss	Abitur	Kommunale Räte	Privatwirtschaft/Management
18	Badum, Lisa	Die Grünen	Masterabschluss	Abitur	Parteiamt auf Bundes- oder Landesebene	Politik
19	Baehrens, Heike	SPD	Masterabschluss	Keine Daten	Kommunale Räte	NGO/Vereinstätigkeit
20	Baerbock, Annalena	Die Grünen	Masterabschluss	Abitur	Parteiamt auf Bundes- oder Landesebene	Politik
21	Bahr, Ulrike	SPD	Staatsexamen	Abitur	Parteiamt auf Bundes- oder Landesebene	NGO/Vereinstätigkeit
22	Baradari, Nezahat	SPD	Staatsexamen	Abitur	Kommunales Führungsamt	NGO/Vereinstätigkeit
24	Barley, Katarina	SPD	Promotion	Abitur	Parteiamt auf Bundes- oder Landesebene	Richteramt
25	Barnett, Doris	SPD	Staatsexamen	Abitur	Kommunales Führungsamt	Soziale Arbeit und Kinderbetreuung
27	Bartol, Sören	SPD	Masterabschluss	Abitur	Kommunales Führungsamt	Politik

Figure 2: A Perfect Setup project: Analyzing the biographies of 709 politicians of the German Bundestag.

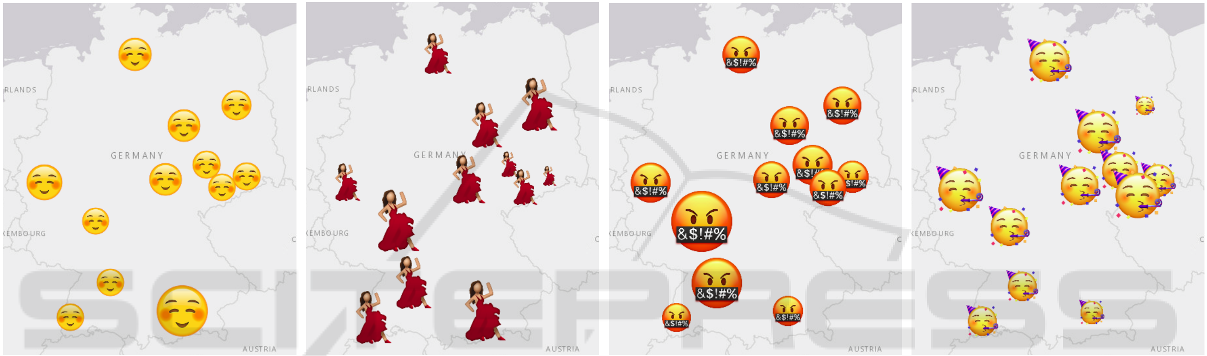


Figure 3: A Perfect Setup project: Favored music styles in 10 German cities regarding the music features *happy, dancable, explicit,* and *energetic*.

Another project group involving one computer science and three journalism students focused on the comparison of favored music features in Eastern vs. Western German cities. Therefore, the Spotify playlists[2] of ten cities were analyzed according to the music features *happy, dancable, explicit* and *energetic*. The mean value from a city's playlist for each of those features is mapped to the range $[0,1]$ that is used to scale an emoji icon that reflects the corresponding music feature best. The resulting maps are shown in Figure 3. As the students extracted the playlists manually, they restricted their observation to few German cities. They admitted that an appropriate answer to the given comparative analysis interest is hard to find: the number of chosen cities was subjective, the equal amount of representatives from East and West did not stand for the population ratio, and the number of songs in the playlists differed. However, the maps revealed some interesting information on city level. Music heard in Frankfurt is typically dancable and explicit, but less happy and energetic. In Saxony, people listen to energetic music that is less happy, dancable and explicit.

4.2 The Real Humanities Scholar

When too many computer science students participated in the course, I prepared a number of digital humanities projects that are usually driven by research interests from humanities scholars. Therefore, I asked my collaboration partners if they could come up with a task that fits into the practical training without overstraining the students. Moreover, I asked them whether they could serve as collaboration partners during the practical training to discuss prototypes and results. The projects had a higher success rate as opposed to *The Perfect Setup* as my collaboration partners are experienced in participating digital humanities projects. Very good results were achieved when a profound research query met an engaged computer science student, also leading to publishing the outcomes (Meinecke and Jänicke, 2018; Khulusi et al., 2020).

In two projects, students worked together with a

[2] https://spotifymaps.github.io/musicalcities/

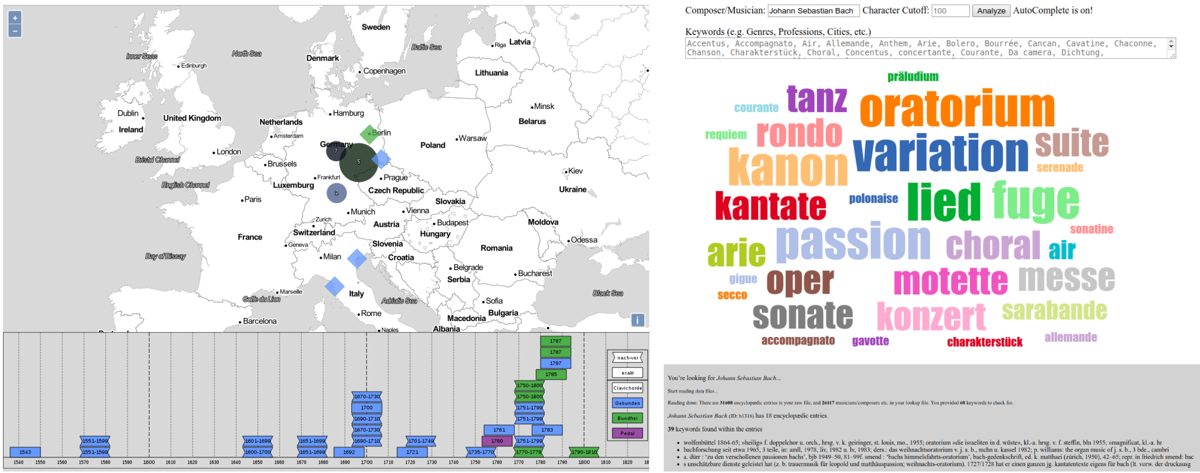


Figure 4: Real Humanities Scholar projects: Comparative visualization of geospatial-temporal metadata of instruments (left), and tag cloud visualization supporting the interactive exploration of a composer's music genres (right).

professor from the university's musicology department with whom I cooperate very successful since 2015 in various projects (Jänicke et al., 2016; Jänicke and Focht, 2017; Khulusi et al., 2019). The first research interest focused on comparatively exploring geospatial-temporal metadata of clavichord instruments in the catalogue of the Musical Instruments Museum Leipzig. Next to mapping the clavichord type to color, a visual design to communicate uncertain datings and different geospatial granularities was developed (see Figure 4, left). The second research interest was an analysis of music genres of the works of composers. Based on the musiXplora (Khulusi et al., 2020) containing biographical information about around 30,000 musicians, fulltext biographies were scanned for a pre-defined list of genre types (such as opera, rondo, concert, etc.) and made available in an interactive tag cloud. Entering the name of a composer, related genre tags are shown. The Information Seeking Mantra is fulfilled as when clicking a genre tag, the according text fragments can be close read. An example for genre tags of Johann Sebastian Bach is shown in Figure 4 (right).

4.3 The Fake Humanities Scholar

As the number of project ideas from my collaboration partners was not exhausive enough, I—being a computer scientist who worked in interdisciplinary projects for ten years—acted myself as the humanities scholar of some projects. Therefore, I prepared project ideas that could generate interesting insights into popular data sets. The project results in this category were typically full-scale implementations of the Information Seeking Mantra (Shneiderman, 1996) serving undirected exploration purposes rather than focused research interests.

The IMDb database[3] suited perfectly well in this constellation as it provides various easily understandable information with different types, and only few sophisticated visualizations (Vlachos and Svonava, 2013; Alam and Jianu, 2016; Auber et al., 2003) have been proposed to explore film history in the spirit of the Filmfinder (Ahlberg and Shneiderman, 1994). Related student projects in my course included developing visual designs to explore movie relationships, biographies of persons such as actors or directors, and film locations. A screenshot of the tool developed to browse film locations, which had to be geocoded, is shown in Figure 5. Each circle refers to one or multiple locations, and when choosing a location, related movies can be explored. The bar chart encodes genre with color, and a bar stands for the number of movies filmed at the chosen location in a certain year.

4.4 The Helper

Especially in the last year with too few computer science students in the course, humanities students required technical support. Many of those were journalism students with only very basic computer science skills. Partially, I trained them how to apply existing tools and visualizations to their data. Also, students took charge of services provided by the university for computer science students of early semesters. Typical for those projects was a very high quality of the data source being composed, and that visualizations were mostly adapted to fit this data. In one project, the students developed a database of German hip hop musicians, and they generated a list of relations among

[3] https://www.imdb.com/interfaces/

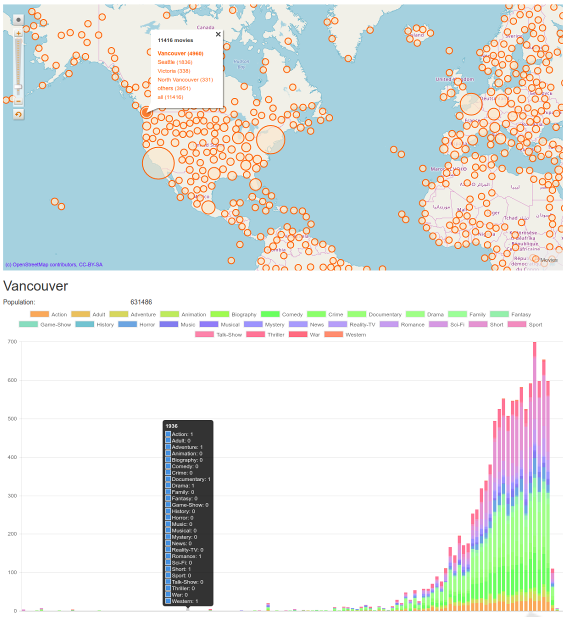


Figure 5: A Fake Humanities Scholar project: Exploring movie locations.

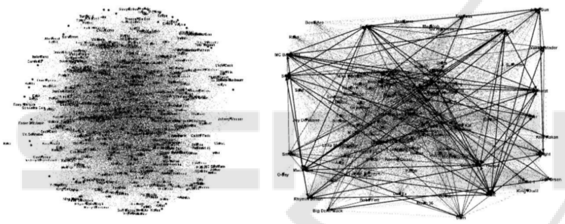


Figure 6: A Helper project: Social network visualization of German hip hop musicians visualized with Gephi.

them. A close relation was defined when two musicians were participating the same band, a loose relation was defined when two musicians having a similar age were living in the same city. Gephi (Bastian et al., 2009) was used for visualizing different subsets of the graph as shown in Figure 6.

Another group focused on the history of German Michelin restaurants. They were able to generate an aestehtic visual output in the form of an interactive map accompanied with a time slider that allows to explore changes throughout the years (see Figure 7). The star size encodes the number of Michelin stars (1, 2, or 3) and a star is highlighted in orange once the number of stars change. When transforming historical data that was provided in lists of Michelin restaurants per year, the students had to encode uncertainties as some restaurants were already closed leading to geocoding errors. As the main focus of the project was comparing the coverage of Michelin restaurants in Eastern and Western Germany, in such cases, cities were chosen instead of exact positions keeping global relationships stable. To support the comparison task,

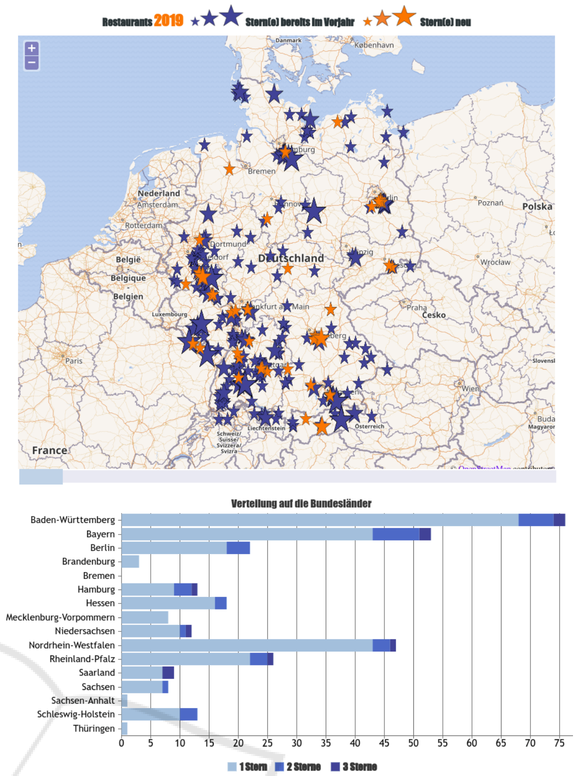


Figure 7: A Helper project: Geospatial-temporal analysis of German Michelin restaurants.

numbers of Michelin restaurants could be compared by state using a horizontal bar chart.

In one case in the first two years, I needed to replace the computer science students of a project as they resigned from the course. The project focused on how the Merseburg incantations are referred to in the media. The collected data was very well structured and together with humanities students this lead to a poster presentation at the digital humanities conference in 2017 (Jänicke, 2017).

5 REFLECTION

In the three years I used different means to evaluate the course, ranging from standardized to content-related questionnaires. In addition, I received valuable feedback in discussions. Together with my experiences gained, I reflect on important aspects of designing a related course.

Group Heterogeneity. The focus of the theoretical course contents should be adjusted according to the needs of the participating students. In the last year, when the majority of students had a background in journalism or digital humanities, I frequently discussed visualizations published in media. However, some topics of the course were seen as too deeply discussed (e.g., close reading, or temporal data visualizations based on space-time cube operations (Bach et al., 2015)). On the other hand, computer science students, being the majority in the first two years, liked the comprehensive overview of related projects from the digital humanities as an application domain. Thus, on the one hand, working with a homogeneous group makes it easier to adapt to the students' needs, on the other hand, a heterogeneous group ensures that students with different backgrounds work together.

Student Projects. Students were seemingly more engaged when they were able to choose their own project topics. To longer the course took during the semester, the more they were able to generate ideas on their own—especially, journalism students brought forth projects that were interesting from a technical data analysis perspective while being relevant to society too. I judged if their ideas fit into the course, and if so, we discussed the extent of the result to be expected. Adaptions in this regard were made when I recognized that a group performed better or worse as expected when working on the projects. As projects were carried out on the basis of Munzner's nested model (Munzner, 2009), which inheres iterative evaluation due to the involvement of domain experts in the visualization design process, the evaluation of visualizations was not explicitly discussed.

Grading. Projects were not graded, but a successful project was necessary to join an oral exam. Students presented their project results in front of the group, and they were asked to contextualize their project in related works from visualization and digital humanities. I further engaged them to reflect on collaboration experiences gained during their projects. In three years, a total of four students resigned from the course. Out of 18 projects, 13 were assessed as successful, the students who presented insufficient results received additional project time. In this second iteration 4 out of 5 projects were assessed as successful. All students passed the subsequent oral exam.

Data Retrieval and Modeling. Many projects faced data acquisition problems. On the one hand, this relates to limited data processing skills, on the other hand, to the short project durations. While in one project numbers needed to be manually extracted from old documents for which OCR failed, other projects suffered from the long response time of data deliveries—seemingly all valuable lessons learned for future projects. Though students who start with already given data sets overall gained better project results concerning visualization quality, the challenges of compiling and modeling a data set for a given research inquiry is a valuable task, especially for (digital) humanities students. In addition, the exchange on data modeling supports the interdisciplinary discourse among students having different backgrounds.

Data Transformation. A broad overview of data transformation techniques should be provided to convey all aspects of the visual text analysis process sufficiently. This was remarked by students as my limited lecture time allowed only the discussion of few related methods such as Named Entity Recognition (NER). Ideally, students should gain pre-requisites in text processing before partaking the course.

Hands-on Sessions. Students without a technical background require practical hands-on sessions throughout the theoretical part of the course. On the one hand, they wish to apply existing visualization tools to data sets—a common work practice of digital humanities scholars and journalists—, on the other hand, the programming of sample snippets was seen valuable before beginning the practical training in student projects. For the former, I suggest sessions with standard tools like Voyant[4], CollateX[5] or Gephi[6] in order to foster understanding of the underlying research questions and to discuss potential visualization-related improvements.

Collaboration. Many students described the value of the project work similar to the following remark: *"The development of visualization has brought me a lot further, because I can now imagine the implementation of such projects much more concretely."* Students with a non-technical background remarked they would not have been able to design related visualizations without the knowledge gained in the course. Most of the students joined an interdisciplinary project for the first time, and the course taught them strategies how to approach them. Some students valued the relevance of the project work for their future, also they saw how the undertaken steps could be applied to support other data analysis tasks. Lastly, the exchange with other groups facing similar tasks was important. On the contrary, students faced typical callenges of interdisciplinary work, especially in the perfect setup projects.

[4]https://voyant-tools.org/
[5]https://collatex.net/
[6]https://gephi.org/

An Ideal Course. A 5 ECTS course is too short to teach all aspects of the visual text analysis process (Jänicke et al., 2017), especially for students having a non-technical background. I suggest two 5 ECTS courses in subsequent semesters with the following objectives:

- **Semester 1—Contents:** visual text analysis overview, introduction to digital humanities, traditional vs. digitally supported workflows, text mining basics, standard data transformation techniques (e.g., tokenization, stemming, named entity recognition, part-of-speech tagging)
- **Semester 1—Project:** data retrieval and data modeling, preferably driven by an own project idea
- **Semester 2—Contents:** visualization design, nested model, Information Seeking Mantra, close and distant reading visualization techniques
- **Semester 2—Project:** iterative development of a tool to support visual text analyses, preferably based on a data set generated in the first semester

Though the spread throughout two semesters increases the likeliness of project groups breaking apart or loosing members, this way, students gain more time to generate data sets according to their own interests. If the first project does not deliver an appropriate data set, ready-to-use data sets can be used in the second semester. As the course pursues that students experince interdisciplinary digital humanities projects, only "perfect setup" and "real humanities scholar" projects should be conducted. To avoid other constellations, the ratios of partaking (digital) humanities and computer science students should both range around 50%, in case of a lower ratio of humanities students, the number of real humanities scholar projects should compensate this imbalance.

6 SUMMARY

The results illustrated in the figures are—without any changes—the visual designs as presented by the groups on the last course day. Even though most visualizations leave room for improvements due to some inappropriate design decisions, I was impressed by the quality of project outcomes considering the fact that many students had neither computer science nor visualization skills before joining the course.

When designing a course like this, one needs to be able to take different roles. For me, it was much easier to take the helper role as my capabilities as a computer scientist were requested. But my digital humanities experience helped me also to join in as a fake humanities scholar. Nevertheless, having a series of project ideas offered by real humanities scholars participating as partners during the project developments is invaluable as it guarantees the steepest learning curve for students with a computer science background.

REFERENCES

Ahlberg, C. and Shneiderman, B. (1994). Visual information seeking using the filmfinder. In *Conference companion on Human factors in computing systems*, pages 433–434. ACM.

Aigner, W., Miksch, S., Schumann, H., and Tominski, C. (2011). *Visualization of Time-Oriented Data*. Springer Publishing Company, Incorporated, 1st edition.

Alam, S. S. and Jianu, R. (2016). Analyzing eye-tracking information in visualization and data space: from where on the screen to what on the screen. *IEEE transactions on visualization and computer graphics*, 23(5):1492–1505.

Andrews, T. L. and Macé, C. (2013). Beyond the tree of texts: Building an empirical model of scribal variation through graph analysis of texts and stemmata. *Literary and Linguistic Computing*, 28(4):504–521.

Auber, D., Chiricota, Y., Jourdan, F., and Melançon, G. (2003). Multiscale visualization of small world networks. In *IEEE Symposium on Information Visualization 2003*, pages 75–81.

Bach, B., Dragicevic, P., Archambault, D., Hurter, C., and Carpendale, S. (2015). A Descriptive Framework for Temporal Data Visualizations Based on Generalized Space-Time Cubes. *Computer Graphics Forum*, 36(6):36–61.

Bastian, M., Heymann, S., Jacomy, M., et al. (2009). Gephi: an open source software for exploring and manipulating networks. *ICWSM*, 8:361–362.

Börner, K., Eide, O., Mchedlidze, T., Rehbein, M., and Scheuermann, G. (2019). Network Visualization in the Humanities (Dagstuhl Seminar 18482). *Dagstuhl Reports*, 8(11):139–153.

Brehmer, M., Lee, B., Bach, B., Riche, N. H., and Munzner, T. (2017). Timelines Revisited: A Design Space and Considerations for Expressive Storytelling. *IEEE Transactions on Visualization and Computer Graphics*, 23(9):2151–2164.

Burke, B. (2012). A Close Look at Close Reading: Scaffolding Students with Complex Texts.

Busa, R. (1980). *The Annals of Humanities Computing: The Index Thomisticus*. Raccolta delle opere di Roberto Busa. North Holland Publishing Company.

Cao, N. and Cui, W. (2016). *Introduction to Text Visualization*. Atlantis briefs in artificial intelligence. Atlantis Press, Paris.

Cheema, M. F., Jänicke, S., and Scheuermann, G. (2016). AnnotateVis: Combining Traditional Close Reading with Visual Text Analysis. In *Workshop on Visualization for the Digital Humanities, IEEE VIS 2016, Baltimore, Maryland, USA*.

Collins, C., Viegas, F. B., and Wattenberg, M. (2009). Parallel tag clouds to explore and analyze faceted text corpora. In *2009 IEEE Symposium on Visual Analytics Science and Technology*, pages 91–98.

Dohn, N. and Dolin, J. (2015). *Research-based teaching*, pages 43–63. Samfundslitteratur, 1. edition.

Dykes, J., Keefe, D. F., Kindlmann, G., Munzner, T., and Joshi, A. (2010). Perspectives on Teaching Data Visualization. *IEEE VisWeek 2010*.

Haentjens Dekker, R., van Hulle, D., Middell, G., Neyt, V., and van Zundert, J. (2014). Computer-supported collation of modern manuscripts: Collatex and the beckett digital manuscript project. *Literary and Linguistic Computing*, 30(3):452–470.

Healey, M. (2005). Linking research and teaching exploring disciplinary spaces and the role of inquiry-based learning. *Reshaping the University: New Relationships between Research, Scholarship and Teaching*, pages 67–78.

Hinrichs, U., Alex, B., Clifford, J., Watson, A., Quigley, A., Klein, E., and Coates, C. M. (2015). Trading Consequences: A Case Study of Combining Text Mining and Visualization to Facilitate Document Exploration. *Digital Scholarship in the Humanities*, 30(suppl_1):i50–i75.

Jänicke, S. (2016). Valuable Research for Visualization and Digital Humanities: A Balancing Act. In *Workshop on Visualization for the Digital Humanities, IEEE VIS 2016, Baltimore, Maryland, USA*.

Jänicke, S. (2017). On the Impact of the Merseburg Incantations. In *Conference Abstracts of the Digital Humanities 2017*.

Jänicke, S. and Focht, J. (2017). Untangling the social network of musicians. In *Conference Abstracts of the Digital Humanities 2017*.

Jänicke, S., Focht, J., and Scheuermann, G. (2016). Interactive visual profiling of musicians. *IEEE transactions on visualization and computer graphics*, 22(1):200–209.

Jänicke, S., Franzini, G., Cheema, M. F., and Scheuermann, G. (2017). Visual Text Analysis in Digital Humanities. *Computer Graphics Forum*, 36(6):226–250.

Jänicke, S., Geßner, A., Franzini, G., Terras, M., Mahony, S., and Scheuermann, G. (2015). TRAViz: A Visualization for Variant Graphs. *Digital Scholarship in the Humanities*, 30(suppl_1):i83–i99.

Jänicke, S. and Wrisley, D. J. (2017). Interactive Visual Alignment of Medieval Text Versions. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 127–138.

Jockers, M. L. (2013). *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press, Champaign, IL, USA, 1st edition.

Kehoe, A. and Gee, M. (2013). eMargin: A Collaborative Textual Annotation Tool. *Ariadne*, 71.

Kerren, A., Stasko, J. T., and Dykes, J. (2008). Teaching Information Visualization. pages 65–91. Springer.

Khulusi, R., Jänicke, S., Kusnick, J., and Focht, J. (2019). An Interactive Chart of Biography. *Pacific Visualization Symposium (PacificVis), 2019 IEEE*.

Khulusi, R., Kusnick, J., Focht, J., and Jänicke, S. (2020). musiXplora:Visual Analysis of a Musicological Encyclopedia. In *Proceedings of the 11th International Conference on Information Visualization Theory and Applications (IVAPP)*.

Meinecke, C. and Jänicke, S. (2018). Visual Analysis of Engineer's Biographies and Engineering Branches. In *LEVIA 2018: Leipzig Symposium on Visualization in Applications*.

Mesmer, H. A. and Rose-McCully, M. (2018). A closer look at close reading: Three under-the-radar skills needed to comprehend sentences. *The Reading Teacher*, 71(4):451–461.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., , Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., and Aiden, E. L. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books. 331(6014):176–182.

Moretti, F. (2005). *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso.

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

Munzner, T. (2014). *Visualization Analysis and Design*. CRC press.

NYT (2019). New York Times: How Every Member Got to Congress. https://www.nytimes.com/interactive/2019/01/26/opinion/sunday/paths-to-congress.html (Retrieved 2019-07-08).

Propp, V. (2010). *Morphology of the Folktale*, volume 9. University of Texas Press.

Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Visual Languages, Proceedings*, pages 336–343.

Sinclair, J. and Cardew-Hall, M. (2008). The folksonomy tag cloud: when is it useful? *Journal of Information Science*, 34(1):15–29.

Spence, R. (2007). *Information Visualization: Design for Interaction*. Pearson/Prentice Hall.

Tamassia, R. (2007). *Handbook of Graph Drawing and Visualization (Discrete Mathematics and Its Applications)*. Chapman & Hall/CRC.

Viegas, F. and Wattenberg, M. (2008). TIMELINES: Tag Clouds and the Case for Vernacular Visualization. *interactions*, 15(4):49–52.

Vlachos, M. and Svonava, D. (2013). Recommendation and visualization of similar movies using minimum spanning dendrograms. *Information Visualization*, 12:85–101.

Ware, C. (2012). *Information Visualization: Perception for Design*. Elsevier.